Characterizing uncertainty in species distribution models derived from interpolated weather station data

M. Fernández,[1],† H. Hamilton,[2] and L. M. Kueppers[1,3]

[1]Environmental Systems Graduate Group, University of California, 5200 North Lake Road, Merced, California 95343 USA
[2]Marine Conservation Institute, 454 Bolinas Road, Fairfax, California 94930 USA
[3]School of Natural Sciences and Sierra Nevada Research Institute, University of California,
5200 North Lake Road, Merced, California 95343 USA

Abstract.   Species distribution models (SDMs) are used to generate hypotheses regarding the potential distributions of species under different environmental conditions, such as forecasts of species range shifts in response to climate change and predictions of invasive species range expansions. However, an accurate description of species' geographic ranges as a function of the environment requires that species observations and climatic variables are measured at the same spatial and temporal resolution, which is usually not the case. Weather station data are interpolated and these resulting continuous data layers are incorporated into SDMs, often without any uncertainty assessment. Here we quantify the effects of three unrelated but complementary aspects of uncertainty in weather station interpolations on SDM performance using MaxEnt. We examine the influence of topographic heterogeneity, interannual variability, and distance to station on the over- and under-prediction of modeled North American bird distributions. Our species observations are derived from presence-absence information for 20 bird species with well-known distributions. These three metrics of uncertainty in interpolated weather station data have varying contributions to over- and under-prediction errors in SDMs. Topographic heterogeneity had the highest contribution to omission errors; the lowest contribution to commission errors was from Euclidean distance to station. The results confirm the importance of establishing an appropriate relational basis in time and space between species and climatic layers, providing key operational criteria for selection of species observations fed into SDMs. Our findings highlight the importance of identifying weather stations locations used in interpolated products, which will allow a characterization of some aspects of uncertainty and identification of regions where users need to be particularly careful when making a decision based on a SDM.

Key words: climate-based niche models; gridded climate; uncertainty.

† E-mail: miguel.fernandez.trigoso@gmail.com

INTRODUCTION

An attempt to understand species distributions, the fundamental unit of biogeography (Angert 2009, Lomolino 2010), has stimulated the development of tools to model the geographic distribution of organisms as a function of environmental factors. These models are used not only to understand distributions under contemporary environmental conditions, but also to predict whether or not a species might find suitable habitat outside the boundaries of

its current distribution (e.g., Vaclavik and Meentemeyer 2012) and to investigate the response of a species to projected future climates or reconstructed paleo-climates (e.g., Fordham et al. 2012, Stigall 2012). Species distribution modeling, also known as ecological or environmental niche modeling, has undergone an exponential growth in popularity and applications in recent years (Elith and Leathwick 2009) and is now a frequently used method in multiple fields such as ecology, evolution, conservation biology, epidemiology and agriculture.

As with any model, the output of species distribution models (SDMs) is dependent on the quality of data upon which they are built. Two key sources of SDM input data include: (1) species observations in nature, such as georeferenced point occurrences, and (2) environmental variables, such as high-resolution gridded climate layers. Multiple correlations between environmental parameters and known locations of species occurrences are constructed using geostatistical algorithms, defining an n-dimensional space that represents the climatic requirements of a species. This inferred multidimensional space can then be projected back into geographic space to produce a map of the species' potential distribution. While the quality of species observations and their effect on SDMs have been extensively documented (Soberon et al. 2000, Graham et al. 2008, Hortal et al. 2008, Loiselle et al. 2008, Fernández et al. 2009, Lobo et al. 2010, Feeley and Silman 2011, Naimi et al. 2011), the deficiencies and biases of environmental variables have seldom been considered (Peterson and Nakazawa 2008, McInerny and Purves 2011, Synes and Osborne 2011), despite the key role they play in the process of building, evaluating and calibrating SDMs.

SDM applications that project models to a space or time other than that from which they were created often use interpolations based on weather station data (e.g., Gonzáles et al. 2010, Fernández et al. 2012), in which the spatial pattern of uncertainties is non-uniform and highly variable (Johnson et al. 2000), and as such, can lead to misinterpretation of spatial and temporal accuracy by users (Beale and Lennon 2012). Worldclim (Hijmans et al. 2005), a gridded climatic dataset that consists of monthly mean temperature and precipitation values averaged over the 1950 to 1999 period at 1 km^2 spatial resolution, is one example of an interpolated dataset widely used in SDMs where the degree of uncertainty associated with individual cell values for a particular climatic variable is influenced by three elements: (1) spatial variability, (2) temporal variability, and (3) the density of available observations (Zhang and Goodchild 2002). The first two elements are considered to be intrinsic characteristics of the parameter to be estimated, while the last one is considered to be a characteristic attributable to the observation system. Although these three unrelated but complementary elements represent the core of uncertainty characterization of interpolations for weather station data, their effect on SDM performance has not been quantified.

Recent studies suggest that poor model performance, in part, can be attributed to high levels of uncertainty in the environmental data (Kriticos and Leriche 2010, Beale and Lennon 2012, Kamino et al. 2012); however the relationship between model omission and commission errors and the degree of uncertainty in the interpolated environmental input layers has not yet been addressed in the literature. The aim of this study is to determine whether SDM performance can be directly attributed to any of these three aspects of uncertainty. Explicitly accounting for the underlying uncertainty in the weather station interpolated data, we investigate three hypotheses: (1) SDM omission and commission errors are more often found in regions with high levels of spatial variability, (2) SDM performance is degraded by the mismatch between the scale of climatic variability used to create the model and the scale at which species distributions respond, and (3) SDM omission and commission errors are expected to be higher in regions with relatively low density of weather stations. These expectations are tested using a combination of high quality species occurrence data and novel gridded datasets that include estimates of environmental uncertainty. Our results suggest that biogeographers can benefit from increased attention to the variability and uncertainty in gridded spatial climate data when developing and applying SDMs.

Table 1. Species list.

Species	Distribution comments
Callipepla squamata†	Range fluctuates in response to variability in winter rainfall (Giuliano and Lutz 1993).
Dendragapus obscurus†	Prefers high lands in the winter. Occurs in pine and fir forest habitats from sea level to 3,600 m (Johnsgard 1988).
Tympanuchus cupido†	Extirpated from much of the range in U.S. Native prairie is preferred, but also adapted to cropland (Schroeder and Braun 1993).
Centrocercus urophasianus†	Adapted to winter extremes. Distribution is reduced as a result of loss of sagebrush habitat. Current distribution estimated at 56% of pre-settlement (Drut et al. 1994).
Columba fasciata	Moves seasonally to areas higher or lower than normal range. Timing of breeding a factor of food availability (Howell and Webb 1995).
Buteo regalis†	Distribution and density closely associated with cycles of prey abundance (NatureServe 2012).
Picoides borealis†	Cooperative breeder influenced by loss of habitat, requires >80 ha of continuous habitat. Dependent of fire-maintained, old-growth pine forest (BirdLife 2012).
Picoides nuttallii†	Endemic species to California and Baja California. Confined to oak woodlands (NatureServe 2012).
Picoides albolarvatus†	Fire suppression and fragmentation has contributed to range decline in the northern part of the distribution. Biology and ecology remains unstudied (IUCN 2012).
Melanerpes lewis†	Strongly associated with fire-maintained old-growth ponderosa pine. Patchy distribution, inconsistent from year-to-year (Saab and Dudley 1998).
Calypte anna‡	Moves to low elevations in the winter. The only hummingbird that spends the winter in northern climates (Johnsgard 1983).
Selasphorus platycercus‡	Some individuals have moved into urban and suburban areas of southwestern due to hummingbird feeders (Calder 1994).
Selasphorus sasin	Apparent expansion in breeding range due to availability of non-native flowers (Johnsgard 1983).
Pyrocephalus rubinus†	Northern populations move south in the winter. Can be found between 0 and 3,000 m in elevation (NatureServe 2012).
Aphelocoma californica†	Can be found in scrub-brush, boreal forests and temperate forests. Well adapted to suburban areas (NatureServe 2012).
Calamospiza melanocorys‡	Arrives until late May to the northern edge of its range (NatureServe 2012).
Limnothlypis swainsonii	Summer and winter distribution. One of the least observed of North American birds if it weren't for its loud song (BirdLife 2012).
Vermivora luciae‡	Incomplete information on breeding ecology. Arrives and departs early from breeding grounds perhaps to evade much of the summer heat (BirdLife 2012).
Vermivora virginiae‡	Limited information on distribution and habitat preferences (NatureServe 2012).
Dendroica caerulescens‡	Male common in forest at lower to mid-elevations, female uses shrubbier habitat at higher elevations. Mortality from exposure to cold or rainy weather (BirdLife 2012).

† Year round distribution.
‡ Winter and breeding distribution.

METHODS

Species occurrence data

We selected a subset of twenty species of birds from distribution data compiled by the North American Bird Breeding Survey (BBS; Sauer et al. 2006). The BBS dataset is a well-vetted, standardized, spatially balanced, long-term source of bird species occurrences (Sauer et al. 2006). Importantly, it is a close approximation to a true presence-absence, multi-species, observational dataset available at a continental scale. From the full list, twenty species were selected based on the criteria that their breeding distributions are largely determined by climate (Table 1), increasing the likelihood their distributions can be reasonably modeled using primarily climate variables.

For each species, observation data points representing multi-year survey routes (Sauer et al. 2003) were split into presence and absence. In order to avoid a subjective decision in the placement of the break between presence vs. absence across all survey years, two complementary approaches were applied. First, species observations in transect location maps were plotted as histograms, supporting the detection of naturally occurring breaks in the data. Secondly, BBS range maps were compared to an independent source of species range descriptions (BirdLife International and Natureserve 2011). Following Fernández et al. (2012), the independent range maps were transformed into a raster format that matched the BBS map's spatial resolution. The maps were compared using ArcGIS Version 10 Spatial Analyst, looking for the value in the classification of each BBS continuous map that provided the highest value of similarity among them. The two approaches

Table 2. List of variables used in the analysis.

Code	Variable description	Source
Bioclimatic layer	Annual mean temperature (°C)	1
Bioclimatic layer	Mean diurnal temp range (°C)	1
Bioclimatic layer	Isothermality	1
Bioclimatic layer	Temperature seasonality	1
Bioclimatic layer	Max temperature of warmest month (°C)	1
Bioclimatic layer	Min Temperature of coldest month (°C)	1
Bioclimatic layer	Temperature annual range (°C)	1
Bioclimatic layer	Mean temperature of wettest quarter (°C)	1
Bioclimatic layer	Mean temperature of driest quarter (°C)	1
Bioclimatic layer	Mean temperature of warmest quarter (°C)	1
Bioclimatic layer	Mean temperature of coldest quarter (°C)	1
Bioclimatic layer	Annual precipitation (mm)	1
Bioclimatic layer	Precipitation of wettest month (mm)	1
Bioclimatic layer	Precipitation of driest month (mm)	1
Bioclimatic layer	Coefficient of variation for annual precipitation	1
Bioclimatic layer	Precipitation of wettest quarter (mm)	1
Bioclimatic layer	Precipitation of driest quarter (mm)	1
Bioclimatic layer	Precipitation of warmest quarter (mm)	1
Bioclimatic layer	Precipitation of coldest quarter (mm)	1
Uncertainty layer	Topographic heterogeneity	2
Uncertainty layer	Standard deviation from mean annual temperature	3
Uncertainty layer	Coefficient of variation from annual total precipitation	3
Uncertainty layer	Euclidean distance to closest temperature weather station	2, 4 & 5
Uncertainty layer	Euclidean distance to closest precipitation weather station	2, 4 & 5

Note: Sources are 1, Worldclim; 2, SRTM; 3, ClimVar; 4, FAOCLIM-2; 5, GHCN.

agreed for all species. We therefore defined presences as a survey route location point along which a particular bird species had been recorded during at least one of the ten years that the route was visited (1994 to 2003); routes without at least one positive record were considered absences.

Climatic gridded data

Nineteen climate layers, at a resolution of 1 km^2, were obtained from Worldclim Version 1.4 (Hijmans et al. 2005). These variables, commonly referred to as bioclimatic layers (Table 2), represent biologically relevant aspects of temperature and precipitation. This gridded climatic dataset, which provides one of the finest spatial resolutions relative to other similar products at a global extent (e.g., Purvis et al. 2011, Roura-Pascual et al. 2011, Zelazowski et al. 2011), was chosen for multiple reasons. First, for North America, Worldclim was generated from interpolated weather station data obtained from the Global Historical Climatology Network (GHCN; Vose et al. 1992), and the World-wide Agroclimatic Database from the Food and Agriculture Organization (FAOCLIM-2 2005), datasets to which we also had access. Second, this climatic dataset is available only as an averaged product

for the period of 1950 to 1999 and, therefore, does not account for interannual climatic variability. Third, Worldclim does not provide the user with an assessment of the quality of the information or uncertainty characterization in the data. Finally, this climatic dataset has been cited 1,534 different times since its publication (ISI 2012), and constitutes a critical resource for studies in diverse scientific fields including ecology, conservation, paleobiology, public health, anthropology and developmental biology (e.g., Daszak et al. 2012, de Oliveira et al. 2012, Kamilar et al. 2009, Levsen et al. 2012, Rosell et al. 2012).

Uncertainty layers

To test our three hypotheses, we generated three gridded datasets that represent three different aspects of uncertainty in interpolated climatic data layers. The first dataset provided a metric of spatial variability, a factor known to contribute to biodiversity at the landscape level (Kreft and Jetz 2007). Spatial variability was quantified as topographic heterogeneity, measured by the number of unique elevation values within 25 km^2 of the target pixel derived from the 1 km^2 spatial resolution Shuttle Radar Topographic Mission digital elevation model (SRTM-DEM, Farr et al. 2007), using a Python script to

iterate ArcGIS Version 10 Zonal Statistics tool.

The second uncertainty dataset was an index of temporal climate variation, ClimVar (Fernández et al. 2012), calculated for weather stations that had a record of at least 30 years. This dataset was chosen because it represents a fine resolution spatial characterization of interannual climatic variability, which represents the largest temporal fluctuation in the climatic system, as compared to daily, intraseasonal and interdecadal variability (Ghil 2002). ClimVar is based on the same combined sources of weather stations (Table 2) used to create Worldclim. The specific ClimVar layers used here are one standard deviation of mean annual temperature and the coefficient of variation of annual total precipitation.

The third dataset provided a measure of the density of available information for interpolation. This dataset was based on the 6,499 stations recording monthly average temperature and 8,671 stations recording precipitation for the continental United States, the same weather station data used to produce ClimVar and WorldClim. Using the SRTM-DEM as reference for cell size, cell center position, and elevation value, we calculated a new gridded layer where each cell value reflected the combination of vertical and horizontal distance (i.e., Euclidean distance) from the center of the cell to the closest temperature or precipitation weather station (these were rarely equivalent). The Marine Geospatial Ecology Tools Version 0.8a44 and the Proximity Toolset in ArcGIS Version10 were used to determine the closest weather station and measure the distances as follows:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (1)$$

where $x_2 - x_1$ represents the longitudinal distance between any point in the reference shapefile (x_1) and the nearest weather station (x_2), $y_2 - y_1$ represents the latitudinal distance between any point in the reference shapefile (y_1) and the nearest weather station (y_2), and $z_2 - z_1$ represents the difference in elevation from any point in the reference shapefile (z_1) and the elevation at the nearest weather station (z_2).

Species distribution modeling

Continuous SDMs were generated using MaxEnt Version 3.3.3e (Phillips et al. 2006), a machine learning algorithm that uses the principle of maximum entropy to derive a set of rules correlating environmental variables and species occurrences to estimate the potential geographic distribution of a target species. MaxEnt was chosen because of its well established performance relative to alternative niche modeling techniques (Elith et al. 2006, Elith and Leathwick 2009, but see Li et al. 2011), and its capacity to deal with multicollinearity in the environmental variables, by considering redundant information without penalizing models by overfitting (Phillips et al. 2006, Phillips and Dudik 2008, Elith et al. 2011).

In order to avoid spatial autocorrelation (Peterson et al. 2011), all the localities that included presences and absences for each species were split into three subsets: 60% for model training, 20% for threshold selection and 20% for model evaluation (Fig. 1) based on spatially structured partitioning and random selection of the data (modified from Daszak et al. 2012). Only the presence portion of the 60% subset was used to train each SDM. The first 20% subset, which included presences and absences, was used to select the threshold value applied to convert the continuous SDM output into a binary map. A cutoff value that maximizes sensitivity and specificity was computed based on the area under the receiver operating characteristic (ROC) curves (Fielding and Bell 1997) using SigmaPlot Version 11.0. The remaining 20% subset, which also included presences and absences, was contrasted with the resulting MaxEnt binary map using a Python script to detect observations that were over- or under-predicted by the model. This process was repeated one hundred times, each time randomly sampling with replacement a new combination of 60–20–20. This analysis produced two combined datasets that include all of the localities that were identified as false positives (commission errors) or false negatives (omission errors) by at least one of the bootstrapped SDMs for each species. These combined datasets were then used as the reference locations for which associated values of uncertainty were extracted (Fig. 1), using the Geospatial Modeling Environment Version 0.7.2 RC2 Isectpntrst tool (Beyer 2012).
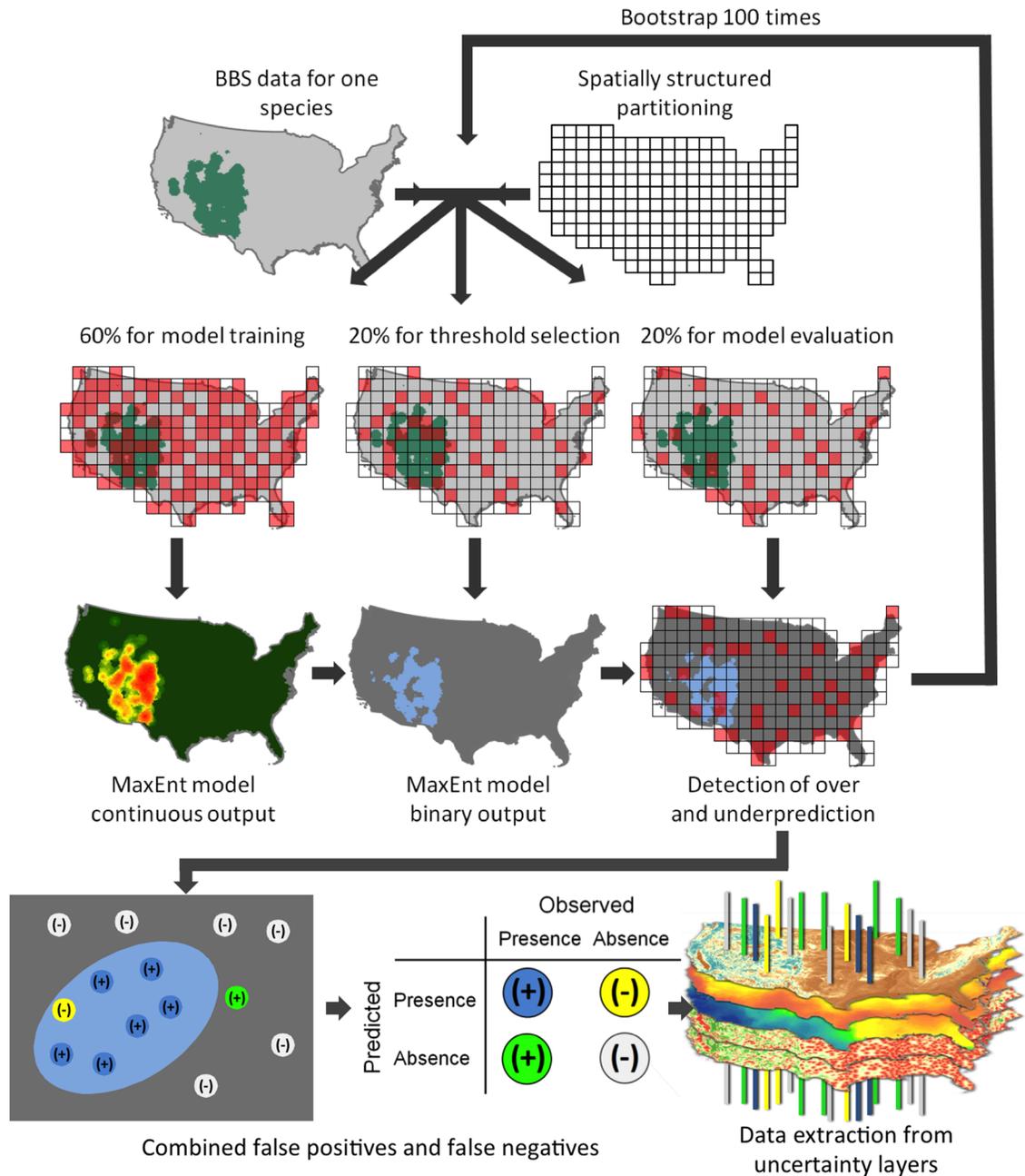
Fig. 1. Flowchart describing spatially structured partitioning of the data and sampling strategy. All localities were randomly split into three subsets: 60% for model training, 20% for threshold selection and 20% for model evaluation. The presence portion of the 60% subset was used to train a SDM. The first 20% subset was used to select the threshold. Cutoff values that maximize sensitivity and specificity were computed based on the area under the ROC curve. The second 20% subset was contrasted with the resulting MaxEnt output to detect observations that were over- or under-predicted by the model. This process was repeated for each species to produce 100 bootstrapped iterations. The results were two combined datasets that include all the localities that were identified as false positives or false negatives by at least one of the bootstrapped SDMs for each species. These datasets were then used as the reference locations for which associated values of uncertainty were extracted.

Data analysis

To understand the relationship between performance errors in the SDMs and the three sets of layers that represent different aspects of uncertainty in the interpolated station data, we compared the uncertainty values associated with over- and under-predictions to a null distribution using the two-sample Kolmogorov-Smirnov test (K-S) with Stata Version 11.2. The two-sample K-S test is commonly used to assess whether two independent samples come from an identical distribution, making no assumptions about the normality of the data. The two-sample K-S statistic D represents a measure of the maximum difference between the cumulative distribution functions for each sample (Conover 1999). For each species' withheld evaluation data, two different sets of 100 localities were randomly resampled with replacement (bootstrap, Efron 1982) 10,000 times. Using the two-sample K-S test, their associated values of uncertainty were compared to produce the underlying probability density for D under the null distribution. From this distribution the mean value, D_{null}, was calculated (Fig. 2a).

To compare the correctly predicted absences to the incorrectly predicted absences (i.e., false negatives), two subsets of 100 localities obtained from the withheld evaluation data were contrasted using the two-sample K-S test. The process was bootstrapped 10,000 times to obtain the underlying probability density for D, in the case of the commission errors. From this distribution the mean value, D_{ov}, was calculated (Fig. 2b).

To compare the correctly predicted presences to the incorrectly predicted presences (i.e., false positives), two subsets of 100 localities obtained from the withheld evaluation data were contrasted using the two-sample K-S test. The process was also bootstrapped 10,000 times to obtain the underlying probability density for D, this time for the omission errors. From this distribution the mean value, D_{un}, was calculated (Fig. 2c).

For each distribution, the mean value calculated for the resulting bootstrapped parameter (i.e., D_{ov} and D_{un}) was subtracted from the mean value calculated from its corresponding null distribution (i.e., D_{null}). The absolute magnitudes of the differences were used to determine the strength of the association between commission or omission errors and each of the uncertainty layers
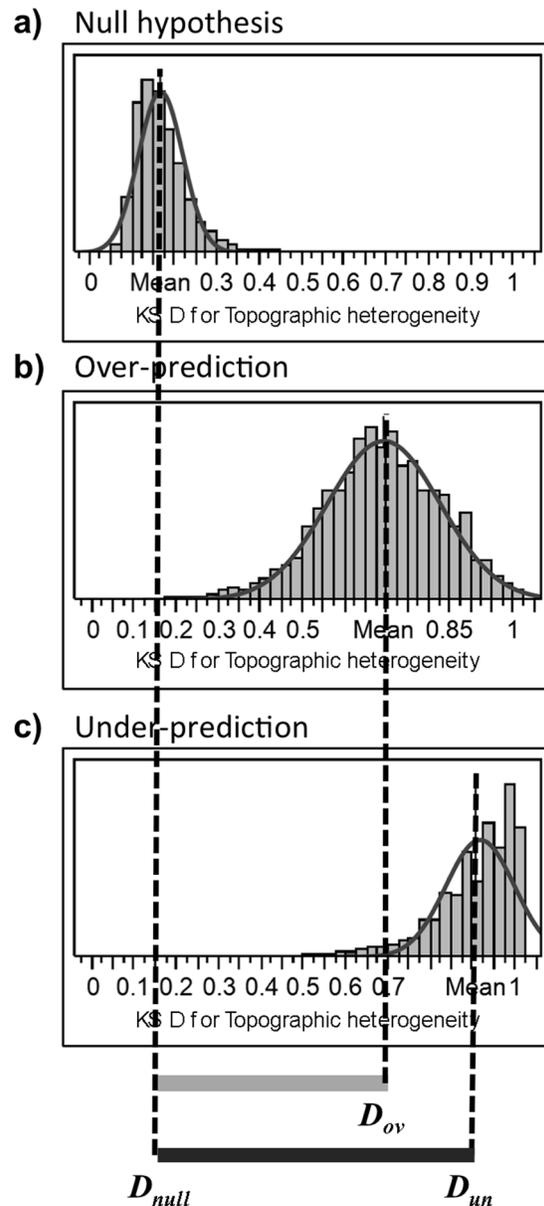

Fig. 2. Normalized mean differences were obtained by subtracting the average value of the 10,000 bootstrap null distribution (D_{null}) for the Kolmogorov-Smirnov test (D) from the mean value from the distributions obtained by bootstrapping the over-predicted (X_{ov}) localities and the under-predicted (X_{un}) localities for *Vermivora luciae* (shown) and for all other species.

Table 3. Two-sample Kolmogorov-Smirnov test results for topographic heterogeneity

	Topographic heterogeneity									
	Overprediction					Underprediction				
Species	Mean Obs. Coef.	Boot. SE	Q (2.5%)	Q (97.5%)	Sig. 0.05	Mean Obs. Coef.	Boot. SE	Q (2.5%)	Q (97.5%)	Sig. 0.05
A. californica	0.490	0.139	0.244	0.796	...	0.873	0.129	0.535	1	*
B. regalis	0.553	0.161	0.256	0.898	...	0.805	0.187	0.408	1	*
C. melanocorys	0.578	0.197	0.250	0.950	...	0.450	0.146	0.229	0.816	...
C. squamata	0.497	0.177	0.223	0.908	...	0.795	0.197	0.371	1	*
C. anna	0.737	0.109	0.505	0.937	*	0.903	0.122	0.500	1	*
C. urophasianus	0.568	0.120	0.351	0.818	*	0.875	0.146	0.459	1	*
C. fasciata	0.681	0.108	0.462	0.884	*	0.866	0.152	0.459	1	*
D. obscurus	0.646	0.097	0.451	0.831	*	0.852	0.163	0.449	1	*
D. caerulescens	0.494	0.135	0.261	0.810	...	0.778	0.204	0.375	1	*
L. swainsonii	0.517	0.181	0.242	0.949	...	0.853	0.156	0.490	1	*
M. lewis	0.558	0.112	0.344	0.789	*	0.931	0.086	0.657	1	*
P. albolarvatus	0.727	0.110	0.496	0.915	*	0.896	0.134	0.480	1	*
P. borealis	0.579	0.173	0.274	0.939	...	0.877	0.157	0.459	1	*
P. nuttallii	0.753	0.116	0.500	0.948	*	0.851	0.173	0.439	1	*
P. rubinus	0.312	0.114	0.152	0.599	...	0.910	0.103	0.602	1	*
S. platycercus	0.552	0.120	0.330	0.798	*	0.881	0.127	0.510	1	*
S. sasin	0.772	0.137	0.472	0.990	*	0.937	0.063	0.776	1	*
T. cupido	0.497	0.137	0.250	0.787	...	0.799	0.201	0.381	1	*
V. luciae	0.691	0.134	0.418	0.947	*	0.915	0.081	0.707	1	*
V. virginiae	0.454	0.118	0.222	0.696	...	0.891	0.126	0.551	1	*

(Fig. 2, Tables 3, 4 and 5). The same approach was repeated for each uncertainty layer and each species, totaling 200 comparisons. Finally, confidence intervals were defined for D under the null distribution, the commission errors, and the omission errors, based on the percentile method (Dixon 1993), where the intervals were calculated and compared directly from the frequency distributions of the bootstrapped statistics ($Q2.5\%$ and $Q97.5\%$) to define statistical significance.

RESULTS

The results of our analysis indicate that false negatives and false positives, in that order, were significantly ($p < 0.05$) associated (1) with regions of high topographic heterogeneity for 95% and 50% of the species (Fig. 3a), (2) with regions of high interannual precipitation variability for 75% and 70% of the species (Fig. 3b, left panel) and with regions of high interannual temperature variability for 70% and 55% of the species (Fig. 3b, right panel), and (3) with regions located further away from precipitation stations for 95% and 20% of the species (Fig. 3c, left panel) and with regions located further away from temperature stations for 95% and 25% of the species (Fig. 3c, right panel), when compared to

their corresponding null distributions.

The comparisons based on the precipitation weather stations and those based on the temperature stations for interannual variability were weakly correlated (Fig. 4a). However, when an outlier was not included in the regression, the correlation for false negatives was stronger ($r^2 = 0.74$; Fig. 4a, right panel, dotted line). Correlations between precipitation and temperature for Euclidean distance uncertainty were high for both false positives ($r^2 = 0.99$; Fig. 4b, left panel) and for false negatives ($r^2 = 0.97$; Fig. 4b, right panel).

Among the three datasets characterizing uncertainty in interpolated climatic layers, interannual precipitation variability showed the highest values associated with false positive errors in the SDM, with a species average relative distance ($|D_{null} - D_{ov}|$) of 0.45 (where 0 represents no effect and 1 represents the highest effect). Euclidean distance to station showed the lowest species average relative distance ($|D_{null} - D_{ov}|$) of 0.26, associated with false positive errors for both temperature and precipitation based layers (Fig. 5). For SDM false negative errors, the dataset that showed the highest associated value was topographic heterogeneity with a species average relative distance ($|D_{null} - D_{un}|$) of 0.68; followed by interannual variability with a species average

Table 4. Two-sample Kolmogorov-Smirnov test results for interannual variability.

Species	Variable	Interannual variability									
		Overprediction					Underprediction				
		Mean Obs. Coef.	Boot. SE	Q (2.5%)	Q (97.5%)	Sig. 0.05	Mean Obs. Coef.	Boot. SE	Q (2.5%)	Q (97.5%)	Sig. 0.05
A. californica	Precip	0.660	0.129	0.379	0.914	*	0.676	0.197	0.316	1.000	*
	Temp	0.560	0.115	0.330	0.772	*	0.685	0.199	0.323	0.990	*
B. regalis	Precip	0.526	0.137	0.286	0.828	...	0.615	0.190	0.296	0.990	*
	Temp	0.390	0.140	0.179	0.726	...	0.606	0.199	0.267	0.970	...
C. melanocorys	Precip	0.577	0.128	0.337	0.889	*	0.800	0.083	0.611	0.929	*
	Temp	0.452	0.150	0.234	0.816	...	0.613	0.123	0.390	0.888	*
C. squamata	Precip	0.694	0.148	0.369	0.938	*	0.592	0.192	0.276	0.980	...
	Temp	0.544	0.150	0.260	0.854	...	0.615	0.200	0.280	0.980	...
C. anna	Precip	0.790	0.167	0.417	0.990	*	0.610	0.176	0.296	0.990	*
	Temp	0.689	0.138	0.394	0.959	*	0.653	0.199	0.296	0.990	*
C. urophasianus	Precip	0.558	0.117	0.335	0.783	*	0.606	0.190	0.286	0.990	...
	Temp	0.377	0.123	0.178	0.641	...	0.600	0.192	0.027	0.980	...
C. fasciata	Precip	0.584	0.146	0.293	0.882	*	0.585	0.189	0.278	0.980	...
	Temp	0.591	0.093	0.392	0.756	*	0.607	0.203	0.275	0.990	...
D. obscurus	Precip	0.448	0.129	0.220	0.720	...	0.603	0.191	0.285	0.990	...
	Temp	0.342	0.105	0.167	0.572	...	0.609	0.196	0.276	0.990	...
D. caerulescens	Precip	0.653	0.134	0.388	0.884	*	0.691	0.184	0.327	1.000	*
	Temp	0.519	0.102	0.337	0.732	*	0.657	0.188	0.316	0.990	*
L. swainsonii	Precip	0.499	0.160	0.248	0.818	...	0.645	0.178	0.316	0.990	*
	Temp	0.734	0.109	0.459	0.929	*	0.700	0.194	0.316	0.990	*
M. lewis	Precip	0.569	0.131	0.302	0.825	*	0.608	0.191	0.286	0.990	...
	Temp	0.390	0.112	0.195	0.625	...	0.627	0.204	0.281	0.990	...
P. albolarvatus	Precip	0.609	0.146	0.337	0.919	*	0.622	0.180	0.299	0.990	*
	Temp	0.569	0.111	0.333	0.763	*	0.650	0.195	0.302	0.990	*
P. borealis	Precip	0.433	0.109	0.265	0.684	...	0.613	0.182	0.298	0.990	*
	Temp	0.788	0.121	0.480	0.949	*	0.644	0.195	0.295	0.990	*
P. nuttallii	Precip	0.859	0.117	0.574	1.000	*	0.603	0.177	0.299	0.990	*
	Temp	0.663	0.118	0.389	0.889	*	0.634	0.190	0.296	0.980	*
P. rubinus	Precip	0.578	0.150	0.289	0.902	...	0.699	0.190	0.337	1.000	*
	Temp	0.771	0.065	0.642	0.902	*	0.723	0.198	0.344	0.990	*
S. platycercus	Precip	0.611	0.125	0.359	0.853	*	0.626	0.193	0.296	0.990	*
	Temp	0.442	0.125	0.211	0.691	...	0.634	0.201	0.292	0.990	*
S. sasin	Precip	0.836	0.166	0.454	1.000	*	0.656	0.193	0.316	1.000	*
	Temp	0.759	0.146	0.378	0.970	*	0.674	0.200	0.313	0.990	*
T. cupido	Precip	0.394	0.127	0.202	0.705	...	0.612	0.181	0.296	0.990	*
	Temp	0.412	0.093	0.266	0.625	...	0.629	0.189	0.292	0.990	*
V. luciae	Precip	0.878	0.121	0.542	1.000	*	0.715	0.202	0.337	1.000	*
	Temp	0.739	0.144	0.408	0.960	*	0.709	0.201	0.323	0.990	*
V. virginiae	Precip	0.706	0.084	0.533	0.862	*	0.667	0.181	0.333	0.990	*
	Temp	0.386	0.118	0.178	0.628	...	0.683	0.195	0.326	0.990	*

relative distance ($|D_{null} - D_{un}|$) of 0.47 for the temperature weather station based layer and 0.48 for the precipitation weather station based layer. Euclidean distance showed the lowest species average relative distance of ($|D_{null} - D_{un}|$) of 0.45, associated with false negatives for both temperature and precipitation based layers (Fig. 5).

DISCUSSION

Our goal was to determine the effects of three different metrics of uncertainty associated with climate data geo-processing on the performance of species distribution models. Specifically, we were interested in understanding the uncertainty associated with weather station-based interpolations and its effect on omission and commission errors. We found evidence for the influence of all three of the tested sources of uncertainty on SDM performance: topographic heterogeneity, interannual variability and distance to the closest weather station.

Our results are consistent with each of our expectations; however, there are some clear differences in the degree of association between each of the uncertainty layers we analyzed and the observed errors of omission and commission. We found that, among the three uncertainty

Table 5. Two-sample Kolmogorov-Smirnov test results for Euclidean distance.

| | | Euclidean distance | | | | | | | | | |
| | | Overprediction | | | | | Underprediction | | | | |
Species	Variable	Mean Obs. Coef.	Boot. SE	Q (2.5%)	Q (97.5%)	Sig. 0.05	Mean Obs. Coef.	Boot. SE	Q (2.5%)	Q (97.5%)	Sig. 0.05
A. californica	Precip	0.339	0.109	0.176	0.594	...	0.937	0.191	0.430	1.000	*
	Temp	0.341	0.108	0.178	0.592	...	0.841	0.187	0.438	1.000	*
B. regalis	Precip	0.362	0.129	0.180	0.677	...	0.863	0.187	0.459	1.000	*
	Temp	0.361	0.130	0.184	0.680	...	0.833	0.199	0.418	1.000	*
C. melanocorys	Precip	0.456	0.171	0.212	0.889	...	0.455	0.156	0.216	0.798	...
	Temp	0.452	0.170	0.255	0.859	...	0.441	0.147	0.217	0.773	...
C. squamata	Precip	0.429	0.165	0.200	0.888	...	0.946	0.117	0.546	1.000	*
	Temp	0.426	0.165	0.200	0.857	...	0.952	0.106	0.592	1.000	*
C. anna	Precip	0.444	0.153	0.212	0.814	...	0.963	0.105	0.576	1.000	*
	Temp	0.439	0.152	0.211	0.808	...	0.951	0.119	0.525	1.000	*
C. urophasianus	Precip	0.298	0.099	0.152	0.532	...	0.947	0.107	0.566	1.000	*
	Temp	0.300	0.099	0.150	0.527	...	0.919	0.129	0.515	1.000	*
C. fasciata	Precip	0.347	0.115	0.168	0.602	...	0.912	0.147	0.512	1.000	*
	Temp	0.356	0.117	0.170	0.613	...	0.901	0.156	0.500	1.000	*
D. obscurus	Precip	0.298	0.096	0.147	0.515	...	0.827	0.199	0.423	1.000	*
	Temp	0.300	0.095	0.149	0.513	...	0.834	0.194	0.429	1.000	*
D. caerulescens	Precip	0.594	0.123	0.362	0.857	*	0.792	0.216	0.378	1.000	*
	Temp	0.602	0.125	0.365	0.866	*	0.808	0.212	0.418	1.000	*
L. swainsonii	Precip	0.641	0.164	0.333	0.960	*	0.838	0.201	0.490	1.000	*
	Temp	0.638	0.163	0.326	0.948	*	0.854	0.198	0.500	1.000	*
M. lewis	Precip	0.291	0.091	0.152	0.505	...	0.957	0.099	0.606	1.000	*
	Temp	0.293	0.091	0.153	0.502	...	0.930	0.128	0.531	1.000	*
P. albolarvatus	Precip	0.402	0.127	0.202	0.691	...	0.991	0.043	0.898	1.000	*
	Temp	0.389	0.126	0.189	0.674	...	0.951	0.117	0.531	1.000	*
P. borealis	Precip	0.617	0.147	0.333	0.899	*	0.969	0.097	0.586	1.000	*
	Temp	0.609	0.152	0.330	0.908	*	0.972	0.088	0.667	1.000	*
P. nuttallii	Precip	0.479	0.158	0.232	0.857	...	0.983	0.060	0.776	1.000	*
	Temp	0.476	0.156	0.237	0.838	...	0.969	0.086	0.667	1.000	*
P. rubinus	Precip	0.423	0.130	0.197	0.695	...	0.913	0.162	0.500	1.000	*
	Temp	0.408	0.126	0.190	0.677	...	0.921	0.155	0.500	1.000	*
S. platycercus	Precip	0.300	0.098	0.150	0.526	...	0.895	0.171	0.500	1.000	*
	Temp	0.308	0.100	0.158	0.537	...	0.863	0.187	0.485	1.000	*
S. sasin	Precip	0.562	0.189	0.260	1.000	...	0.960	0.102	0.592	1.000	*
	Temp	0.563	0.175	0.296	1.000	*	0.971	0.080	0.667	1.000	*
T. cupido	Precip	0.584	0.127	0.333	0.837	*	0.871	0.192	0.480	1.000	*
	Temp	0.580	0.127	0.333	0.837	*	0.889	0.184	0.490	1.000	*
V. luciae	Precip	0.462	0.157	0.227	0.869	...	0.894	0.159	0.500	1.000	*
	Temp	0.463	0.156	0.233	0.847	...	0.936	0.111	0.586	1.000	*
V. virginiae	Precip	0.274	0.088	0.142	0.479	...	0.858	0.193	0.480	1.000	*
	Temp	0.280	0.089	0.143	0.484	...	0.873	0.187	0.500	1.000	*

components, the highest degree of association occurred between false negative errors and topographic heterogeneity (Fig. 5), indicating that in areas of high topographic heterogeneity, SDMs are more likely to under-predict than over-predict a species' distribution. This finding can be explained by how the models classify a particular pixel as species presence vs. absence. For example, if an area situated in the Great Plains in Kansas, where climatic layers have relatively low spatial variability, is compared to a second area located on the Eastern slope of the Andes where variables change abruptly over relatively short distances, errors will diverge. In the first example the model will tend to over-predict because the climatic similarity from one pixel to the next is high and decreases slowly, making it easier for the model to fit a curve that explains the relationship between observations and the environmental layers. In the second example, the climatic similarity from one pixel to the next has the potential to change dramatically, thus making it harder for the model to fit a curve that explains the complexity in the environmental layer without raising the degree of the polynomial approximation. In this case, the model choice will be the most parsimonious solution, which results in an under-prediction. These results suggest that a
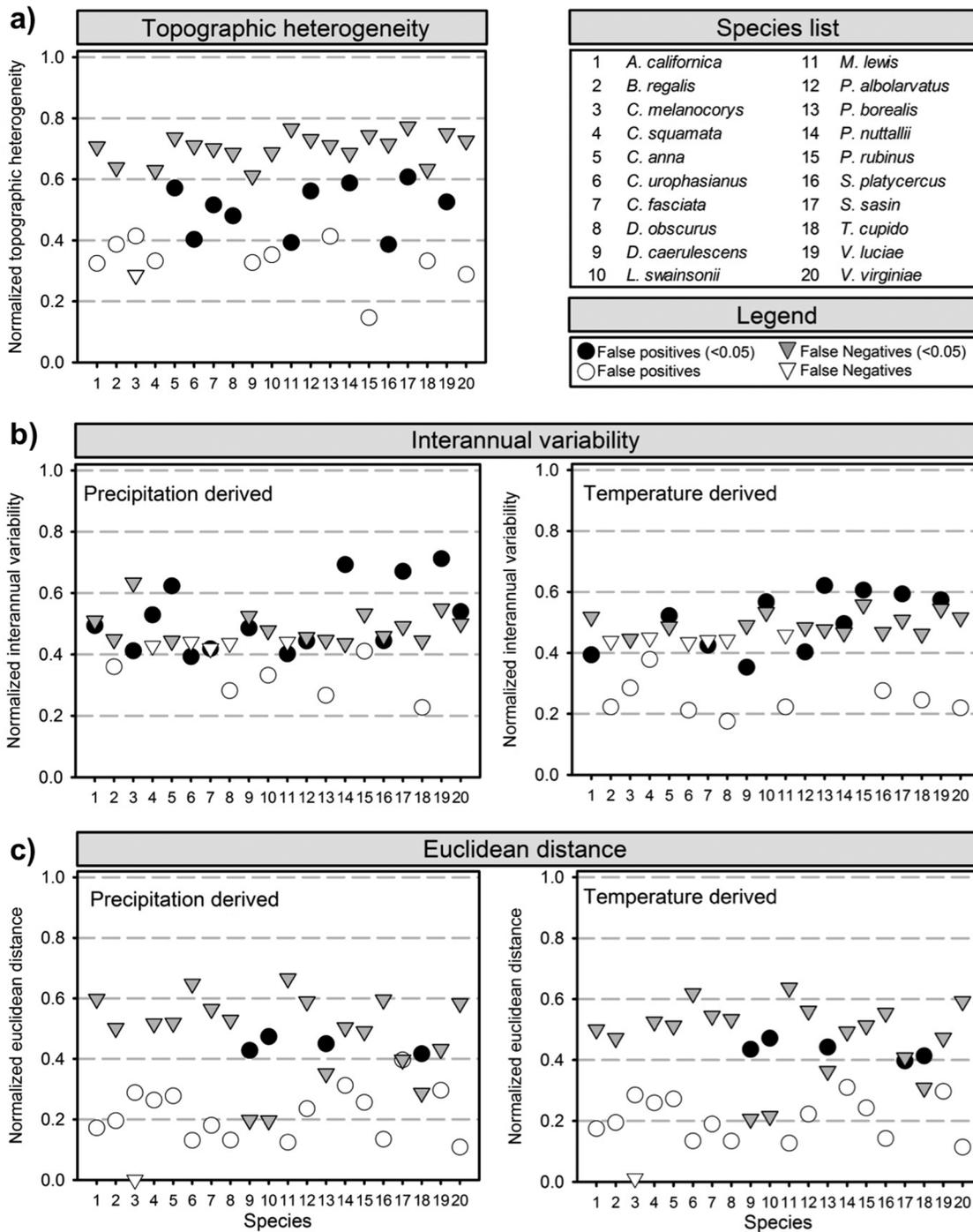
Fig. 3. Species-specific normalized mean differences for each uncertainty layer. False positives are marked with circles and false negatives are marked with triangles. Statistically significant values are shown with filled markers.
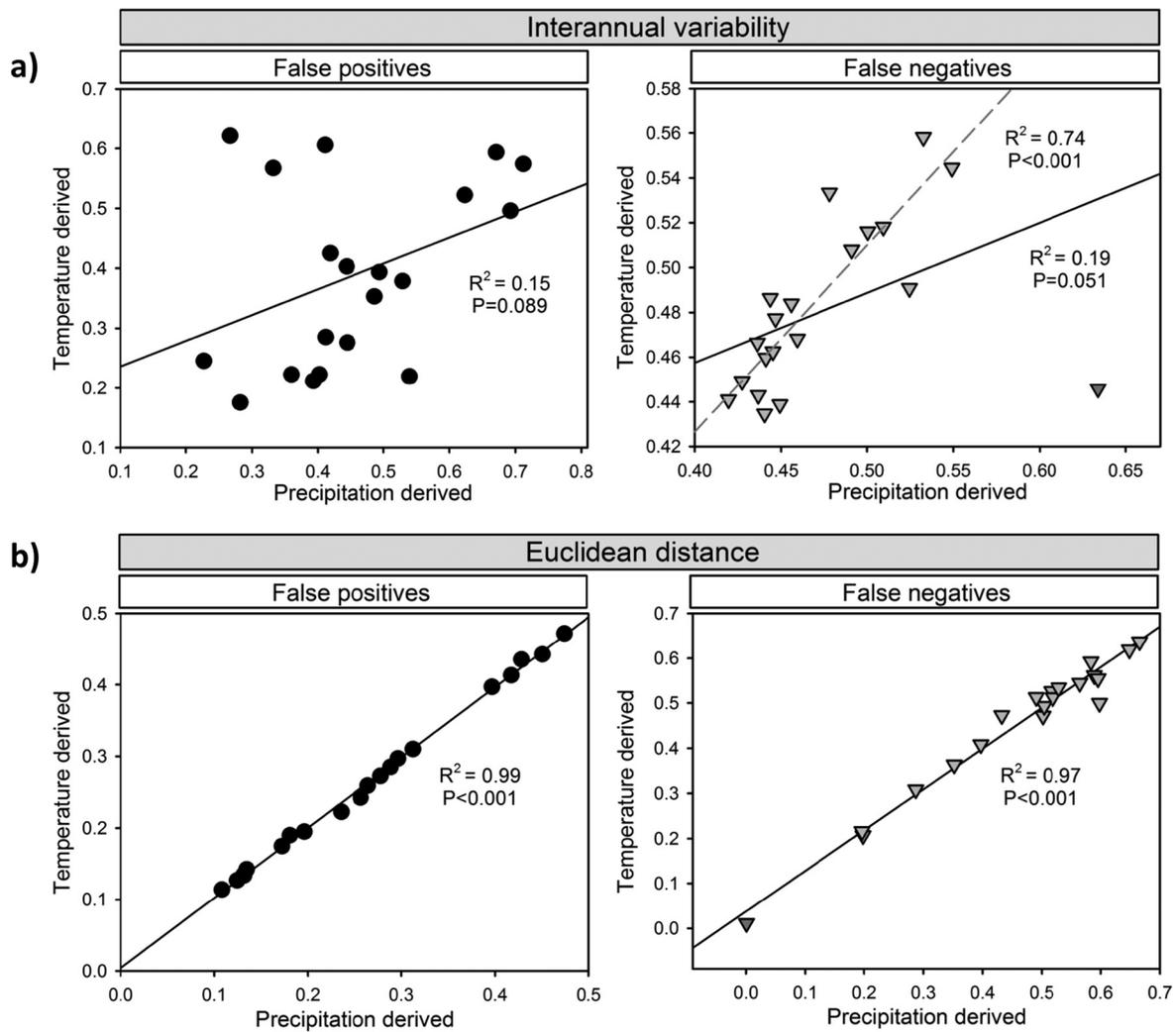
Fig. 4. Comparison of temperature and precipitation results based on the species-specific normalized mean differences ($|D_{null} - D_{un}|$ and $|D_{null} - D_{ov}|$) for interannual variability and Euclidean distance. Note that for Euclidean distance the results have a high degree of similarity between temperature-derived and precipitation-derived uncertainty layers.

model of a species that is distributed in areas of high topographic heterogeneity will have lower performance than one of a species that comes from areas of relatively low topographic heterogeneity, and that poor performance is more likely to result in under-prediction of the species' actual range.

The second largest degree of association was between interannual climatic variability and false negatives (Fig. 5), which highlights the importance of establishing the appropriate temporal relationships between species observations and

environmental layers as a step towards improving model performance. The processes that condition a species distribution operate at different spatial and temporal scales (e.g., Wiens 1989). Here we use interannual climatic variability, a metric that is highly synchronized with ecological processes that affect the distribution of the taxa under study. However, for some species this strategy can have implications that influence the interpretation of our results. For example, artificially influenced resource availability (e.g., bird feeders) can provide inaccurate cues regarding
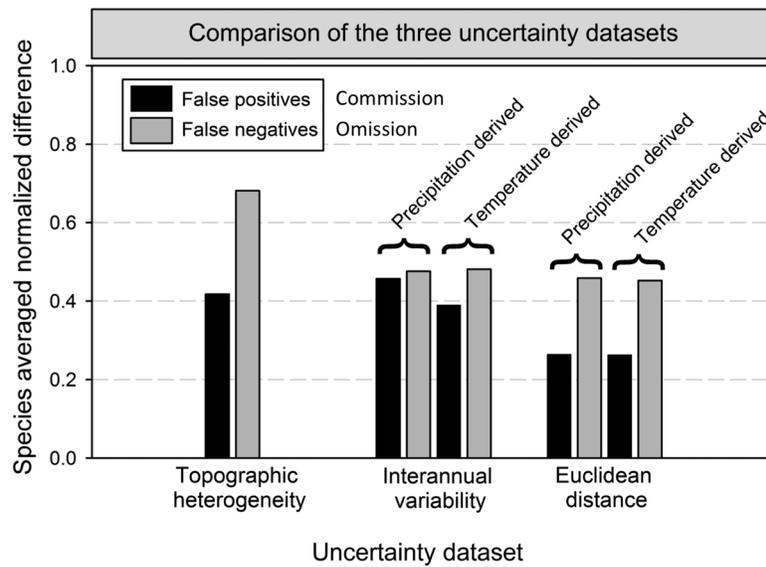
Fig. 5. Results for normalized mean differences averaged across all species. Averaged values corresponding to false positives (commission errors) are shown in black and averaged values corresponding to false negatives (omission errors) are shown in grey. Note that results for interannual variability and Euclidean distance have two results, the first corresponding to precipitation and the second one corresponding to temperature.

habitat quality (Robb et al. 2008), causing a temporal and/or spatial mismatch between the species and the natural availability of the resource. Also, natural climatic oscillations occur at multiple temporal scales, and organisms are adapted to cope with this variability at some scales better than others. For example, plants in xeric environments are well adapted to high variability in temperature at short time scales (i.e., diurnal), while a change in the pattern of interannual precipitation for the same system can have profound impacts on levels of physiological stress (Freas and Kemp 1983). To directly incorporate the effect of natural climatic variability on SDMs requires a direct connection between each particular species observation and the year when the observation was made. This is not always possible due to the lack of high-resolution gridded time series climate data for most of the world. When time-series climatic data is not available, we recommend that model confidence should be based on the life cycles of the taxa under scrutiny. For example, perennial plant species or animal species that do not migrate might be less sensitive to interannual variability than annual or migratory species.

The assumption that climatic similarity be-

tween two points in space is a function of the distance between them does not hold under every circumstance. Climatic similarity between two points will also depend on their difference in elevation; the Euclidean distance metric we used here integrates both vertical and horizontal distance to capture the effect on SDM performance of relative weather station density in topographically similar areas. Our results for this distance-based analysis show a clear difference in how omission vs. commission errors respond (Fig. 5). As an example, we can compare India, with a robust network of highly dense precipitation weather stations (0.001 stations/km^2) recording over many decades, to its neighbor Myanmar, where the density of weather stations (0.00002 stations/km^2) is two orders of magnitude less. The Euclidean distance alone is not enough to characterize the uncertainty in the interpolations of climatic parameters in both countries. It is also critical to know where these weather stations are located in relationship to the underlying environmental variability. Regions in the north of India will require a considerably larger number of weather stations to accurately characterize local precipitation patterns due to their high topographic heterogeneity. Conversely,

South Myanmar may be climatically well characterized by the current density of weather stations. We advocate for the inclusion and sharing of weather station locations used to build interpolated climatic products. Open access to this information will allow users to develop their own uncertainty metrics, and to identify regions where they need to be particularly careful when interpreting the results of SDMs.

Our results are based on the assumption that the spatial distributions of the twenty bird species we analyze here are mainly determined by climate. However, for any taxon, it is unlikely that climatic variables alone will shape their realized distribution. Although examining whether climate determines these species' ranges is not the goal of this study, we acknowledge that non-climatic factors can also be responsible for a false positive or a false negative. Alternative factors responsible for absence data predicted as presence are not simple (see Lobo et al. 2010); and possibilities can be grouped into two categories: (1) species related factors (e.g., locality climatically favorable but dispersal barriers prevent occurrence, interspecific interactions, local extinctions, or limited resources), and (2) extrinsic factors (e.g., incomplete surveys and biased information). Because we chose to work with a multiyear observational dataset, our results should be less affected by these extrinsic factors, leaving species-related issues potentially contributing to errors. In the case of presence data predicted as absence, our analysis suggests that the underlying climatic data is not only unsuitable, but also incorrectly characterized by the interpolation due to high uncertainty. However, the underlying climate may be unsuitable and correctly characterized by the interpolation, requiring alternative explanations for why the observation was recorded as a presence. Such a result could be attributed to source-sink dynamics, transient occupancy observed by chance, or even artificial food availability. This will not change the implicit SDM assumption that documented species observations always represent suitable habitat. In other words, the MaxEnt algorithm assumes that a species will always choose the appropriate habitat. However, it is possible for an individual of a species to err in selecting climatically suitable habitat. For example, juvenile birds will have less experience in choosing locations for a nesting site; also, late arrivers will have fewer options in site selection. These outliers under normal circumstances will also have a lower chance of successfully breeding (Martin and Roper 1988). Indeed, part of the process of natural selection is individuals making mistakes in the selection of suitable habitat. These ecological mistakes, essential in the process of evolution and currently overlooked by SDM practice, have the potential to be applied to our understanding of how a species will respond to climate change.

The reasons behind our choice of taxonomic group to test our hypotheses were mainly based on species occurrence data quality and availability, which also limits inference and application to other taxonomic groups. Yet we believe that the same principles and mechanisms explored here apply also to other organisms, and that the results can be cautiously extrapolated to other taxonomic groups. However, underlying quality and accuracy of the gridded information used as environmental layers should not be the only direction to look for sources of over- and under-prediction errors and ways to improve the models. Coastal redwood (*Sequoia sempervirens*), as an example of a relict plant species, has its suitable bioclimatic envelope restricted to a narrow 50 km belt in the coast of California. However, some parts of its actual current distribution may be better explained by factors of land use change rather than climate (Pyke 2004).

Attributing the omission and commission errors in areas that were predicted by the model as presence due to climate alone will be incorrect. A simple yet difficult-to-achieve recommendation for resource managers applying the results of SDMs is to asses all the potential sources of uncertainty and to focus on characterization of the sources that provide the largest amount of error. However, integrated tools to aid resource managers in evaluating multiple sources uncertainty, while needed, are not yet available.

Finally, a useful SDM is not only precise, but also accurate. While past literature that deals with SDM environmental layer uncertainty focuses on model precision (e.g., Kriticos and Leriche 2010) here we ask how divergent results can be between known and modeled distributions if a parameter is uncertain. We quantified

the relationship between omission and commission errors in the predictions and the degree of uncertainty in the interpolated environmental input layers. We attribute decrease in SDM performance to the three aspects of uncertainty evaluated here; however, not all of them were identified as equally important sources of over- and under-prediction errors in SDMs. Our results confirm the importance of establishing appropriate relationships in time and space between species and environmental layers. Uncertainty characterizations for environmental layers can provide operational criteria for the selection of species observations fed into SDMs, and help identify conditions where users can weigh their degree of confidence when making decisions based on a SDM.

Acknowledgments

Literature Cited

Angert, A. L. 2009. The niche, limits to species' distributions, and spatiotemporal variation in demography across the elevation ranges of two monkeyflowers. Proceedings of the National Academy of Sciences USA 106:19693–19698.

Beale, C. M. and J. J. Lennon. 2012. Incorporating uncertainty in predictive species distribution modelling. Philosophical Transactions of the Royal Society B 367:247–258.

Beyer, H. 2012. Geospatial modelling environment. Spatial Ecology, Toronto, Ontario, Canada.

BirdLife. 2012. BirdLife International Species factsheets. BirdLife International, Cambridge, UK.

Calder, W. A. 1994. When do hummingbirds use torpor in nature? Physiological Zoology 67:1051–1076.

Conover, W. J. 1999. Practical nonparametric statistics. John Wiley & Sons, New York, New York, USA.

Daszak, P., C. Zambrana-Torrelio, T. Bogich, M. Fernández, J. Epstein, K. Murray, and H. Hamilton. 2012. Interdisciplinary approaches to understanding disease emergence: The past, present, and future drivers of *Nipah* virus emergence. Proceedings of the National Academy of Sciences USA 110:3681–3688.

de Oliveira, G., M. B. Araújo, T. F. Rangel, D. Alagador, and J. A. F. Diniz-Filho. 2012. Conserving the Brazilian semiarid (Caatinga) biome under climate change. Biodiversity and Conservation 21:2913–2926.

Dixon, P. 1993. The bootstrap and the jackknife: describing the precision of ecological indices. Pages 290–318 *in* S. M. Scheiner and J. Gurevitch, editors. Design and analysis of ecological experiments. Oxford University Press, New York, New York, USA.

Drut, M. S., J. A. Crawford, and M. A. Gregg. 1994. Brood habitat use by sage grouse in Oregon. Western North American Naturalist 54:170–176.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.

Elith, J., H. Graham, P. Anderson, M. Dudik, S. Ferrier, A. Guisan, J. Hijmans, F. Huettmann, R. Leathwick, and A. Lehmann. 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. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. Yates. 2011. A statistical explanation of MaxEnt for ecologists. Ecology 45:1372–1381.

FAOCLIM-2. 2005. World-wide agroclimatic database. Food and Agriculture Organization, Rome, Italy.

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, and L. Roth. 2007. The shuttle radar topography mission. Reviews of Geophysics 45:RG2004.

Feeley, K. J. and M. R. Silman. 2011. Keep collecting: accurate species distribution modelling requires more collections than previously thought. Diversity and Distributions 17:1132–1140.

Fernández, M., S. Blum, S. Reichle, Q. Guo, B. Holzman, and H. Hamilton. 2009. Locality uncertainty and the differential performance of four common niche-based modeling techniques. Biodiversity Informatics 6:36–52.

Fernández, M., H. Hamilton, O. Alvarez, and Q. Guo. 2012. Does adding multi-scale climatic variability

improve our capacity to explain niche transferability in invasive species? Ecological Modelling 246:60–67.

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

Fordham, D. A., H. Resit Akçakaya, M. B. Araújo, J. Elith, D. A. Keith, R. Pearson, T. D. Auld, C. Mellin, J. W. Morgan, and T. J. Regan. 2012. Plant extinction risk under climate change: are forecast range shifts alone a good indicator of species vulnerability to global warming? Global Change Biology 18:1357–1371.

Freas, K. and P. R. Kemp. 1983. Some relationships between environmental reliability and seed dormancy in desert annual plants. Journal of Ecology 71:211–217.

Ghil, M. 2002. Natural climatic variability. Pages 544–549 in T. Munn, editor. Encyclopedia of global environmental change. John Wiley & Sons, Chichester, UK.

Giuliano, W. M., and R. S. Lutz. 1993. Quail and rain: what's the relationship. Proceedings of the National Quail Symposium 3:64–68.

González, C., O. Wang, S. E. Strutz, C. González-Salazar, V. Sánchez-Cordero, and S. Sarkar. 2010. Climate change and risk of leishmaniasis in North America: predictions from ecological niche models of vector and reservoir species. PLoS Neglected Tropical Diseases 4:e585.

Graham, C. H., J. Elith, R. J. Hijmans, A. Guisan, A. Townsend Peterson, and B. A. Loiselle. 2008. The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology 45:239–247.

Hijmans, R., S. Cameron, J. Parra, P. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

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.

Howell, S. and S. Webb. 1995. A guide to the birds of Mexico and northern Central America. Oxford University Press, New York, New York, USA.

ISI. 2012. Institute for Scientific Information. http://thomsonreuters.com/products_services/science/science_products/a-z/web_of_science

Johnsgard, P. A. 1983. The grouse of the world. University of Nebraska Press, Lincoln, Nebraska, USA.

Johnson, G. L., C. Daly, G. H. Taylor, and C. L. Hanson. 2000. Spatial variability and interpolation of stochastic weather simulation model parameters. Journal of Applied Meteorology 39:778–796.

Kamilar, J. M. 2009. Environmental and geographic correlates of the taxonomic structure of primate communities. American Journal of Physical Anthropology 139:382–393.

Kamino, L. H. Y., J. R. Stehmann, S. Amaral, P. De Marco Jr, T. F. Rangel, M. F. de Siqueira, R. De Giovanni, and J. Hortal. 2012. Challenges and perspectives for species distribution modelling in the neotropics. Biology Letters 8:324–326.

Kreft, H. and W. Jetz. 2007. Global patterns and determinants of vascular plant diversity. Proceedings of the National Academy of Sciences USA 104:5925–5930.

Kriticos, D. J. and A. Leriche. 2010. The effects of climate data precision on fitting and projecting species niche models. Ecography 33:115–127.

Levsen, N. D., P. Tiffin, and M. S. Olson. 2012. Pleistocene speciation in the genus Populus (Salicaceae). Systematic Biology 61:401–412.

Li, W., Q. Guo, and C. Elkan. 2011. Can we model the probability of presence of species without absence data? Ecography 34:1096–1105.

Lobo, J. M., A. Jiménez Valverde, and J. Hortal. 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography 33:103–114.

Loiselle, B. A., P. M. Jorgensen, T. Consiglio, I. Jimenez, J. G. Blake, L. G. Lohmann, and O. M. Montiel. 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? Journal of Biogeography 35:105–116.

Lomolino, M. V. 2010. Biogeography. Fourth edition. Sinauer Associates, Sunderland, Massachusetts, USA.

Martin, T. E. and J. J. Roper. 1988. Nest predation and nest-site selection of a western population of the Hermit Thrush. Condor 90:51–57.

McInerny, G. J. and D. W. Purves. 2011. Fine-scale environmental variation in species distribution modelling: regression dilution, latent variables and neighbourly advice. Methods in Ecology and Evolution 2:248–257.

Naimi, B., A. K. Skidmore, T. A. Groen, and N. A. S. Hamm. 2011. Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modelling. Journal of Biogeography 38:1497–1509.

NatureServe. 2012. NatureServe Explorer: an online encyclopedia of life. Version 7.1. NatureServe, Arlington, Virginia, USA.

Peterson, A. and Y. Nakazawa. 2008. Environmental data sets matter in ecological niche modelling: an example with Solenopsis invicta and Solenopsis richteri. Global Ecololgy and Biogeography 17:135–144.

Peterson, A. T., J. Soberon, R. G. Pearson, R. P.

Anderson, E. Martinez-Meyer, M. Nakamura, and M. B. Araújo. 2011. Ecological niches and geographic distributions (MPB-49). Princeton University Press, Princeton, New Jersey, USA.

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.

Purvis, A., S. A. Fritz, J. Rodríguez, P. H. Harvey, and R. Grenyer. 2011. The shape of mammalian phylogeny: patterns, processes and scales. Philosophical Transactions of the Royal Society B 366:2462–2477.

Pyke, C. R. 2004. Habitat loss confounds climate change impacts. Frontiers in Ecology and the Environment 2:178–182.

Robb, G. N., R. A. McDonald, D. E. Chamberlain, and S. Bearhop. 2008. Food for thought: supplementary feeding as a driver of ecological change in avian populations. Frontiers in Ecology and the Environment 6:476–484.

Rosell, J. A., M. E. Olson, R. Aguirre-Hernandez, and F. J. Sanchez-Sesma. 2012. Ontogenetic modulation of branch size, shape, and biomechanics produces diversity across habitats in the Bursera simaruba clade of tropical trees. Evolution and Development 14:437–449.

Roura-Pascual, N., C. Hui, T. Ikeda, G. Leday, D. M. Richardson, S. Carpintero, X. Espadaler, C. Gómez, B. Guénard, and S. Hartley. 2011. Relative roles of climatic suitability and anthropogenic influence in determining the pattern of spread in a global invader. Proceedings of the National Academy of Sciences USA 108:220–225.

Saab, V. A. and J. G. Dudley. 1998. Responses of cavity-nesting birds to stand-replacement fire and salvage logging in ponderosa pine/Douglas-fir forests of southwestern Idaho. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Boise, Idaho, USA.

Sauer, J. R., J. E. Fallon, and R. Johnson. 2003. Use of North American Breeding Bird Survey data to estimate population change for bird conservation regions. Journal of wildlife Management 67:372–389.

Sauer, J. R., J. E. Hines, and J. Fallon. 2006. The North American breeding bird survey, results and analysis 1966-2007. United States Geological Survey, Patuxent Wildlife Research Center, Laurel, Madison, USA.

Schroeder, M. A., and C. E. Braun. 1993. Partial migration in a population of greater prairie-chickens in northeastern Colorado. Auk 110:21–28.

Soberon, J. M., J. B. Llorente, and L. Oñate. 2000. The use of specimen-label databases for conservation purposes: an example using Mexican Papilionid and Pierid butterflies. Biodiversity and Conservation 9:1441–1466.

Stigall, A. L. 2012. Using ecological niche modelling to evaluate niche stability in deep time. Journal of Biogeography 39:772–781.

Synes, N. W. and P. E. Osborne. 2011. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change. Global Ecology and Biogeography 20:904–914.

Vaclavik, T. and R. K. Meentemeyer. 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion. Diversity and Distributions 18:73–83.

Vose, R. S., R. L. Schmoyer, P. M. Steurer, and T. C. Peterson. 1992. The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. Oak Ridge National Lab, Carbon Dioxide Information Analysis Center, Oak Ridge, Tennessee, USA.

Wiens, J. A. 1989. Spatial scaling in ecology. Functional Ecology 3:385–397.

Zelazowski, P., Y. Malhi, C. Huntingford, S. Sitch, and J. B. Fisher. 2011. Changes in the potential distribution of humid tropical forests on a warmer planet. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 369:137–160.

Zhang, J. and M. F. Goodchild. 2002. Uncertainty in geographical information. Taylor & Francis, London, UK.