ECOGRAPHY

Software note

Achieving higher standards in species distribution modeling by leveraging the diversity of available software

Jamie M. Kass[1,2], Adam B. Smith[3], Dan L. Warren[4,5], Sergio Vignali[6], Sylvain Schmitt[7,8], Matthew E. Aiello-Lammens[9], Eduardo Arlé[10,11], Ana Márcia Barbosa[12], Olivier Broennimann[13], Marlon E. Cobos[14], Maya Guéguen[15], Antoine Guisan[13], Cory Merow[16], Babak Naimi[17], Michael P. Nobis[18], Ian Ondo[19,20], Luis Osorio-Olvera[21,22], Hannah L. Owens[23,24], Gonzalo E. Pinilla-Buitrago[25], Andrea Sánchez-Tapia[26], Wilfried Thuiller[15], Roozbeh Valavi[27], Santiago José Elías Velazco[28,29], Alexander Zizka[30] and Damaris Zurell[31]

[1]Macroecology Laboratory, Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi, Japan
[2]Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan
[3]Missouri Botanical Garden, Saint Louis, MO, USA
[4]Environmental Science and Informatics Section, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan
[5]Gulbali Institute, Charles Sturt University, Albury, NSW, Australia
[6]Division of Conservation Biology, Institute of Ecology and Evolution, University of Bern, Bern, Switzerland
[7]CIRAD, UPR Forêts et Sociétés, Montpellier, France
[8]Forêts et Sociétés, Univ Montpellier, CIRAD, Montpellier, France
[9]Department of Environmental Studies and Science, Pace University, Pleasantville, NY, USA
[10]School of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel
[11]Macroecology and Society, German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Leipzig, Germany
[12]Research Centre for Geospatial Sciences (CICGE), Porto University, Alameda do Monte da Virgem, Vila Nova de Gaia, Portugal
[13]Department of Ecology and Evolution and Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland
[14]Department of Ecology and Evolutionary Biology and Biodiversity Institute, University of Kansas, Lawrence, KS, USA
[15]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LECA, Laboratoire d'Ecologie Alpine, Grenoble, France
[16]Eversource Energy Center and Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA
[17]Quantitative Biodiversity Dynamics (QBD), Department of Biology, University of Utrecht, Utrecht, the Netherlands
[18]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland
[19]Royal Botanic Gardens, Kew, Richmond, UK
[20]United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC), Cambridge, UK
[21]Laboratorio de Ecoinformática de la Biodiversidad, Departamento de Ecología de la Biodiversidad, Instituto de Ecología, Unidad Mérida, Yucatán, Universidad Nacional Autónoma de México, Ciudad de México, México
[22]Laboratorio Nacional Conahcyt de Biología del Cambio Climático, CONAHCyT, Ciudad de México, México
[23]Center for Global Mountain Biodiversity, Globe Institute, University of Copenhagen, Copenhagen, Denmark
[24]Florida Museum of Natural History, University of Florida, Gainesville, FL, USA
[25]PhD Program in Biology, The Graduate Center and Department of Biology, City College of New York, City University of New York, New York, NY, USA

Ecography
2024: e07346
doi: 10.1111/ecog.07346

Subject Editor:
Michael Krabbe Borregaard
Editor-in-Chief:
Robert D. Colwell
Accepted 4 October 2024

NORDIC SOCIETY OIKOS

www.ecography.org

[26]Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
[27]CSIRO Environment, Clayton South, VIC, Australia
[28]Instituto de Biología Subtropical, Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad Nacional de Misiones, Puerto Iguazú, Misiones, Argentina
[29]Department of Geography, San Diego State University, San Diego, CA, USA
[30]Biodiversity of Plants Lab, Department of Biology, Philipps-University Marburg, Marburg, Germany
[31]Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany

Correspondence: Jamie M. Kass (kass@tohoku.ac.jp)

The increasing online availability of biodiversity data and advances in ecological modeling have led to a proliferation of open-source modeling tools. In particular, R packages for species distribution modeling continue to multiply without guidance on how they can be employed together, resulting in high fidelity of researchers to one or several packages. Here, we assess the wide variety of software for species distribution models (SDMs) and highlight how packages can work together to diversify and expand analyses in each step of a modeling workflow. We also introduce the new R package 'sdmverse' to catalog metadata for packages, cluster them based on their methodological functions, and visualize their relationships. To demonstrate how pluralism of software use helps improve SDM workflows, we provide three extensive and fully documented analyses that utilize tools for modeling and visualization from multiple packages, then score these tutorials according to recent methodological standards. We end by identifying gaps in the capabilities of current tools and highlighting outstanding challenges in the development of software for SDMs.

Keywords: code, habitat, niche, programming, R package, reproducibility

Introduction

Software for ecology and evolution has expanded dramatically over recent years (Farley et al. 2018) and is often released as open-source packages for the R programming language (www.r-project.org). This expansion is particularly striking for species distribution models (SDMs), which estimate species' responses to environmental variables and can predict their geographic distributions for unsampled locations and times (Franklin 2010, Guisan et al. 2017, Peterson et al. 2011). Packages in R for building, evaluating, and applying SDMs have expanded steadily over the past 20 years as the field has grown (Guisan and Zimmerman 2000, Valavi et al. 2022). Researchers can now conduct complete analyses from data acquisition and geospatial data manipulation to modeling and post-processing of results within a single programming environment. In 2023 on CRAN, the main network of servers maintaining up-to-date R package versions, approximately 40 packages relating to various aspects of SDMs were available (Sillero et al. 2023), with many more referencing SDMs on Github (>200 results; Supporting information).

Why are there so many packages for SDMs? The field is consolidating around a set of methodological standards (Araújo et al. 2019, Zurell et al. 2020), and researchers from different groups are developing tools to achieve them. This results in multiple approaches to address similar issues, and growing coding literacy among ecologists means new methods are typically accompanied by novel software implementations. Additionally, SDM analysis involves multiple linked tasks necessitating a variety of tools, including downloading data from public repositories, tabular data management and cleaning, manipulation of geospatial data, and statistical modeling. However, there is still a tendency for researchers, including package developers, to conduct their analyses using a single SDM package (Ahmed et al. 2015), which restricts the available tools for a given study and can result in an underestimation of methodological uncertainty arising from different approaches to the same problem. In addition, it is often difficult to identify the most appropriate methods from the wealth of options, especially for researchers new to the field. Removing these barriers would improve SDMs by better utilizing available tools in current packages and the growing trove of those yet to be developed.

Representing a broad sample of developers of R packages for SDMs, we first assess the SDM software ecosystem, highlighting ways packages can work together to diversify and expand analyses. We focus on packages available on CRAN and Github before May 2023, but packages for SDMs are in constant development and keeping up-to-date can be a laborious exercise. In an effort to catalog, annotate, and explore packages for SDMs, we present the new R package 'sdmverse 1' (https://github.com/sylvainschmitt/sdmverse; https://doi.org/10.5281/zenodo.13927581), a community-moderated package bibliography explained in more detail below. Additionally, to demonstrate how different packages can be combined to improve SDM workflows, we provide three fully worked and documented analyses featuring frequent topics in the field (https://doi.org/10.6084/m9.figshare.27312903), then score them based on recent methodological standards (Araújo et al. 2019). We end by identifying gaps in the capabilities of current tools and suggest possible ways forward by highlighting outstanding challenges in the development of SDM software.

Species distribution modeling workflow

We structure our assessment of existing software by following a typical SDM workflow (Fig. 1), which has been described at length in books (Franklin 2010, Peterson et al. 2011,

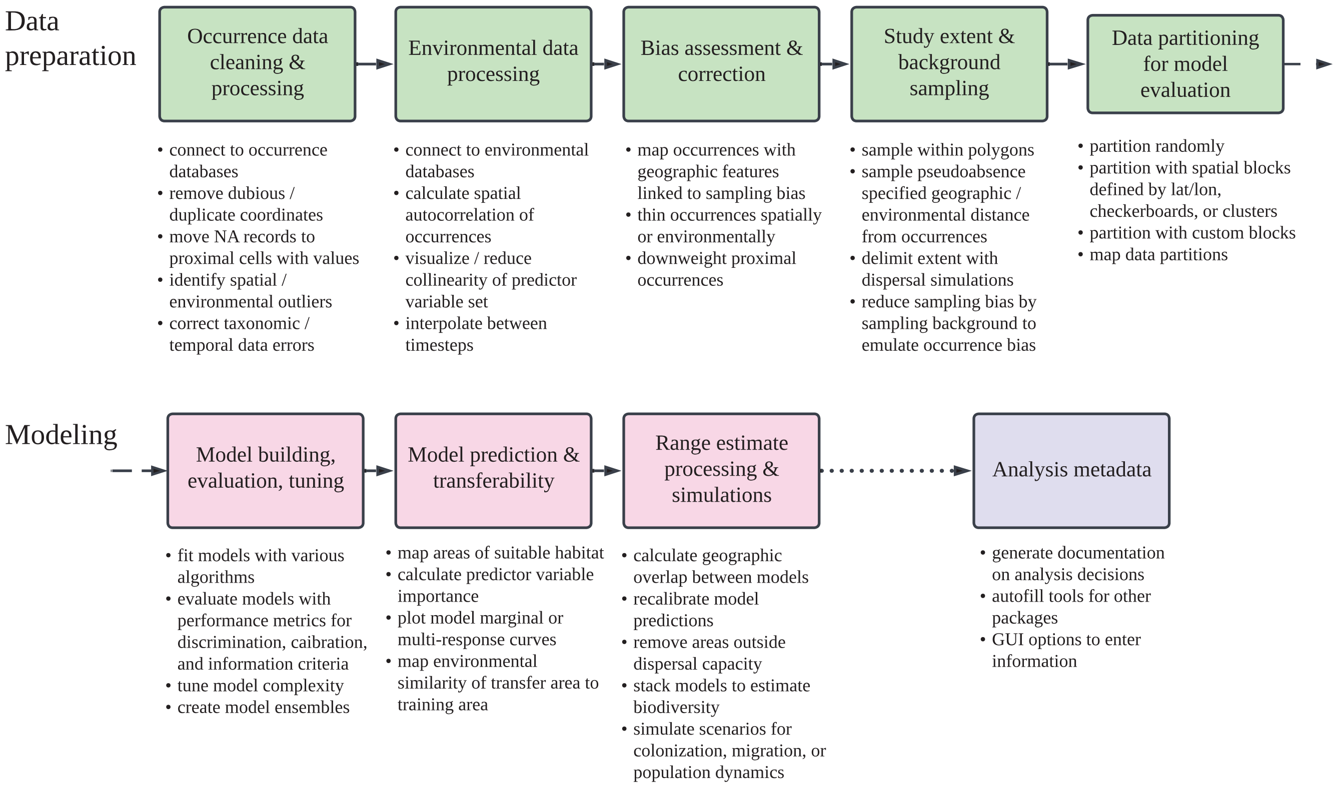

Figure 1. Species distribution model (SDM) workflow as described in this paper, separated into steps within two main phases: "Data preparation" (green) and "Modeling" (pink). Analysis metadata (purple) is collated after the modeling phase is complete. A summary of analyses, visualizations, and processes enabled by R packages is found below each step.

Guisan et al. 2017), software notes (Naimi and Araújo 2016, Golding et al. 2018, Kass et al. 2018, Cobos et al. 2019), and step-by-step guides (Sillero et al. 2021, Peterson et al. 2022). This paper covers a representative variety of tools but is not meant to be a comprehensive survey of R packages for SDMs. We focus on workflows for correlative SDMs for modeling individual species, and thus exclude biophysical and mechanistic models, as well as joint and multispecies models. We also mention several of the many packages that focus on retrieving environmental data or conducting general spatial analyses. Citations and title descriptions for packages noted in the main text appear in Table 1, and URLs for package tutorials can be found in the Supporting information.

Data preparation

Occurrence data cleaning and processing

Occurrence data are widely available via online repositories hosting biodiversity records from museums, herbaria, research studies, and citizen science initiatives. However, most require curation before they are fit for use in SDMs (Anderson et al. 2020). Several R packages connect to these databases through APIs for downloading occurrence data (e.g. 'rgbif' [Chamberlain et al. 2023], 'BIEN' [Maitner et al. 2018], or 'robis' [Provoost and Bosch 2022]).

Although large volumes of data are available, they are often prone to sampling bias and geographical, temporal, and taxonomical errors (Marcer et al. 2022) that can result in spurious estimates of species' niches and distributions (Zizka et al. 2020). Several packages are tailored for cleaning and classifying species' occurrence records. 'CoordinateCleaner' flags potentially erroneous records by checking proximity to well-known geographic features (e.g., centroids for administrative regions and institutions), identifying spatial outliers, and finding coordinates with suspicious values or formatting problems. Relatedly, 'bRacatus' provides a continuous index of the validity and biogeographical status of occurrences based on range maps or checklists. Other packages have functions to move occurrence records to the closest predictor variable grid cell when coordinate uncertainty or georeferencing errors place them in 'impossible' locations (e.g., a terrestrial organism in the ocean; 'ellipsenm', 'kuenm'), and for removing records with missing or duplicated values ('ENMTools', 'fuzzySim', 'modleR', 'sdm', 'SDMtune'). Some have specialized filtering tools to identify environmental outliers ('flexsdm'), estimate influence of positional uncertainty on model output ('usdm'), and flag potential taxonomic or temporal data errors ('bdc').

Environmental data processing

Gridded data representing environmental predictor variables such as interpolated climate reconstructions, land use, land cover, topography, soil, geology, marine bathymetry, flow/currents, and salinity are available from a wide variety of open databases, and packages are available to access many of these

Table 1. A collection of R packages for species distribution modeling mentioned in the main text with their citations and title descriptions. Package titles were acquired in November 2023 from either CRAN or Github. Some titles were reworded slightly for clarity. When more than one reference existed for a package (e.g. two or more papers), the most recent one is listed. URLs for package tutorials are in the Supporting information.

Package	Citation	Package title
bamm	Soberón and Osorio-Olvera 2023	Species distribution models as a function of biotic, abiotic and movement factors (BAM)
bdc	Ribeiro et al. 2022	Biodiversity data cleaning
biomod2	Thuiller et al. 2009	Ensemble platform for species distribution modeling
blockCV	Valavi et al. 2019	Spatial and environmental blocking for K-fold and LOO cross-validation
bRacatus	Arlé et al. 2021	A method to estimate the accuracy and biogeographical status of georeferenced biological data
CAST	Meyer et al. 2023	'caret' applications for spatial-temporal models
CoordinateCleaner	Zizka et al. 2019	Automated cleaning of occurrence records from biological collections
dismo	Hijmans et al. 2023c	Species distribution modeling
dsmextra	Bouchet et al. 2020	A toolkit for extrapolation assessments in density surface models
ecospat	Di Cola et al. 2017	Spatial ecology miscellaneous methods
ellipsenm	Cobos et al. 2023	Ecological niche characterizations using ellipsoids
embarcadero	Carlson 2020	Species distribution models with BART
ENMeval	Kass et al. 2021	Automated tuning and evaluations of ecological niche models
enmSdmX	Smith et al. 2023	Species distribution modeling and ecological niche modeling
ENMTML	Andrade et al. 2020	Create and process ecological niches, including several pre- and post-processing methods
ENMTools	Warren et al. 2021	Analysis of niche evolution using niche and distribution models
flexsdm	Velazco et al. 2022	Tools for data preparation, fitting, prediction, evaluation, and post-processing of species distribution models
fuzzySim	Barbosa 2015	Fuzzy similarity in species distributions
grinnell	Machado-Stredel et al. 2021	Dispersal simulations based on ecological niches
KISSMig	Nobis and Normand 2014	A keep it simple species migration model
kuenm	Cobos et al. 2019	Detailed development of ecological niche models using Maxent
megaSDM	Shipley et al. 2022	Integrating dispersal and time-step analyses into species distribution models
MigClim	Engler et al. 2012	Implementing dispersal into species distribution models
MinBAR	Rotllan-Puig and Traveset 2021	Determining the minimal background area for species distribution models
modEvA	Barbosa et al. 2013	Model evaluation and analysis
modleR	Sánchez-Tapia et al. 2020	A workflow for ecological niche models
ntbox	Osorio-Olvera et al. 2020	From getting biodiversity data to evaluating species distribution models in a friendly GUI environment
occCite	Owens et al. 2021	Querying and managing large biodiversity occurrence datasets
predicts	Hijmans 2023a	Spatial prediction tools
rangeModelMetadata	Merow et al. 2019	Provides templates for metadata files associated with species range models
RangeShiftR	Malchow et al. 2021	An R package for individual-based simulation of spatial eco-evolutionary dynamics and species' responses to environmental changes
sampbias	Zizka et al. 2021	Evaluating geographic sampling bias in biological collections
sdm	Naimi and Araújo 2016	Species distribution modeling
SDMtune	Vignali et al. 2020	Species distribution model selection
ShinyBIOMOD	Ondo 2023	A 'shiny' interface to 'biomod2' (summary of package description)
spThin	Aiello-Lammens et al. 2015	Functions for spatial thinning of species occurrence records for use in ecological models
SSDM	Schmitt et al. 2017	Stacked species distribution modeling
terra	Hijmans 2023b	Spatial data analysis
usdm	Naimi et al. 2014	Uncertainty analysis for species distribution models
voluModel	Owens and Rahbek 2023	Modeling species distributions in three dimensions
wallace	Kass et al. 2023	A modular platform for reproducible modeling of species niches and distributions

in R (e.g., 'geodata' [Hijmans et al. 2023d] or 'sdmpredictors' [Bosch and Fernandez 2023]). The 'terra' package and its predecessor 'raster' provide many tools for working with vector and raster data in R. Some packages automate raster processing for SDM analyses (e.g., 'SSDM' can crop a set of rasters to a common extent and resample them to the coarsest resolution). Other useful functions for manipulating predictor variables can interpolate between timesteps ('enmSdmX') and investigate spatial autocorrelation with respect to the occurrence data ('ecospat').

Environmental variables tend to have high collinearity (Jiménez-Valverde et al. 2009), which can lead to high variance in model coefficients (i.e., sensitivity to small changes in input data) and challenges in interpreting variable importance (Smith and Santos 2020). Several approaches can be taken to address correlated predictor variables: making diagnostic plots ('ENMTools'), removing redundant variables (pairwise: 'ENMTML', 'flexsdm', 'modleR', 'ntbox'; sequential: 'fuzzySim', 'SDMtune', 'usdm'), or reducing variable dimensionality through ordination ('ENMTML', 'ENMTools', 'flexsdm', 'kuenm', 'ntbox'). As selecting appropriate predictor variables is complex and may require multiple steps, these tools can help users make better informed decisions.

Bias assessment and correction

Since occurrence data are typically compiled from studies or collection efforts without robust sampling designs, they often reflect spatial sampling bias (Erickson and Smith 2021). Sampling bias can arise for many reasons, such as imperfect detection or site accessibility. Ignoring these biases can result in models that have high predictive accuracy against the same (biased) data yet misrepresent species' true distributions and environmental tolerances.

As an initial assessment, the geographical distributions of occurrence data and features that may cause spatial biases can be explored using visualization tools in the 'sampbias' package. Spatial thinning reduces geographic clustering by removing occurrence records (and optionally pseudo-absence or background records) within some geographical distance of other records either measured directly ('ecospat', 'ellipsenm', 'enmSdmX', 'ENMTML', 'flexsdm', humboldt, 'ntbox', 'spThin') or dictated by a spatial grid ('biomod2', 'dismo', 'ENMTools', 'fuzzySim'). Spatial-grid thinning is much quicker than distance-based methods, so the former can precede the latter for larger datasets. Alternatively, environmental thinning implemented in 'flexsdm' reduces clustering in environmental space to address biases related to sampled environments, or occurrence records can be downweighted by proximity (Stolar and Nielsen 2015) with 'enmSdmX'.

Model training extent and background sampling

When modeling species' distributions from presence-only data, background (or pseudo-absence) samples are typically taken across an area defined as the 'training extent', representing the area theoretically accessible to the species (Barve et al. 2011). Choosing this extent carefully is important, as the model can make spurious associations if the extent includes regions with suitable conditions that have been inaccessible to the species (Soberón 2007).

Random sampling of grid cells to generate background points can be performed easily in R, but there are many other tools for sampling points following more specialized strategies. Random samples within spatial buffers of a specified distance around occurrences, or within polygons describing habitat types, can ensure the model training extent excludes areas too far from observed presences or isolated by dispersal barriers ('biomod2', 'ellipsenm', 'flexsdm', 'ENMTML', 'MinBAR', 'modleR', 'SSDM'). Some packages can also restrict sampling to areas within a specified environmental distance ('biomod2', 'ENMTML', 'flexsdm'). Model training extents can also be delimited based on dispersal simulations to better represent areas accessible to the species ('grinnell'), or alternatively on the extent that optimizes model fit ('MinBAR'). 'voluModel' can be used to generate training extents using alpha hulls for 3-dimensional systems (e.g., marine environments) and sample both horizontally across the study region and through depth-structured environmental data.

Data partitioning for model evaluation

In order to evaluate models, we must first decide which data are used to assess model accuracy. As independent testing data (i.e., data separate from those used for model training) are rare, SDM accuracy is often assessed with cross-validation. This involves partitioning the occurrence records into *k* folds (i.e., subsets), then sequentially building models using all folds but one and evaluating them on the withheld fold. Alternatively, *n* randomly selected occurrence records can be sequentially withheld for validation in a process called either leave-n-out or delete-d jackknife ('ENMeval', 'flexsdm', 'modleR', 'sdm', 'SSDM'). Small sample-size challenges robust assessment of model accuracy, and for these cases pooling solutions can be used ('ecospat'; Collart and Guisan 2023).

Random data partitions tend to overestimate model performance when predictions are made to new areas or times. One way to help address this issue is to partition data into 'blocks' according to some structure (e.g., spatial, temporal, environmental; Roberts et al. 2017) that increases statistical independence between training and validation data. Spatial partitions can be implemented in a 'checkerboard' pattern where data (occurrence and/or background records) are grouped by an array of squares ('blockCV', 'ENMeval', 'flexsdm'), along latitudinal or longitudinal lines ('biomod2', 'blockCV', 'ellipsenm', 'ENMeval', 'flexsdm'), or with spatial clustering algorithms ('biomod2', 'blockCV', 'enmSdmX'). Users can alternatively define their own partitions ('biomod2', 'blockCV', 'ENMeval') based on other blocking structures. Lastly, visualizations of blocked data partitions on maps can aid interpretation of evaluation results ('blockCV', 'ENMeval').

Modeling

Model building, evaluation and tuning

There are many decisions to make when modeling species' niches and distributions, including choice of algorithm(s), model performance metrics, and methods for tuning model complexity. Several packages we highlight have model-building functions for one algorithm, but many are structured to build models with multiple algorithms. The packages 'biomod2', 'ENMeval', 'flexsdm', 'sdm', and 'SSDM' additionally have custom objects for user-specified algorithms and settings, which allow other package functions to use them in subsequent steps.

There is a wide variety of accuracy metrics used to evaluate SDMs (Guisan et al. 2017), but two main categories are discrimination (how well the model distinguishes between occupied sites and background) and calibration (how well model predictions correlate with the observed proportions of occupied sites). Multiple discrimination metrics can be calculated by most packages that build models, and options exist to weight sites for biases in validation data ('enmSdmX'). Calibration metrics include Miller's slope ('modEvA'; Pearce and Ferrier 2000), the continuous Boyce index ('biomod2', 'ecospat', 'ENMeval', 'enmSdmX', 'ENMTools', 'modEvA'; Hirzel et al. 2006), and expected and maximum calibration error (ECE and MCE; 'ENMTools'). Information criteria (typically AIC or BIC) are based on model likelihood values of the full dataset (not cross-validation) and include a penalty on the number of predictors to downweight overfit models (Warren and Seifert 2011) — some packages use these criteria as an option for model tuning, explained below ('biomod2', 'ENMeval', 'ENMTools', 'enmSdmX', 'kuenm', 'SDMtune'). The significance of accuracy metrics can also be determined by building null models and comparing empirical values to null distributions (Bohl et al. 2019) with 'dismo', 'ENMeval', or 'ENMTools'.

Many of the packages we assess here that build and evaluate models contain functions for model tuning, or evaluating ranges of hyperparameters to achieve optimal model complexity for the training data (Merow et al. 2014). This is implemented in various ways in multiple packages ('biomod2', 'ENMeval', 'enmSdmX', 'ENMTools', 'flexsdm', 'kuenm', 'SDMtune'). 'SDMtune' additionally provides graphical tools to assess model performance using different tuning approaches: random, grid-based, or with a genetic algorithm. Alternatively, model ensembles can combine multiple algorithms to produce a single consensus prediction, which can be weighted by model performance, for example. Ensembles can be used to assess variance across algorithms (Thuiller et al. 2019) by quantifying and mapping algorithm uncertainty, and ensembling many bivariate models can increase performance when occurrence data are scarce ('ecospat', 'flexsdm'; Breiner et al. 2015). Extensive options for ensemble modeling are available in 'biomod'2 and 'flexsdm', but other packages also include ensemble options ('ecospat', 'ENMTML', 'fuzzySim', 'modleR', 'sdm', 'SSDM').

Model prediction and transferability

Visualizing aspects of model prediction behavior through diagnostic plots can highlight issues that may not be readily discernible with evaluation metrics. Biological realism of correlative model predictions can be evaluated with diagnostic plots of model responses along environmental gradients ('biomod2', 'dismo', 'ENMTools', 'predicts', or 'SDMtune') — this is especially important for visualizing how models transfer to non-analog conditions (Guevara et al. 2018). Other tools expand plotting flexibility and have other visual features, such as two- or three-dimensional response plots ('biomod2', 'embarcadero', 'ENMTools', 'flexsdm'), which can help visualize interactions between variables. The 'sdm' and 'SSDM' packages also provide interactive plots to explore the outcomes of fitted models (e.g., metrics of performance and plots, response curves, and variable importance). Related to these visual tools, assessments of variable importance can inform selection of variables for modeling and inference into drivers of species' niches and distributions ('biomod2', 'ENMTools', 'predicts', 'sdm', 'SDMtune', 'SSDM'; Smith and Santos 2020).

When models are transferred to other times and/or places, exploration of environmental similarity maps can highlight areas where predictions can be unreliable due to extrapolation. The multivariate environmental similarity surface (MESS; Elith et al. 2010) approach quantifies the 'novelty' of transfer conditions compared to reference conditions (e.g. the model training extent; 'dismo', 'ENMeval', 'modEvA', 'predicts', 'voluModel'), but this metric can be heavily influenced by individual variables with extreme values. Other variations use different definitions of novelty and thus measure the degree of extrapolation differently. Mobility-oriented parity (MOP) allows comparisons of reference and transfer conditions that are restricted to environmental ranges closer to those of the training extent ('ENMTML', 'kuenm'; Owens et al. 2013). Extrapolation detection (ExDet) discriminates between extrapolation in univariate ranges and novel covariate combinations ('dsmextra', 'ecospat'; Mesgaran et al. 2014). Area of applicability (AOA) weights the multivariate predictor space by variable importance ('CAST'; Meyer and Pebesma 2021). The Shape algorithm calculates the multivariate distance relative to the environmental dispersion of the training data and has an adjustable threshold to control binary discrimination between acceptable and unacceptable degrees of extrapolation ('flexsdm'; Velazco et al. 2024).

Analyses and simulations using distribution estimates as inputs

Geographic predictions of potential distribution can be used as inputs for analyses to refine range estimates or to simulate population dynamics or dispersal. For example, model recalibration, or rescaling continuous suitability predictions to better fit the observed frequency of occurrences given the environment, is available in 'ENMTools' via functions that use the 'CalibratR' package (Schwarz and Heider 2019). Model predictions outside a species' dispersal capability can be interpreted as potential distribution, but if current species' range delineations are the research objective, methods that use spatial variables as model predictors or as geographical constraints in post-processing can remove inaccessible areas ('ENMTML', 'flexsdm', 'MSDM'; Mendes et al 2020). When projecting SDMs to past or future climate scenarios, spatial agreement between different scenarios can be calculated ('biomod2', 'kuenm'), as well as the speed and direction a species tracking suitable environmental conditions must move in response to climate change (i.e., biotic velocity; 'enmSdmX'). Model predictions for multiple species can be stacked in different ways to make estimates of biodiversity ('fuzzySim', 'SSDM'), and the resulting community composition estimates per grid cell can be corrected with biotic filters to avoid overprediction of richness (i.e., SESAM approach in 'ecospat' and 'SSDM'; D'Amen et al. 2015). Community properties (e.g., species richness, composition) of stacked SDMs can be evaluated without thresholding by comparing them directly to observed occurrence patterns ('ecospat', 'SSDM'; Scherrer et al. 2020). Relatedly, geographic overlap between species can be calculated to delineate areas of potential sympatry ('ENMeval', 'ENMTools', 'fuzzySim', 'SSDM').

An increasing number of packages enable dispersal simulations using SDM predictions as proxies for either dispersal potential or carrying capacity (Zurell et al. 2016). If detailed data are available on species' demography and dispersal, then population models (e.g., 'RangeShiftR') can be used to simulate complex population dynamics based on SDM predictions. When such data are scarce, models with fewer assumptions can conduct simple, stochastic simulations of colonization and extinction ('KISSMig'), integrate biotic and movement constraints (the BAM approach *sensu* Soberón and Peterson 2005) via cellular automata simulations ('bamm', 'MigClim'), or weight predictions by dispersal distance ('megaSDM').

Options for package integration and accessibility with graphical user interfaces

Graphical user interfaces (GUIs) composed with the R package 'shiny' (Chang et al. 2023) provide users an alternative to writing code, thus increasing accessibility to the tools we have discussed and providing real-time visualization. The package 'ntbox' implements a model-selection workflow for ellipsoid models and features tools for assessing extrapolation risk, transforming environmental variables to principal components, performing geospatial operations, and making 3D visualizations of ellipsoid niche estimates in environmental space. The packages 'ShinyBIOMOD' and 'SSDM' offer graphical interfaces to run a range of modeling tools for building ensemble models, comparing their performance, and estimating uncertainties in their predictions. The package 'wallace' runs the Wallace EcoMod application, which steps through a full SDM workflow and includes downloading and processing occurrence data and environmental rasters, model tuning, and transferring predictions to new locations and times (Kass et al. 2023). It also offers guidance text and references, metadata generation, and tools to download code that reproduces the analysis. These 'shiny' applications present solutions to integrate multiple packages and mobilize their tools within a single interactive workflow. For example, 'ShinyBIOMOD' includes other R packages to address problems like multicollinearity among environmental predictors and spatial sampling biases in occurrence data. Wallace EcoMod 2.0 features a streamlined system for integrating user-authored 'modules' that expand functionality, and the application showcases modules from different research groups (Kass et al. 2023).

Analysis metadata

Although code and data are increasingly made available with published manuscripts, recording metadata about the modeling process is crucial because it facilitates reproducibility by documenting data sources and specifications, model parameterizations, and other details ('ENMeval', 'ENMTML', 'modleR', 'sdm'). To this end, the package 'occCite' returns documentation on original occurrence data providers served through data repositories such as GBIF, which can be time-consuming to locate manually (Owens et al. 2021). Recently proposed metadata standards for SDMs (Merow et al. 2019, Zurell et al. 2020) help researchers follow best practices and guidelines for the field (Araújo et al. 2019) — these metadata can be documented via code-based and interactive tools. Range model metadata standards (RMMS) provides a metadata dictionary that can be populated using automated tools ('rangeModelMetadata'; Merow et al. 2019). RMMS objects are generated by 'wallace', and other packages can include wrapper functions to update them ('ENMeval'). The overview, data, model, assessment and prediction (ODMAP) framework (Zurell et al. 2020) integrates RMMS; these metadata can be completed using a 'shiny' web application (https://odmap.wsl.ch/).

Use of multiple packages enhances breadth and robustness of analysis

Integrating tools from different packages helps researchers expand workflows to address more key methodological steps of SDM analyses. We developed the community-driven package 'sdmverse' to help researchers understand how different packages are related and what unique attributes they have. Package maintainers can upload their package's metadata via the GitHub interface and contribute to 'sdmverse' after an open, web-based review process. The 'sdmverse' package allows users to explore package metadata dynamically via a 'shiny' application, or to visualize their relationships graphically with tile plots or dendrograms. We used plotting tools in 'sdmverse' to categorize and cluster SDM packages into methodological groups based on their functions in order to map their relationships with one another, and we found they clustered into five easily interpretable groups (Fig. 2). Packages focused on modeling analyses represent the largest group, including those more specialized (e.g., for evaluation or model tuning) and those more general (e.g., with tools for more workflow steps, including GUIs). The smaller groups focus on post-processing of SDM results, data processing, data partitioning, and analysis metadata.

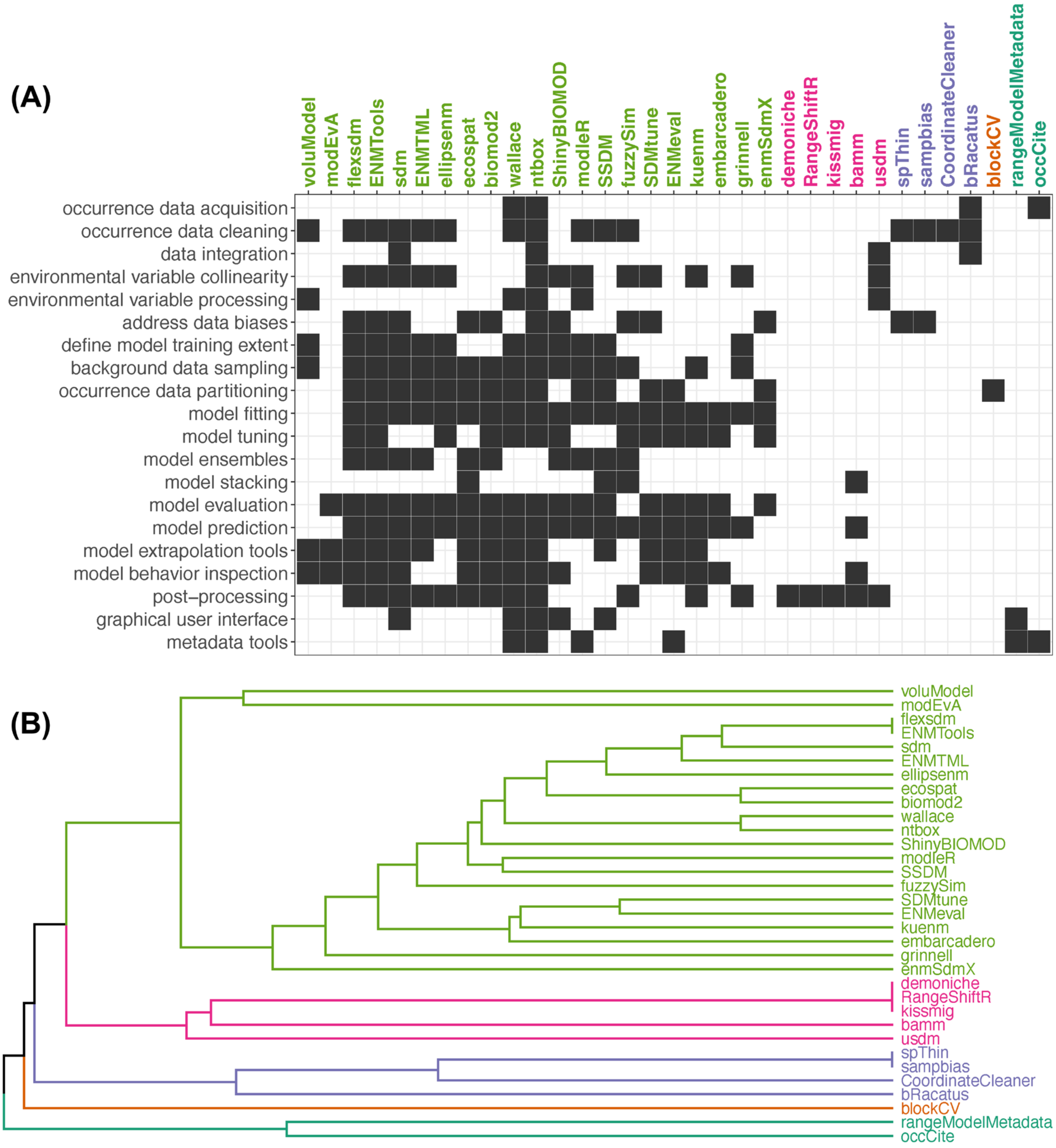

Figure 2 Visualizations of characteristics for R packages that focus on species distribution models (SDMs) and the relationships between them, made with the 'sdmverse' package (downloaded August 2023). A) Tile plot of R packages and the methodological categories for SDMs discussed in this manuscript that their tools address. B) Dendrogram showing hierarchical clusters for these packages. Colors correspond to five binary (Jaccard) clusters displayed in B) based on package characteristics in A). Cluster number was chosen to best represent broad groupings of package types: from bottom, metadata (teal), data partitioning (brown), data processing (purple), post-processing (pink), and modeling analyses (green). The largest group, modeling analyses, includes packages with more comprehensive toolsets (a subset being GUIs) and those more focused on model evaluation and tuning. Future work should focus on the production of more packages and/or supplemental functions for smaller clusters to expand available tools.

To showcase the utility of pluralistic use of R packages for SDMs, we provide three comprehensive analysis vignettes representing frequently researched topics in the field (code: https://doi.org/10.5281/zenodo.14044044, pdf: https://doi.org/10.6084/m9.figshare.27312903). The first vignette predicts future distributional shifts due to climate change for the moose *Alces alces* in North America. The second models native and potentially invasive global distributions of the small Indian mongoose *Urva auropunctata*. The third estimates the distribution of a poorly known plant species with few known occurrence records, *Asclepias scaposa*, and calculates conservation range metrics based on potential distribution maps.

To assess how use of pluralistic workflows helps achieve SDM standards, we scored each analysis vignette based on the standards from Araújo et al. (2019), who grade achievements into four classes: gold ('aspirational'), silver ('cutting-edge approaches' that involve 'imperfect (but best available) data' and address uncertainty and bias), bronze ('minimum currently acceptable practices'), and deficient. For each methodological standard category, we summarize which packages contributed to achieving the score and the reasoning behind these decisions, and we display the results graphically (Supporting information). These vignettes were created solely for demonstration and are not meant to represent comprehensive analyses, so some categories we did not directly address were rated as 'deficient'. However, in spite of this, all vignettes were able to achieve at least bronze (and often silver) for most categories due to the use of multiple packages.

Future directions

The tremendous gains in SDM software achieved over the past two decades have opened new avenues for improvement and innovation. Below, we discuss some outstanding challenges for existing software and suggest future directions for tool development.

Methodological uncertainty

As the diversity of software presents many choices that are equally defensible, exploration, quantification, and mapping of methodological uncertainty remain key, cross-cutting needs (Jansen et al. 2022). Among the categories we used to assess package functionality (Fig. 2), the median number of packages covering any one task was 14.5 (min: 5, max: 21). As each package implements these tasks in different ways, users with the same overall goal (e.g., modeling a particular invasive species) can produce different outputs. We covered tools that quantify uncertainty from georeferencing of occurrence data, algorithms (with ensemble models), and model transfer (with environmental similarity metrics), but we lack tools to help us jointly consider different sources of SDM uncertainty. In general, more packages dedicated to methodological uncertainty for SDMs are needed, ideally those that can combine different kinds of uncertainty to produce maps or other visualizations of aggregated model confidence.

Data curation

Even with existing tools, data cleaning often necessitates a considerable amount of manual checking, as automated solutions based on heuristic rulesets can be too conservative or liberal in their filtering. Further, species records lacking coordinate information (or those with unacceptable spatial uncertainty) cannot be used in modeling unless geocoding based on ancillary locality information is employed, which can be a labor-intensive process (Kass et al. 2022). With the advent of AI-based large language models that can perform georeferencing and geocoding tasks (Gougherty and Clipp 2024), we foresee great potential for the development of automated tools for the curation and generation of species occurrence data. For example, such models could also help flag potentially erroneous records based on general queries (e.g., 'Show me specimens that may be cultivated.').

Bias correction

Approaches to bias correction would benefit from translating cutting-edge methods into more accessible software solutions. For example, data integration approaches that combine systematically sampled, 'structured' data (which are expensive to acquire, so limited in extent) with widely available opportunistic, 'unstructured' data can reduce sampling biases (Isaac et al. 2020). Likewise, biases arise from decisions of individual collectors and observers, so accounting for collector identity can correct for bias (Erickson and Smith 2021), but this is hindered by the challenge of parsing the different ways collectors' names can be reported in biodiversity databases.

Model evaluation and transfer

Current solutions for SDM validation typically employ cross-validation averages, which may be insufficient for understanding final model performance. Especially regarding block cross-validation for model transfer applications, we may be more interested in the withheld validation group that performs most poorly and what these particular conditions are. Interactive tools to examine cross-validation results in more depth with exploratory visualizations would help us better assess confidence in model transfers. Further, model performance against present-day data may not correspond to performance in different times or places (Santini et al. 2021). Software solutions that enlist different data types reflecting different aspects of distributions (e.g., genetic, dendroecological, fossil) could provide alternative assessments of model accuracy and thus more confidence in model predictions (Fordham et al 2014).

Post-modeling steps

Most existing simulation software that uses SDM predictions emphasizes dispersal constraints on potential distributions, but biotic interactions and adaptive evolution also constrain species' responses to environmental change. Likewise, species' phenotypes and genotypes vary through time, and can interact to influence distributions (Bocedi et al. 2022). Software for incorporating these processes are needed.

Package redundancy and maintenance

With the incipient recognition of software products as key research outputs, package development increasingly accompanies research, and this can result in functional redundancy. While ample opportunities exist for creation of new software, we recommend first assessing if there are existing solutions or if they can be improved via collaborative platforms like GitHub or with the 'sdmverse' package. As package developers, we acknowledge our own efforts could be improved in this light, but we also appreciate that some replication can avoid dependencies on other packages that may not be maintained in the future. After all, varying funding opportunities and competing priorities mean that continuous maintenance is not always guaranteed. To enable better integration across packages to reduce redundancy, an obvious need is for tools that enable coercion of package-specific R objects to other types.

Conclusions

The burgeoning array of software packages developed for SDMs can be daunting, but integrating different tools can make analyses more robust by addressing data errors, biases, and uncertainty in various ways; more interpretable by employing different visualization tools; and better at achieving methodological standards. In our collective experience, we are often asked, 'Which package should I use for my SDM analysis?' Here, we show that relying on any one package considerably limits the breadth of analyses available, and that integrating multiple tools in a single script is not only tractable but recommended. Above all, we encourage developers who adhere to their own software to broaden their toolboxes and seek synergies in analysis. A commitment to plurality can help us bound methodological uncertainty and ultimately answer pivotal questions in ecology and evolution.

To cite 'sdmverse' or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for 'version 1.0':

Kass, J. M. et al. 2024. Achieving higher standards in species distribution modeling by leveraging the diversity of available software. – Ecography 2024: e07346 (ver. 1.0).

Acknowledgements – We want to thank all the developers of R packages for species distribution models for helping to create such a vibrant and diverse toolbox for the community that continues to advance the field.
Funding – JMK was supported in part by the Japan Society for the Promotion of Science Postdoctoral Fellowships for Foreign Researchers Program and JSPS Grants-In-Aid for Scientific Research (KAKENHI) (20F20774). ABS was supported by the Alan Graham Fund in Global Change. MEA-L and GEP-B acknowledge funding from NSF DBI 1661510 and NASA 80NSSC18K0406. EA acknowledges funding by the Azrieli Foundation's International Postdoctoral Fellowship, in association with the Belmaker Lab – Marine Ecology and Biodiversity. CM was supported by grants from NSF BoCP 2225078 and NASA 80NSSC22K0883. LO-O acknowledges partial support from PAPIIT project IA202824 and CONACyT-Ciencia de Frontera CF-2023-I-115. HLO was supported under EU Horizon 2020 Marie Skłodowska-Curie grant agreement no. 891702 and research grant no. 25925 from Villum Fonden. DZ was supported by German Research Foundation (DFG) grant no. ZU 361/6-1.

Author contributions

Jamie M. Kass: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Resources (equal); Software (equal); Supervision (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead). **Adam B. Smith**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Resources (equal); Software (equal); Supervision (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Dan L. Warren**: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Resources (equal); Software (equal); Supervision (supporting); Visualization (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Sergio Vignali**: Conceptualization (supporting); Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Writing – original draft (supporting); Writing – review and editing (supporting). **Sylvain Schmitt**: Conceptualization (supporting); Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Writing – original draft (supporting); Writing – review and editing (supporting). **Matthew E. Aiello-Lammens**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Eduardo Arlé**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Ana Márcia Barbosa**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Olivier Broennimann**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Marlon E. Cobos**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Maya Guéguen**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Antoine Guisan**: Methodology (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Cory Merow**: Methodology (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Babak Naimi**: Methodology (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Michael P. Nobis:** Methodology (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Ian Ondo**: Methodology (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Luis Osorio-Olvera**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Hannah L. Owens**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Gonzalo E. Pinilla-Buitrago**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Andrea Sánchez-Tapia**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Wilfried Thuiller**: Methodology (supporting);

Writing – original draft (supporting); Writing – review and editing (supporting). **Roozbeh Valavi**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Santiago José Elías Velazco**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Alexander Zizka**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Damaris Zurell**: Methodology (supporting); Resources (supporting); Writing – original draft (supporting); Writing – review and editing (supporting).

Transparent peer review

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ecog.07346.

Data availability statement

The 'sdmverse' package is under continuing development on the public GitHub repository https://github.com/sylvainschmitt/sdmverse, and the version described in the manuscript is archived on Zenodo (https://doi.org/10.5281/zenodo.13927581).

The vignettes described in the manuscript are available as .pdf files on Figshare (https://doi.org/10.6084/m9.figshare.27312903), and the code and data to build and run the vignettes and figures from this paper are available on the public Github repository https://github.com/jamiemkass/sdm-pkg-plurality-tutorials and archived on Zenodo: https://doi.org/10.5281/zenodo.14044044 (Kass et al. 2024).

Supporting information

The Supporting information associated with this article is available with the online version.

References

Ahmed, S. E., McInerny, G., O'Hara, K., Harper, R., Salido, L., Emmott, S. and Joppa, L. N. 2015. Scientists and software – surveying the species distribution modelling community. – Divers. Distrib. 21: 258–267.

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B. and Anderson, R. P. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. – Ecography 38: 541–545.

Anderson, R. P., Araújo, M. B., Guisan, A., Lobo, J. M., Martínez-Meyer, E., Peterson, A. T. and Soberón, J. M. 2020. Optimizing biodiversity informatics to improve information flow, data quality, and utility for science and society. – Front. Biogeogr. 12: e47839.

Andrade, A. F. A. de, Velazco, S. J. E. and De Marco Júnior, P. 2020. ENMTML: an R package for a straightforward construction of complex ecological niche models. – Environ. Modell. Softw. 125: 104615.

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., Zimmermann, N. E. and Rahbek, C. 2019. Standards for distribution models in biodiversity assessments. – Sci. Adv. 5: eaat4858.

Arlé, E., Zizka, A., Keil, P., Winter, M., Essl, F., Knight, T., Weigelt, P., Jiménez-Muñoz, M. and Meyer, C. 2021. bRacatus: a method to estimate the accuracy and biogeographical status of georeferenced biological data. – Methods Ecol. Evol. 12: 1609–1619.

Barbosa, A. M. 2015. fuzzySim: applying fuzzy logic to binary similarity indices in ecology. – Methods Ecol. Evol. 6: 853–858.

Barbosa, A. M., Real, R., Muñoz, A.-R. and Brown, J. A. 2013. New measures for assessing model equilibrium and prediction mismatch in species distribution models. – Divers. Distrib. 19: 1333–1338.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J. and Villalobos, F. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. – Ecol. Modell. 222: 1810–1819.

Bocedi, G., Palmer, S. C. F., Malchow, A.-K., Zurell, D., Watts, K. and Travis, J. M. J. 2022. RangeShifter 2.0: an extended and enhanced platform for modeling special eco-evolutionary dynamics and species' responses to environmental change. – Ecography 44: 1453–1462.

Bohl, C., Kass, J. M. and Anderson, R. P. 2019. A new null model approach to quantify performance and significance for ecological niche models of species distributions. – J. Biogeogr. 46: 1101–1111.

Bosch, S. and Fernandez, S. 2023. sdmpredictors: species distribution modelling predictor datasets. – https://CRAN.R-project.org/package=sdmpredictors.

Bouchet, P. J., Miller, D. L., Roberts, J. J., Mannocci, L., Harris, C. M. and Thomas, L. 2020. dsmextra: extrapolation assessment tools for density surface models. – Methods Ecol. Evol. 11: 1464–1469.

Breiner, F. T., Guisan, A., Bergamini, A. and Nobis, M. P. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. – Methods Ecol. Evol. 6: 1210–1218.

Carlson, C. J. 2020. embarcadero: species distribution modelling with Bayesian additive regression trees in r. – Methods Ecol. Evol. 11: 850–858.

Chamberlain, S., Barve, V., Mcglinn, D., Oldoni, D., Desmet, P., Geffert, L. and Ram, K. 2023. rgbif: interface to the global biodiversity information facility API. – https://CRAN.R-project.org/package=rgbif.

Chang, W. et al. 2023. shiny: web application framework for R. – https://CRAN.R-project.org/package=shiny.

Cobos, M. E., Peterson, A. T., Barve, N. and Osorio-Olvera, L. 2019. kuenm: an R package for detailed development of ecological niche models using Maxent. – PeerJ 7: e6281.

Cobos, M. E., Osorio-Olvera, L., Soberón, J. and Peterson, A. T. 2023. Ellipsenm: ecological Niche's characterizations using ellipsoids. – https://github.com/marlonecobos/ellipsenm.

Collart, F. and Guisan, A. 2023. Small to train, small to test: dealing with low sample size in model evaluation. – Ecol. Inform. 75: 102106.

D'Amen, M., Dubuis, A., Fernandes, R. F., Pottier, J., Pellissier, L. and Guisan, A. 2015. Using species richness and functional traits predictions to constrain assemblage predictions from stacked species distribution models. – J. Biogeogr. 42: 1255–1266.

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Pellissier, L., Mateo, R. G., Hordijk, W., Salamin, N. and Guisan, A. 2017. ecospat: an R package to support spatial analyses and modeling of species niches and distributions. – Ecography 40: 774–787.

Elith, J., Kearney, M. and Phillips, S. 2010. The art of modelling range-shifting species. – Methods Ecol. Evol. 1: 330–342.

Engler, R., Hordijk, W., Guisan, A., Engler, R., Hordijk, W. and Guisan, A. 2012. The MIGCLIM R package – seamless integration of dispersal constraints into projections of species distribution models. – Ecography 35: 872–878.

Erickson, K. D. and Smith, A. B. 2021. Accounting for imperfect detection in data from museums and herbaria when modeling species distributions: combining and contrasting data-level versus model-level bias correction. – Ecography 44: 1341–1352.

Farley, S. S., Dawson, A., Goring, S. J. and Williams, J. W. 2018. Situating ecology as a big-data science: current advances, challenges, and solutions. – BioScience 68: 563–576.

Fordham, D. A., Brook, B. W., Moritz, C. and Nogués-Bravo, D. 2014. Better forecasts of range dynamics using genetic data. – Trends Ecol. Evol. 29: 436–443.

Franklin, J. 2010. Mapping species distributions: spatial inference and prediction. – Cambridge Univ. Press.

Golding, N., August, T. A., Lucas, T. C. D., Gavaghan, D. J., van Loon, E. E. and McInerny, G. 2018. The zoon R package for reproducible and shareable species distribution modelling. – Methods Ecol. Evol. 9: 260–268.

Gougherty, A. V. and Clipp, H. L. 2024. Testing the reliability of an AI-based large language model to extract ecological information from the scientific literature. – NPJ Biodiversity 3: 13.

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

Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – Ecol. Modell. 135: 147–186.

Guisan, A., Thuiller, W. and Zimmermann, N. E. 2017. Habitat suitability and distribution models: with applications in R. – Cambridge Univ. Press.

Hijmans, R. 2023a. predicts: spatial prediction tools. – https://CRAN.R-project.org/package=predicts.

Hijmans, R. J. 2023b. terra: spatial data analysis. – https://CRAN.R-project.org/package=terra.

Hijmans, R. J., Phillips, S., Leathwick, J. and Elith, J. 2023c. Dismo: species distribution modeling. – https://CRAN.R-project.org/package=dismo.

Hijmans, R. J., Barbosa, M., Ghosh, A. and Mandel, A. 2023d. geodata: download geographic data. – https://CRAN.R-project.org/package=geodata.

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C. and Guisan, A. 2006. Evaluating the ability of habitat suitability models to predict species presences. – Ecol. Modell. 199: 142–152.

Isaac, N. J. B., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., Henrys, P. A., Jarvis, S., Lahoz-Monfort, J., Pagel, J., Pescott, O. L., Schmucki, R., Simmonds, E. G. and O'Hara, R. B. 2020. Data integration for large-scale models of species distributions. – Trends Ecol. Evol. 35: 56–67.

Jansen, J., Woolley, S. N. C., Dunstan, P. K., Foster, S. D., Hill, N. A., Haward, M. and Johnson, C. R. 2022. Stop ignoring map uncertainty in biodiversity science and conservation policy. – Nat. Ecol. Evol. 6: 828–829.

Jiménez-Valverde, A., Nakazawa, Y., Lira-Noriega, A. and Peterson, A. T. 2009. Environmental correlation structure and ecological niche model projections. – Biodivers. Inform. 6: 28–35.

Kass, J. M., Vilela, B., Aiello-Lammens, M. E., Muscarella, R., Merow, C. and Anderson, R. P. 2018. Wallace: a flexible platform for reproducible modeling of species niches and distributions built for community expansion. – Methods Ecol. Evol. 9: 1151–1156.

Kass, J. M., Muscarella, R., Galante, P. J., Bohl, C. L., Pinilla-Buitrago, G. E., Boria, R. A., Soley-Guardia, M. and Anderson, R. P. 2021. ENMeval 2.0: redesigned for customizable and reproducible modeling of species' niches and distributions. – Methods Ecol. Evol. 12: 1602–1608.

Kass, J. M. et al. 2022. The global distribution of known and undiscovered ant biodiversity. – Sci. Adv. 8: eabp9908.

Kass, J. M., Pinilla-Buitrago, G. E., Paz, A., Johnson, B. A., Grisales-Betancur, V., Meenan, S. I., Attali, D., Broennimann, O., Galante, P. J., Maitner, B. S., Owens, H. L., Varela, S., Aiello-Lammens, M. E., Merow, C., Blair, M. E. and Anderson, R. P. 2023. *wallace* 2: a shiny app for modeling species niches and distributions redesigned to facilitate expansion via module contributions. – Ecography 2023: e06547.

Kass, J. M. et al. 2024. Data from: Achieving higher standards in species distribution modeling by leveraging the diversity of available software. – Zenodo Digital Repository, https://doi.org/10.5281/zenodo.14044044.

Machado-Stredel, F., Cobos, M. E. and Peterson, A. T. 2021. A simulation-based method for selecting calibration areas for ecological niche models and species distribution models. – Front. Biogeogr. 13: e48814.

Maitner, B. S. et al. 2018. The bien R package: a tool to access the botanical information and ecology network (BIEN) database. – Methods Ecol. Evol. 9: 373–379.

Malchow, A. K., Bocedi, G., Palmer, S. C. F., Travis, J. M. J. and Zurell, D. 2021. RangeShiftR: an R package for individual-based simulation of spatial eco-evolutionary dynamics and species' responses to environmental changes. – Ecography 44: 1443–1452.

Marcer, A., Chapman, A. D., Wieczorek, J. R., Xavier Picó, F., Uribe, F., Waller, J. and Ariño, A. H. 2022. Uncertainty matters: ascertaining where specimens in natural history collections come from and its implications for predicting species distributions. – Ecography 2022: e06025.

Mendes, P., Velazco, S. J. E., Andrade, A. F. A. de and De Marco, P. 2020. Dealing with overprediction in species distribution models: how adding distance constraints can improve model accuracy. – Ecol. Modell. 431: 109180.

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

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

Mesgaran, M. B., Cousens, R. D. and Webber, B. L. 2014. Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. – Divers. Distrib. 20: 1147–1159.

Meyer, H. and Pebesma, E. 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. – Methods Ecol. Evol. 12: 1620–1633.

Meyer, H., Milà, C. and Ludwig, M. 2023. CAST: ‘caret’ applications for spatial-temporal models. – https://CRAN.R-project.org/package=CAST.

Naimi, B. and Araújo, M. B. 2016. Sdm: a reproducible and extensible R platform for species distribution modelling. – Ecography 39: 368–375.

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K. and Toxopeus, A. G. 2014. Where is positional uncertainty a problem for species distribution modelling? – Ecography 37: 191–203.

Nobis, M. P. and Normand, S. 2014. KISSMig – a simple model for R to account for limited migration in analyses of species distributions. – Ecography 37: 1282–1287.

Ondo, I. 2023. ShinyBIOMOD: an R application for modelling species distribution. – https://gitlab.com/IanOndo/shinybiomod.

Osorio-Olvera, L., Lira-Noriega, A., Soberón, J., Peterson, A. T., Falconi, M., Contreras-Díaz, R. G., Martínez-Meyer, E., Barve, V. and Barve, N. 2020. ntbox: an R package with graphical user interface for modelling and evaluating multidimensional ecological niches. – Methods Ecol. Evol. 11: 1199–1206.

Owens, H. L. and Rahbek, C. 2023. voluModel: modelling species distributions in three-dimensional space. – Methods Ecol. Evol. 14: 841–847.

Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C. M., Myers, C. E. and Peterson, A. T. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. – Ecol. Modell. 263: 10–18.

Owens, H. L., Merow, C., Maitner, B. S., Kass, J. M., Barve, V. and Guralnick, R. P. 2021. occCite: tools for querying and managing large biodiversity occurrence datasets. – Ecography 44: 1228–1235.

Pearce, J. and Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. – Ecol. Modell. 133: 225–245.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M. and Araújo, M. B. 2011. Ecological niches and geographic distributions (MPB-49). – Princeton Univ. Press.

Peterson, A. T. et al. 2022. ENM2020: a free online course and set of resources on modeling species’ niches and distributions. – Biodivers. Inform. 17: 1–9.

Provoost, P. and Bosch, S. 2022. robis: ocean biodiversity information system (OBIS) client. – https://CRAN.R-project.org/package=robis.

Ribeiro, B. R., Velazco, S. J. E., Guidoni-Martins, K., Tessarolo, G., Jardim, L., Bachman, S. P. and Loyola, R. 2022. bdc: a toolkit for standardizing, integrating and cleaning biodiversity data. – Methods Ecol. Evol. 13: 1421–1428.

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., Wintle, B. A., Hartig, F. and Dormann, C. F. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. – Ecography 40: 913–929.

Rotllan-Puig, X. and Traveset, A. 2021. Determining the minimal background area for species distribution models: MinBAR package. – Ecol. Modell. 439: 109353.

Sánchez-Tapia, A., Mortara, S. R., Rocha, D. S. B., Barros, F. S. M., Gall, G. and Siqueira, M. F. de 2020. modleR: a modular workflow to perform ecological niche modeling in R. – https://www.biorxiv.org/content/10.1101/2020.04.01.021105v1.

Santini, L., Benítez-López, A., Maiorano, L., Čengić, M. and Huijbregts, M. A. J. 2021. Assessing the reliability of species distribution projections in climate change research. – Divers. Distrib. 27: 1035–1050.

Scherrer, D., Mod, H. K. and Guisan, A. 2020. How to evaluate community predictions without thresholding? – Methods Ecol. Evol. 11: 51–63.

Schmitt, S., Pouteau, R., Justeau, D., de Boissieu, F. and Birnbaum, P. 2017. ssdm: an R package to predict distribution of species richness and composition based on stacked species distribution models. – Methods Ecol. Evol. 8: 1795–1803.

Schwarz, J. and Heider, D. 2019. GUESS: projecting machine learning scores to well-calibrated probability estimates for clinical decision-making. – Bioinformatics 35: 2458–2465.

Shipley, B. R., Bach, R., Do, Y., Strathearn, H., McGuire, J. L. and Dilkina, B. 2022. megaSDM: integrating dispersal and timestep analyses into species distribution models. – Ecography 2022: e05450.

Sillero, N., Arenas-Castro, S., Enriquez-Urzelai, U., Vale, C. G., Sousa-Guedes, D., Martínez-Freiría, F., Real, R. and Barbosa, A. M. 2021. Want to model a species niche? A step-by-step guideline on correlative ecological niche modelling. – Ecol. Modell. 456: 109671.

Sillero, N., Campos, J. C., Arenas-Castro, S. and Barbosa, A. M. 2023. A curated list of R packages for ecological niche modelling. – Ecol. Modell. 476: 110242.

Smith, A. B. and Santos, M. J. 2020. Testing the ability of species distribution models to infer variable importance. – Ecography 43: 1801–1813.

Smith, A. B., Murphy, S. J., Henderson, D. and Erickson, K. D. 2023. Including imprecisely georeferenced specimens improves accuracy of species distribution models and estimates of niche breadth. – Global Ecol. Biogeogr. 32: 342–355.

Soberón, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. – Ecol. Lett. 10: 1115–1123.

Soberón, J. and Peterson, A. T. 2005. Interpretation of models of fundamental ecological niches and species’ distributional areas. – Biodivers. Inform. 2.

Soberón, J. and Osorio-Olvera, L. 2023. A dynamic theory of the area of distribution. – J. Biogeogr. 50: 1037–1048.

Stolar, J. and Nielsen, S. E. 2015. Accounting for spatially biased sampling effort in presence-only species distribution modelling. – Divers. Distrib. 21: 595–608.

Thuiller, W., Lafourcade, B., Engler, R. and Araújo, M. B. 2009. BIOMOD - a platform for ensemble forecasting of species distributions. – Ecography 32: 369–373.

Thuiller, W., Guéguen, M., Renaud, J., Karger, D. N. and Zimmermann, N. E. 2019. Uncertainty in ensembles of global biodiversity scenarios. – Nat. Commun. 10: 1446.

Valavi, R., Elith, J., Lahoz-Monfort, J. J. and 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 Ecol. Evol. 10: 225–232.

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J. and Elith, J. 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. – Ecol. Monogr. 92: e01486.

Velazco, S. J. E., Rose, M. B., de Andrade, A. F. A., Minoli, I. and Franklin, J. 2022. flexsdm: an R package for supporting a comprehensive and flexible species distribution modelling workflow. – Methods Ecol. Evol. 13: 1661–1669.

Velazco, S. J. E., Rose, M. B., De Marco Jr, P., Regan, H. M. and Franklin, J. 2024. How far can I extrapolate my species distribution model? Exploring shape, a novel method. – Ecography 2024: e06992.

Vignali, S., Barras, A. G., Arlettaz, R. and Braunisch, V. 2020. SDMtune: an R package to tune and evaluate species distribution models. – Ecol. Evol. 10: 11488–11506.

Warren, D. L. and Seifert, S. N. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. – Ecol. Appl. 21: 335–342.

Warren, D. L., Matzke, N. J., Cardillo, M., Baumgartner, J. B., Beaumont, L. J., Turelli, M., Glor, R. E., Huron, N. A., Simões, M., Iglesias, T. L., Piquet, J. C. and Dinnage, R. 2021. ENMTools 1.0: an R package for comparative ecological biogeography. – Ecography 44: 504–511.

Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V. and Antonelli, A. 2019. CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. – Methods Ecol. Evol. 10: 744–751.

Zizka, A. et al. 2020. No one-size-fits-all solution to clean GBIF. – PeerJ: e9916.

Zizka, A., Antonelli, A. and Silvestro, D. 2021. Sampbias, a method for quantifying geographic sampling biases in species distribution data. – Ecography 44: 25–32.

Zurell, D., Thuiller, W., Pagel, J., Cabral, J. S., Münkemüller, T., Gravel, D., Dullinger, S., Normand, S., Schiffers, K. H., Moore, K. A. and Zimmermann, N. E. 2016. Benchmarking novel approaches for modelling species range dynamics. – Global Change Biol. 22: 2651–2664.

Zurell, D. et al. 2020. A standard protocol for reporting species distribution models. – Ecography 43: 1261–1277.