Comparing sample bias correction methods for species distribution modeling using virtual species

RICHARD INMAN1,2†, JANET FRANKLIN1,3, TODD ESQUE1,2, AND KENNETH NUSSEAR1,4

1School of Geographical Sciences and Urban Planning, Arizona State University, PO Box 875302, Tempe, Arizona 85287 USA
2U.S. Geological Survey, Western Ecological Research Center, Henderson Field Station, 160 N. Stephanie Street, Henderson, Nevada 89074 USA
3Department of Botany and Plant Sciences, University of California–Riverside, 900 University Ave, Riverside, California 92521 USA
4Department of Geography, University of Nevada–Reno, 1664 N. Virginia Street, Reno, Nevada 89557 USA

Citation: Inman, R., J. Franklin, T. Esque, and K. Nussear. 2021. Comparing sample bias correction methods for species distribution modeling using virtual species. Ecosphere 12(3):e03422. 10.1002/ecs2.3422

Abstract. A key assumption in species distribution modeling (SDM) with presence-background (PB) methods is that sampling of occurrence localities is unbiased and that any sampling bias is proportional to the background distribution of environmental covariates. This assumption is rarely met when SDM practitioners rely on federated museum records from natural history collections for geo-located occurrences due to inherent sampling bias found in these collections. We use a simulation approach to explore the effectiveness of three methods developed to account for sampling bias in SDM with PB frameworks. Two of the methods rely on careful filtering of observation data—geographic thinning (G-Filter) and environmental thinning (E-Filter)—while a third, FactorBiasOut, creates selection weights for background data to bias locations toward areas where the observation dataset was sampled. While these methods have been assessed previously, evaluation has emphasized spatial predictions of habitat potential. Here, we dig deeper into the effectiveness of these methods by exploring how sampling bias not only affects predictions of habitat potential, but also our understanding of niche characteristics such as which explanatory variables and response curves best represent species–environment relationships. We simulate 100 virtual species ranging from generalist to specialist in their habitat preferences and introduce geographic and environmental bias at three intensity levels to measure the effectiveness of each correction method to (1) predict true probability of occurrence across a study area, (2) recover true species–environment relationships, and (3) identify true explanatory variables. We find that the FactorBiasOut most often showed the greatest improvement in recreating known distributions but did no better at correctly identifying environmental covariates or recreating species–environment relationships than G-Filter or E-Filter methods. Narrow niche species are most problematic for biased calibration datasets, such that correction methods can, in some cases, make predictions worse.

Key words: background weights; environmental space; FactorBiasOut; geographic space; MaxEnt; sampling bias; spatial filtering; spatial thinning; species distribution modeling; virtual species.

Received 5 May 2020; revised 6 November 2020; accepted 16 November 2020; final version received 12 January 2021.
Corresponding Editor: Debra P. C. Peters.
Copyright: © 2021 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
† E-mail: rdinman@asu.edu

INTRODUCTION

Understanding the distributions of organisms and their habitat is a fundamental question in conservation ecology and is needed for implementing landscape conservation plans. When implementing such plans, resource managers are often tasked with resource inventory or
monitoring. This has become essential for land managers who must respond to dramatic changes in species composition due to changing climates (Travis 2003, Rosenzweig et al. 2008, Butchart et al. 2010, Rosenberg et al. 2019). In addition to local declines and extirpations caused by changing climates, surface-disturbing land uses (e.g., urbanization, transportation corridors, military training, agriculture, recreational activities, and energy development) are altering vast areas of terrestrial landscapes (Leu et al. 2008), fragmenting habitat (Opdam and Wascher 2004, Walston et al. 2016), and disrupting landscape linkages (Andrews 1990, Schneider 2009). Approaches that can help quantify changes in distributions of habitat from these compounding stressors are useful to land managers responsible for managing landscapes when sustainable environmental, social, and economic outcomes are desired (Freeman et al. 2015).

One such methodology to quantify species distributions is species distribution modeling (SDM), also referred to as ecological niche modeling (ENM), which is being used with increasing frequency in land-use planning and conservation (Schwartz 2012, Guisan et al. 2013). SDM is a quantitative modeling approach that relates locations of species occurrences to environmental covariates hypothesized to influence or define habitat potential (Franklin 2010). While SDM encompasses different statistical methods and approaches, one method has emerged as especially user-friendly and is now used widely: Maxent (Phillips et al. 2006). Since its publication in 2006, the original paper has been cited 6796 times in the Web of Science (accessed April 2020), with well over 1300 papers using “maxent” and “conservation” as keywords. Due to its user-friendly interface and accessibility, Maxent is arguably the most commonly used method for SDM in conservation and land-use planning. For example, biodiversity assessments for the UN Paris Agreement on climate change used Maxent to estimate distributions for over 80,000 species of five terrestrial taxa—plants, amphibians, reptiles, birds, and mammals (Warren et al. 2013, 2018).

In addition to its ease of use, the Maxent algorithm does not require knowledge of locations where organisms are absent, which can be difficult to obtain. Instead, a presence-background (PB) modeling approach is used that compares environmental conditions at locations where a species has been observed to environmental conditions across a study area (background). Methods such as Maxent using this approach can also leverage the ever-growing wealth of global biodiversity observation information available from online repositories such as the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/), whereas presence-absence data require costly surveys.

While PB methods benefit from availability of data and software, they rely on several assumptions that are not always acknowledged in practice (Elith and Leathwick 2009). One key assumption of PB methods is that sampling effort is uniform across the landscape such that environmental conditions are sampled proportionally to their relative availability (Hastie and Fithian 2013) and that a species’ niche is sampled over the full range of environmental conditions in which it occurs (Phillips et al. 2009). This assumption is rarely met when observations are obtained from any of the growing number of biodiversity repositories providing geo-located species occurrence records because these collections generally represent opportunistic and haphazard sampling efforts (Newbold 2010) that introduce sampling bias to the locality information for thousands of species. Random or stratified-random sampling designs can be used to collect observations in a manner free of sampling bias (Edwards et al. 2006), but the high cost of conducting these surveys often precludes their use and has resulted in the reliance on widely available biodiversity observations from federated repositories such as GBIF for SDM (Newbold 2010).

Two sources of sampling bias are problematic for SDM: (1) incomplete sampling and (2) over-sampling; when unknown, both result in unquantifiable spatial heterogeneity in the sampling intensity across a landscape and its environmental space (Freitag et al. 1998, Bystriakova et al. 2012, Guillera-Arroyo et al. 2015). The former occurs when not all parts of the environmental space where a species can occur (occupied niche space) are sampled, thereby leaving some combinations of environmental conditions under-sampled and results in few observations in areas that are in fact occupied (Hefley et al. 2013). Incomplete sampling may occur when...
certain areas are inaccessible due to administrative reasons such as private property or governmental regulations but may also rise from imperfect or nondetection when occupied areas are sampled, yet no observations are obtained (Kramer-Schadt et al. 2013, Hefl ey et al. 2013). Imperfect detection becomes problematic when knowledge of the true sampling effort is unavailable, thus yielding incomplete sampling. Oversampling occurs when some occupied regions of environmental space are sampled at higher intensities than others (Bean et al. 2012, Gelfand et al. 2012), thereby causing model coefficients to become more representative of areas that have been heavily sampled than of a species’ entire range (Hortal et al. 2008, Hijmans 2012, Guillera-Arroita et al. 2015). Areas proximal to roadways and access routes, for example, may be sampled more heavily than remote areas (Freitag et al. 1998, Kadmon et al. 2004), and administrative designations, such as national parks, wildlife refuges, or other locations of interest, may introduce unknown heterogeneity in sampling intensity by being sampled preferentially.

These two sources of sampling bias can manifest in both geographic and environmental spaces. The latter, environmental sampling bias, occurs when incomplete or over-sampling of the environmental space of the true distribution causes covariance between the sampling probability and the environmental variables used to estimate a species’ distribution (Phillips et al. 2009). Environmental sampling bias can lead to errors in model fitting (Hortal et al. 2008, Yackulic et al. 2012) and is therefore the primary concern for SDM (Guillera-Arroita et al. 2015). However, geographic sampling bias can often cause environmental bias in the observational dataset (Hortal et al. 2008, Bystrriakova et al. 2012). Moreover, it is geographic sampling bias that is readily apparent to practitioners and has been the focus of many studies (Lobo and Tognelli 2014, Aiello-Lammens et al. 2015, Fernández and Nakamura 2015, Soley-Guardia et al. 2016, Qiao et al. 2017). We therefore address these two sources of sampling bias because both can occur in museum records because these opportunistic datasets rarely stem from systematic sampling regimes (Newbold 2010).

We use a simulation approach to explore the effects of sampling bias on SDM in the Maxent PB framework and to compare three bias correction methods that are often used but have not been systematically compared across a robust set of virtual species. Where these methods have been evaluated previously (e.g., Kramer-Schadt et al. 2013, Varela et al. 2014, Fourcade et al. 2014, Stolar and Nielsen 2015, Gábor et al. 2019), emphasis has been on spatial predictions of habitat potential. Here, we dig deeper into their use by exploring how sampling bias not only affects spatial predictions, but also our understanding of fundamental niche characteristics such as which explanatory variables and species–environment relationships best represent a species’ true niche. We simulate 100 virtual species ranging in landscape prevalence and niche breadth from habitat specialists to generalists and introduce geographic and environmental sampling bias at three intensity levels designed to mimic bias levels found in real-world situations to measure the effectiveness of each correction method to (1) predict the true habitat potential across our study area, (2) recreate true species–environment relationships, and (3) identify true explanatory variables.

**Materials and Methods**

**Study area**

We use environmental maps of an arid, interior southwestern region of the United States to define the environmental domain and correlates of realistic, but simulated, species distributions. Our study area covered 918,557 km² and incorporated a broad range of ecosystems ranging from Southern and Central California Chaparral and Oak Woodlands to the southwestern Mojave, Sonoran, and Chihuahua deserts. We chose this area for the availability of a rich set of environmental data (e.g., Inman et al. 2014) describing a broad range of physiographic (e.g., landform type, surface texture, and geologic character) and climatic (e.g., temperature, precipitation norms, and extremes) environmental conditions available to develop distributions for virtual species that could plausibly inhabit this region.

**Virtual species**

We developed 100 virtual species distributions using the package virtualspecies (Leroy et al. 2015) in R 3.3.2 (R Core Team 2020). Each species’ niche was defined using two to five randomly
selected environmental covariates from a suite of ten uncorrelated explanatory covariates spanning physiographic and climatic constraints. These raster layers represented environmental conditions hypothesized to influence the distribution of many actual species in the region (Inman et al. 2014), and were generalized to a spatial resolution of 1 km, a grain size typical of widely available environmental covariates (e.g., Amatulli et al. 2018). We provide environmental covariates and all code used to generate virtual species in Data S1. For simplicity, we set the occupied niches of these species equal to their fundamental niches, precluding the need to simulate predation or competition, and assumed that these virtual species were in equilibrium with their environment and therefore exhibited stable population sizes.

In order to create realistic habitat preferences and distributions, we ensured that species–environment relationships (hereafter “response curves”) could not contradict one another on the landscape (e.g., result in species living simultaneously in the hottest and coldest portions of the study area) using the realistic.sp option, wherein mechanistic response curves were created sequentially (Fig. 1). In the first step, a single environmental covariate was selected, and a corresponding response curve was randomly generated from a linear, quadratic, logistic, or Gaussian function to create a map of habitat potential values based only on that single environmental descriptor. Prior to generating additional response curves, habitat potential values were linearly rescaled between 0 and 1 to identify high habitat potential areas with values >0.5. For each additional explanatory variable, a response curve was randomly generated and used to predict new habitat potential values as the sum of the response curve functions within the high habitat potential areas, and linearly rescaled between 0 and 1 (step 2 in Fig. 1). This process was iterated until each of the two to five randomly selected environmental covariates were included, with the distribution of true habitat potential values for each virtual species set as the habitat potential values resulting from the sum of all response curve functions (step 3 in Fig. 1). We aimed to create a range of specialist and generalist species by modifying the shape parameters for each of the response curve functions to ensure that some species occupied rare environments, while others occurred widely across the study area. Occupied areas were selected with a probabilistic approach rather than thresholding, which has been shown to introduce distortion when attempting to recover response curves (Meynard et al. 2019). Cells that were occupied were selected with the probability:

$$P(i) = \frac{1}{1 + e^{-\frac{x_{ij}}{\theta}}}$$

(1)

where $$x_{ij}$$ is the habitat potential value for cell $$i$$ for $$x$$ explanatory variable (Meynard et al. 2019). We calculated the prevalence of occupied areas for each species, which spanned 0.05% to over 97% of the study area, with mean landscape habitat potential values ranging from 0.13 to 0.94.

**Rarity and niche breadth**

We hypothesized that the breadth of each species’ niche and their landscape prevalence would affect how biased sampling distributions degraded SDM performance. Specifically, we predicted that rare specialists would show erratic responses to differing levels of sampling bias because, while rare species may require fewer samples to adequately estimate their distributions (Franklin et al. 2009), they may also be missed entirely or sampled sparsely at extreme levels of sampling bias. In cases of the latter, we assumed that SDM performance would be severely degraded. We measured rarity by landscape prevalence as the occupied proportion of the study area, and niche breadth with a novel approach quantifying the uniqueness of the environmental conditions defining each species’ geographic distribution. Metrics of niche breadth can range from volumetric measurements of n-dimensional hypervolumes (e.g., Blonder et al. 2014), to reduced principal component axes of occupied regions of environmental space (e.g., Saupe et al. 2015), to counts of unique habitat types presumed to be occupied (e.g., Nürnberg and Aberhan 2013). Our approach to quantifying the environmental uniqueness of occupied habitat draws on the Mahalanobis distances in environmental space because we were interested in how unique the occupied habitat was from the rest of the study area, and we knew the true
environmental variables determining each species’ distribution. The niche breadth value for each species was defined as the median of the squared Mahalanobis distance of all occupied cells:

\[ D^2 = (x_{occ} - \mu_{sa})^\top \Sigma^{-1} (x_{sa} - \mu_{sa}) \]  

where \( x_{occ} \) is the matrix of explanatory variables used to define the species’ niche over all occupied cells, \( x_{sa} \) is the matrix of the same explanatory variables over all cells in the study area, \( \mu_{sa} \) is a vector of variable means of the study area, and \( \Sigma \) is the covariance matrix of \( x_{sa} \). Occupied cells with high \( D^2 \) values have greater environmental distance from all other cells and are therefore more unique, indicative of specialist species. \( D^2 \) values for this region and environmental space can range from near 0 to well over 400, a completely unrealistic value representing the most specialist species possible in our study area, namely a species occurring on a single grid cell with the most unique environmental conditions. We therefore rescaled \( D^2 \) into niche breadth

Fig. 1. Multi-step virtual species creation. In the first step, a single environmental covariate was selected and a corresponding response curve was randomly generated from a linear, quadratic, logistic, or Gaussian function to create a map of habitat potential values based only on that single environmental descriptor. Habitat potential values were linearly rescaled between 0 and 1 to identify high habitat potential areas as those with values greater than 0.5. For each additional explanatory variable (step 2), a response curve was randomly generated and used to generate new habitat potential values as the sum of the response curve functions within the high habitat potential areas and linearly rescaled between 0 and 1 after each step. This step was iterated until each of the two to five randomly selected environmental covariates were included, with the distribution of true habitat potential values for each virtual species set as the habitat potential values resulting from the sum of all response curve functions (step 3). Occupied areas were then selected with a probabilistic approach using Eq. 1.
values as the inverse of the realistic minimum and maximum possible values in our study area (0 and 10) with the equation:

\[
\text{nich} \text{e breadth} = 1 - \frac{D^2}{10}
\]

This index ranges from 0 to 1, wherein a niche breadth value of 0 indicates extreme uniqueness of environmental variables at occupied cells, and results from a median-squared Mahalanobis distance value of 10, the maximum realistic value in our study area. In contrast, a niche breadth value of one indicates the most generalist species possible, namely one that occupies the most diverse range of environmental conditions such as cells with the Mahalanobis distances near zero.

**Sampling bias**

We introduced environmental and geographic sampling bias into observation datasets of the virtual species to explore how each affects the performance of SDM (Fig. 2). We aimed to mimic sampling bias that might occur under a wide array of conditions, and therefore did not replicate biased sampling from a single source, such as proximity to roads. Instead, we used a generic clustering approach to create biased observation datasets for each species at three levels of bias intensity (low, medium, and high) that might be found in datasets obtained from haphazard sampling inherent in PB modeling approaches (Ponder et al. 2001). These levels ranged from almost none (nearly indistinguishable from randomly sampled observations) to an extreme level greater than that expected in most observation datasets. Bias was introduced with inclusion weights, which were used to preferentially sample a portion of occupied cells for inclusion to each biased observation dataset. The schemas used to set inclusion weights were designed to mimic patterns that might arise from a wide array of biased or haphazard sampling, such as sampling more often near roads in some areas, or not sampling in difficult to access areas such as military lands or private property.

Inclusion weights for the environmentally biased datasets were defined by identifying 15 unique clusters in the multivariate space of the ten possible explanatory variables. Clusters were identified using a partitioning-around-medoids algorithm in the package cluster (Maechler et al. 2016) in R 3.3.2 (R Core Team 2020) and mapped to geographic space. Inclusion weights were assigned to occupied cells in each cluster based on the total area of each cluster with the equation:

\[
P(i) = \left( \frac{1}{\text{cluster area}(i)} \right)^a
\]

where \(a = 1, 3, \text{or} 5 \) for the low, medium, or high bias scenarios, and cluster area is the proportion of the study area that is covered by the mapped cluster containing cell \((i)\). This ensured that mapped clusters with larger areas were less likely to be selected. Inclusion weights were linearly rescaled between 0 and 1 prior to use and were used to randomly select N occupied cells according to their inclusion weights. The degree of sampling bias for each level was assessed with the random skewers method of covariance similarity (Cheverud 1996) in the package phytools (Revell 2012) in R 3.3.2 (R Core Team 2020), by measuring the collinearity between two random selection vectors, each operating on a single covariance matrix (Cheverud and Marroig 2007). This method is used in evolutionary biology and genetics to estimate similarity among genetic covariance matrices (Roff et al. 2012). Here, we use it to estimate the relatedness between the covariance matrix of environmental explanatory variables at locations in the biased observation datasets and the covariance matrix from all true occupied cells, with the most extreme biased datasets showing the least relatedness. The random skewer covariance similarity coefficient among the 100 virtual species ranged from 0.586 to 0.999 for each of the four levels of environmentally biased datasets, and each class (none, low, medium, and high bias) was significantly different when accounting for repeated measures of the species \((F_{3,273} = 31.19; P < 0.0001; \text{Fig. 3A})\).

Inclusion weights for the geographically biased observation datasets were calculated using a spatially clustered sampling schema by randomly seeding 5, 8, or 10 cluster centers for the low, medium, and high intensities of bias, respectively. For each level of intensity, we created a Gaussian density kernel raster using the equal split method of Okabe et al. (2009) with a bandwidth of 500, 250, or 100 km for the 5, 8, or
10 cluster centers, respectively, in GRASS GIS (Neteler et al. 2012). The inclusion weight for each occupied cell was defined by rescaling each Gaussian density kernel raster between 0 and 1. This schema ensures that occupied cells closest to the randomly seeded cluster were selected with higher probabilities than those further away. We created biased observation datasets by randomly selecting occupied cells according to their inclusion weights. The number of cells selected \( N \) for each species was determined by \( p \), the prevalence of occupied cells across the study area, where:

\[
N = p \times 500
\]

We limited the maximum number of observations to 500 to ensure a realistic sample size for each biased observation dataset. In order to compare the degree of bias in both the environmentally biased and geographically biased datasets, we used the random skewer method to assess the degree of environmental sampling bias in each dataset. Environmental bias among the unbiased and geographically biased observation datasets for the 100 virtual species ranged from 0.647 to 0.999, and each class (none, low, medium, and high bias) was significantly different when accounting for repeated measures of the 100 virtual species \( F_{3,272} = 31.19; P < 0.0001; \) Fig. 3B). Ripley’s K function (Ripley 1976) also confirmed that the differences in spatial bias among the levels were significant when accounting for repeated measures across the 100 virtual species \( F_{3,272} = 121.5716; P < 0.0001; \) Fig. 3C), and was assessed as the mean difference between theoretical \( K(r) \) and border-corrected \( K(r) \) Ripley’s K density function for
inhomogeneous point processes of the observation datasets out to a maximum distance of 350 km.

Bias correction methods

We consider three commonly used methods for addressing biased observational datasets in PB SDM (Fig. 2): geographic filtering (G-Filter), environmental filtering (E-Filter), and background weight correction (FactorBiasOut). We implemented the G-Filter method of bias correction using a sampling mesh with equally sized rectangular cells, each 225 km² in area (15 km per side), resulting in 4119 unique sampling areas throughout the study area. We limited observations in each sampling area to two observations per 225 km², a rate that has been effective in minimizing sampling bias for other desert species (Inman et al. 2014). This method results in a spatially thinned point set with an effective sampling density of 1 observation per 112 km², or a thinning distance of approximately 10 km between observations. We implemented the E-Filter correction by sampling observations at a uniform rate within multiple multivariate clusters identified in the n-dimensional environmental space defined by the explanatory variables for a given model specification. We draw on the same clustering algorithm used to introduce environmental sampling bias into the observation datasets, but since the degree of multivariate clustering or number of clusters in an observation dataset is rarely known, we estimate an optimal number by maximizing the average silhouette width with package fpc (Hennig 2015) in R 3.3.2 (R Core Team 2020). In each cluster, a random sample with a size equal to the minimum number of observations in any cluster was selected, and these were pooled to create an environmentally thinned observation dataset. This method ensures an equal sampling intensity across environmental space, though not necessarily across geographic space.

We implement background weight correction with the FactorBiasOut algorithm (Dudik et al. 2005) in Maxent (Phillips et al. 2006). FactorBiasOut estimates the combined distribution of the biased sampling distribution (biased survey effort; σ) and the true species distribution (π).
and factors out $\sigma$, which is assumed to be known and represented as an auxiliary variable (Phillips et al. 2009), often in the form of a bias grid (Elith et al. 2011, Merow et al. 2013). This method relies on knowledge of the biased sampling distribution ($\sigma$), which, in practice, is rarely known. However, because the biased observation dataset is a sample of $\sigma$ (i.e., the observations are sampled with the biased sampling distribution), $\sigma$ can be estimated when the observation dataset is large (Phillips et al. 2009). We create a bias grid as an estimate of $\sigma$ by creating a kernel density raster of each biased observation dataset. The bandwidth for each kernel was estimated using cross-validation to minimize mean square error (Baddeley et al. 2015).

Models and performance criteria

Because it is very widely used by SDM practitioners, we used MaxEnt version 3.4.0 (Phillips et al. 2018) with the default settings to create distribution models for each virtual species. Newer versions of Maxent software (e.g., version 3.4.0 and later) implement a new function to produce an estimate of occurrence probability: the complementary log–log function (Phillips et al. 2017). We used this transformation as an indication of habitat potential and allow inclusion of all feature classes (linear, quadratic, product, threshold, and hinge) to assess the three bias correction methods in their ability to (1) predict true habitat potential (distributions) across the study area, (2) recover the shape of true response curves, and (3) identify true explanatory variables (Fig. 2).

To address the first and second objectives, we created a single model for each of the 100 virtual species specified with the true explanatory variables because for these analyses we assumed that the true explanatory variables were known. While this is never truly known in the real world, rare is it that ecologists lack hypotheses about the environmental drivers of their study organisms’ geographic distribution. Here, we use the true predictors of each virtual species’ distribution to focus on the effects of sampling bias and the commonly used methods to address it.

Predicting distributions

We assessed the ability of each bias correction method to correctly estimate the true habitat potential for each virtual species using the expected fraction of shared presences (Godsoe 2013) by comparing the estimated habitat potential to the true habitat potential across all cells with the equation:

$$ESP = \frac{2\sum_j P_{1(j)}P_{2(j)}}{\sum_j (P_{1(j)} + P_{2(j)})}$$

where $P_{1(j)}$ denotes the true habitat potential at location ($j$) and $P_{2(j)}$ denotes the prediction generated from the model at location ($j$). ESP is a modified Sorenson similarity index (Sorenson 1948) used to compare predicted probabilities that each species is present at a given cell rather than relying on presence/absence information (Godsoe 2013). An ESP value of one indicates perfect agreement between the two maps, while a value of zero indicates complete geographic separation (Godsoe 2013).

Recovering response curve shape

The ability to recreate the form of the true response curves was assessed with the Spearman rank correlation coefficient between the true relationship and that estimated by the model for each explanatory variable. We used the nonparametric rank correlation coefficient because of the nonlinear relationships found between many of the true and estimated response curves. We measured collinearity between the marginal response curves estimated by Maxent for each explanatory variable by holding all other variables at their mean for the study area (Phillips et al. 2006) and those used to create each corresponding virtual species.

Identifying explanatory variables

We address the third objective by doing variable selection with 10 randomly selected virtual species to determine whether sampling bias or any of the bias correction methods affect the ability to correctly identify the true drivers of a species’ geographic distribution. Variable selection is a key step in SDM development, yet few studies explicitly consider the choice of potential environmental covariates, opting instead to include widely available climatic variables with little justification (Bell and Schlaepfer 2016). We consider a plausible set of candidate models with unique
combinations of two to five of the ten possible explanatory variables, resulting in a set of 534 models considered for each species.

The two filtering methods were applied in a replicated fashion because each involves random thinning to remove observations. We replicated the G-Filter and E-Filter methods 20 times each to generate a set of 20 filtered training datasets for each biased observation dataset. We estimated a Maxent model with each of the replicated filtered datasets for each of the candidate model specifications. The background weight correction was implemented using FactorBiasOut with the bootstrap option, also with 20 replications. For each of the 80 biased observation datasets (two bias types, four bias intensities, 10 virtual species) and four bias correction methods (no correction, G-Filter, E-Filter, and FactorBiasOut), we selected the single best performing combination of variables with the average area under the receiver operating characteristic curve value (Fielding and Bell 1997) across the 20 replicates. AUC is commonly used for a jackknife test on variable importance and provides a measure of a model’s ability to discriminate between presence and absence localities, independent of an arbitrary cutoff threshold (Cumming 2000), although it has been criticized for its sensitivity to geographic extent, among other factors (Lobo et al. 2008). However, because our models were evaluated using the same geographic extent and because our goal was to select a single model with which to evaluate multiple bias correction methods, we use AUC as an evaluation metric.

We evaluated each bias correction method’s ability to identify the correct explanatory variables with the Jaccard similarity coefficient using the equation:

$$J = \frac{(A \cap B)}{(A \cup B)} \quad (7)$$

where $A$ is the set of true explanatory variables, $B$ is the set of explanatory variables selected by the model, and $A \cap B$ is the count of correctly identified explanatory variables and $A \cup B$ is the count of unique explanatory variables in the true virtual species model and those in the best model selected by a mean AUC for a given bias correction method. Higher $J$ values indicate a greater ability to discriminate the correct set of predictors. For all three objectives and their respective performance metrics (ESP, Spearman’s correlation coefficient, Jaccard’s similarity), we test for loss in performance due to each type and level of bias intensity, as well as a gain in performance due to the three correction methods. We use mixed models to account for random effects among species and denote differences where significant.

**Observing sample bias in practice**

In order to provide real-world context for our low, medium, high, and no bias levels of geographic bias, we assess the degree of geographic bias that might be expected in PB species locality datasets obtained from biodiversity databases such as the Global Biodiversity Information Facility (GBIF; www.gbif.org). Sampling effort in these datasets is often unknown, and therefore, we use Ripley’s $K$ function for inhomogeneous point processes to check for spatial clustering. We did not check for environmental bias because this depends on the environmental explanatory variables and geographic study region considered for SDM and can only be assessed after these have been identified. We obtained locality records for nine species of diverse flora and fauna, including Squamata, Aves, Rodentia, Asterales, and Zygophyllales from GBIF with the package rgbif (Chamberlain and Boettiger 2017) in R 3.3.2 (R Core Team 2020), which are provided in Appendix S1.

**RESULTS**

**Predicting distributions**

When using unbiased observations, ESP scores measuring the similarity between virtual species’ true and estimated distributions ranged from 0.177 to 0.756, with a mean of 0.487 (SD = 0.128) across 100 virtual species. Species with high landscape prevalence showed greater agreement between the true and estimated distributions (Fig. 4A), as did generalists (Fig. 4B). When bias was introduced, ESP scores were reduced, with mean scores across both types of bias for the 100 virtual species of 0.461 (SD = 0.128), 0.442 (SD = 0.125), and 0.397 (SD = 0.140), for the low, medium, and high bias levels, respectively. Loss of prediction accuracy was greatest for the high environmental bias level, equating to an average reduction of 26% in ESP scores across all species.
However, not all species showed reduced ESP scores when bias was introduced; ~20% resulted in better ESP scores when bias was introduced, and nearly half of these were at the low bias level. Species that were predicted better with biased observation datasets tended to be very rare, with low landscape prevalence (Fig. 5A). We found no effect of landscape prevalence on the loss of ESP scores when bias was introduced ($F_{1,90} = 0.400; P = 0.529$), though rare species showed the greatest variability in loss of ESP scores (Fig. 5A). We did, however, find an effect of niche breadth at the high bias level only ($F_{1,90} = 5.382; P = 0.023$), suggesting that specialist species were more susceptible to sampling bias than were the most generalist species in our study area (Fig. 5B).

Bias correction methods were usually able to increase prediction accuracy under both environmental bias and geographic bias (Fig. 6A and 6B, respectively), though the E-Filter method resulted in additional loss of prediction accuracy 43% of the time. This loss was nearly equally distributed among the three bias levels, suggesting that the E-Filter correction method was poor at compensating for any level of bias. The environmentally biased datasets saw greater improvement than the geographically biased ones (Fig. 6A vs. 6B), especially with the FactorBiasOut correction method. Overall, FactorBiasOut provided the greatest improvements of the three methods across both types of bias ($F_{2, 1560} = 331.856; P < 0.001$). FactorBiasOut was able to improve ESP scores in 96% of cases, whereas the G-Filter and E-Filter only improved ESP scores 82% and 55% of the time, respectively. On average, all three correction methods provided 2% greater improvements with the environmentally biased datasets than with the geographically biased datasets ($F_{1,127} = 29.269; P < 0.001$), suggesting that while high environmental bias caused the greatest reduction in ESP scores, it was also improved the most by correction methods. We found an interaction between prevalence and the bias correction methods ($F_{2, 1558} = 26.776; P < 0.001$; Fig. 7A), as well as an interaction between niche breadth and bias correction methods ($F_{2, 1558} = 83.259; P < 0.001$; Fig. 7B), with specialist and rare species showing greater improvements with the FactorBiasOut method, but not with the other two correction methods.

**Recovering true response curve shape**

On average, the correlation between true and estimated response curves was 0.781 (SD = 0.383) when no sampling bias was present, though Spearman's rank correlation coefficients ranged from −0.999 to 0.999. While some
response curves were estimated poorly, the majority were estimated reasonably well; fewer than 14% of correlation coefficients were below 0.5 for our simulations where the true explanatory variables are known. It was rare for the estimated response curve to be completely wrong, that is, negative correlation occurred only 5% of the time. Under conditions of no sampling bias, we found no apparent relationship between the correlation of the true and estimated response curves with landscape prevalence ($F_{1,91} = 0.305; P = 0.582$), or niche breadth ($F_{1,91} = 0.148; P = 0.701$), though the variability of correlation coefficients was greater among species with wide niches.

When environmental bias was introduced, the mean of the correlations between actual and estimated response curves decreased significantly ($F_{3, 1315} = 9.286; P < 0.001$) to 0.725 (SD = 0.437), 0.682 (SD = 0.493), and 0.617 (SD = 0.525), with proportional reductions of 30%, 37%, and 45% for the low, medium, and high levels of environmental bias, respectively (Table 2). Geographic bias caused fewer overall proportional changes in correlation between actual and estimated response curves with gains of 8%, and 1% and a loss of 19% for the low, medium, and high levels of bias, respectively (Table 2). There was no apparent relationship between decreases in correlation coefficients and landscape prevalence ($F_{1,90} = 0.786; P = 0.377$), with no discernible difference between rare and widespread species. Generally, bias correction methods did not improve estimation of response curves across any of the bias levels ($F_{3, 10440} = 2.050; P = 0.105$), though the FactorBiasOut correction method made recovering true response curve shapes worse across all levels of bias with the environmentally biased datasets ($F_{1, 4904} = 4.355; P = 0.037$; Table 2).

**Identifying true explanatory variables**

In the absence of sampling bias, the ability of SDMs to identify the correct explanatory variables for the randomly selected ten virtual species was poor (mean Jaccard Index scores $J = 0.62; SD = 0.22$). The true set of explanatory variables was selected only 15% of the time. When bias was introduced, $J$ scores were reduced, with losses of 0.06 (environmental) and 0.10 (geographic) for the low bias levels to 0.14

| Bias type   | Correction method | Original Percent change |
|-------------|-------------------|-------------------------|
| High bias   | None              | 0.361 (0.149) $-25.6$ | 24.5 |
|             | FactorBiasOut     | 21.7 (14.6)             |
|             | G-Filter           | 5.2 (7.0)               |
|             | E-Filter           | 4.7 (13.7)              |
| Geographic | None              | 0.432 (0.122) $-11.8$ | 15.9 |
|             | FactorBiasOut     | 11.3 (7.6)              |
|             | G-Filter           | 2.7 (2.6)               |
|             | E-Filter           | 0.5 (8.7)               |
| No bias     | None              | 0.432 (0.123) $-10.0$ | 18.6 |
|             | FactorBiasOut     | 15.1 (13.7)             |
|             | G-Filter           | 2.6 (4.0)               |
|             | E-Filter           | 2.0 (8.2)               |
| No bias     | None              | 0.451 (0.126) $-7.8$  | 14.0 |
|             | FactorBiasOut     | 11.8 (8.9)              |
|             | G-Filter           | 1.4 (1.8)               |
|             | E-Filter           | 2.0 (7.7)               |
| No bias     | None              | 0.466 (0.128) $-4.1$  | 14.9 |
|             | FactorBiasOut     | 9.7 (9.8)               |
|             | G-Filter           | 1.1 (2.1)               |
|             | E-Filter           | 0.4 (7.5)               |
| No bias     | None              | 0.457 (0.127) $-6.8$  | 16.4 |
|             | FactorBiasOut     | 10.8 (9.8)              |
|             | G-Filter           | 0.9 (1.6)               |
|             | E-Filter           | 3.1 (8.9)               |
| No bias     | None              | 0.487 (0.128)           |

Notes: Models were calibrated with and without environmental and geographic sampling bias. Mean and standard deviation (SD) ESP without bias (original) and percent change after bias were introduced or correction methods were applied (% change) are shown.
and 0.10 (geographic) for the high bias levels (Table 3). Reductions in $J$ up to 22% were found in the high environmental bias level; the correct set of explanatory variables was selected only 3% of the time. Models of specialist species did somewhat better at identifying the correct variables than did models of generalist species, though this effect was only significant at the $\alpha = 0.1$ level ($F_{1,18} = 3.927; P = 0.063$). Bias correction methods rarely improved the ability to select the correct explanatory variables, and we found no difference among the three correction methods in their ability to improve $J$ scores ($F_{2,168} = 0.377, P = 0.686$). Similarly, we found no relationship between landscape prevalence ($F_{2,166} = 0.721; P = 0.488$) or niche breadth ($F_{2,166} = 1.065; P = 0.347$) and the three methods’ ability to improve $J$ scores.

**Observing sampling bias in practice**

We found that none of the GBIF observation datasets tested here were geographically unbiased and that most produced Ripley’s $K$ function scores indicative of low or medium geographic bias levels defined in this study. Only two species, Western shovel-nosed snake *Chionactis occipitalis*, and little pocket mouse *Perognathus longimembris*, yielded geographic bias equivalent to our most extreme high bias observation datasets (Fig. 8). Both of these species of conservation concern are well-studied and represented by over 1500 observations; we suspect each suffers from extreme sample bias wherein observations are locally dense along roads and near monitoring sites.

**DISCUSSION**

We use a simulation study with virtual species to compare three sampling bias correction methods that are often used in presence-only SDM but have not been systematically compared across a diverse set of virtual species. We found that the objective for which SDM is used had a large influence on the ability of these methods to compensate for biased observation data. When the intent of SDM is to explain ecological processes, such as identifying environmental factors influencing a species’ distribution or characterizing their response curve shapes, the bias correction methods examined here do little to compensate for biased observation data. Bias correction methods did, however, improve the accuracy of mapped predictions of habitat potential, which is often the goal of SDM studies. For those studies focused on predicting spatial patterns of habitat potential...
rather than understanding the mechanisms driving those patterns, the FactorBiasOut background weight correction method is well-suited. We found that FactorBiasOut outperformed the geographic filtering (G-Filter) and environmental filtering (E-Filter) bias correction methods across all levels and types of bias.

**Predicting distributions**

SDM is most often used for spatial prediction of distributions, and in this case, the use of bias correction methods is clearly recommended. Previous work with virtual species has shown contrasting results, however, with some studies suggesting that G-Filter methods may outperform the FactorBiasOut method in some instances (Kramer-Schadt et al. 2013, Fourcade et al. 2014, Stolar and Nielsen 2015), but others showing E-Filter methods as superior to G-Filter methods when using climatic variables (Varela et al. 2014), or that E-Filtering can lead to mixed results depending on environmental gradients (Gábor et al. 2019) and may not necessarily provide improvements. Each of these studies considered a differing number of virtual species, with at least one, but no more than ten considered. Yet, other work suggests that the FactorBiasOut is superior, especially when target-group background data are incorporated (e.g., Syfert et al. 2013, Fourcade et al. 2014, Vollering et al. 2019). The target-group background approach incorporates an estimate of the biased sampling distribution (σ) from pooled

Fig. 6. All three correction methods were able to increase prediction ability with environmentally (A) and geographically (B) biased datasets in most species, though the FactorBiasOut correction method (FO) showed the greatest improvement with an average 11% greater increase in ESP than the other two methods: E-Filter (EF) and G-Filter (GF) methods overall. EF resulted in poorer (lower) ESP scores, especially for datasets that were geographically biased (B).
observations of multiple species within a target group of similar taxa whose locality data were collected in the same way as the modeled species, and has shown promise for reducing sampling bias (Phillips and Dudik 2008, Elith et al. 2011, Ranc et al. 2016). The benefit of pooling observations from multiple taxa is an increase in the sample size used to estimate \( \sigma \), thereby improving its precision (Phillips and Dudik 2008). However, this method assumes that \( \sigma \) is the same across all species of the target group, which may not be true for

Table 2. Spearman’s rank correlation coefficients between estimated response curves and true species–environment relationships with and without sampling bias.

| Bias type        | Correction method | Original | Percent change |
|------------------|------------------|----------|---------------|
| High bias        |                  |          |               |
| Environmental    | None             | 0.617 (0.525) | −44.7 |
|                  | FactorBiasOut    | −13.6    |               |
|                  | G-Filter         | 9.0      |               |
|                  | E-Filter         | 28.9     |               |
| Geographic       | None             | 0.575 (0.530) | −19.0 |
|                  | FactorBiasOut    | 10.4     |               |
|                  | G-Filter         | 71.8     |               |
|                  | E-Filter         | 5.8      |               |
| No bias          |                  |          |               |
| Medium bias      |                  |          |               |
| Environmental    | None             | 0.682 (0.493) | −36.8 |
|                  | FactorBiasOut    | −83.7    |               |
|                  | G-Filter         | −0.7     |               |
|                  | E-Filter         | 5.6      |               |
| Geographic       | None             | 0.660 (0.496) | 1.0  |
|                  | FactorBiasOut    | 9.9      |               |
|                  | G-Filter         | 10.4     |               |
|                  | E-Filter         | 14.8     |               |
| No bias          |                  |          |               |
| Low bias         |                  |          |               |
| Environmental    | None             | 0.725 (0.437) | −29.8 |
|                  | FactorBiasOut    | −6.8     |               |
|                  | G-Filter         | −3.2     |               |
|                  | E-Filter         | 6.9      |               |
| Geographic       | None             | 0.658 (0.497) | 7.9  |
|                  | FactorBiasOut    | 0.6      |               |
|                  | G-Filter         | 9.3      |               |
|                  | E-Filter         | −142.7   |               |
| No bias          |                  |          |               |
| No bias          | Environmental    | None     |               |
|                  | FactorBiasOut    |         |               |
|                  | G-Filter         |         |               |
|                  | E-Filter         |         |               |
| Geographic       | None             |         |               |
|                  | FactorBiasOut    |         |               |
|                  | G-Filter         |         |               |
|                  | E-Filter         |         |               |
| No bias          | None             | 0.781 (0.383) |      |

Notes: Mean and standard deviation (SD) of correlation coefficients across 100 virtual species assess the ability to represent mechanistic determinants of species distributions when the true explanatory variables are known. Correlation coefficients for models without bias correction (original) and the percent change from the original value (% change) are shown. Many species represented in biodiversity repositories composed of aggregate survey efforts (Newbold 2010).
Our results suggest that the FactorBiasOut method consistently outperformed both of the two filtering methods in their ability to accurately estimate geographic distributions. While our results may contradict some previous findings, they are not too dissimilar. For example, Kramer-Schadt et al. (2013) found that a G-Filter method resulted in lower omission errors than the FactorBiasOut method, but those differences were within 7%, a margin that was found only in a single, wide-ranging species. Similarly, other work has shown considerable variability in the performance of correction methods among differing levels of sampling bias, with the FactorBiasOut method outperforming other methods in a single virtual species, but not two (actual) species (Fourcade et al. 2014). To address this ambiguity, we simulated 100 virtual species with niche characteristics ranging from rare specialists to common generalists and found that the difference between methods was greatest for rare specialists across all levels of sampling bias.

### Inferring niche characteristics

We found that while bias correction methods improved spatial predictions for virtual species, sampling bias greatly reduced the ability to identify the shape of species response curves or select the true explanatory variables using PB SDM, and bias correction methods did not greatly improve this ability. This result highlights differences between predictive and inferential applications of SDM and of statistical ecology generally.

The emphasis on predictive models over explanation has been a trend in ecology over the past several decades (Clark et al. 2001, Evans et al. 2012, Houlanan et al. 2017, Maris et al. 2018) and has led to substantial developments in predictive ecology, such as machine learning methods (Christin et al. 2019, Rammer and Seidl 2019). In general, machine learning methods are better suited for prediction and less so at identifying and explaining causal relationships (Shmueli 2010). This difference is exacerbated when making predictions under novel or no-analog conditions (Williams and Jackson 2007, Dietze et al. 2018) where machine learning methods favor model complexity motivated by model fit over transferability and ecological theory (Sinclair et al. 2010, Bell and Schlaepfer 2016). As a correlative statistical approach, PB SDM can be used much like other statistical methods for explanatory or predictive modeling, though for the former, it is important to include only ecologically relevant and plausible environmental covariates (e.g., Guisan and Zimmermann 2000, Bell and...

### Table 3. Jaccard Index agreement scores across a random subset of 10 virtual species for identifying correct explanatory variables.

| Bias type | Correction method | Jaccard Index |
|-----------|-------------------|---------------|
| High bias |                   |               |
| Environmental | None | 0.48 (0.193) |
|             | FactorBiasOut | 0.48 (0.140) |
|             | G-Filter | 0.48 (0.139) |
|             | E-Filter | 0.48 (0.169) |
| Geographic | None | 0.52 (0.215) |
|             | FactorBiasOut | 0.48 (0.215) |
|             | G-Filter | 0.52 (0.193) |
|             | E-Filter | 0.57 (0.177) |
| No bias |                   |               |
| Medium bias |                   |               |
| Environmental | None | 0.48 (0.169) |
|             | FactorBiasOut | 0.52 (0.169) |
|             | G-Filter | 0.48 (0.169) |
|             | E-Filter | 0.46 (0.190) |
| Geographic | None | 0.54 (0.212) |
|             | FactorBiasOut | 0.50 (0.216) |
|             | G-Filter | 0.58 (0.22) |
|             | E-Filter | 0.50 (0.236) |
| No bias |                   |               |
| Low bias |                   |               |
| Environmental | None | 0.56 (0.227) |
|             | FactorBiasOut | 0.54 (0.232) |
|             | G-Filter | 0.56 (0.227) |
|             | E-Filter | 0.53 (0.231) |
| Geographic | None | 0.52 (0.193) |
|             | FactorBiasOut | 0.58 (0.148) |
|             | G-Filter | 0.52 (0.193) |
|             | E-Filter | 0.52 (0.193) |
| No bias |                   |               |
| No bias |                   |               |
| Environmental | None | 0.62 (0.224) |
|             | FactorBiasOut |               |
|             | G-Filter |               |
|             | E-Filter |               |
| Geographic | None |               |
|             | FactorBiasOut |               |
|             | G-Filter |               |
|             | E-Filter |               |
| No bias | None |               |

Notes: Agreement scores were calculated for the top performing species distribution model selected by area under the receiver/operator curve (AUC). Mean and standard deviation (SD) agreement scores measure the proportion of the true explanatory variables identified by the model with the highest AUC score.
Schlaepfer 2016), and to keep models simple when variable selection is the focus (e.g., Warren and Seifert 2011).

However, due to the inherent analyses of species–environment relationships that serve as the statistical underpinnings of spatial predictions in SDM (Guisan and Zimmermann 2000), PB SDM has often been leveraged to study niche traits and niche characteristics. One niche characteristic often explored is climatic niche conservatism (Wiens and Graham 2005), the condition wherein changes in climate may cause shifts in geographic distributions over evolutionary time-scales. Some formal tests, such as niche equivalency (Warren et al. 2008), have been based on SDM and have been used extensively to investigate niche conservatism in related extant taxa (e.g., Strubbe et al. 2014), though these tests compare spatial predictions of habitat potential between species to infer niche similarity rather than direct comparisons of response curves or selected variables. It seems then that using SDM for identifying the drivers of species distributions (e.g., variable selection) may be better accomplished with unbiased presence and absence data (Guisan and Zimmermann 2000) or in combination with mechanistic models incorporating biophysical ecology and functional traits (Kearney et al. 2010a, b) rather than PB SDM, even with bias correction.

Effect of niche characteristics on sensitivity to sampling bias

In the absence of sampling bias, our simulations confirm previously well-established relationships between sample size (landscape prevalence in our case) and the ability to estimate true distributions (Hernandez et al. 2006, Wisz et al. 2008, Proosdij et al. 2015). However, relationships between niche breadth and the effects of sampling bias are not well-understood. Previous work has suggested that the occupied niche may be more difficult to distinguish from background environmental space in generalists with broad tolerances than for specialists (Brotons et al. 2004, Elith et al. 2006, Franklin et al. 2009); on the other hand, recent work has shown that specialist species are more susceptible to sampling noise because of their often reduced sample sizes (Soltan and Safi 2017). If distributions of specialists are in fact more easily estimated than generalists, this would suggest that specialists may be less susceptible to biased sampling than generalists because their smaller niches are likely to be sampled more completely than larger ones. However, in previous studies, niche breadth is often conflated with landscape prevalence (or rarity), assuming that generalists with wide niches are more common than specialists.

Species with wide niches may in fact be rare if their realized and fundamental niches are substantially differentiated by biotic factors such as predation, disease, or competition (Hutchinson 1991). Studies using SDM rarely incorporate biotic factors, and instead often mix observational occurrence records from realized niches with environmental data describing fundamental niche properties, especially when scenopoetic variables are the only type considered (Soberón and Nakamura 2009). Thus, the type of niche being quantified (e.g., fundamental, potential, or realized) is often not well-defined and can influence perception of niche breadth. Our method used occupied environmental uniqueness to
define niche breadth, and with this definition, we found evidence that specialists with narrow niches are more susceptible to high intensities of biased sampling than generalists (Fig. 5B), which we also found for actual species locality data (Fig. 8). We found no evidence of an effect of landscape prevalence on susceptibility to sampling bias, and no interaction between landscape prevalence and niche breadth, suggesting that any susceptibility to sampling bias was due to niche breadth alone, not rarity. Our results suggest that it is niche breadth, not landscape prevalence, that may influence sensitivity to sampling bias, except in cases of extreme rarity. Here, biased sampling actually improved predictions of habitat potential in a few, very rare, species. By design, our study sampled each virtual species proportionally to landscape prevalence, resulting in small sample sizes for rare species. However, rare species that are well-studied may be heavily sampled (e.g., Fig. 8); in these cases, they may be more easily modeled than generalists, as previously suggested.

**Practical notes about implementing bias correction methods**

Of the three bias correction methods compared here, FactorBiasOut and G-Filter are used most often. These methods are rather straight forward to implement, though each requires tuning with parameters and little guidance is provided in the literature. The FactorBiasOut method implemented in Maxent requires a bias grid, or prior representing hypothesized sampling intensity across the study area (Merow et al. 2013). Where documented, these grids are often created as the Gaussian kernel densities of observations as recommended by Elith et al. (2011), though the choice of standard deviation (SD) for the kernel is not often reported. This choice influences the distance at which adjacent cells affect estimated densities, wherein a small SD equates to a local influence among neighboring cells. For G-Filter methods, the choice of thinning grid size or thinning parameter will influence the local density of observations. The consequences of different choices of these parameters are not well-understood. Elith et al. (2011) suggest SD might reflect some property of a species’ dispersal ability to capture the possibility of animal movement between surveys. However, the haphazard nature of opportunistic sampling inherent in biodiversity databases likely confounds the interpretation of SD, and therefore, we chose to use a cross-validation approach based on the spatial pattern of observations to estimate an optimal value for SD. This method minimizes the mean squared error in the Gaussian kernel density field at multiple SD values and identifies an optimum value for a given observation dataset (Diggle et al. 1998). The trade-offs and implications of using statistical measures over biological knowledge are not well-understood; additional work is needed to provide guidelines for selecting SD and developing bias grids.

The G-Filter method also relies on selecting a spatial parameter, namely the bandwidth used to identify areas with high observation densities. Previous studies have filtered observations based on nearest neighbor distances, such as removing observations that are within a certain distance of one another (e.g., Kramer-Schadt et al. 2013, Boria et al. 2014), while others use spatial grids to filter observations from cells with high sample densities (e.g., Vandergast et al. 2013). The decision for the bandwidth size is rarely discussed, though Veloz et al. (2009) use a novel method of exploring spatial autocorrelation in model residuals to determine an optimal value. Our selection of bandwidth and choice of filtering methods were based on previous work to reduce sampling bias in aggregated observations for the southwestern United States (Vandergast et al. 2013, Inman et al. 2014). As with the FactorBiasOut method, additional research is needed to explore differences among alternative implementations of the G-Filter method.

In contrast to the G-Filter method, the E-Filter method is less well represented in the literature. This is likely because it is more tedious to implement and is dependent on explanatory variables specified in any given model. Whereas G-Filter thinning occurs entirely in geographic space and results in a single dataset that can be used to calibrate multiple model specifications, the E-Filter method results in a separate dataset for each model specification. This precludes the use of multi-model inference or complexity metrics, such as Akaike information criterion (Burnham and Anderson 2002) or Bayesian information criterion (Schwarz 1978). Model complexity is not often discussed, but has been shown to be an
important factor affecting predictions across geographic regions and time periods (Warren and Seifert 2011, Syfert et al. 2013, Merow et al. 2014, Bell and Schlaepfer 2016).

Conclusions

Due to the proliferation and availability of geolocated species observations, SDM has become a primary tool for tasks such as designing conservation and monitoring programs, evaluating the efficacy of proposed land management actions, and developing recovery planning for threatened or endangered species. In these applications, biased observations are all but assured. Bias correction methods may offer improvement in estimating spatial distributions at low and moderate levels of bias, but we suggest practitioners be wary of using model fit to identify explanatory variables or infer ecological meaning from response curves when using SDM in PB frameworks.

Acknowledgments

Any use of trade, firm, or product names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Literature Cited

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

Amatulli, G., S. Domisch, M.-N. Tuanmu, B. Parmientier, A. Ranipeta, J. Malczyk, and W. Jetz. 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Scientific Data 5:815–915.

Andrews, A. 1990. Fragmentation of habitat by roads and utility corridors: a review. Australian Zoologist 26:130–141.

Baddeley, A., E. Rubak, and R. Turner. 2015. Spatial point patterns: methodology and applications with R. Chapman and Hall/CRC Press, London, UK.

Bean, W. T., R. Stafford, and J. S. Brashares. 2012. The effects of small sample size and sample bias on threshold selection and accuracy assessment of species distribution models. Ecography 35:250–258.

Bell, D. M., and D. R. Schlaepfer. 2016. On the dangers of model complexity without ecological justification in species distribution modeling. Ecological Modelling 330:50–59.

Blonder, B., C. Lamanna, C. Violle, and B. J. Enquist. 2014. The n-dimensional hypervolume. Global Ecology and Biogeography 23:595–609.

Boria, R. A., L. E. Olson, S. M. Goodman, and R. P. Anderson. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecological Modelling 275:73–77.

Brotons, L., W. Thuiller, M. Araujo, and A. Hirzel. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography 27:437–448.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference. Springer Verlag, New York, New York, USA.

Butchart, S. H. M., et al. 2010. Global biodiversity: indicators of recent declines. Science 328:1164–1168.

Bystrakova, N., M. Peregrym, R. H. J. Erkens, O. Bezsmertna, and H. Schneider. 2012. Sampling bias in geographic and environmental space and its effect on the predictive power of species distribution models. Systematics and Biodiversity 10:305–315.

Chamberlain, S. A., and C. Boettiger. 2017. R Python, and Ruby clients for GBIF species occurrence data. PeerJ Preprints 5:e3304v1.

Cheverud, J. M. 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (Saguinus oedipus) and saddle-back (S. fuscicollis) tamarins. Journal of Evolutionary Biology 9:5–42.

Cheverud, J. M., and G. Marroig. 2007. Research article comparing covariance matrices: random skewers method compared to the common principal components model. Genetics and Molecular Biology 30:461–469.

Christin, S., É. Hervet, and N. Lecomte. 2019. Applications of deep learning in ecology. Methods in Ecology and Evolution 10:1632–1644.

Clark, J. S., et al. 2001. Ecological forecasts: an emerging imperative. Science 293:657–660.

Cumming, G. S. 2000. Using between-model comparisons to fine-tune linear models of species ranges. Journal of Biogeography 27:441–455.

Dietze, M. C., et al. 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences of the United States of America 115:1424–1432.

Diggle, P. J., J. A. Tawn, and R. A. Moyeed. 1998. Model-based geostatistics. Applied Statistics-Journal of the Royal Statistical Society Series C 47:299–326.

Dudik, M., S. J. Phillips, and R. E. Schapire. 2005. Correcting sample selection bias in maximum entropy density estimation. Pages 323–330 in Y. Weiss, B. Scholkopf and J. C. Platt, editors. Proceedings of the 18th International Conference on Neural Information Processing Systems. MIT Press, Cambridge, Massachusetts, USA.
Edwards, T. C., D. R. Cutler, N. E. Zimmermann, L. Geiser, and G. G. Moisen. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. Ecological Modelling 199:132–141.

Elith, J., et al. 2006. Novel methods improve prediction of species’ distributions from occurrence data. Ecography 29:129–151.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40:677–697.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43–57.

Evans, M. R., K. J. Norris, and T. G. Benton. 2012. Predictive ecology: systems approaches. Philosophical Transactions of the Royal Society B: Biological Sciences 367:163–169.

Fernández, D., and M. Nakamura. 2015. Estimation of spatial sampling effort based on presence-only data and accessibility. Ecological Modelling 299:147–155.

Fielding, A., and J. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49.

Fourcade, Y., J. O. Engler, D. Rödder, and J. Secondi. 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 University Press, Cambridge, UK.

Franklin, J., K. E. Wejnert, S. A. Hathaway, C. J. Rochester, and R. N. Fisher. 2009. Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. Diversity and Distributions 15:167–177.

Freeman, O. E., L. A. Duguma, and P. A. Minang. 2015. Operationalizing the integrated landscape approach in practice. Ecology and Society 20:24.

Freitag, S., C. Hobson, H. C. Biggs, and A. S. van Jaarsveld. 1998. Testing for potential survey bias: the effect of roads, urban areas and nature reserves on a southern African mammal data set. Animal Conservation 1:119–127.

Gábor, L., V. Moudráy, V. Barták, and V. Lecours. 2019. How do species and data characteristics affect species distribution models and when to use environmental filtering? International Journal of Geographical Information Science 00:1–18.

Gelfand, A. E., S. K. Sahu, and D. M. Holland. 2012. On the effect of preferential sampling in spatial prediction. Environmetrics 23:565–578.

Godsoe, W. 2013. Inferring the similarity of species distributions using Species’ Distribution Models. Ecography 37:130–136.

Guillera-Arroita, G., J. L. Lahoz-Monfort, J. Elith, A. Gordon, H. Kujala, P. E. Lentini, M. A. McCarthy, and R. Tingley B. A. Wintle. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography 24:276–292.

Guisan, A., et al. 2013. Predicting species distributions for conservation decisions. Ecology Letters 16:1424–1435.

Guisan, A., and N. Zimmermann. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147–186.

Hastie, T., and W. Fithian. 2013. Inference from presence-only data; the ongoing controversy. Ecography 36:864–867.

Hefley, T. J., A. J. Tyre, D. M. Baasch, and E. E. Blankenship. 2013. Nondetection sampling bias in marked presence-only data. Ecology and Evolution 3:5225–5236.

Hennig, C. 2015. fpc: flexible Procedures for Clustering. R package version 2.2-9. https://CRAN.R-project.org/package=fpc

Hernandez, P. A., C. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29:773–785.

Hijmans, R. J. 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93:679–688.

Hortal, J., A. Jiménez-Valverde, J. F. Gómez, J. M. Lobo, and A. Baselga. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos 117:847–858.

Houlanahan, J. E., S. T. McKinney, T. M. Anderson, and B. J. McGill. 2017. The priority of prediction in ecological understanding. Oikos 126:1–7.

Hutchinson, G. E. 1991. Population studies: animal ecology and demography. Bulletin of Mathematical Biology 53:193–213.

Inman, R. D., K. E. Nussear, T. C. Esque, A. G. Vander gast, S. A. Hathaway, D. A. Wood, K. R. Barr, and R. N. Fisher. 2014. Mapping habitat for multiple species in the Desert Southwest. Open-File Report 2014–1134. U.S. Geological Survey, Reston, Virginia, USA.

Kadmon, R., O. Farber, and A. Danin. 2004. Effect of roadside bias on the accuracy of predictive maps
produced by bioclimatic models. Ecological Applications 14:401–413.

Kearney, M. R., B. A. Wintle, and W. P. Porter. 2010a. Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. Conservation Letters 3:203–213.

Kearney, M., S. J. Simpson, D. Raubenheimer, and B. Helmuth. 2010b. Modelling the ecological niche from functional traits. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 365:3469–3483.

Kramer-Schadt, S., et al. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. Diversity and Distributions 19:1366–1379.

Lobo, J. M., and M. F. Tognelli. 2014. Exploring the human footprint in the west: a large-scale analysis of anthropogenic impacts. Ecological Applications 18:1119–1139.

Leroy, B., C. N. Meynard, C. Bellard, and F. Courchamp. 2015. virtualspecies, an R package to generate virtual species distributions. Ecography 39:599–607.

Leu, M., S. E. Hanser, and S. T. Knick. 2008. The human footprint in the west: a large-scale analysis of anthropogenic impacts. Ecological Applications 18:1119–1139.

Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17:145–151.

Lobo, J. M., and M. F. Tognelli. 2014. Exploring the effects of quantity and location of pseudo-absences and sampling biases on the performance of distribution models with limited point occurrence data. Journal for Nature Conservation 1–7.

Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik. 2016. cluster: cluster Analysis Basics and Extensions. R package version 2.1.0. https://CRAN.R-project.org/package=cluster

Maris, V., P. Huneman, A. Coreau, S. Kéfi, R. Pradel, and V. Devictor. 2018. Prediction in ecology: promises, obstacles and clarifications. Oikos 127: 171–183.

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

Merow, C., M. J. Smith, and J. A. S. Jr. 2013. A practical guide to MaxEnt for modeling species’ distributions: What it does, and why inputs and settings matter. Ecography 36:1058–1069.

Meynard, C. N., B. Leroy, and D. M. Kaplan. 2019. Testing methods in species distribution modelling using virtual species: What have we learnt and what are we missing? Ecography 42:2021–2036.

Neteler, M., M. H. Bowman, M. Landa, and M. Metz. 2012. GRASS GIS: a multi-purpose open source GIS. Environmental Modelling and Software 31:124–130.

Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Progress in Physical Geography 34:3–22.

Nürnberg, S., and M. Aberhan. 2013. Habitat breadth and geographic range predict diversity dynamics in marine Mesozoic bivalves. Paleobiology 39:360–372.

Okabe, A., T. Satoh, and K. Sugihara. 2009. A kernel density estimation method for networks, its computational method and a GIS-based tool. International Journal of Geographical Information Science 23:7–32.

Opdam, P., and D. Wascher. 2004. Climate change meets habitat fragmentation: linking landscape and biogeographical scale levels in research and conservation. Biological Conservation 117:285–297.

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

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

Phillips, S. J., and M. Dudik. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161–175.

Phillips, S. J., M. Dudik, J. Elith, C. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19: 181–197.

Phillips, S. J., M. Dudik, and R. E. Schapire. 2018. Maxent software for modeling species niches and distributions (Version 3.4.1). http://biodiversityinformatics.amnh.org/open_source/maxent/

Ponder, W. F., G. A. Carter, P. Flemons, and R. R. Chapman. 2001. Evaluation of museum collection data for use in biodiversity assessment. Conservation Biology: the Journal of the Society for Conservation Biology 15:648–657.

Qiao, H., A. T. Peterson, L. Ji, and J. Hu. 2017. Using data from related species to overcome spatial sampling bias and associated limitations in ecological niche modelling. Methods in Ecology and Evolution 8:1804–1812.

R Core Team. 2020. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rammer, W., and R. Seidl. 2019. Harnessing deep learning in ecology: an example predicting bark beetle outbreaks. Frontiers in Plant Science 10:1327.

Ranc, N., L. Santini, C. Rondinini, L. Boitani, F. Poitevin, A. Angerbjorn, and L. Maiorano. 2016. Performance tradeoffs in target-group bias correction for species distribution models. Ecography 40:1076–1087.

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Ripley, B. D. 1976. The second-order analysis of stationary point processes. Journal of Applied Probability 13:255–266.

Roff, D. A., J. M. Prokolla, I. Krams, and M. J. Rantala. 2012. There is more than one way to skin a G matrix. Journal of Evolutionary Biology 25:1113–1126.

Rosenberg, K. V., et al. 2019. Decline of the North American avifauna. Science 366:120–124.

Rosenzweig, C., et al. 2008. Attributing physical and biological impacts to anthropogenic climate change. Nature 453:353–357.

Saupe, E. E., H. Qiao, J. R. Hendricks, R. W. Portell, S. J. Hunter, J. Soberón, and B. S. Lieberman. 2015. Niche breadth and geographic range size as determinants of species survival on geological time scales. Global Ecology and Biogeography 24:1159–1169.

Schneider, M. F. 2009. Habitat loss, fragmentation and predator impact: spatial implications for prey conservation. Journal of Applied Ecology 38:1–16.

Schwartz, M. W. 2012. Using niche models with climate projections to inform conservation management decisions. Biological Conservation 155:149–156.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461–464.

Shmueli, G. 2010. To explain or to predict? Statistical Science 25:289–310.

Sinclair, S., M. White, and G. Newell. 2010. How useful are species distribution models for managing biodiversity under future climates? Ecology and Society 15.

Soberón, J., and M. Nakamura. 2009. Niches and distributional areas: concepts, methods, and assumptions. Proceedings of the National Academy of Sciences of the United States of America 106:19644–19650.

Soley-Guardia, M., E. E. Gutiérrez, D. M. Thomas, J. Ochoa-G, M. Aguilera, and R. P. Anderson. 2016. Are we overestimating the niche? Removing marginal localities helps ecological niche models detect environmental barriers. Ecology and Evolution 6:1267–1279.

Sorensen, T. 1948. A method of establishing groups of equal amplitudes in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter 5:1–34.

Soultan, A., and K. Safi. 2017. The interplay of various sources of noise on reliability of species distribution models hinges on ecological specialisation. PLOS ONE 12:e0187906–e187920.

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

Stubbie, D., O. Beauchard, and E. Matthysen. 2014. Niche conservatism among non-native vertebrates in Europe and North America. Ecography 38:321–329.

Syfert, M. M., M. J. Smith, and D. A. Coomes. 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. PLOS ONE 8:e55158.

Travis, J. 2003. Climate change and habitat destruction: a deadly anthropogenic cocktail. Proceedings of the Royal Society of London Series B-Biological Sciences 270:467–473.

van Proosdij, A. S. J., M. S. M. Sosef, J. J. Wieringa, and N. Raes. 2015. Minimum required number of specimen records to develop accurate species distribution models. Ecography 39:542–552.

Vandergast, A. G., R. D. Inman, K. R. Barr, K. E. Nuss-ear, T. C. Esque, S. A. Hathaway, D. A. Wood, P. A. Medica, J. W. Breinholt, C. L. Stephen, A. D. Gottscho, S. B. Marks, W. B. Jennings, and R. N. Fisher. 2013. Evolutionary hotspots in the Mojave Desert. Diversity 5:293–319.

Varela, S., R. P. Anderson, R. García-Valdés, and F. Fernández-González. 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography 37:1084–1091.

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

Vollering, J., R. Halvorsen, I. Auestad, and K. Rydgren. 2019. Bunching up the background betters bias in species distribution models. Ecography 42:1717–1727.

Walston, L. J., K. E. Rollins, K. E. LaGory, K. P. Smith, and S. A. Meyers. 2016. A preliminary assessment of avian mortality at utility-scale solar energy facilities in the United States. Renewable Energy 92:405–414.

Warren, D. L., R. E. Glor, and M. Turelli. 2008. Environmental niche equivalency versus conservatism:
quantitative approaches to niche evolution. Evolution 62:2868–2883.
Warren, D. L., and S. N. Seifert. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecological Applications 21:335–342.
Warren, R., et al. 2013. Quantifying the benefit of early climate change mitigation in avoiding biodiversity loss. Nature Climate Change 3:678–682.
Warren, R., J. Price, J. VanDerWal, S. Cornelius, and H. Sohl. 2018. The implications of the United Nations Paris Agreement on climate change for globally significant biodiversity areas. Climatic Change 147:395–409.
Wiens, J., and C. Graham. 2005. Niche Conservatism: integrating Evolution, Ecology, and Conservation Biology. Annual Review of Ecology, Evolution, and Systematics 36:519–539.
Williams, J. W., and S. T. Jackson. 2007. Novel climates, no-analog communities, and ecological surprises. Frontiers in Ecology and the Environment 5: 475–482.
Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan, and N. P. S. D. W. Group†. 2008. Effects of sample size on the performance of species distribution models. Diversity and Distributions 14:763–773.
Yackulic, C. B., R. Chandler, E. F. Zipkin, J. A. Royle, J. D. Nichols, E. H. C. Grant, and S. Veran. 2012. Presence-only modelling using MAXENT: When can we trust the inferences? Methods in Ecology and Evolution 4:236–243.

**Supporting Information**

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3422/full