|  |  |  |  |
| --- | --- | --- | --- |
| **What to report** | **Why is this important?** | **Example** | **What are we doing?** |
| Source of occurrence data | Reporting occurrence data sources allows one to assess data quality and trace/correct any possible issues that may be detected. | “Species distribution records were collected from the Ocean Biogeographic Information System (OBIS; [http://iobis.org](http://iobis.org/), accessed February 2016), from the Global Biodiversity Information Facility (GBIF; [http://gbif.org](http://gbif.org/), accessed January 2016), the Reef Life Survey (RLS; [http://reeflifesurvey.com](http://reeflifesurvey.com/), accessed February 2016) and for a few species via personal communications.” | -Species distributions records were obtained from museum/voucher collections at UI  - If we’d like to use GBIF, I have a script! I was able to get 500 occurrence data points for Ecuador that we can filter |
| Download data; version of data source | Databases and datasets change over time. | “Occurrences were downloaded from GBIF.org on 28 January 2016” | - We will use the most up-to-date data available  - currently no plan to use data from GBIF or other sources |
| Basis of records | Databases comprise many different types of data (data collected opportunistically, structured or repeated surveys, comprehensive checklists, by scientists or citizen scientists) | “Before fieldwork, we obtained locality information from *C. canescens* herbarium specimens and online biodiversity databases such as the Southwest Environmental Information Network and the Rocky Mountain Herbarium (University of Wyoming). In addition, the Rocky Mountain Herbarium and the Colorado State University Herbarium were visited to examine potentially misidentified specimens from outlying portions of the species’ distribution.” | Samples were collected over the past 20 years opportunistically and as part of both structured and repeated surveys. |
| Spatial extent | Spatial extent of occurrences is crucial for interpretation of model predictions, including whether potential sink populations are included, whether sampling is biased or whether records outside the native range are used. | We integrated missing countries by obtaining occurrences from the literature. To increase the accuracy of the analysis, we excluded the following records: (1) localities for which we were not able to obtain precise coordinates; [...] (4) record of *M. bourneti* in the Canary Islands, due to taxonomical issues currently unresolved” | * We excluded localities without precise coordinates * We excluded localities with unclear taxonomic identification * We could include locality information from other sources, but we are not missing major spatial areas |
| Temporal range | Environments can change over time, thus the timestamp of occurrence records is crucial for linking them to the relevant environmental conditions experienced by the species | “Although the sightings dataset extended over 257 years, 79% of sightings occurred between 2000 and 2015. Therefore only this subset of 5,419 sightings was retained for further analysis. These sightings were divided into each quarter of the year (Jan–Mar, Apr–May, Jun–Aug and Sep–Dec) and matched with recent climate data available through online data sharing platforms.” | * Our sampling is only over the last ~20 years |
| Duplicate records | Duplicated coordinates can bias model training | “…including removing duplicate species records from the same grid square.” | * Duplicate coordinates have been removed for each species |
| Spatial/environmental outliers; errors | Outliers or errors may lead to model errors | “Finally, we plotted all the points on maps and excluded any point falling far outside the proven distribution described in Krapovickas et al. (2007).” | * We will remove outlier points on the distribution maps based on expert opinion |
| Spatial and coordinate uncertainty | Coordinates may not be exact, different degrees of specificity, georeferenced based on description | ““For this study, precise locality coordinates for *P. solenopsis* were not available, so the district-level occurrence data published by Nagrare et al. (2009) were used (*n* = 42 records). The centroid method may be acceptable if the target scale of prediction is global but may not be appropriate at national, state or finer scales.” | * Our coordinates are exact or very close to exact * No estimation or coordinates based on descriptions |
| Sampling bias | Unequal sampling of distribution may cause the model to overfit conditions associated with such samples | “To reduce the effects of sampling bias, we spatially filtered the occurrence dataset to ensure that no two localities were within 10 km of one another.” | * We definitely have a sampling bias |
| Spatial autocorrelation | Non-independent spatial distribution of occurrences | “In order to account for autocorrelation in the observations, models were also fitted in which contagion (see below: spatial interpolators) was included as an autocovariate term in the initial variable set (AGLM)…” | * We can try to account for this * I’m a little confused about what this means! |

**Environmental Data**

|  |  |  |  |
| --- | --- | --- | --- |
| Source | Reporting the source of environmental data | “Climate data comprised the 19 BIOCLIM variables available from WorldClim (Hijmans et al., 2005) at a resolution of 2.5 arc-min. Elevation data were obtained from the Digital Elevation Model at PRISM (Precipitation-elevation Regressions on Independent Slopes Model; Daly et al., 1994) at 2.5 arc-min resolution. | -We will report this  - Climate data comprised of the 19 BIOCLIM variables at a resolution of 0.5 arc-min |
| Download date; version of data source | Data and databases are not static — they change over time. | ““Out of the available WorldClim data ([http://www.worldclim.org](http://www.worldclim.org/)), we used the 19 bioclimatic variables, which express 11 temperature and 8 precipitation metrics at about 1-km resolution (WorldClim version 1.4)” | -We can do this |
| Spatial resolution | Environmental data usually have various spatial resolutions that need to be reconciled for model training. Also, the decision of spatial resolution is both a technical and an ecological issue. | “Four static variables were derived from the digital elevation model (DEM) of the EMODnet Bathymetry portal: depth (the DEM); slope and curvature, calculated using DEM Surface Tools for ArcGIS 10.2; distance to the nearest 200 m bathymetric line, calculated using QGis 2.12. Curvature was used as a proxy of sea bottom roughness, providing an estimate of sea floor relief, which can influence some cetacean species (Lindsay et al., 2016). All static variables were calculated at a spatial resolution of 0.5 × 0.5 km.” | - Is the spatial resolution not just the 0.5 arc-min? That converts to ~1 x 1 km |
| Temporal range | The temporal range (time period across which the variable was measured and averaged) is needed to determine the temporal match or mismatch with species’ occurrences | “We summarized occurrence of passerine bird species at BBS routes in the conterminous U.S. during historical (1977–1979) and recent (2012–2014) periods. Land use covariates were the proportion of the buffer surrounding each route in developed and conservation or low human use classes based on the 1974 and 2012 versions of the U.S. conterminous wall-to-wall anthropogenic land use trends dataset (NWALT; Falcone, 2015).” | - The bioclim variables are from August 2020 |

**Model calibration**

|  |  |  |  |
| --- | --- | --- | --- |
| Modeling domain | The geographic domain of a model has to be specified | “In the second approach, locality data were overlaid on terrain base maps in ArcGIS 10.2 (Environmental Systems Research Institute, 2011) together with a world ecoregions layer. These were used to identify breaks in habitat and ecological regions in topographically homogeneous areas. [...] Restricting calibration areas to regions bounded by significant abiotic barriers and known or hypothesized dispersal distances yielded more accurate models and reduced these errors Thus, in our study, Ms were constrained by deep valleys, the crests of mountains (and ~other distinct features likely to act as barriers to species distributions” | - Nothing like terrain base maps was used |
| Number of background data | Background data are assumed to represent the environmental composition of species’ accessible area, thus the optimal number of background data may depend on the extent of the study area and resolution of environmental data, as well as computation capacity. | “For each geographical background we selected 10,000 random cells that did not hold a species presence record (or all available cells if fewer than 10,000 were available).” | - For each species, the background points are the points from all the other species |
| Sampling method for background data | Random selection of background data has been used as the default strategy in some algorithms, but new methods have been developed for different purposes. | “We used Maxent with default settings, except that we applied a targeted background sampling to reduce the influence of sample selection bias (Phillips et al., 2009) by using 666 vertebrate fossil site localities throughout New Zealand as background points.” | - Sampling method for background data is the same as presence data, since it’s just points of other species |
| Variable selection | Selection of variables is biologically and/or statistically relevant | “Four ‘bioclimatic’ layers were used to calibrate models: mean temperature of the warmest quarter, mean temperature of the coldest quarter, precipitation of the wettest quarter, and precipitation of the driest quarter. These four layers were chosen because they represent the climatic extremes that often constrain species distributions and because most other bioclimatic layers are derived from different combinations of or are tightly correlated with these variables (Root, 1988).” | - 4 bioclimatic layers were used: Mean Diurnal Range, Isothermality, Temperature Annual Range, and Annual Precipitation. These were chosen because they were not tightly correlated with other variables and are important to snails. |
| Name | Name of model | “ENM was performed using the maximum entropy approach as implemented in MAXENT 3.3.3k” | - GLM implemented in the R package ENMTools |
| Version of algorithm and software | Algorithm, settings and libraries change overtime | “Folds were set at random and other parameters were left as default in the gbm R package (version 2.1.1). Runs on R version 3.3.2.” | - Runs on R version 4.0.2 using the ENMTools package and the enmtools.glm command with an ENMTools species object |
| Parameterization | Parameter settings can influence results | “Selecting the best settings for the regularization multipliers and number of feature classes, which determine the model complexity, requires quantitative evaluation (Merow et al., 2013). The optimal model parameters were tuned using the function ENMevaluate in the package ‘ENMeval’ (Muscarella et al., 2014) for R. Within ENMevaluate, we jackknifed each species presence record and evaluated models with the following feature classes: linear, quadratic, and hinge, and the following values of regularization multipliers: 0.75, 1, 1.25, 1.5.” | - The GLM equation includes both linear and polynomial terms to allow for more complex suitability distributions |

**Model transfer and evaluation**

|  |  |  |  |
| --- | --- | --- | --- |
| Evaluation index | Proper understanding of model performance requires the use of model evaluation indices. | “We evaluated the performance of the models by three different methods using an independent dataset of occurrences for model evaluation: (a) an omission error test [...] (b) the binomial cumulative probability [...] (c) the partial receiver operating characteristic [...]” | - ENMTools reports an AIC for each model. It also reports AUC for model fit (training data), environment space model fit (training data), model fit (test data), and environment space model fit (test data). AIC is our best bet though (Warren and Seifert, 2011). |
| Threshold for evaluation index | Calculation of some evaluation indices requires a threshold. The threshold will vary by study because there is no single, default method for choosing a threshold. | “A threshold to convert continuous predicted probabilities into a binomial output was estimated for each model run, using the threshold value that maximized specificity (true negative rate) and sensitivity (true positive rate) over the evaluation dataset predictions (Liu, Newell, & White, 2016). Using this threshold, two metrics of predictive performance were derived: the sensitivity of models when predicting ARGOS tracking locations (“sensitivity. ARGOS”, in% correctly classified as presences), and the true statistic skill when predicting the evaluation datasets (“TSS”; Allouche, Tsoar, & Kadmon, 2006).” | - Not entirely sure what this one means! |
| Dataset used to evaluate models | Evaluation of a model is usually based on another independent dataset, or part of the dataset not used in model training. The choice of dataset can influence the evaluation results and the subsequent interpretations. | All these methods used observed presences as input with a 70% random sample for model development and the remaining 30% sample for model evaluation.” | -I hadn’t considered this. I wonder if it’s necessary for our project though? |
| Format/transformation | The raw model predictions are sometimes transformed (for example, logistic transformation) via different methods under different assumptions. | “[...] we used the logistic output format [...]” | - No transformation used |
| Threshold | Often, the model predictions are in continuous format, which is subsequently transformed into a binary prediction under a particular threshold. Researchers have proposed different ways of thresholding for different purposes and under varied assumptions. | ““We repeated this procedure 20 times for each algorithm and used the Lowest Present Threshold values (Pearson et al., 2007) to transform each map in binary.” | - We have a suitability gradient, but we consider the most suitable cells (have a score >= 0.80) for choosing a good point to visit |
| Novelty of projected environments relative to training environments | Transferring a model across space and/or time may lead to extrapolation if the projected environments are novel compared with training environments. Quantification of novel environments could help understand the uncertainties associated with model predictions. | “To assess the effect of model extrapolation on values of predictor variables lying outside the training range, that is, projecting models on non-analogous climates (cf. Nogues-Bravo, 2009), we conducted a multivariate environmental similarity surfaces (MESS) analysis, following Elith et al. (2011).” | - I wonder if we don’t have to worry about this because the extrapolation is suitability and not environmental variables? |
| Collinearity shift between training and projected environments | Transferring a model outside training data may be affected by differences in collinearity structure between training and projection environments, which can lead to degraded prediction performance. Therefore, quantification of collinearity shift or any steps towards correcting for it should be specified. | “We compared the correlation matrix of the 6 variables in the training region to the average of the correlation matrices of present and future climate layers in the projected area (Tables [S2](https://www.nature.com/articles/s41559-019-0972-5#MOESM1) & [S3](https://www.nature.com/articles/s41559-019-0972-5#MOESM3) in Supplement [2](https://www.nature.com/articles/s41559-019-0972-5#MOESM1)). The highest absolute change of *r* was 0.3 for bio4 and bio17, and *r* increased above the 0.7 threshold for 2 pairs of variables (−0.78 for bio3 and bio4; 0.71 for bio16 and bio17; Supplement [2](https://www.nature.com/articles/s41559-019-0972-5#MOESM1)).” | - We are not predicting changes in the bioclim layers, so this is not applicable |
| Extrapolation strategy | Model extrapolation is statistically challenging. Different extrapolation strategies can lead to very different model predictions, therefore the choice of extrapolation, even the default setting of an algorithm, should be provided. | “Five replicates of each model were conducted with no clamping or extrapolation and with all the default ‘features’ used.” | - No clamping or extrapolation, used only the default ENMTools GLM settings |