

ECOGRAPHY

Research article

Species distribution models affected by positional uncertainty in species occurrences can still be ecologically interpretable

Lukáš Gábor¹✉, Walter Jetz^{2,3}, Alejandra Zarzo-Arias^{1,4,5}, Kevin Winner^{2,3}, Scott Yanco^{2,3,6}, Stefan Pinkert^{2,3,7}, Charles J. Marsh^{2,3}, Matthew S. Rogan^{2,3}, Jussi Mäkinen^{2,3,8}, Duccio Rocchini^{1,9}, Vojtěch Barták¹, Marco Malavasi¹⁰, Petr Balej¹ and Vítězslav Moudrý¹

¹Dept of Spatial Sciences, Faculty of Environmental Sciences, Czech Univ. of Life Sciences Prague, Prague, Czech Republic

²Dept of Ecology and Evolutionary Biology, Yale Univ., New Haven, CT, USA

³Center for Biodiversity and Global Change, Yale Univ., New Haven, CT, USA

⁴Univ. de Oviedo, Asturias, Spain

⁵Dept of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), Madrid, Spain

⁶Max Planck – Yale Center for Biodiversity Movement and Global Change, New Haven, CT, USA

⁷Dept of Conservation Ecology, Univ. of Marburg, Marburg, Germany

⁸Research Centre for Ecological Change, Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, Univ. of Helsinki, Helsinki, Finland

⁹BIOME Lab, Dept of Biological, Geological and Environmental Sciences, Alma Mater Studiorum Univ. of Bologna, Bologna, Italy

¹⁰Dept of Chemistry, Physics, Mathematics and Natural Sciences, Univ. of Sassari, Sassari, Italy.

Correspondence: Lukáš Gábor (gabor.lucas@gmail.com)

Ecography

2023: e06358

doi: [10.1111/ecog.06358](https://doi.org/10.1111/ecog.06358)

Subject Editor: Cory Merow

Editor-in-Chief: Miguel Araújo

Accepted 09 February 2023



Species distribution models (SDMs) have become a common tool in studies of species–environment relationships but can be negatively affected by positional uncertainty of underlying species occurrence data. Previous work has documented the effect of positional uncertainty on model predictive performance, but its consequences for inference about species–environment relationships remain largely unknown. Here we use over 12 000 combinations of virtual and real environmental variables and virtual species, as well as a real case study, to investigate how accurately SDMs can recover species–environment relationships after applying known positional errors to species occurrence data. We explored a range of environmental predictors with various spatial heterogeneity, species' niche widths, sample sizes and magnitudes of positional error. Positional uncertainty decreased predictive model performance for all modeled scenarios. The absolute and relative importance of environmental predictors and the shape of species–environment relationships co-varied with a level of positional uncertainty. These differences were much weaker than those observed for overall model performance, especially for homogenous predictor variables. This suggests that, at least for the example species and conditions analyzed, the negative consequences of positional uncertainty on model performance did not extend as strongly to the ecological interpretability of the models. Although the findings are encouraging for practitioners using SDMs to reveal generative mechanisms based on spatially uncertain data, they



www.ecography.org

suggest greater consequences for applications utilizing distributions predicted from SDMs using positionally uncertain data, such as conservation prioritization and biodiversity monitoring.

Keywords: birds, ecological modeling, location error, niche models, species–environment relationship

Introduction

Species occurrences are increasingly being recorded in online, public, global databases such as GBIF (www.gbif.org), eBird (www.ebird.org), or iNaturalist (www.inaturalist.org), where scientists and the general public worldwide share field observations. However, whereas the number of records in these databases is constantly growing, many observations are characterized by substantial uncertainty in the occurrence location (Moudrý and Devillers 2020). Such uncertainty poses problems for analyses aimed at revealing species–environment relationships because the environmental conditions at recorded sites could differ from those at true locations.

Species distribution model (SDMs) are a widely used class of ecological models that use occurrence data to estimate species–environment relationships (Ferrier et al. 2017) and allow researchers to predict the relative probability of occurrence across unsampled areas of a study region. SDMs have broad utility in ecology (Elith and Leathwick 2009, Franklin 2010, Guisan et al. 2013, Zurell et al. 2019) and have been successfully used to identify critical habitats (Volis and Tojibaev 2021), delineate suitable locations for relocations (Segal et al. 2021), or assess the potential impacts of climate change (Santini et al. 2021). SDMs are also frequently used to infer the importance of environmental variables defining the species niche (Moudrý and Šimová 2013, Bradie and Leung 2017, Lecours et al. 2020, Li and Kou 2021, Smith and Santos 2020) and to determine the shapes of species responses to the environment (Austin et al. 2006, Hargreaves et al. 2014, Lee-Yaw et al. 2016, Dvorský et al. 2017, Bazzichetto et al. 2018). However, despite methodological advances improving the performance of SDMs over the last two decades (Phillips et al. 2006, Varela et al. 2014, Graham et al. 2019, Tassarolo et al. 2021), they remain sensitive to the spatial accuracy of occurrence data used in model fitting (Visscher 2006, Moudrý and Šimová et al. 2012, Moudrý et al. 2017, Araújo et al. 2019, Byaraktarov et al. 2020, Isaac et al. 2020, Etherington et al. 2021, Gábor et al. 2022, Moudrý et al. 2023).

Maximum Entropy-based SDMs estimate a response curve in environmental space which discriminates between observed occurrences and ‘background’ samples that do not contain occurrence information (Fig. 1). Positional uncertainty describes the magnitude of error in the locations of occurrence records. In some cases, it quantifies the likelihood of a mismatch between the true environmental variables’ values and the assigned value. Even if a spatial error does not lead to directional bias in environmental space, increased sampling error can decrease predictive model performance and even bias the slope of the response curves; or the estimations of variable importance (Fig. 1; Johnson and Gillingham 2008, Fernandez et al. 2009, Osborne and Leitão 2009, Hefley et al.

2014, Fernandes et al. 2019). The magnitude of positional error in environmental variables measurements may be amplified in highly heterogeneous or structured landscapes where spatial autocorrelation (SAC) in environmental variables is relatively low (Naimi et al. 2011, Naimi et al. 2014). Moreover, even uniform spatial error can create persistent bias in measurements of environmental variables, depending on the spatial structure of the relevant variable. For example, uniformly distributed spatial error for occurrences of a mountaintop-dwelling species would always lead to estimates of elevation that are biased to lower elevations than reality. Such a bias in even one environmental variable could reduce overall model predictive performance and bias the estimated response curve and variable importance (Fig. 1).

Even more troubling, common strategies for mitigating the effects of positional uncertainty on SDMs have recently

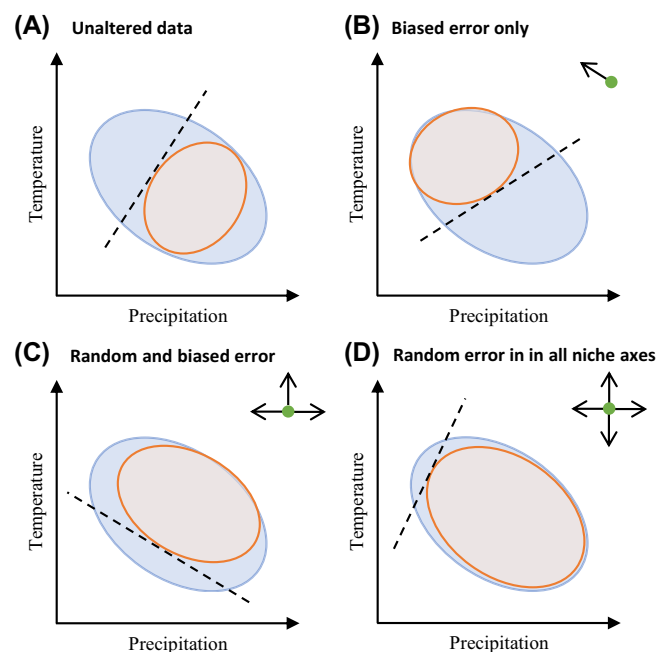


Figure 1. Heuristic illustration of the effect of positional error (depicted in environmental space) on response curve estimation. Panel (A) shows the area of species occurrences (orange) relative to area of the sampled background points (blue) without positional uncertainty. The ‘true’ response curve that would be estimated from these data would result in an approximate discrimination threshold which is represented by the dashed line. The response curve would differ from the ‘true’ response curve when spatial error leads to (B) persistent bias in environmental space, (C) persistent bias and unbiased sampling error in environmental space, and even (D) random sampling error in environmental space (i.e. without directional bias). Arrows in the upper-right of each panel indicate the directions of shift for the presences (green point).

been shown to be ineffective. For example, Gábor et al. (2020) demonstrated that increased sample sizes do not reduce the negative effects of positional uncertainty. Similarly, Smith et al. (2023) showed that discarding data with high positional uncertainty limits our ability to determine species' distribution and climatic niche tolerances properly. In particular, they demonstrated that using only accurate data dramatically reduces range size estimates and overestimates exposure to climate change. Recently, Gábor et al. (2022) concluded that coarsening the analysis grain to compensate for positional error did not improve model performance, and recommended the development of models with the finest possible analysis grain and as close to the response grain as possible, even when available species occurrences suffer from positional errors.

Although previous studies confirmed the effect of positional error on the model predictive performance (Graham et al. 2008, Johnson and Gillingham 2008, Fernandez et al. 2009, Osborne and Leitão 2009, Naimi et al. 2011, 2014, Hefley et al. 2014, Tulowiecki et al. 2015, Gueta and Carmel 2016, Mitchell et al. 2017, Soutan and Safi 2017, Fernandes et al. 2019, Gábor et al. 2022), especially when specialist species are modeled (Visscher 2006, Gábor et al. 2020), the question about how positional uncertainty in species occurrences affects models' parameter estimation (species–environment relationships inference) remain largely unexplored. Therefore, in this study, we explored the extent to which parameter estimation is affected by positional uncertainty. Specifically, we investigated the influence of positional error on variable importance and the shape of the response curves. We hypothesized that increasing positional uncertainty would lead to decreased model predictive performance, and imprecise variable importance and response curves, with more pronounced effects for species with narrow niche and heterogeneous variables.

Material and methods

We used a virtual species approach across two workflows (Fig. 2), which allowed us to know the true underlying occurrence location and thus enabled us to characterize relative bias in parameter estimates (Zurell et al. 2010, Moudry 2015, Meynard et al. 2019), as well as specify various spatial autocorrelation levels (SAC; Naimi et al. 2011, 2014) in the environmental variables.

We simulated 12 560 combinations of virtual and real environmental data and virtual species to investigate our assumptions and fitted over 628 000 models. Simulations were divided into two workflows and a third workflow investigated a real species. In Workflow 1, we combined virtual variables with different levels of SAC and virtual species with varying widths of niches and number of occurrences. For these scenarios, models were fitted with only one variable (Fig. 2). Thanks to this, we got a simplified yet detailed insight into how various levels of positional uncertainty affects a model's ability to properly detect species response to the environment across various SAC, niche width, and sample size.

Additionally, to mimic real SDMs situations, we combined real environmental variables with virtual species with different niche widths and sample sizes, and fitted models with multiple environmental variables (Workflow 2; Fig. 2). This allowed us to explore our assumptions with more model complexity. Moreover, using numerous environmental variables to fit the models, we tested how positional uncertainty affects a models' ability to properly detect the most influential variables (i.e. those used to generate virtual species).

Finally, we tested our assumptions using real environmental variables and real species (band-tailed pigeon; Workflow 3; Fig. 2). Our simulations showed that the model parameter estimation is negatively affected across various species niches (note, however, that the magnitude varies). Therefore, we hypothesized, considering the number of occurrences ($n=111$) and the fact that the species is widely spread across the western part of the USA, that positional uncertainty will bias model response curves and variable importance.

In Workflow 1, the artificial study area was given by the extent of the virtual landscape (200×200 cells; see details below). Virtual species in Workflow 2 used Spain (except islands) as a study area, whereas the band-tailed pigeon was modeled for the USA (Fig. 2; Workflow 3).

Occurrence data

Workflow 1

We generated artificial occurrences using the 'virtualspecies' package (Leroy et al. 2016, ver. 1.5) in the statistical software R (ver. 4.1.0, www.r-project.org) with three steps: 1) define (virtual) species–environment relationships, 2) project range into geographic space, and 3) sample occurrence data from simulated range. We used a normal distribution to define the response of virtual species to the virtual environmental variable. To simulate species with different niche widths, we used the same mean (0.005) and varied standard deviation from 0.005 up to 0.2 using a logarithmically spaced sequence. In total, we generated 25 species with various niche widths. We then projected habitat suitability across our study area to define the probability distribution of occurrences. In the final step, we used a probabilistic simulation approach and logistic function with $\alpha=-0.05$ (controls the slope of the logistic curve) and $\beta=0.3$ (the point of inflection of the logistic curve, i.e. the value of the environmental gradient at which the probability of occurrence is 50%), as recommended in prior studies, to convert the habitat suitability raster to a randomized binary presence–absence raster (Meynard and Kaplan 2012, 2013, Meynard et al. 2019). This allowed us to generate virtual species with gradual responses to the environment that mimic the real species, as demonstrated by Meynard and Kaplan (2012, 2013). Subsequently, we sampled 20, 100, 300, and 1000 species occurrences.

Workflow 2

To generate virtual species occurrences for Workflow 2, we used two environmental variables with various SAC (elevation – high SAC, aspect – low SAC). To simulate species with

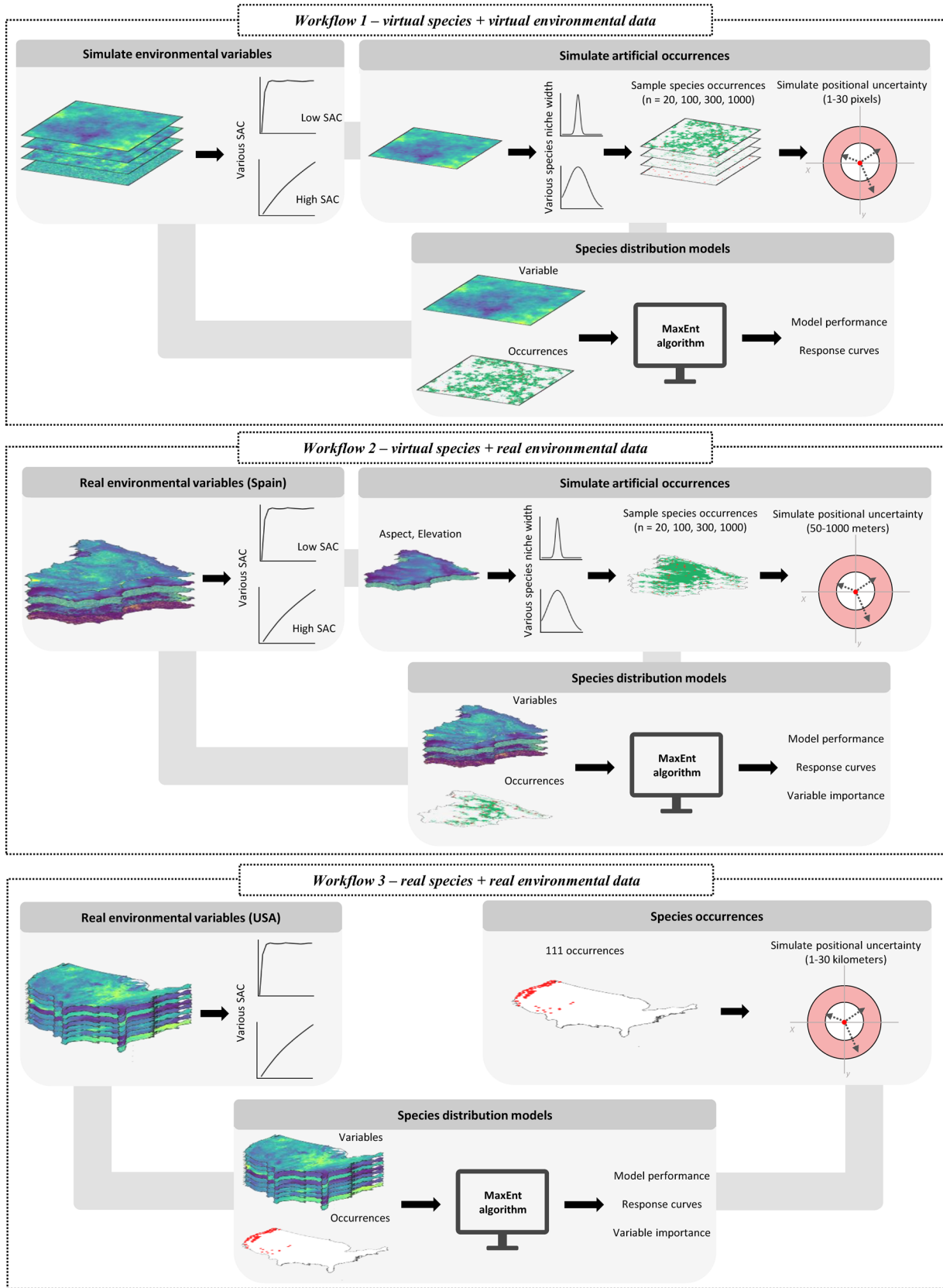


Figure 2. General modeling process for all three workflows. Each experiment was repeated 50 times.

different niche widths, we used normal distributions with a mean of 1000 m and standard deviation of 100–500 m for elevation, and a mean of 100° and standard deviation of 10–100° for aspect. This allowed us to generate three species with various niche widths (narrow, medium, wide). We then projected habitat suitability across our study area to define the probability distribution of occurrences. We used a logistic function with $\alpha = -0.05$ and $\beta = 0.3$ to convert the habitat suitability raster to a randomized binary presence–absence raster. We sampled 20, 100, 300, and 1000 species occurrences.

Workflow 3

Occurrences for the band-tailed pigeon, a species with high detection probability (Keppie and Braun 2000), were extracted from the North American Breeding Bird Survey (BBS, Sauer et al. 2017), a long-term collection of over 4800 survey routes distributed across North America. Each survey route consists of 50 points count locations distributed ~ 0.8 km apart and sampled for 3 min. We considered only routes from the study region (the USA, Fig. 2) and retained only those routes where we assumed there was high certainty that the species was present. First, we discarded data sampled before the year 2000 and then kept only those routes with at least ten years of samples post-2000. We considered occurrences from a minimum of five years of samples as presences. Routes where the species was detected but on fewer occasions, and therefore presence status was unsure, were removed from the analysis. The final dataset contained 111 presences.

Environmental variables

Naimi et al. (2011, 2014) showed that SAC in environmental variables affects the degree to which positional uncertainty creates mismatches between true and measured environmental variables values. Therefore, we generated artificial environmental variables and selected real environmental variables that spanned different degrees of SAC (see variograms in Supporting information).

Workflow 1

We generated artificial environmental variables using the R 'gstat' package (ver. 2.0-9, www.r-project.org) and unconditional simulation over a regular grid of 200 × 200 cells. Unconditional simulation allows for a generation of environmental variables with different SAC, where the level of SAC is defined by a variogram (Dungan 2002, Naimi et al. 2011). We used an exponential variogram with the same sill parameter of 0.025 for all simulations. To simulate variables across different SAC levels, we scaled the range parameters from 1 (low SAC, high heterogeneity) to 49 (high SAC, low heterogeneity) by increments of 2 to a total of 25 virtual environmental variables. Only one variable was used to generate virtual species and subsequently model the species distribution (Fig. 2).

Workflow 2

We chose five environmental variables to construct models for Workflow 2 (Supporting information). Two of the

variables were related to habitat characteristics: grassland coverage and forest coverage (<http://centrodedescargas.cnig.es/>; Spain National Center for Geographic Information), and three were related to topography: topography wetness index, aspect and elevation (<http://centrodedescargas.cnig.es/>; Spain National Center for Geographic Information). We used elevation and aspect that serve as a proxy for temperature (Müller and Brandl 2009, Coops et al. 2010, Vierling et al. 2011, Work et al. 2011, Vogeler et al. 2014), and a topography wetness index that is a proxy for water availability (Petrosselli et al. 2013, Reif et al. 2018, Title and Bemmels 2018). The topography wetness index was derived from the elevation model (SAGA-GIS ver. 2.1.4; Conrad et al. 2015). All environmental variables were resampled from an original resolution of 25 × 25 m (elevation) or 20 × 20 m (all other variables) to 50 × 50 m cell resolution using the mean values of the original data (Moudrý et al. 2019) for modeling purposes (Supporting information). Only elevation and aspect were used to generate virtual species, while all variables were used to fit models.

Workflow 3

For the band-tailed pigeon, we selected nine variables that reflect fine to coarse scales of spatial and temporal variation. Four variables were related to climate: mean annual temperature, seasonality of precipitation, growing season precipitation (CHELSA ver. 1.2; Karger et al. 2017), and the inter-annual variation of cloud cover (EarthEnv; Wilson and Jetz 2016). Two variables were related to vegetation productivity: spatial heterogeneity of enhanced vegetation index (EVI) (EarthEnv; Tuanmu and Jetz 2015) and the mean enhanced EVI for winter, derived from MODIS (Didan et al. 2015). Two variables were related to soil characteristics: proportion of soil silt content and soil clay content (SoilGrids ver. 2; Poggio et al. 2021). The final variable, the terrain ruggedness index (EarthEnv; Amatulli et al. 2018), represented topographical variation. All variables were resampled to a 1 × 1 km cell size from their native projections (see Supporting information for further details and provenance).

We used variance inflation factor analysis (VIF; 'usdm' package, ver. 1.1-18, www.r-project.org) to identify potential multicollinearity issues between our environmental variables. Multicollinearity between predictors can negatively affect SDMs by causing unstable parameter estimates and biased test statistics (Belsley 1991, Chatfield 1995, Dormann et al. 2013). All VIF values indicated low multicollinearity (< 3). Thus we did not exclude any variables on this basis (Zuur et al. 2010).

Simulating positional uncertainty in occurrence data

Positional error in species occurrences may vary depending on the data source and original collection method (e.g. geographic coordinates or written description). Whereas for occurrences gathered with GNSS (Global Navigation Satellite System) the positional uncertainty may range from a couple up to tens of meters, occurrences gathered with older

technologies or those georeferenced from museum databases may have positional uncertainty of up to tens of kilometers (Moudry and Devillers 2020). Therefore, to mimic the range of positional uncertainty in real datasets, we shifted occurrences in a (uniform) random direction according to four different scenarios. As the resolution of environmental variables used in SDMs was different for both virtual species (1×1 pixel respectively 50×50 m) and for band-tailed pigeon (1×1 km), we shifted occurrences in a random direction by drawing a shift distance from a uniform distribution from the following distances: S1: 1–2 pixels, S2: 2–5 pixels, S3: 5–10 pixels, S4: 10–30 pixels (Workflow 1); S1: 50–100 m, S2: 100–250 m, S3: 250–500 m, S4: 500–1500 m (Workflow 2); S1: 1–2 km, S2: 2–5 km, S3: 5–10 km, S4: 10–30 km (Workflow 3; see Supporting information). If the original data points were shifted outside of the study area, the shift was recalculated until the new coordinates were located within the boundaries of the study area.

Model fitting and evaluation

We built SDMs in the statistical software R (package ‘sdm’ ver. 1.0-98, www.r-project.org; Naimi and Araújo 2016) using the MaxEnt modeling method (Phillips et al. 2006), a presence-background method often adopted in ecological studies (Rodríguez et al. 2019, Santamarina et al. 2019, Ancillotto et al. 2020, El-Gabbas et al. 2021, Boral and Moktan 2021, Ellis-Soto et al. 2021, Venne and Currie 2021, Zarzo-Arias et al. 2022). We used 10 000 randomly sampled background points and default model settings (Phillips and Dudík 2008), except that we set the beta parameter to 0.5 and restricted used features. Only hinge features were allowed for virtual species (Workflows 1 and 2). Although hinge features might lead to model overfitting, we used them as our virtual species response to the environment was defined using a normal distribution (Elith et al. 2010). For band-tailed pigeon, we sampled background points only in the extent of species occurrences (western coast of USA; VanDerWal et al. 2009, Barve et al. 2011, Merow et al. 2013) and used quadratic features to avoid overfitting (Austin 2007).

We used a variety of discrimination metrics to evaluate predictive model performance. We used the Sorensen index (SI), recommended for SDMs evaluation using presence-only occurrences (Li and Guo 2013, Leroy et al. 2018). SI ranges from 0 to 1, where 0 means that none of the predictions matched any observation, and 1 means that predictions perfectly fit observations without any false positive or false negative (Leroy et al. 2018). We also calculated overprediction (OPR, Barbosa et al. 2013) and underprediction (UPR, Fielding and Bell 1997) rates to explore whether positional uncertainty led to a consistent over- or underprediction bias. The OPR measures the percentage of predicted presences corresponding to false presences, whereas UPR measures the percentage of actual presences not predicted by the model (Fielding and Bell 1997, Barbosa et al. 2013, Leroy et al. 2018). In addition, we computed the true skill statistic (TSS, Allouche et al. 2006), despite recent criticisms about its use

due to prevalence dependency (Lobo et al. 2008, Jiménez Valverde 2012, Leroy et al. 2018). We explored TSS in addition to SI as it is still widely applied in ecological studies (Fern et al. 2020, Holder et al. 2020, Eduardo et al. 2022, Sanguet et al. 2022). TSS ranges from -1 to $+1$, where $+1$ indicates perfect agreement and values of zero or less indicate random performance (Allouche et al. 2006).

We ran SDMs using fivefold cross-validation (Merow et al. 2013), where species occurrences and background points were divided randomly into fivefolds, and each fold was retained for model testing while the other fourfolds were used for model training. We repeated each experiment 50 times, and evaluations represent averages of the 50 repetitions.

We evaluated each predictor variable’s importance and visualized predicted responses to the environmental variables to explore the effect on inference about generative mechanisms. To estimate variable importance, we used a leave-one-out sensitivity analysis method which calculates the improvement in the model performance with the inclusion of each variable compared to when the variable is excluded (AUCtest; Murray and Conner 2009). Response curves were automatically generated by the ‘sdm’ package (Naimi and Araújo 2016) using the ‘evaluation strip’ approach. This approach visualizes species responses for used environmental variables by including data frames that show the distribution of observed presence point locations within the environmental range investigated by the evaluation strips (Kindt 2018; detailed in Elith et al. 2005).

Results

Model predictive performance

Note that here we present only results for the SI to simplify the presentation of the results, but results for the TSS qualitatively followed the same pattern (Supporting information).

In general, in Workflow 1 where points were not shifted (hereafter unaltered), models achieved excellent performances for species with narrow niche widths (SI > 0.9 , OPR and UPR < 0.03 ; Fig. 3, Supporting information). However, predictive performance generally decreased with increasing species niche width (SI decreased on average by 0.53, while OPR and UPR increased on average by 0.57, respectively by 0.5). Predictive performance generally decreased with increasing positional error in occurrence data. Where the level of SAC was low and the sample size small, the more pronounced was the negative effect of positional error in species occurrences (Fig. 3, Supporting information). In Workflow 2, unaltered models achieved very good model performances (SI > 0.86 , OPR < 0.04 , UPR < 0.19 , Fig. 3, Supporting information). Again, performance decreased with increasing niche width and with introducing positional error (Fig. 3, Supporting information).

Unaltered models for band-tailed pigeon (Workflow 3) achieved very good model performance (SI achieved on average 0.86, OPR 0.13, and UPR 0.15) and, once more,

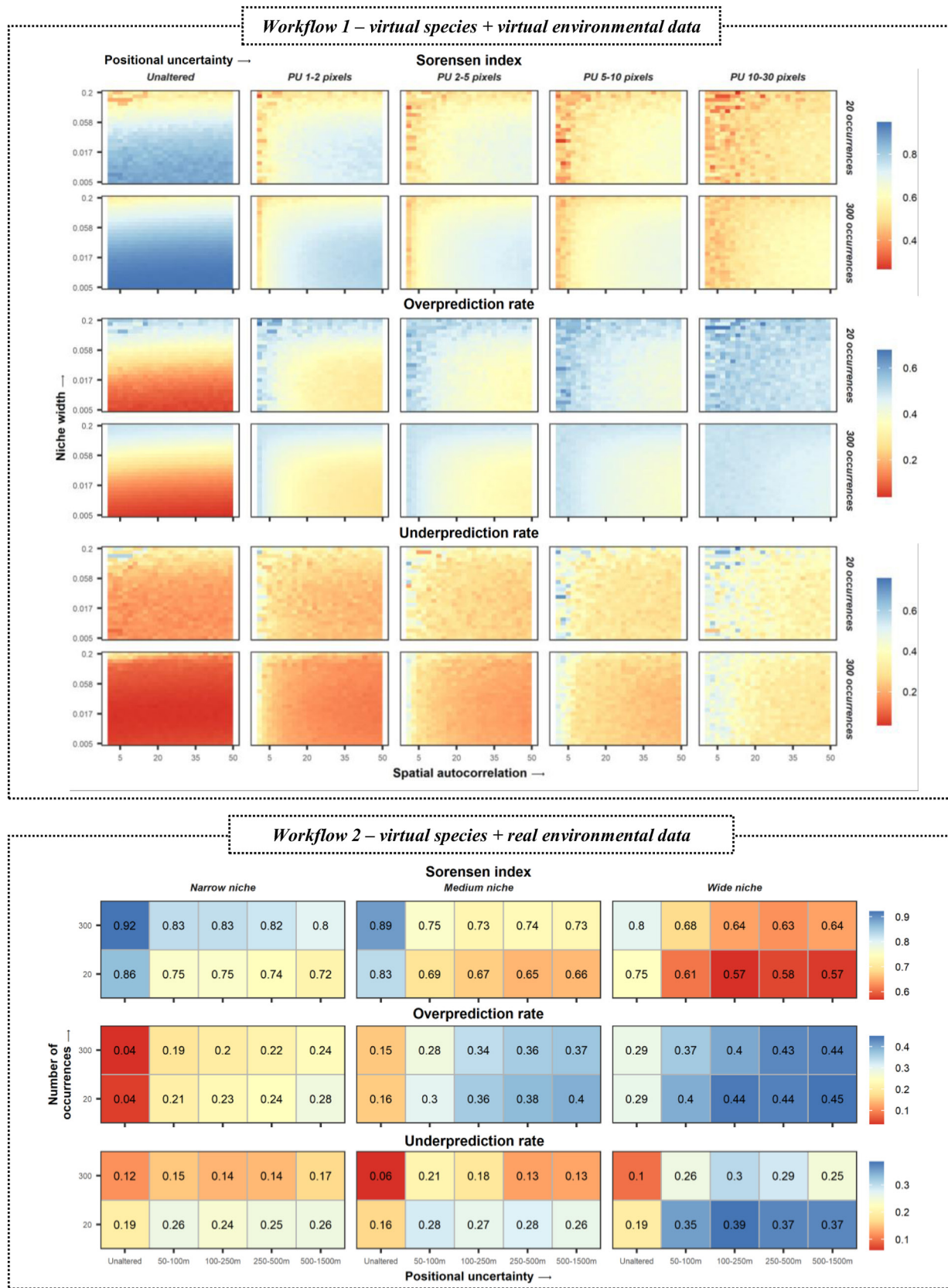


Figure 3. Change in model performance, measured through the Sorensen index, overprediction rate and underprediction rate for Workflows 1 and 2. See Supporting information for plots of resulting changes for all scenarios. Values represent averages of the 50 repetitions.

positional uncertainty led to decreases in model performance (Fig. 4). Compared to virtual species data, the decrease in model performance was, however, lower (SI for Workflow 2 virtual species decreased on average over 0.3, versus an average of 0.03 for the real species; Fig. 3, 4, Supporting information).

The general increase in over- and underprediction rates across all workflows implies that models fit to data with positional error tended to overpredict and, at the same time, underpredict species habitat suitability. Therefore, using positionally uncertain data might be highly risky for some ecological applications (e.g. nature conservation).

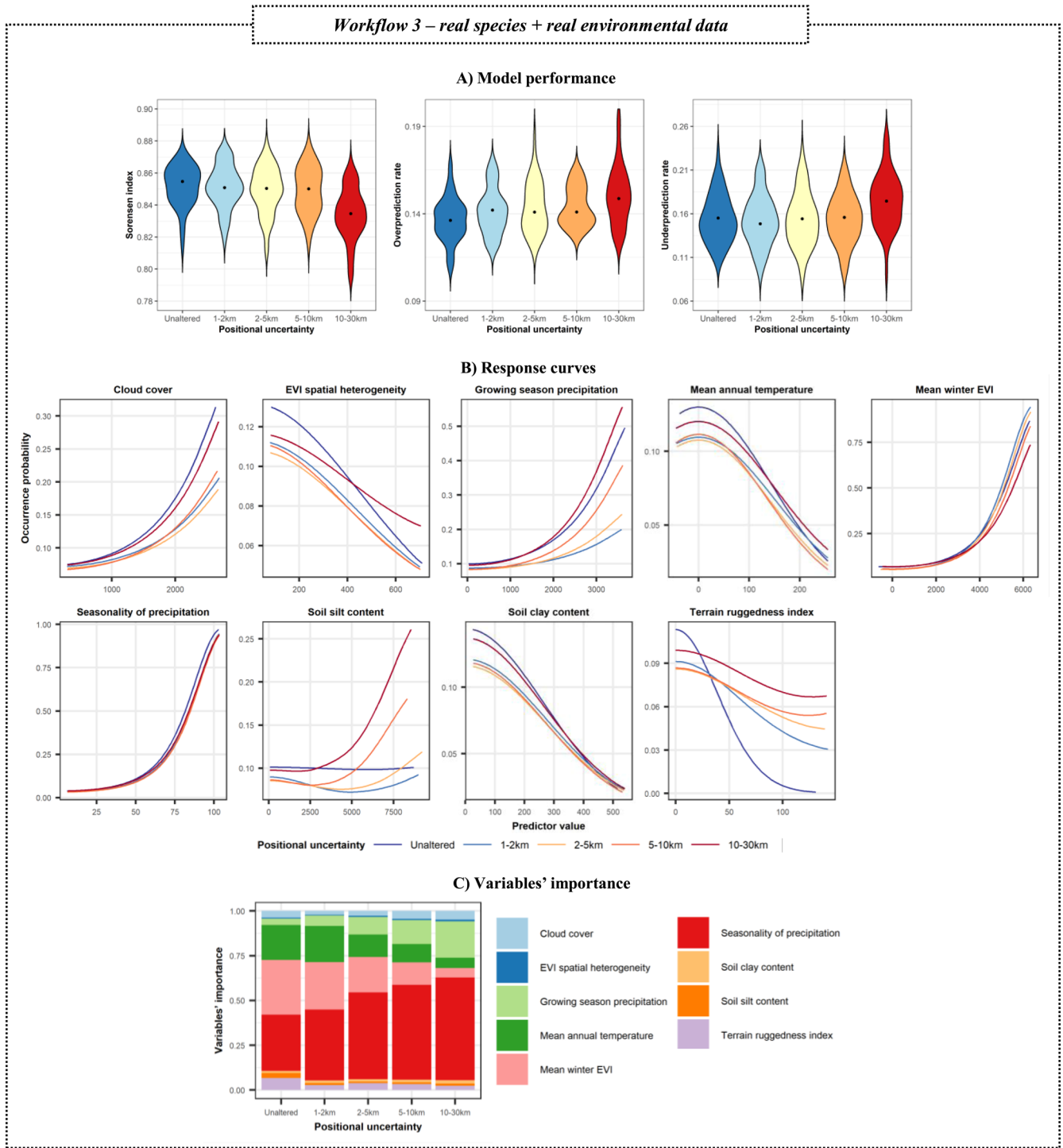


Figure 4. Resulting performance metrics (A), response curves (B) and variables' importance (C) of unaltered and altered models for all scenarios with real species and real environmental data. Values represent averages of the 50 repetitions.

Parameter estimation

Variable importance

For Workflow 2, the models correctly – across all modeled scenarios – detected the aspect and elevation, which were used to generate virtual species as the most influential variables (only these variables were used to generate virtual species; Fig. 5, Supporting information). Increasing sample size increased the estimated importance of aspect and elevation. On the other hand, as niche width increased, models estimated the greater importance of other variables.

For band-tailed pigeon (Workflow 3), the most influential variables were mean winter EVI (16.5%) and seasonality of precipitation (17%), followed by mean annual temperature (10.1%), with other variables below 10% (terrain ruggedness, soil clay content, growing season precipitation, EVI spatial heterogeneity, soil silt content, cloud cover; Fig. 4).

Positional errors led to changes in variable importance. Workflow 2 models correctly inferred the most influential variables regardless of the degree of positional error, although in the high-error scenario, the absolute importance of aspect decreased in importance by 41.2% while the importance of elevation increased by almost 32% (Fig. 5, Supporting information). For variables with minor importance, we generally observed only small changes (i.e. < 4.4% change) to their importance.

In Workflow 3, models correctly inferred the seasonality of precipitation as the most influential variable independently of positional error. As error increased, we observed a decrease in the importance of mean winter EVI and mean annual temperature (by 13.6 and 6.1%, respectively), and an increase in the importance of seasonality of precipitation (by 13.1%), as well as growing season precipitation (by 10.8%), becoming the second most influential variable (Fig. 4).

Response curves

In Workflow 1, unaltered models accurately recovered the true mean response except for species with wide niches and 20 occurrences, where models failed to recover the response mean (Fig. 6, Supporting information). The estimated standard deviation, however, varied considerably across different scenarios. Where sample size was lowest (20 occurrences), models overestimated the standard deviation and thus tended to overpredict the probability of suitable habitat. The standard deviation estimation was significantly improved with increasing sample size and was, on average, better for scenarios with higher levels of SAC (homogenous variables). This pattern was independent of species niche width (Fig. 6, Supporting information).

Therefore, where there was positional error and small sample sizes (20 occurrences), models were unable to accurately estimate either the response mean or standard deviation

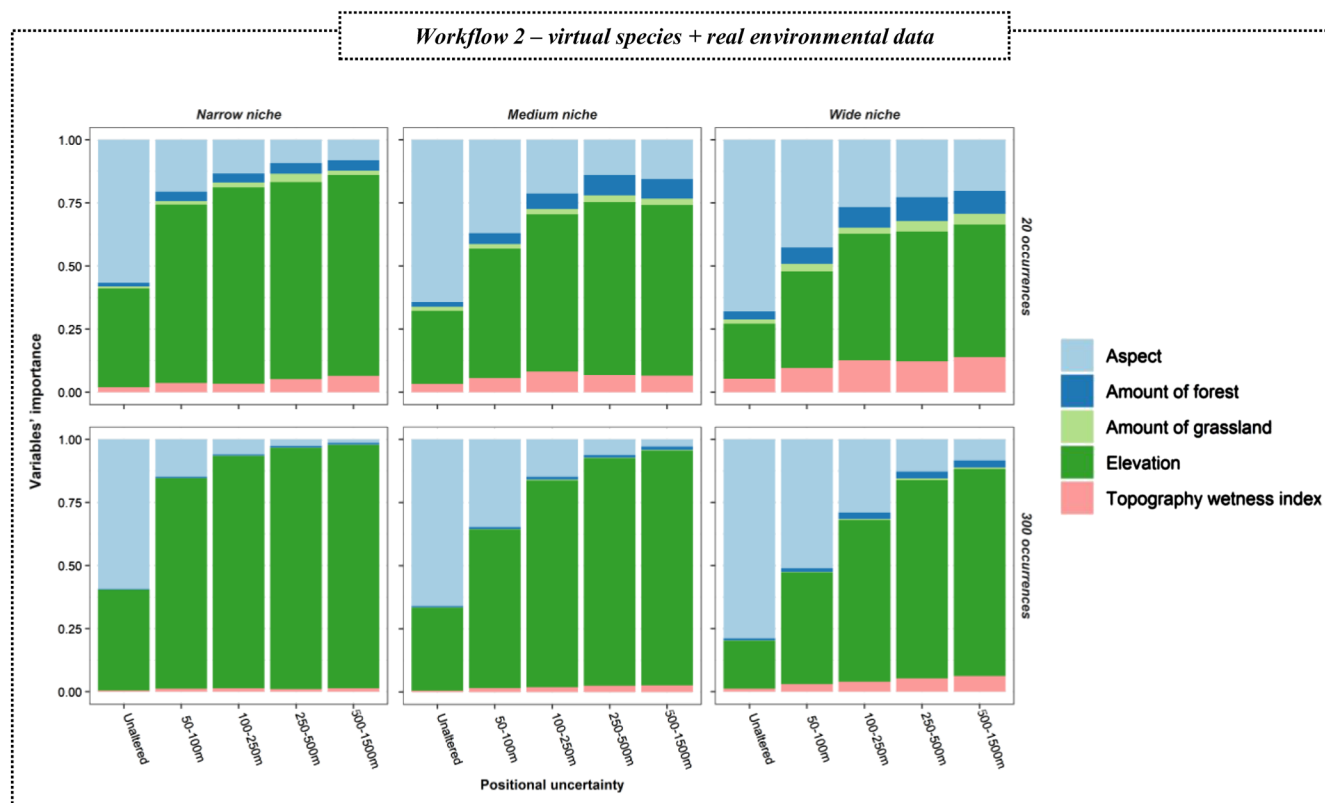


Figure 5. Comparison of the change in variables' importance between models generated with positionally accurate presences (Unaltered) and models built with various positional error in the data across various sample sizes. Values represent averages of the 50 repetitions. See Supporting information for plots of variables' importance for all scenarios.

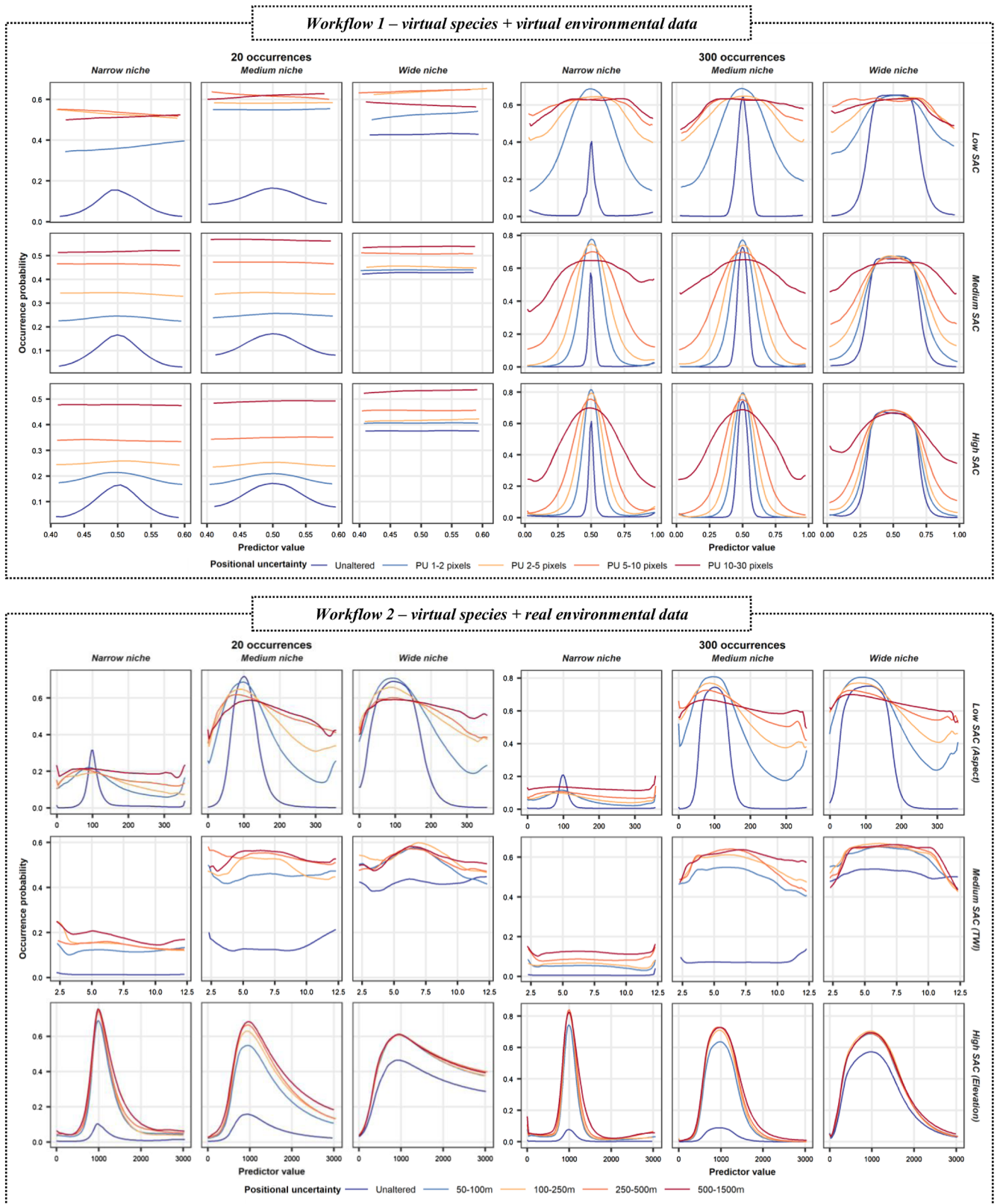


Figure 6. Variation of selected environmental response curves across models with unaltered and altered data, and various sample sizes for Workflows 1 and 2. See Supporting information for plots of resulting response curves for all scenarios. Response curves represent averages of the 50 repetitions.

across all SAC types. This ability was improved with increasing sample size and level of SAC (Fig. 6, Supporting information). Across all species niche widths, where positional error was pronounced, models were better able to estimate response means and standard deviations when sample sizes were large and SAC levels high. However, even with the largest sample size (1000 occurrences) and the highest level of SAC, the models overestimated the standard deviation (Fig. 6, Supporting information).

For Workflow 2, the unaltered models were able to recover responses to both aspect and elevation, which were used to generate virtual species. This was independent of the modeled scenario (Fig. 6, Supporting information). When introducing positional error, models could still capture the approximate response to the elevation with a relatively high SAC level. However, although models recovered the correct shape of the response curve, standard deviation increased. In contrast, for aspect (low SAC level), the models developed with positional uncertainty failed to recover the correct response curve, even when larger sample sizes were used (Fig. 6, Supporting information). Note that Workflow 2 models could estimate the response even with the smallest sample size. This is in contrast to Workflow 1, potentially due to greater model complexity. These results support our assumptions that models developed with data containing positional uncertainty might be able to detect species responses for variables with high SAC levels, but fail to detect meaningful responses for variables with low SAC levels.

The positional error also affected response curves for the band-tailed pigeon. The largest changes to response curves and over- and underprediction tended to occur with the most heterogeneous variables, for example the terrain ruggedness index (Fig. 4).

Discussion

In our study, we used combinations of virtual and real environmental variables with different SAC levels, sets of virtual species with a variety of niche widths, and one real species to explore how positional error in species occurrence data can affect model performance and its ecological interpretability. Specifically, we investigated the ability of SDMs to appropriately detect species' responses to the environment and variable importance using various scenarios with artificially applied positional error.

Our results show that positional uncertainty in species occurrences leads to a decrease in model predictive performance across all combinations of species niche widths, sample sizes, and SAC levels of the environmental variables, but that the magnitude of the negative impact of positional uncertainty varied for different combinations and depending on the distance that points were shifted. The negative influence was most pronounced for species with a narrow niche and scenarios with more heterogeneous environmental variables. This is consistent with previous studies, which concluded that more accurate occurrence data

generally yielded better-performing SDMs (Visscher 2006, Johnson and Gillingham 2008, Osborne and Leitao 2009, Tulowiecki et al. 2015, Mitchell et al. 2017, Soutan and Safi 2017, Fernandes et al. 2019). It is important to highlight that the magnitude of the negative effect of positional uncertainty varied across prior studies. This can be explained by using environmental variables with different heterogeneity (Naimi et al. 2011, 2014) or by using species with varying niche width (Gábor et al. 2020).

Our models for real species were less affected by positional uncertainty than models for virtual species. This could possibly be explained by the spatial error already embedded in the real species data, meaning the 'unaltered' scenario actually presents some minimal (but unknown) level of error, as well as errors in the environmental layers. Additionally, observations of transient individuals merely passing through unsuitable habitats could contribute to this finding, although we attempted to filter out such cases. There may also be spatial and/or environmental biases in the data, such as disproportionate sampling efforts in locations where specific behaviors take place (e.g. water sources), where species detectability is increased (e.g. open areas), or in areas with greater accessibility (e.g. near roads, walking trails; Kramer-Schadt et al. 2013, Fourcade et al. 2014).

On the other hand, our results showed that models built with even positionally inaccurate data may still be ecologically interpretable. The absolute and relative importance of environmental predictors and the shape of species–environmental relationships co-varied with the level of positional uncertainty. But these differences were much weaker than those observed for overall model performance. This indicates that low model performance does not necessarily lead to low capacity to infer which variables drive species distributions and the strength of those drivers. It is important to note that sample size and the SAC level of environmental variables play an important role here. In general, the higher the sample size and the higher the level of SAC in environmental variables the better were models able to recover response curves and detect the importance of the environmental variables (Fig. 4–6).

On the other hand, in the case of environmental variables with low SAC level (high heterogeneity), positional error obscured the main patterns (e.g. aspect in Workflow 2 or terrain ruggedness index in Workflow 3; Fig. 4, 6). Our results suggest that, at least for the example species, positionally inaccurate records may still prove useful for assessing the relative importance of environmental variables in generating species distributions and for determination of the shapes of species responses. Thus, for some purposes, positionally inaccurate records need not be discarded (as is common practice; Watcharamongkol et al. 2018, Gueta and Carmel 2016). This finding is particularly fortuitous because discarding positionally uncertain occurrence data can limit our ability to estimate range sizes and overestimates exposure to climate change (Smith et al. 2023).

Drawing methodological conclusions based on real data is difficult since the true underlying population distribution is unknown, as are data deficiencies that could potentially affect

results (Winner et al. 2018, Meynard et al. 2019, Šímová et al. 2019, Vollerling et al. 2019, Mendes et al. 2020, Somveille et al. 2020, Yanco et al. 2020, Grimmett et al. 2021, Inman et al. 2021, Jiménez-Valverde 2021). On the other hand, simulated datasets simplify the real world, and their results should be interpreted cautiously (Wunder et al. 2008, Zurell et al. 2010, Meynard et al. 2019). Indeed, our results show that a virtual species approach may show different results than those using a real species. For example, our virtual species simulations showed a rapid decrease in model performance with increasing positional error, whereas the band-tailed pigeon showed only a slight decrease in model performance. We strongly recommend that future studies should follow a growing trend and combine simulations and real species data when studying methodological questions (Fithian et al. 2015, Guélat and Kéry 2018, Mertes and Jetz 2018, Renner et al. 2019).

Although this study provides extensive insights that are optimistic about the potential utility of SDMs, caution is warranted in generalizing these results, and further research is needed. For example, future studies could explore whether our findings are robust to different MaxEnt settings, various modeling techniques, response, and analysis grain and different types of data uncertainty (e.g. spatial bias rather than a random error). In addition, within global aggregation databases, spatial uncertainty may not be uniformly distributed. Analyses that characterize the patterns of spatial uncertainty within these databases would allow researchers to identify situations wherein models are likely to fail.

Acknowledgements – The authors greatly appreciate the contribution of the subject editor and both reviewers. The first author would also like to recognize the never-ending support from his former PhD supervisor Vítězslav Moudrý (co-author) during the often challenging times I created when I was a PhD student. If you are reading this and thinking about a PhD, you should definitely contact him. It's hard to find a better mentor and friend.

Funding – This research was funded by the Technological grant agency of the Czech Republic (grant no. SS02030018 DivLand) and by OP RDE Improvement in Quality of the Internal Grant Scheme at CZU, reg. no. CZ.02.2.69/0.0/0.0/19_073/0016944 (grant no. 43/2021). In addition, this paper was made possible by generous support from the Fulbright-Masaryk program sponsored by US and Czech governments, which provided Lukáš Gábor with the opportunity to conduct research at Yale University.

Author contributions

Lukáš Gábor: Conceptualization (equal); Data curation-Supporting, Formal analysis (lead); Funding acquisition (lead); Methodology (lead); Software (lead); Supervision (equal); Validation (lead); Visualization (equal); Writing – original draft (lead); Writing – review and editing (equal). **Walter Jetz:** Conceptualization (equal); Formal analysis (equal); Methodology (lead); Supervision-Supporting, Visualization (equal); Writing – review and editing (equal). **Alejandra Zarzo-Arias:** Conceptualization (equal); Data curation (lead); Formal analysis-Supporting, Investigation

(equal); Methodology-Supporting, Software-Supporting, Visualization (equal); Writing – original draft (equal); Writing – review and editing (equal). **Kevin Winner:** Methodology (equal); Visualization (equal); Writing – review and editing (equal). **Scott Yanco:** Methodology-Supporting, Visualization (equal); Writing – original draft (equal); Writing – review and editing (equal). **Stefan Pinkert:** Validation-Supporting, Visualization (equal); Writing – original draft-Supporting, Writing – review and editing (equal). **Charles J. Marsh:** Conceptualization (equal); Data curation (lead); Visualization (equal); Writing – original draft-Supporting, Writing – review and editing (equal). **Matthew S. Rogan:** Formal analysis-Supporting, Methodology (lead); Writing – review and editing (equal). **Jussi Mäkinen:** Visualization (equal); Writing – original draft-Supporting, Writing – review and editing (equal). **Duccio Rocchini:** Conceptualization-Supporting, Methodology-Supporting, Writing – original draft-Supporting, Writing – review and editing (equal). **Vojtěch Barták:** Formal analysis-Supporting, Visualization (equal); Writing – original draft-Supporting. **Marco Malavasi:** Conceptualization-Supporting, Methodology-Supporting, Writing – review and editing (equal). **Petr Balej:** Investigation-Supporting, Visualization-Supporting, Writing – review and editing (equal). **Vítězslav Moudrý:** Conceptualization (lead); Formal analysis-Supporting, Funding acquisition-Supporting, Methodology (lead); Supervision (lead); Visualization-Supporting, Writing – original draft (lead); Writing – review and editing (lead).

Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06358>.

Data availability statement

Links to used R codes and step-by-step guidelines are available from the Zenodo Digital Repository: <https://zenodo.org/record/7650826> (Gábor et al. 2023).

Supporting information

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

References

- Allouche, O., Tsoar, A. and Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). – *J. Appl. Ecol.* 43: 223–232.
- Amatulli, G., Domisch, S., Tuanmu, M. N., Parmentier, B., Rani-peta, A., Malczyk, J. and Jetz, W. 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. – *Sci. Data* 5: 1–15.
- Ancillotto, L., Bosso, L., Smeraldo, S., Mori, E., Mazza, G., Herkt, M., Galimberti, A., Ramazzotti, F. and Russo, D. 2020. An

- African bat in Europe, *Plecotus gaisleri*: biogeographic and ecological insights from molecular taxonomy and species distribution models. – *Ecol. Evol.* 10: 5785–5800.
- 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.
- Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. – *Ecol. Modell.* 200: 1–19.
- Austin, M. P., Belbin, L., Meyers, J. A. A., Doherty, M. D. and Luoto, M. 2006. Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. – *Ecol. Modell.* 199: 197–216.
- Barbosa, M. A., Real, R., Muñoz, A. R. and Brown, J. A. 2013. New measures for assessing model equilibrium and prediction mismatch in species distribution models. – *Diversity 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.
- Bazzichetto, M., Malavasi, M., Barták, V., Acosta, A. T. R., Moudrý, V. and Carranza, M. L. 2018. Modeling plant invasion on Mediterranean coastal landscapes: an integrative approach using remotely sensed data. – *Landsc. Urban Plan.* 171: 98–106.
- Belsley, D. A. 1991. A guide to using the collinearity diagnostics. – *Comput. Sci. Econ. Manage.* 4: 33–50.
- Boral, D. and Moktan, S. 2021. Predictive distribution modeling of *Swertia bimaculata* in Darjeeling-Sikkim eastern Himalaya using MaxEnt: current and future scenarios. – *Ecol. Processes.* 10: 1–16.
- Bradie, J. and Leung, B. 2017. A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. – *J. Biogeogr.* 44: 1344–1361.
- Chatfield, C. 1995. Model uncertainty, data mining and statistical inference. – *J. R. Stat. Soc. Ser. A* 158: 419–444.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhrner, J. 2015. System for automated geoscientific analyses (SAGA) ver. 2.1.4. – *Geosci. Model Dev.* 8: 1991–2007.
- Coops, N. C., Duffe, J. and Koot, C. 2010. Assessing the utility of lidar remote sensing technology to identify mule deer winter habitat. – *Can. J. Remote Sens.* 36: 81–88.
- Didan, K., Munoz, A. B., Solano, R. and Huete, A. 2015. MODIS vegetation index user's guide (MOD13 series). University of Arizona: Vegetation Index and Phenology Lab, 35.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D. and Lautenbach, S. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. – *Ecography* 36: 27–46.
- Dungan, J. L. 2002. Conditional simulation: an alternative to estimation for achieving mapping objectives. – In: Stein, A., van der Meer, F. and Gorte, B. (eds), *Spatial statistics for remote sensing*. – Springer, pp. 135–152.
- Dvorský, M., Macek, M., Kopecký, M., Wild, J. and Doležal, J. 2017. Niche asymmetry of vascular plants increases with elevation. – *J. Biogeogr.* 44: 1418–1425.
- Eduardo, A. A., Liparini, A., Martinez, P. A., Gouveia, S. F. and Riul, P. 2022. Assessing multitemporal calibration for species distribution models. – *Ecol. Inform.* 71: 101787.
- El-Gabbas, A., Van Opzeeland, I., Burkhardt, E. and Boebel, O. 2021. Static species distribution models in the marine realm: the case of baleen whales in the Southern Ocean. – *Diversity Distrib.* 27: 1536–1552.
- Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – *Annu. Rev. Ecol. Evol. Syst.* 40: 677–697.
- Elith, J., Ferrier, S., Huettmann, F. and Leathwick, J. 2005. The evaluation strip: a new and robust method for plotting predicted responses from species distribution models. – *Ecol. Modell.* 186: 280–289.
- Elith, J., Kearney, M. and Phillips, S. 2010. The art of modelling range-shifting species. – *Methods Ecol. Evol.* 1: 330–342.
- Ellis-Soto, D., Merow, C., Amatulli, G., Parra, J. L. and Jetz, W. 2021. Continental-scale 1 km hummingbird diversity derived from fusing point records with lateral and elevational expert information. – *Ecography* 44: 640–652.
- Fern, R. R., Morrison, M. L., Grant, W. E., Wang, H. and Campbell, T. A. 2020. Modeling the influence of livestock grazing pressure on grassland bird distributions. – *Ecol. Processes* 9: 1–11.
- Fernandes, R. F., Scherrer, D. and Guisan, A. 2019. Effects of simulated observation errors on the performance of species distribution models. – *Diversity Distrib.* 25: 400–413.
- Fernandez, M., Blum, S., Reichle, S., Guo, Q., Holzman, B. and Hamilton, H. 2009. Locality uncertainty and the differential performance of four common niche-based modeling techniques. – *Biodivers. Inform.* 6: 36–52.
- Ferrier, S., Jetz, W. and Scharlemann, J. 2017. Biodiversity modelling as part of an observation system. – In: Walter, M. and Scholes, R. J. (eds), *The GEO Handbook on Biodiversity Observation Networks*. Springer, pp. 239–257.
- Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – *Environ. Conserv.* 24: 38–49.
- Fithian, W., Elith, J., Hastie, T. and Keith, D. A. 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. – *Methods Ecol. Evol.* 6: 424–438.
- Fourcade, Y., Engler, J. O., Rödder, D. and Secondi, J. 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. – *PLoS One* 9: e97122.
- Franklin, J. 2010. *Mapping species distributions: spatial inference and prediction*. – Cambridge Univ. Press.
- Gábor, L., Moudrý, V., Lecours, V., Malavasi, M., Barták, V., Fogl, M., Šimová, P., Rocchini, D. and Václavík, T. 2020. The effect of positional error on fine scale species distribution models increases for specialist species. – *Ecography* 43: 256–269.
- Gábor, L., Jetz, W., Lu, M., Rocchini, D., Cord, A., Malavasi, M., Zarzo-Arias, A., Barták, V. and Moudrý, V. 2022. Positional errors in species distribution modelling are not overcome by the coarser grains of analysis. – *Methods Ecol. Evol.* 13: 2289–2302.
- Gábor, L., Jetz, W., Zarzo-Arias, A., Winner, K., Yanco, S., Pinkert, S., Marsh, C. J., Rogan, M. S., Mäkinen, J., Rocchini, D., Barták, V., Malavasi, M., Balej, P. and Moudrý, V. 2023. Data from: Species distribution models affected by positional uncertainty in species occurrences can still be ecologically interpreted

- able. – Zenodo Digital Repository, <https://zenodo.org/record/7650826>.
- Graham, C. H., Elith, J., Hijmans, R. J., Guisan, A., Townsend Peterson, A., Loiselle, B. A. and NCEAS Predicting Species Distributions Working Group. 2008. The influence of spatial errors in species occurrence data used in distribution models. – *J. Appl. Ecol.* 45: 239–247.
- Graham, L. J., Spake, R., Gillings, S., Watts, K. and Eigenbrod, F. 2019. Incorporating fine-scale environmental heterogeneity into broad-extent models. – *Methods Ecol. Evol.* 10: 767–778.
- Grimmett, L., Whitsed, R. and Horta, A. 2021. Creating virtual species to test species distribution models: the importance of landscape structure, dispersal and population processes. – *Ecography* 44: 753–765.
- Guélat, J. and Kéry, M. 2018. Effects of spatial autocorrelation and imperfect detection on species distribution models. – *Methods Ecol. Evol.* 9: 1614–1625.
- Gueta, T. and Carmel, Y. 2016. Quantifying the value of user-level data cleaning for big data: a case study using mammal distribution models. – *Ecol. Inform.* 34: 139–145.
- Guisan, A. et al. 2013. Predicting species distributions for conservation decisions. *Ecol. Lett.* 16: 1424–1435.
- Hargreaves, A. L., Samis, K. E. and Eckert, C. G. 2014. Are species' range limits simply niche limits writ large? A review of transplant experiments beyond the range. – *Am. Nat.* 183: 157–173.
- Hefley, T. J., Baasch, D. M., Tyre, A. J. and Blankenship, E. E. 2014. Correction of location errors for presence-only species distribution models. – *Methods Ecol. Evol.* 5: 207–214.
- Holder, A. M., Markarian, A., Doyle, J. M. and Olson, J. R. 2020. Predicting geographic distributions of fishes in remote stream networks using maximum entropy modeling and landscape characterizations. – *Ecol. Modell.* 433: 109231.
- Inman, R., Franklin, J., Esque, T. and Nussear, K. 2021. Comparing sample bias correction methods for species distribution modeling using virtual species. – *Ecosphere* 12: e03422.
- Isaac, N. J., 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.
- Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. – *Global Ecol. Biogeogr.* 21: 498–507.
- Jiménez-Valverde, A., Aragón, P. and Lobo, J. M. 2021. Deconstructing the abundance–suitability relationship in species distribution modelling. – *Global Ecol. Biogeogr.* 30: 327–338.
- Johnson, C. J. and Gillingham, M. P. 2008. Sensitivity of species-distribution models to error, bias, and model design: an application to resource selection functions for woodland caribou. – *Ecol. Modell.* 213: 143–155.
- Karger, D. N., Conrad, O., Böhrner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Wilber, R., Zimmermann, N. E., Linder, H. P. and Kessler, M. 2017. Climatologies at high resolution for the earth's land surface areas. – *Sci. Data* 4: 1–20.
- Keppie, D. M. and Braun, C. E. 2000. Band-tailed pigeon. – *American Ornithologists' Union*.
- Kindt, R. 2018. Ensemble species distribution modelling with transformed suitability values. – *Environ. Modell. Softw.* 100: 136–145.
- Kramer-Schadt, S. et al. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity Distrib.* 1911: 1366–1379.
- Lecours, V., Gábor, L., Edinger, E. and Devillers, R. 2020. Fine-scale habitat characterization of The Gully, the Flemish Cap, and the Orphan Knoll, Northwest Atlantic, with a focus on cold-water corals. – In: Harris, P. T. and Baker, E. (eds), *Seafloor geomorphology as benthic habitat*. Elsevier, pp. 735–751.
- Lee-Yaw, J. A., Kharouba, H. M., Bontrager, M., Mahony, C., Csergő, A. M., Noreen, A. M., Li, Q., Schuster, R. and Angert, A. L. 2016. A synthesis of transplant experiments and ecological niche models suggests that range limits are often niche limits. – *Ecol. Lett.* 19: 710–722.
- Leroy, B., Delsol, R., Hugué, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M. and Bellard, C. 2018. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. – *J. Biogeogr.* 45: 1994–2002.
- Li, Q. and Kou, X. 2021. WiBB: an integrated method for quantifying the relative importance of predictive variables. – *Ecography* 44: 1557–1567.
- Li, W. and Guo, Q. 2013. How to assess the prediction accuracy of species presence–absence models without absence data? – *Ecography* 36: 788–799.
- Lobo, J. M., Jiménez-Valverde, A. and Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. – *Global Ecol. Biogeogr.* 17: 145–151.
- Mendes, P., Velazco, S. J. E., de Andrade, A. F. A. and Júnior, P. D. M. 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. and Silander Jr, J. A. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. – *Ecography* 36: 1058–1069.
- Mertes, K. and Jetz, W. 2018. Disentangling scale dependencies in species environmental niches and distributions. – *Ecography* 41: 1604–1615.
- Meynard, C. N. and Kaplan, D. M. 2012. The effect of a gradual response to the environment on species distribution modeling performance. – *Ecography* 35: 499–509.
- Meynard, C. N. and Kaplan, D. M. 2013. Using virtual species to study species distributions and model performance. – *J. Biogeogr.* 40: 1–8.
- Meynard, C. N., Leroy, B. and Kaplan, D. M. 2019. Testing methods in species distribution modelling using virtual species: what have we learnt and what are we missing? – *Ecography* 42: 2021–2036.
- Mitchell, P. J., Monk, J. and Laurenson, L. 2017. Sensitivity of fine-scale species distribution models to locational uncertainty in occurrence data across multiple sample sizes. – *Methods Ecol. Evol.* 8: 12–21.
- Moudrý, V. 2015. Modelling species distributions with simulated virtual species. – *J. Biogeogr.* 42: 1365–1366.
- Moudrý, V. and Devillers, R. 2020. Quality and usability challenges of global marine biodiversity databases: an example for marine mammal data. – *Ecol. Inform.* 56: 101051.
- Moudrý, V. and Šimová, P. 2012. Influence of positional accuracy, sample size and scale on modelling species distributions: a review. – *Int. J. Geogr. Inf. Sci.* 26: 2083–2095.
- Moudrý, V. and Šimová, P. 2013. Relative importance of climate, topography, and habitats for breeding wetland birds with different latitudinal distributions in the Czech Republic. – *Appl. Geogr.* 44: 165–171.

- Moudrý, V., Komárek, J. and Šímová, P. 2017. Which breeding bird categories should we use in models of species distribution? – *Ecol. Indic.* 74: 526–529.
- Moudrý, V., Lecours, V., Malavasi, M., Misiuk, B., Gábor, L., Gdulová, K., Šímová, P. and Wild, J. 2019. Potential pitfalls in rescaling digital terrain model-derived attributes for ecological studies. – *Ecol. Inform.* 54: 100987.
- Moudrý, V., Keil, P., Cord, A. F., Gábor, L., Lecours, V., Zarzarias, A., Barták, B., Malavasi, M., Rocchini, D., Torresani, M., Gdulová, K., Grattarola, F., Leroy, F., Marchetto, E., Thouverai, E., Prošek, P., Wild, J. and Šímová, P. 2023. Scale mismatches between predictor and response variables in species distribution modelling: a review of practices for appropriate grain selection. – *Progr. Phys. Geogr.* 03091333231156362.
- Müller, J. and Brandl, R. 2009. Assessing biodiversity by remote sensing in mountainous terrain: the potential of LiDAR to predict forest beetle assemblages. – *J. Appl. Ecol.* 46: 897–905.
- Murray, K. and Conner, M. M. 2009. Methods to quantify variable importance: implications for the analysis of noisy ecological data. – *Ecology* 90: 348–355.
- Naimi, B. and Araújo, M. B. 2016. sdm: a reproducible and extensible R platform for species distribution modelling. – *Ecography* 39: 368–375.
- Naimi, B., Skidmore, A. K., Groen, T. A. and Hamm, N. A. 2011. Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. – *J. Biogeogr.* 38: 1497–1509.
- Naimi, B., Hamm, N. A., 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.
- Osborne, P. E. and Leitão, P. J. 2009. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. – *Diversity Distrib.* 15: 671–681.
- Petroselli, A., Vessella, F., Cavagnuolo, L., Piovesan, G. and Schirone, B. 2013. Ecological behavior of *Quercus suber* and *Quercus ilex* inferred by topographic wetness index (TWI). – *Trends Ecol. Evol.* 27: 1201–1215.
- Phillips, S. J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. – *Ecography* 31: 161–175.
- Phillips, S. J., Anderson, R. P. and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Modell.* 190: 231–259.
- Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G., Kempen, B., Ribeiro, E. and Rossiter, D. 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. – *Soil* 7: 217–240.
- Reif, J., Reifová, R., Skoracka, A. and Kuczyński, L. 2018. Competition-driven niche segregation on a landscape scale: evidence for escaping from syntopy towards allotopy in two coexisting sibling passerine species. – *J. Anim. Ecol.* 87: 774–789.
- Renner, I. W., Louvrier, J. and Gimenez, O. 2019. Combining multiple data sources in species distribution models while accounting for spatial dependence and overfitting with combined penalized likelihood maximization. – *Methods Ecol. Evol.* 10: 2118–2128.
- Rodríguez, L., García, J. J., Carreño, F. and Martínez, B. 2019. Integration of physiological knowledge into hybrid species distribution modelling to improve forecast of distributional shifts of tropical corals. – *Diversity Distrib.* 25: 715–728.
- Sanguet, A., Wyler, N., Petitpierre, B., Honeck, E., Poussin, C., Martin, P. and Lehmann, A. 2022. Beyond topo-climatic predictors: does habitats distribution and remote sensing information improve predictions of species distribution models? – *Global Ecol. Conserv.* 39: e02286.
- Santamarina, S., Alfaro-Saiz, E., Llamas, F. and Acedo, C. 2019. Different approaches to assess the local invasion risk on a threatened species: opportunities of using high-resolution species distribution models by selecting the optimal model complexity. – *Global Ecol. Conserv.* 20: e00767.
- Santini, L., Benítez-López, A., Maiorano, L., Čengić, M. and Huijbregts, M. A. 2021. Assessing the reliability of species distribution projections in climate change research. – *Diversity Distrib.* 27: 1035–1050.
- Sauer, J. R., Niven, D. K., Hines, J. E., Ziolkowski, Jr, D. J., Pardieck, K. L., Fallon, J. E., and Link, W. A. 2017. The North American breeding bird survey, results and analysis 1966 – 2015. Version 12.23.2015. USGS Patuxent Wildlife Research Center, Laurel, MD.
- Segal, R. D., Massaro, M., Carlile, N. and Whitsed, R. 2021. Small-scale species distribution model identifies restricted breeding habitat for an endemic island bird. – *Anim. Conserv.* 24: 959–969.
- Šímová, P., Moudrý, V., Komárek, J., Hrach, K. and Fortin, M. J. 2019. Fine scale waterbody data improve prediction of waterbird occurrence despite coarse species data. – *Ecography* 42: 511–520.
- 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.
- Somveille, M., Wikelski, M., Beyer, R. M., Rodrigues, A. S., Manica, A. and Jetz, W. 2020. Simulation-based reconstruction of global bird migration over the past 50,000 years. – *Nat. Commun.* 11: 1–9.
- Soultan, A. and Safi, K. 2017. The interplay of various sources of noise on reliability of species distribution models hinges on ecological specialisation. – *PLoS One* 12: e0187906.
- Tessarolo, G., Ladle, R. J., Lobo, J. M., Rangel, T. F. and Hortal, J. 2021. Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. – *Ecography* 44: 1743–1755.
- Title, P. O. and Bemmels, J. B. 2018. ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. – *Ecography* 41: 291–307.
- Tuanmu, M. N. and Jetz, W. 2015. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. – *Global Ecol. Biogeogr.* 24: 1329–1339.
- Tulowiecki, S. J., Larsen, C. P. and Wang, Y. C. 2015. Effects of positional error on modeling species distributions: a perspective using presettlement land survey records. – *Plant Ecol.* 216: 67–85.
- VanDerWal, J., Shoo, L. P., Graham, C. and Williams, S. E. 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? – *Ecol. Modell.* 220: 589–594.

- Varela, S., Anderson, R. P., García-Valdés, R. and Fernández-González, F. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. – *Ecography* 37: 1084–1091.
- Venne, S. and Currie, D. J. 2021. Can habitat suitability estimated from MaxEnt predict colonizations and extinctions? – *Diversity Distrib.* 27: 873–886.
- Vierling, K. T., Bässler, C., Brandl, R., Vierling, L. A., Weiß, I. and Müller, J. 2011. Spinning a laser web: predicting spider distributions using LiDAR. – *Ecol. Appl.* 21: 577–588.
- Visscher, R. D. 2006. GPS measurement error and resource selection functions in a fragmented landscape. – *Ecography* 29: 458–464.
- Vogeler, J. C., Hudak, A. T., Vierling, L. A., Evans, J., Green, P. and Vierling, K. T. 2014. Terrain and vegetation structural influences on local avian species richness in two mixed-conifer forests. – *Remote Sens. Environ.* 147: 13–22.
- Volis, S. and Tojibaev, K. 2021. Defining critical habitat for plant species with poor occurrence knowledge and identification of critical habitat networks. – *Biodivers. Conserv.* 30: 3603–3611.
- Vollering, J., Halvorsen, R., Auestad, I. and Rydgren, K. 2019. Bunching up the background betters bias in species distribution models. – *Ecography* 42: 1717–1727.
- Watcharamongkol, T., Christin, P. A. and Osborne, C. P. 2018. C4 photosynthesis evolved in warm climates but promoted migration to cooler ones. – *Ecol. Lett.* 21: 376–383.
- Wilson, A. M. and Jetz, W. 2016. Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. – *PLoS Biol.* 14: e1002415.
- Winner, K., Noonan, M. J., Fleming, C. H., Olson, K. A., Mueller, T., Sheldon, D. and Calabrese, J. M. 2018. Statistical inference for home range overlap. – *Methods Ecol. Evol.* 9: 1679–1691.
- Work, T. T., Onge, B. S. and Jacobs, J. M. 2011. Response of female beetles to LIDAR derived topographic variables in Eastern boreal mixedwood forests (Coleoptera, Carabidae). *ZooKeys* 147: 623.
- Wunder, J., Reineking, B., Bigler, C. and Bugmann, H. 2008. Predicting tree mortality from growth data: how virtual ecologists can help real ecologists. – *J. Ecol.* 96: 174–187.
- Yanco, S. W., McDevitt, A., Trueman, C. N., Hartley, L. and Wunder, M. B. 2020. A modern method of multiple working hypotheses to improve inference in ecology. – *R. Soc. Open Sci.* 7: 200231.
- Zarzo-Arias, A., Penteriani, V., Gábor, L., Šímová, P., Grattarola, F. and Moudrý, V. 2022. Importance of data selection and filtering in species distribution models: a case study on the Cantabrian brown bear. – *Ecosphere* 13: e4284.
- Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B. and Grimm, V. 2010. The virtual ecologist approach: simulating data and observers. – *Oikos* 119: 622–635.
- Zurell, D. and Engler, J. O. 2019. Ecological niche modeling. – In: Dunn, P. O. and Møller, A. P. (eds), *Effects of climate change on birds*. Oxford University Press.
- Zuur, A. F., Ieno, E. N. and Elphick, C. S. 2010. A protocol for data exploration to avoid common statistical problems. – *Methods Ecol. Evol.* 1: 3–14.