

LOCALITY UNCERTAINTY AND THE DIFFERENTIAL PERFORMANCE OF FOUR COMMON NICHE-BASED MODELING TECHNIQUES

MIGUEL A. FERNANDEZ^{1,5}, STANLEY D. BLUM², STEFFEN REICHLÉ³, QINGHUA GUO¹, BARBARA HOLZMAN⁴, AND HEALY HAMILTON⁵

1 Sierra Nevada Research Institute, University of California Merced

2 Research Informatics, California Academy of Sciences

3 The Nature Conservancy

4 Department of Geography & Human Environmental Studies, San Francisco State University

5 Center for Biodiversity Research, California Academy of Sciences

Abstract.— We address a poorly understood aspect of ecological niche modeling: its sensitivity to different levels of geographic uncertainty in organism occurrence data. Our primary interest was to assess how accuracy degrades under increasing uncertainty, with performance measured indirectly through model consistency. We used Monte Carlo simulations and a similarity measure to assess model sensitivity across three variables: locality accuracy, niche modeling method, and species. Randomly generated data sets with known levels of locality uncertainty were compared to an original prediction using Fuzzy Kappa. Data sets where locality uncertainty is low were expected to produce similar distribution maps to the original. In contrast, data sets where locality uncertainty is high were expected to produce less similar maps. BIOCLIM, DOMAIN, Maxent and GARP were used to predict the distributions for 1200 simulated datasets (3 species x 4 buffer sizes x 100 randomized data sets). Thus, our experimental design produced a total of 4800 similarity measures, with each of the simulated distributions compared to the prediction of the original data set and corresponding modeling method. A general linear model (GLM) analysis was performed which enables us to simultaneously measure the effect of buffer size, modeling method, and species, as well as interactions among all variables. Our results show that modeling method has the largest effect on similarity scores and uniquely accounts for 40% of the total variance in the model. The second most important factor was buffer size, but it uniquely accounts for only 3% of the variation in the model. The newer and currently more popular methods, GARP and Maxent, were shown to produce more inconsistent predictions than the earlier and simpler methods, BIOCLIM and DOMAIN. Understanding the performance of different niche modeling methods under varying levels of geographic uncertainty is an important step toward more productive applications of historical biodiversity collections.

Key words.— georeferencing; spatial uncertainty; ecological niche modeling; comparative performance; fuzzy kappa

Our maps of species' distributions ultimately derive from primary observations of their occurrence in nature. Because these data are typically sparse in comparison to the complete range of a species, biologists have devised a variety of methods to visualize and analyze species ranges based on field samples. These range from simply plotting occurrence points on maps, to drawing a free-form line around peripheral locality records. Recently, researchers interested in species' distributions have been able to integrate spatial tools and environmental data to produce probability distribution maps that indicate variation in habitat suitability, maps which convey more

information than either point locality maps or outline maps. Known also as ecological niche models (ENM; *sensu* Grinnell 1917), these maps are the result of integrative algorithms embedded in a GIS framework that use the taxonomic and geographic data associated with specimens and/or observations and fine scale environmental data to produce a set of rules that identify the environmental space where the species was collected or observed (Peterson and Vieglais 2001). This environmental space can be projected onto geographic space to identify appropriate conditions where the species may occur, resulting in a modeled distribution.

Despite the fact that these presence-only inferential maps are abstract representation of species ranges, they are still valuable summaries of

correspondence email: mfernandez5@ucmerced.edu.

biogeographic information, and have been applied to a broad range of topics, from theoretical ecology and evolution (Leathwick and Whitehead 2001; Hugall et al. 2002; Graham et al. 2004), to practical uses in conservation (Bustamante 1997; Corsi et al. 1999; Anderson et al. 2002; Raxworthy et al. 2003; Araújo et al. 2004) agriculture, invasive species (Higgins et al. 2000; Welk et al. 2002; Underwood et al. 2004), and human health (Mills and Childs 1998; Peterson and Shaw 2003).

While tremendous progress has been achieved on many aspects of building and evaluating ENM (Guisan and Zimmermann 2000; Pearce and Ferrier 2000; Williams and Hero 2001; Hirzel et al. 2002; Stockwell and Peterson 2002; Brotons et al. 2004; Reese et al. 2005; Barry and Elith 2006; Pearson et al. 2006; Wisz et al. 2008), enhanced frameworks for assessing errors and uncertainties have not been fully developed. Specifically, uncertainty in the organism occurrence data (see Fig. 1) has not been fully explored (Graham et al. 2004; Murphy et al. 2004; Soberon and Peterson 2004; Wiecek et al. 2004; Rowe 2005; Guo et al. 2008). Understanding the susceptibility of ENM methods to the positional error associated with a collection event becomes a critical factor in selecting a method to use in a particular case.

Comparing ENM performance against uncertainty

A predicted species distribution is generally determined by three elements: the algorithm or modeling method, the environmental layers upon which it is based, and the occurrence data. Although researchers have explored how each of these elements contributes separately or together to the overall performance of the technique, as yet, there is no agreement on the influence of uncertainties on ENM. Some studies show that different methods perform surprisingly similarly (Peterson et al. 2007; Peterson et al. 2008), while others studies show that alternative ENM produce highly distinct outputs when predicting species' geographic ranges (Manel et al. 1999; Elith et al. 2006; Pearson et al. 2006; Phillips et al. 2006; Kelly et al. 2007; Peterson et al. 2007; Tsoar et al. 2007; Ortega-Huerta and Peterson 2008). Further research is required to address these discrepancies in model performance. Specifically, standardized and improved parameterization and enhanced evaluation tools are needed to tease apart these differences in modeling outputs (Araújo and

Guisan 2006; Peterson et al. 2008). While few studies have measured the sensitivity of distribution models to grid cell size in the environmental layers (Guisan et al. 2007), others have addressed the effect of remote sensing derived products as alternative environmental layers in ENM (Parra et al. 2004; Roura-Pascual et al. 2004; Peterson et al. 2006; Zimmermann et al. 2007; Bradley and Fleishman 2008; Buermann et al. 2008). Limited studies attempted to incorporate true absence and more meaningful pseudo-absence data in ENM (Manel et al. 2001; Brotons et al. 2004; Engler et al. 2004; Chefaoui and Lobo 2007; Phillips 2008). Numerous tests have also addressed the effect of occurrence data quantity on ENM (Peterson and Cohoon 1999; Stockwell and Peterson 2002; Kadmon et al. 2003; Hernandez et al. 2006; Pearson et al. 2007; Wisz et al. 2008).

However quality in species occurrence data can also have profound consequences in ENM. Localities may be geographically biased, for example, highly correlated with rivers and access roads (Reddy and Davalos 2003), or collected using different sampling intensity and sampling methods (Anderson 2003). Localities that have been retrospectively georeferenced have uncertainty associated with the lack of geographic details in the textual descriptions (Beaman et al. 2004; Rowe 2005; Chapman and Wiecek 2006). More standardized techniques have been developed that allow a better quantification of the positional error of occurrence data (Murphy et al. 2004; Wiecek et al. 2004; Guralnick et al. 2007; Guo et al. 2008). The effect of the positional error on resulting ENM output using different methodologies has been underexplored. Recently, Graham et al. (2008) evaluated how locality uncertainty affects the performance of ten common niche modeling techniques by comparing a control model calibrated using the original accurate data to an error treatment where the positional accuracy of the data was degraded randomly in a radius of 5 km. Even though they demonstrated that model performance can change markedly with increased locality uncertainty. Their single randomization treatment is not sufficient to establish the relationship between the magnitude of locality uncertainty and ENM performance.

Close to 2.5 billion specimens (Duckworth et al. 1993) have been collected and housed in natural

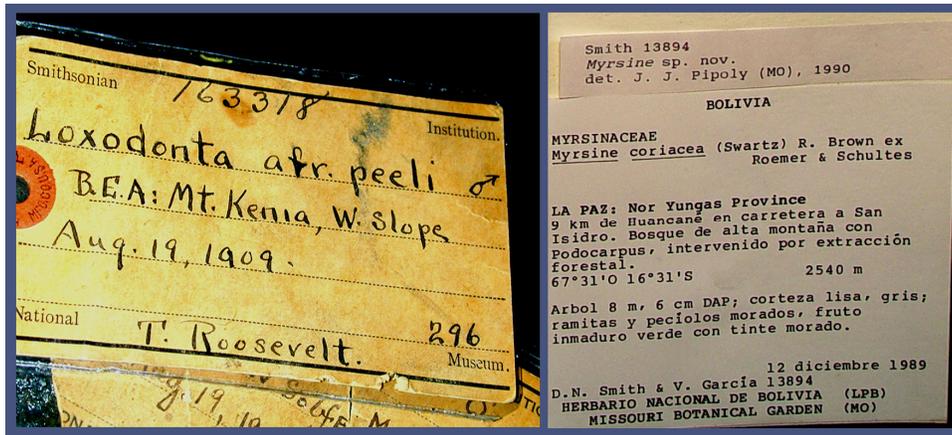


Figure 1. Differences in geographic uncertainty between two labels of specimens housed in natural history museum collections. Left, a vague description of the collection location; right, a much more precise description of the collecting location.

history museums by different collectors, at different times, with different sampling techniques (see Fig. 1). As a consequence, the geographic information associated with specimen collections has very different levels of geographic uncertainty.

Many historical localities were recorded only as textual descriptions, without geographic coordinates, which effectively makes them unavailable to GIS-based analyses. As discussed above, the subsequent interpretation of textual localities as geocoordinates, known as retrospective georeferencing can introduce still greater spatial uncertainty (Proctor 2004; Wieczorek et al. 2004; Rowe 2005). In order to make more effective use of the wealth of biodiversity information stored in natural history museums, it is critical to fully explore the sensitivity of ENM techniques to different levels of geographic uncertainty in the organism occurrence data. Only by quantifying how uncertainty interacts with modeling methods and landscape variability will we be able to understand the reliability of predicted distributions or their suitability to particular uses.

METHODS

Because distribution modeling outputs differ, no simple statistic is available to measure intermodel performance across all approaches (Phillips et al. 2006; Lobo et al. 2008; Peterson et al. 2008). Many commonly used methods give

results as probability surfaces, rather than binary distributions, in which the species is predicted to occur or not occur in a particular grid cell. Evaluating these models directly requires selecting an arbitrary threshold value to create the binary prediction, which might then result in under and over prediction. Given that our primary interest was to assess how accuracy degrades under increasing uncertainty, we chose to measure model performance indirectly, through the consistency of repeated simulations.

In this study we used Monte Carlo simulations and a similarity measure to assess the consistency of predictions across three variables: locality uncertainty, niche modeling method, and species. We created data sets with known levels of locality uncertainty and compared them to an original prediction using a similarity measure, Fuzzy Kappa (discussed further below). We expected data sets where locality uncertainty is low to produce distribution maps that are similar to the original. In contrast, we expected data sets where locality uncertainty is high to produce maps that are less similar or more inconsistent. In addition, we wanted to examine whether the response to uncertainty would differ across modeling methods and whether taxonomic or landscape variability would also influence this sensitivity. Ultimately, we would like to know what degree of data quality, estimated by maximum error distance we can tolerate to produce predictions that are sufficiently accurate.

Species selection criteria

Three species of Bolivian frogs were selected for this analysis: *Oreobates cruralis*, *Leptodactylus elenae*, and *Pleurodema marmoratum*. These species were selected for the following reasons: (a) their geographic ranges are comparable in area; (b) none is narrowly endemic or broadly distributed; and (c) they represent each of the main geographic areas in Bolivia, regions that are expected to have very different landscape characteristics: *O. cruralis* is widely distributed in the Yungas region of Bolivia, *L. elenae* is distributed in the lowlands of Bolivia, commonly associated with savannas; and *P. marmoratum* is restricted to the highlands.

Simulating locality uncertainty by random displacement

In our experiments, an “original data set” is the group of non-repeated collecting localities for each species, expressed as latitude and longitude, either taken by one of us using a GPS (SR), or georeferenced by one of us (MF) (Table 1 and Fig. 2). Even though we only use occurrences with positional uncertainty represented by maximum error estimates of less than 1 km, we note the goal of this exercise was not to evaluate how well the models fit the real distribution of the species, but to test what is the effect of degrading the localities across a broad range of positional accuracies.

An “original ecological niche model” is the output produced by one of the modeling techniques using the original dataset and a set of 19 standard bioclimatic variables derived from Worldclim 1.4 (Hijmans et al. 2005) at a spatial resolution of ~1 km². From this “original dataset” we generated 100 different “new data sets” that simulate an increased level of locality uncertainty using the Random Point Generator ArcView 3.x extension (Jenness Enterprises, 2005), which produces a random selection of points approaching a uniform distribution (see Fig. 3). We used this tool to randomly displace every point in each of the original datasets to a new position within a selected buffer distance. Each of these new 100 points per buffer size and per locality is combined randomly with other generated points for other localities to form a “new dataset”. This new dataset is composed of the same number of point

localities as the “original dataset” but located at different distances within the selected buffer. Therefore, buffer size represents our experimental model of locality uncertainty. This simulation is similar to the point-radius method of retrospective georeferencing described by Wieczorek et al. (2004). This method encompasses a wide variety of processes that contribute to different degrees of uncertainty (see Guo et al. 2008).

Although it is possible to derive a probability density function for each locality (Guo et al. 2008), something other than equi-probable or even, these functions entail prior knowledge of the processes that produced the data, such as assumptions on referenced objects that are used to georeference species localities, and assumptions on spatial relationships that describe the species localities.. While these may be reasonable assumptions, their purpose is to minimize the effect of uncertainty and extract better information from occurrence data. That is not our purpose here. In this study we are measuring the effect of uncertainty, so our goal is to incorporate uncertainty in a reasonable and easily understood way. We chose to represent uncertainty as a circle around the original point, where any point in that area has an equal probability of selection. This is currently a common practice in estimating locality uncertainty in occurrence data derived from retrospective georeferencing (Wieczorek et al. 2004; Guo et al. 2008).

Modeling methods

Four distribution modeling techniques were used to predict the distributions for 1200 simulated datasets (3 species x 4 buffer sizes x 100 randomized data sets). Thus, our experimental design produced a total of 4800 similarity measures, as each of these predicted distributions was compared to the prediction produced from the original data set and corresponding modeling method. Two of the methods we used are based on a climatic envelope concept and presence only localities, BIOCLIM (Busby 1991) and DOMAIN (Carpenter et al. 1993). The other two methods use both presence and pseudo-absence localities, GARP (Stockwell and Peters 1999), which is based on a genetic algorithm, and Maxent (Phillips et al. 2006) which is based on the maximum entropy concept.

Table 1. List of number of non-repeated localities for the original datasets per species

Species Name	Number of localities
<i>Oreobates cruralis</i>	38
<i>Leptodactylus elenae</i>	39
<i>Pleurodema marmoratum</i>	29

BIOCLIM relates occurrence localities to climatic conditions, and produces a single rule that identifies all areas with a similar climate to the locations of the species within a minimal rectilinear “climatic envelope”. In BIOCLIM, user-specified thresholds for each environmental predictor are identified to define the multidimensional environmental space. This can be projected onto landscapes producing a model of appropriate climatic conditions for the species. However, the assumption that species’ distributions are controlled by a defined climatic envelope is largely simplistic. Species ranges in nature are controlled by a complex combinations of factors and unlikely to be a box shape in the environmental space. DOMAIN is a tool based on a point-to-point similarity metric (Gower metric). Similarity between the site of interest and each of the recorded present occurrence locations is calculated by summing the standardized distance between the two points for each predictor variable. The standardization is achieved by dividing the distance by the predictor variable range for the presence sites, equalizing the contribution from each predictor variable. The standardized distance is subtracted from 1 to obtain the complementary similarity (Carpenter et al. 1993). Predictions are not to be interpreted as maps of probability of occurrence, but as a measure of classification confidence. Neither of these two methods provides explanatory power of the relevant factors controlling the species’ distributions, nor statistically quantifies the variance, thus, the accuracy of the predictions is unknown (Stockwell 2006).

GARP is a non-deterministic model that uses a machine learning approach to test several inferential algorithms (e.g. atomic, logistic regression, range rules, and negated range) in an iterative manner to develop multiple sets of rules that will provide multiple solutions given the same input. For each new iteration, GARP divides the

occurrences in: (a) training data, which is used to produce the rules that will define the model, and (b) testing data, which is used to internally evaluate the model based on omission and commission errors. In the next iteration the data is resampled, a new training and testing data set is produced, and the process starts over again. This process is repeated until the program can not create an improved model (Stockwell and Peters, 1999). Since GARP doesn’t produce a single probabilistic output, to deal with this stochasticity, multiple runs can be performed within the same GARP session, producing a chosen number of output prediction maps. GARP reports measures of omission and commission errors for each generated model, and provides the option to select a ‘best subset’ based on these accuracy measures. The predictions for the ‘best subset’ models can be arithmetically combined to produce a final predicted distribution map (Anderson et al. 2003).

Maxent (Phillips et al. 2004), estimates a species niche by finding the probability distribution of maximum entropy, subject to the constraint that the expected value of each environmental variable under this estimated distribution matches its empirical average. Continuous environmental data can also be entered as both quadratic and product features, thereby adding further constraints to the estimation of the probability distribution by restricting it to be within the variance for each environmental predictor and covariance for each pair of environmental predictors. The program starts with a uniform probability distribution, and iteratively alters one weight at a time to maximize the likelihood of reaching an optimum probability distribution. The algorithm is guaranteed to converge, and therefore the outputs are deterministic. Since the traditional implementation of maximum entropy is prone to over-fitting the probability distribution, Maxent actually employs a relaxation method. It does not constrain the estimated distribution to the exact empirical

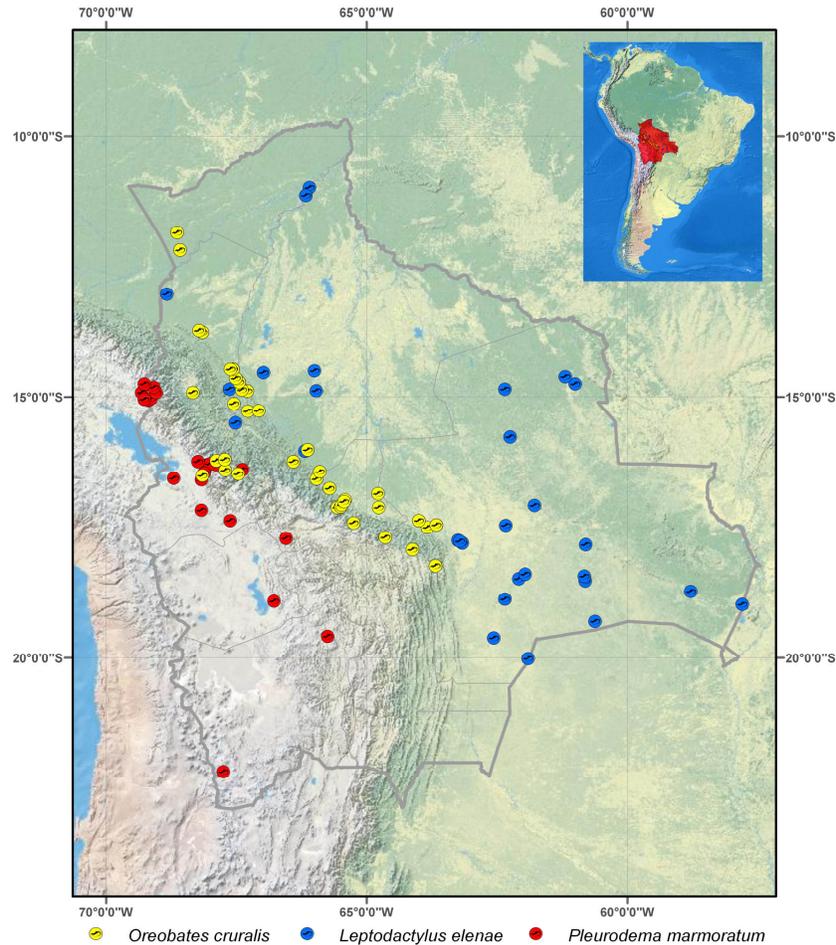


Figure 2. Original species occurrence datasets for the three species included in this study.

average, but to within the empirical error bounds of the average value for a given predictor, in a procedure called ‘regularization’ (Phillips et al. 2004).

Fuzzy Kappa

Every predicted distribution was standardized (rescaled) into an Idrisi Andes compatible grid format in which cell values range from 0 to 100 (see below):

$$x_b = \frac{(x_a - min)}{(max - min)}$$

where x_b is the rescaled value of each cell in the raster layer, x_a is the original value from the model

output, and *min* and *max* are the minimum and maximum values from the model output, respectively.

To compare the predicted distributions of simulated data sets against the original, we used the similarity measure called Fuzzy Kappa (Hagen 2003), implemented in the Map Comparison Kit 3.0 (Visser and de Nijs 2006). Fuzzy Kappa is based on the simple Kappa algorithm, however, it enables the comparison of two maps (both categorical and non-categorical data), and produces a similarity statistic that represents the average similarity of the entire map. The principal benefit of using Fuzzy Kappa over Kappa is that Kappa is based on binary logic, where the result of comparing the values of two corresponding cells is either “equal” or “different.” In contrast, Fuzzy Kappa uses a fuzzy logic where

the measure of similarity is continuous and based on the values of corresponding cells, as well as the distance to similar cells within a buffer defined by the user. This is based on the notion that the fuzzy representation of a cell depends on the cell itself and its neighboring cells with correspondingly lesser weight. This key distinction allows Fuzzy Kappa not only to evaluate differences but actual levels of difference, and models a human assessment of similarity more closely than simple Kappa (Visser and de Nijs 2006) (see Fig. 5).

Fuzzy Kappa is calculated in a similar manner as the traditional Kappa:

$$K_{fuzzy} = \frac{(S - E)}{(1 - E)}$$

where S is the average similarity over all cells based on fuzzy memberships, and E is the expected similarity. The fuzzy membership is used to account for the location error as shown in Fig. 4. In this study, we used the Gaussian distance decay functions to define the fuzzy membership (Visser and de Nijs 2006). Detailed discussion regarding Fuzzy Kappa can be found in Pontius (2000) and Hagen-Zanker et al. (2005).

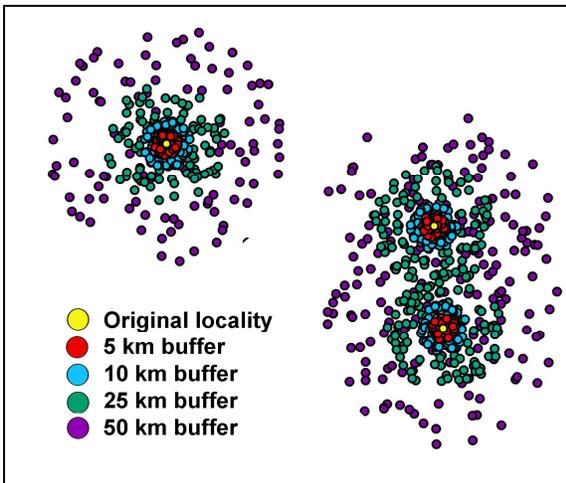


Figure 3. Random localities selected from a buffer zone, emulating different degrees of uncertainty in locality description.

As shown in Figure 4, the top A0 to A4 maps portray the ecological niche models based on the BIOCLIM algorithm and increasingly degraded localities from left (original localities) to right (localities degraded in a buffer of 50 km). The second row of maps portrays the Kappa map comparison based on consecutive comparison of

the original ecological niche model (A0) to each map resulting from increasingly degraded localities (A1, A2, A3 and A4). The bottom row of maps represents the Numerical Fuzzy Kappa map comparison based on consecutive comparison of the original ecological niche model (A2) to each of the maps created with degraded localities. Even though both indexes show a decrease in similarity with increasing buffer size, the value of the Kappa is too sensitive to small differences, and misses some of the basic similarity between the two maps. On the other hand, Fuzzy Kappa is a more conservative index that varies less dramatically when the position of a multi-pixel “object” shifts slightly, which makes it a better tool for measuring the similarity between two maps.

Experimental design

We measured how the similarity of predicted distributions changes in response to buffer size, an experimentally controlled continuous variable, as well as two categorical variables, species and modeling method. The similarity measure, Fuzzy Kappa, varies between zero and one. Our intention was to perform a general linear model (GLM) analysis, which would enable us to measure simultaneously the effect of buffer size, modeling method, and species using a two-way analysis of variance with an ordinary least squares regression, as well as test for interactions among all variables. The full factorial model was specified as:

$$Sp + Mm + Bfr + Sp \times Mm + Sp \times Bfr + Mm \times Bfr + Sp \times Mm \times Bfr$$

where *Sp* is the categorical effect for species, *Mm* is the categorical effect for modeling method, *Bfr* is the covariate, buffer, and interaction terms are specified with a multiplication symbol between the codes for the primary effects. The sample sizes were balanced, with every permutation of treatments evaluated with 100 simulated data sets.

RESULTS

The results of this study can be understood most directly through visualization. Figure 6 shows box plots of similarity measures for the series of buffer sizes within each modeling method and species combination. Several things are evident from this figure. First, large differences exist among the modeling methods; BIOCLIM scores were highest, while Maxent scores were lowest. Second, very

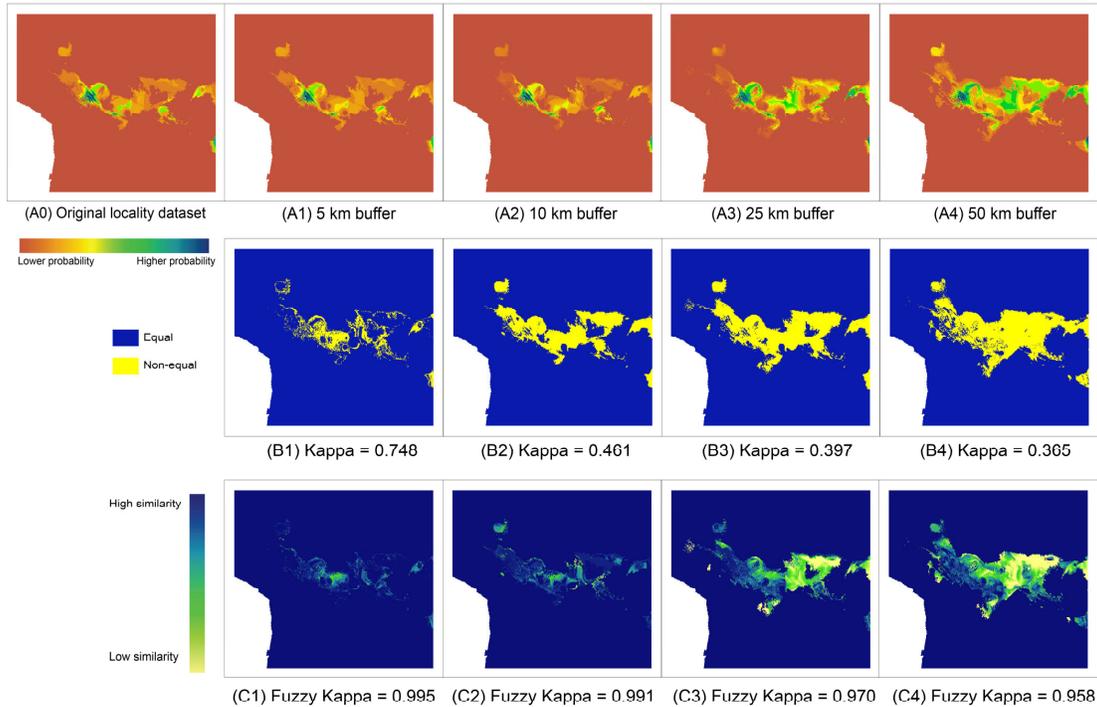


Figure 4. The top (A0 to A4) maps portray the ENM based on BIOCLIM. The second row portrays the Kappa comparison. The bottom row represents the numerical Fuzzy Kappa map comparison.

large differences also exist among the variances across treatment combinations; the largest variance is more than 700 times larger than the smallest. Third, within most combinations of species and modeling method, the mean similarity score tends to decrease with increasing buffer size (i.e., locality uncertainty). Fourth, the variance in similarity tends to increase with buffer size. Fifth, the relationships between similarity and buffer size are not the same across combinations of species and modeling methods; i.e., there appear to be interaction effects between the categorical variables and the covariate. Among the BIOCLIM analyses for example, *O. cruralis* shows a strong relationship between similarity and buffer size, whereas the relationship is weaker in *P. marmoratum*. In contrast, this comparison is reversed in the DOMAIN analyses; *O. cruralis* shows a weaker relationship, while *P. marmoratum* shows a stronger one.

The GLM analysis assumes that deviations from expected are effectively summarized by a normally distributed random variant with equal variance across all treatment levels. Because some

cases show an increase in variance with a decrease in mean similarity, we tested for a correlation between mean similarity and its variance. The Pearson correlation coefficient (r) between mean similarity and variance was -0.278 , which has a probability of 0.028 in a one-tailed test. (We used a one-tailed test because we expected the variance to increase as the mean decreased.) We applied an arcsin transformation in an attempt to reduce this correlation; this transformation is commonly used with measures that range between zero and one. In the transformed data, the correlation (r) was reduced to -0.096 , which has a one-tailed probability of 0.26 . Because the transformed data show a reduced and insignificant correlation, we used the transformed data in our primary analysis. The comparable box plots for the transformed data are shown in Figure 7. While the correlation is reduced, the variances are still strongly heterogeneous across treatment combinations. The largest variance is still more than 130 times larger than the smallest. Consequently, the probability values obtained in the primary analysis below can only be taken as broadly indicative.

The results of our GLM analysis are shown in Table 2. Every term in the model is significant well beyond the commonly used 0.05 level. The fact that the interaction terms are significant means that the primary terms are not additive; the effect of any particular value depends on the values of the other variables. In particular, the rate at which consistency declines with uncertainty (the slope) depends on both the modeling method and the species.

A more detailed view of our results can be seen in Figure 8. These histograms show the distributions of similarity scores for each of the 48 permutations of the primary parameters. We include these graphs because the assumptions of normally distributed error terms and homogenous variances within groups are violated. These histograms show how the distributions of similarity scores change across the experimental variables.

In 9 of the 12 combinations of species by modeling-method (columns of histograms in fig. 8A, B, and C) the distributions are close to normal and have similar variance across buffer-size. In the other three cases, the distributions change markedly with buffer size. The scores for *L. elenae* modeled with DOMAIN are skewed to the left at 5 and 10 km, become flatter at 25 km, and become skewed to the left again at 50 km. At the smallest buffer size, the scores for *P. marmoratum* and DOMAIN cluster toward the upper range with a sparse tail to the left. The maximum and minimum scores, and hence the range, do not change much between 5 and 50 km, but the distribution goes from skewed to flat and the variance gets 100 times larger from the smallest buffer size to the largest. In the *P. marmoratum* and Maxent analyses, similarity scores cluster tightly in the 5 km simulations, while the distribution flattens and the mode decreases at the larger buffer sizes.

DISCUSSION

The range of uncertainty used in this study, 5 to 50 km, is realistic and meaningful in comparison to both the degree of uncertainty that exists in real data and the resolution or scale of various gridded environmental surfaces that are routinely employed in distribution modeling: 1 km to 1/2° cell sizes (Hijmans et al. 2005; Mitchell and Jones

2005). Furthermore, the variable specificity of historical localities introduces geographic uncertainty well within the range of the buffer sizes tested here. Thus these results should help inform users of retrospectively georeferenced data regarding the distribution modeling methods that are most and least sensitive to degree of specimen locality uncertainty.

We expect the difference between environmental space at a given point A and B to be inversely proportional to the distance that separates these two points; in other words, the closer the points in geographic space, the more similar they should be in terms of environmental space (Tobler 1970). As a consequence, points selected from a 50 km buffer should be more different from the original point and from each other than points selected from the 5 km buffer. This environmental space translated into geographic space can have profound consequences in the modeling outputs. One possible outcome is that the area of the predicted distribution will be proportional to the differences among the points used to train the models, in other words, the model will become more general (see Fig. 10). However, comparing predicted areas of suitability has one major difficulty that forms the basis of our choice to use Fuzzy Kappa: the issue of threshold selection. To measure the relationship between predicted area and buffer size, a threshold must be selected and binary outputs must be compared. The relationship between predicted area and buffer size and the issue of threshold selection are two very important elements deserving of further attention that we did not explicitly evaluate in this paper.

In this study, we did not address the issue of spatial autocorrelation explicitly. There are two types of spatial autocorrelation that will influence the effective sample size of localities: 1) the spatial autocorrelation among species occurrence localities, and 2) the spatial autocorrelation within the buffer. Although Dormann et. al. (2007) suggest that differences in parameter estimates and inference between spatial and non-spatial models are small, i.e., (the spatial models accounted for spatial autocorrelation, while the non-spatial models did not), this problem may also depend on the degree of environmental heterogeneity across sampled environmental space.

Table 2. Summary of ecological niche modeling parameters under the four methods used in this study.

	BIOCLIM	DOMAIN	GARP	Maxent
Software used	DIVA GIS 5.4 (Hijmans et al. 2001).	DIVA GIS 5.4 (Hijmans et al. 2001).	Desktop GARP 1.1.3 (Kansas University)	Maxent 2.3 (Phillips et al. 2004)
Removal of duplicated localities	Yes	Yes	Yes	Yes
Outlier detection	No	No	---	---
Parameters details	Percentile used: 0.025	---	Atomic, range, negated range, and logit rules.	Regularization multiplier = 1 Random test % = 0
Internal evaluation	---	Training: 50% localities 20 best-subset models	Training: 50% localities 20 best-subset models	Yes
Outputs rescaled from 0 to 100	Yes	Yes	Yes	Yes

Table 3. ANOVA table for the general linear model analysis of transformed similarity scores.

Source	Partial SS	df	MS	<i>F</i>	Prob > <i>F</i>
Model	306.346	23	13.31940	3188.9	<0.001
Buffer	9.343	1	9.34283	2236.9	<0.001
Species	1.364	2	0.68186	163.3	<0.001
Modeling method	122.772	3	40.92412	9798.0	<0.001
Buffer*Species	0.063	2	0.03171	7.6	<0.001
Buffer* Modeling method	3.099	3	1.03310	247.3	<0.001
Species* Modeling method	5.228	6	0.87128	208.6	<0.001
Buffer*Species* Modeling method	2.752	6	0.45871	109.8	<0.001
Residual	19.948	4776	0.00418		
Total	326.295	4799	0.06799		

The GLM analysis shows that modeling method has the largest effect on similarity scores and uniquely accounts for 40% of the total variance in the model. The second most important factor was buffer size, but it uniquely accounts for only 3% of the model. This may seem like a small percentage, but buffer size interacts with the categorical variables, which obscures the effect of buffer size alone. Further interpretation of the proportion of variance is also ill-advised because buffer size is an experimenter-controlled variable, so the proportion of variance it explains is determined by the range of values we chose as inputs to the simulation.

Among the modeling methods, BIOCLIM and DOMAIN produced distribution maps that were overall more similar to the original maps than either GARP or Maxent. Given the popularity of GARP and Maxent, we were surprised to see these methods produce maps that were significantly less consistent than the two older and simpler methods. BIOCLIM and DOMAIN behaved as expected in showing a decline in consistency with increasing locality uncertainty, but the other two methods were less sensitive to uncertainty, either consistently (GARP), or in two of three cases (Maxent). GARP showed the least sensitivity to locality uncertainty; distributions generated from

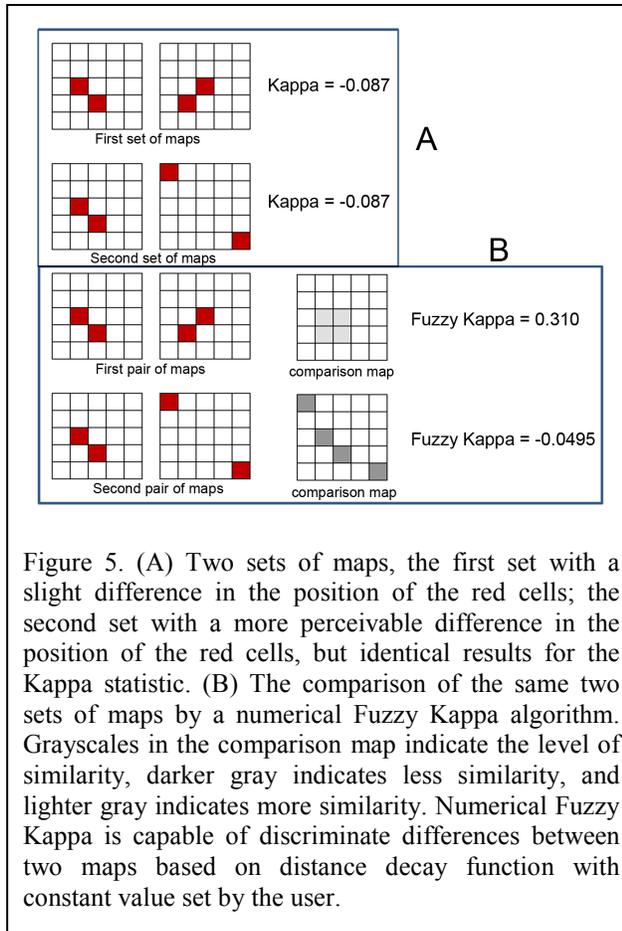


Figure 5. (A) Two sets of maps, the first set with a slight difference in the position of the red cells; the second set with a more perceivable difference in the position of the red cells, but identical results for the Kappa statistic. (B) The comparison of the same two sets of maps by a numerical Fuzzy Kappa algorithm. Grayscales in the comparison map indicate the level of similarity, darker gray indicates less similarity, and lighter gray indicates more similarity. Numerical Fuzzy Kappa is capable of discriminate differences between two maps based on distance decay function with constant value set by the user.

points with up to 50 km of uncertainty were only moderately less similar to the originals than those generated from points with only a maximum of 5 km uncertainty. Maxent distributions showed the lowest consistency and moderate sensitivity to locality uncertainty. Several reasons may contribute to the differences:

1. The BIOCLIM model identifies locations where all environmental factors fall within certain percentiles (e.g., 95%) of the observation records (Busby 1986). Therefore, unless a significant number of extreme large or small values are changed when increasing the buffer size, the locality uncertainty will have relatively little effect on the modeling results.
2. The DOMAIN method assigns a classification value to an unknown site based on the distance of its closest similar site in environmental space. The effort on the locality variation is local, and even extreme values are found, they will only influence some nearby points in environmental space.
3. GARP is based on genetic algorithms, which aim to find exact or approximate solutions to an

optimization or search problem. GARP can be considered a non-parametric machine learning algorithm which normally makes few assumptions about the data distribution, and is more robust to data outliers. However, variation of the Fuzzy Kappa values is greater than that of BIOCLIM and DOMAIN methods. This is due to the fact that variation also comes from the stochastic generation of rule sets for the GARP method and the random sampling of the background area, which will generate slightly different results in each iteration of the GARP model.

4. Maxent is a general-purpose machine learning method. Similar to generalized linear model (GLM) and generalized additive models (GAM), Maxent needs to make certain assumptions on the probability distributions. Exponential models are normally used (Phillips et al. 2006), which could be more sensitive to variation of the training data compared to non-parametric approaches.

Finally, we would like to emphasize that the variable we labeled “species” in these experiments is not actually a simple repetition of the experiment with another taxon, with all other factors equal. The three species selected in this study are all allopatric and come from regions where environmental parameters are expected to change very differently with comparable horizontal displacement or uncertainty. We expected similarity scores based on *O. cruralis* to decline sharply with increasing buffer size, because it is found in the Yungas or eastern Andes where the elevation gradients are steep. We expected *P. marmoratum* from the Andean highlands to exhibit intermediate sensitivity to buffer size, and *L. elenae* from the Amazonian lowlands to show the least sensitivity. Our expectations were never fully born out. In comparison to the other species, *L. elenae* produced the highest scores in the GARP and Maxent analyses, but *P. mamorata* produced the lowest scores in three out of four cases.

CONCLUSIONS

In several respects the results of our simulations were very different from what we expected. Modeling method produced the largest effect; more than the primary experimental treatment of displacing original localities by up to 50 km, more than species differences, and more than topographic heterogeneity.

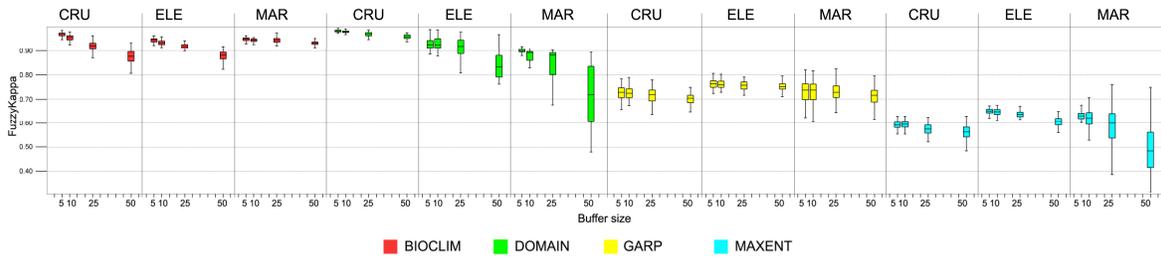


Figure 6. Summary distributions of similarity scores from 48 experiments, each made of 100 simulated data sets. Standard Tukey’s box-plots represent the Fuzzy Kappa similarity scores for the series of buffer sizes within each modeling method and species combination. *P. marmoratum* (MAR); *O. cruralis* (CRU); and *L. elenae* (ELE).

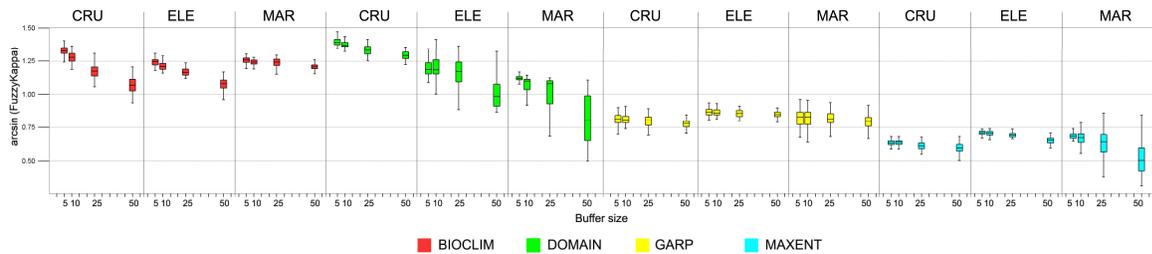


Figure 7. Summary distributions of similarity scores from 48 experiments, each made of 100 simulated data sets. Standard Tukey’s box-plots represent the transformed similarity score arcsin of Fuzzy Kappa for the series of buffer sizes within each modeling method and species combination. *P. marmoratum* (MAR); *O. cruralis* (CRU); and *L. elenae* (ELE).

An inescapable observation is that the newer and currently more popular methods, GARP and Maxent, were shown to produce more inconsistent predictions than the earlier and simpler methods, BIOCLIM and DOMAIN. We do not necessarily interpret this to mean that BIOCLIM and DOMAIN predict distributions more accurately than GARP or Maxent. A method that predicts with higher consistency may not be closer to the true distribution because it could be biased. For example, it might consistently over-predict the true distribution. On the other hand, a single prediction may not be very close to the true distribution if the method is relatively inconsistent. It is worth investigating further why the GARP and Maxent analyses, as we performed them here, gave inconsistent predictions.

Graham et al (2008) conclude similarly that not all modeling techniques are equally influenced by positional error. They suggested that some modeling techniques (Maxent and Boosted Regression Trees) are particularly “robust” to moderate levels of uncertainty in locality data. On

the contrary, our research finds that GARP is the most robust technique to positional error and DOMAIN the most sensitive of the four techniques we evaluated. This contradictory finding may be explained in that Graham et al. (2008) addressed a slightly different but complementary problem. They evaluated the effect of degrading positional accuracy on the capacity of the model to predict accurately an independent dataset, using AUC as a metric. In contrast, our goal was to evaluate how different modeling methods respond to varying levels of degraded positional accuracy. Moreover, Graham et al. (2008) used a single error treatment (5 km), while our study addressed multiple levels of locality uncertainty. Our finding that DOMAIN is the most sensitive method and GARP is the more robust method of the four ENM tested here doesn’t imply that one method is better over others. We aim to provide information model performance relative to one additional source of uncertainty that will assist the user in model selection.

Finally, we sampled only four points along the potentially larger domain of uncertainty values.

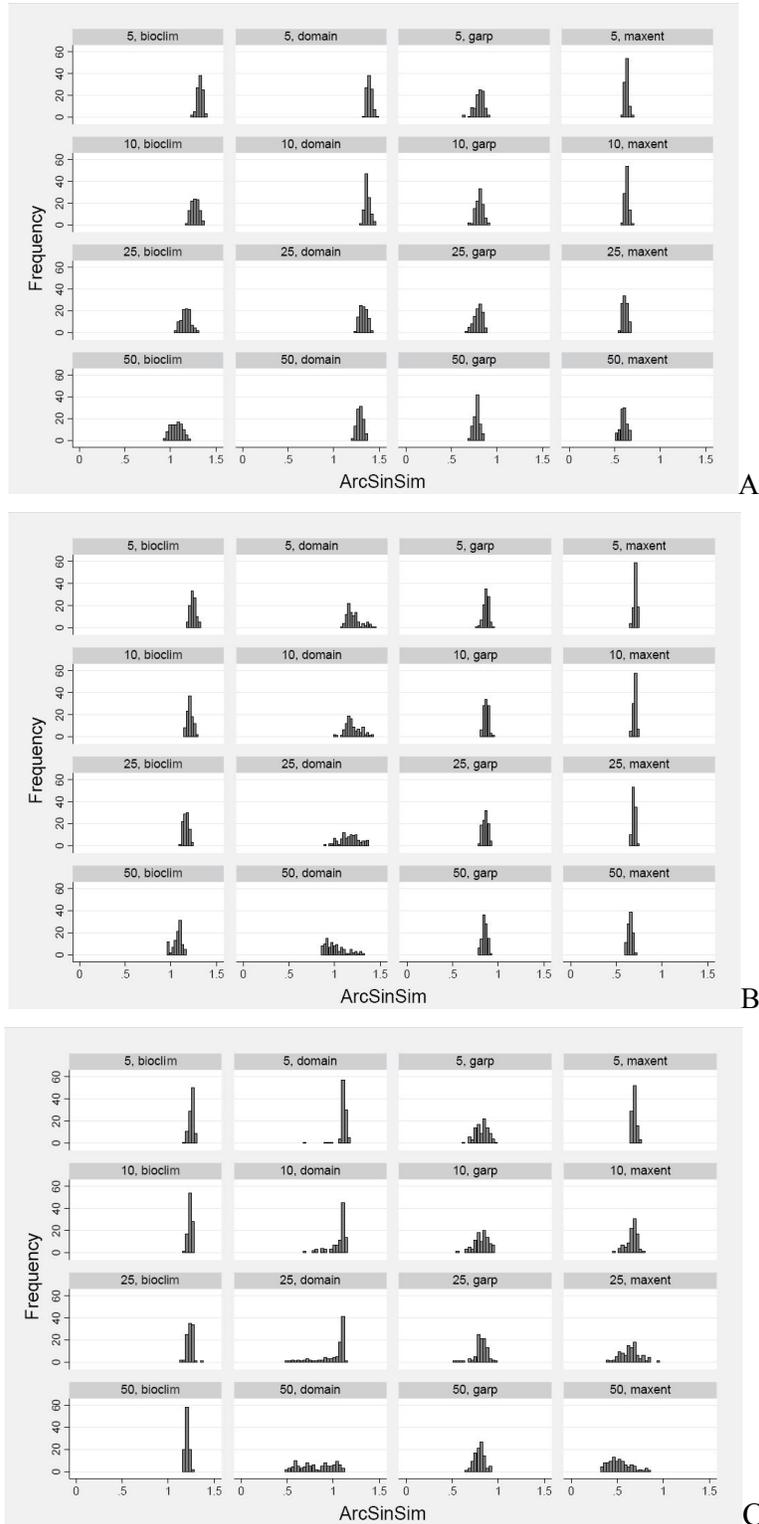


Figure 8. Histograms of similarity scores for the 48 permutations of experimentally controlled primary variables, buffer-size, modeling-method, and species. Arcsin transformed similarity is along the x-axis, frequency is on the y-axis, and the histograms are grouped by buffer-size (rows), modeling-method (columns) and species across pages (a. *O. cruralis*, b. *L. elenae*, and c. *P. marmoratum*). The scaling and range of the axes are the same across all histograms.

Consequently, we cannot evaluate whether the response of consistency to uncertainty is linear or curvilinear. It also remains to be determined what might happen beyond the limits we sampled.

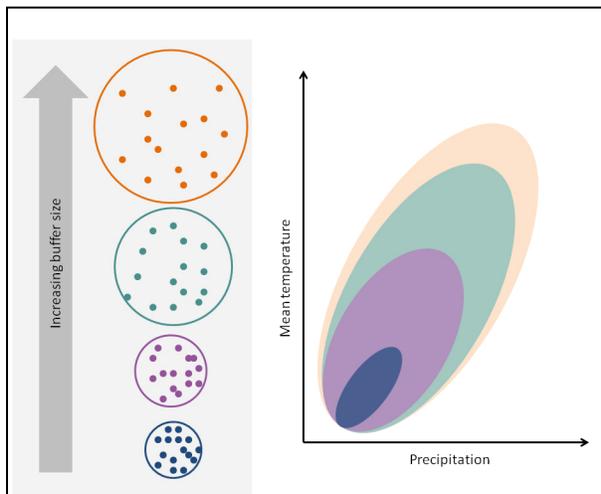


Figure 9. Hypothetical relationship between the buffer size and environmental space. Left figure: increasing uncertainty buffer size, and right figure: the possible change of its environmental space (using temperature and precipitation as example environmental features) due to the increasing uncertainty. Note that the actual shape in the feature space may not be the ellipse shape, and there are situations that don't follow the same trend (e.g. environmental space may not be so homogeneous).

ACKNOWLEDGMENTS

We would like to thank Jerry Davis, Matthew Merrifield, Michel Koo, Kristin Byrd, Simon Ferrier, Catherine Graham, Juan Parra, Robert Hijmans, and Kazuya Naoki for their input on our analytical methods. We also would like to thank Town Peterson and the anonymous reviewers who help to improve this manuscript. This project was funded by the California Academy of Sciences Lakeside Fund for International Students and the WWF Russell E. Train Education for Nature Program.

LITERATURE CITED

- Anderson, R. 2003. Real vs. artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. *Journal of Biogeography* 30:591-605.
- Anderson, R., M. Gomez-Laverde, and A. Peterson. 2002. Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography* 11:131-141.
- Anderson, R., D. Lew, and A. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* 162:211-232.
- Araújo, M., M. Cabeza, W. Thuiller, L. Hannah, and P. Williams. 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology* 10:1618-1626.
- Araújo, M., and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33:1677-1688.
- Barry, S., and J. Elith. 2006. Error and uncertainty in habitat models. *Ecology* 43:413-423.
- Beaman, R., P. Museum, J. Wieczorek, and S. Blum. 2004. Determining space from place for natural history collections. *D-Lib Magazine* 10:1082-9873.
- Bradley, B., and E. Fleishman. 2008. Can remote sensing of land cover improve species distribution modelling? *Journal of Biogeography* 35:1158-1159.
- Brotos, L., W. Thuiller, M. Araújo, and A. Hirzel. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27:437-448.
- Buermann, W., S. Saatchi, T. Smith, B. Zutta, J. Chaves, B. Milá, and C. Graham. 2008. Predicting species distributions across the Amazonian and Andean regions using remote sensing data. *Journal of Biogeography* 35:1160-1176.
- Busby, J. 1986. A biogeoclimatic analysis of *Nothofagus cunninghamii* (Hook.) Oerst. in southeastern Australia. *Austral Ecology* 11:1-7.
- Busby, J. 1991. BIOCLIM-a bioclimate analysis and prediction system. Pp. 64-68 in C. Margules, and M. Austin, eds. *Nature conservation: cost effective biological surveys and data analysis*. CSIRO, Canberra.
- Bustamante, J. 1997. Predictive models for lesser kestrel *Falco naumanni* distribution, abundance and extinction in southern Spain. *Biological conservation* 80:153-160.
- Carpenter, G., A. Gillison, and J. Winter. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation* 2:667-680.
- Chapman, A., and J. Wieczorek. 2006. Guide to best practices for georeferencing. *Global Biodiversity Information Facility*.

- Chefaoui, R., and J. Lobo. 2007. Assessing the conservation status of an Iberian moth using pseudo-absences. *Journal of Wildlife Management* 71:2507-2516.
- Corsi, F., E. Dupre, and L. Boitani. 1999. A large-scale model of wolf distribution in Italy for conservation planning. *Conservation Biology* 13:150-159.
- Dormann, C., J. McPherson, M. Araújo, R. Bivand, J. Bolliger, G. Carl, R. Davies, A. Hirzel, W. Jetz, and W. Kissling. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609-628.
- Duckworth, W. D., H. H. Genoways, and C. L. Rose. 1993. Preserving natural science collections: chronicle of our environmental heritage. National Institute for the Conservation of Cultural Property, Washington, DC.
- Elith, J., C. Graham, R. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. Hijmans, F. Huettmann, J. Leathwick, and A. Lehmann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129.
- Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Ecology* 41:263-274.
- Graham, C., J. Elith, R. Hijmans, A. Guisan, A. T. Peterson, and B. Loiselle. 2008. The influence of spatial errors in species occurrence data used in distribution models. *Journal of Applied Ecology* 45:239-247.
- Graham, C., S. Ron, J. Santos, C. Schneider, and C. Moritz. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution* 58:1781-1793.
- Grinnell, J. 1917. Field tests of theories concerning distributional control. *American Naturalist* 51:115.
- Guisan, A., C. Graham, J. Elith, and F. Huettmann. 2007. Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13:332-340.
- Guisan, A., and N. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.
- Guo, Q., Y. Liu, and J. Wiecek. 2008. Georeferencing locality descriptions and computing associated uncertainty using a probabilistic approach. *International Journal of Geographical Information Science* 22:1067-1090.
- Guralnick, R., A. Hill, and M. Lane. 2007. Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters* 10:663-672.
- Hagen-Zanker, A., B. Straatman, and I. Uljee. 2005. Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science* 19:769-785.
- Hagen, A. 2003. Multi-method assessment of map similarity. *International Journal of Geographical Information Science* 17:235-249.
- Hernandez, P., C. Graham, L. Master, and D. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29:773-785.
- Higgins, S., D. Richardson, and R. Cowling. 2000. Using a dynamic landscape model for planning the management of alien plant invasions. *Ecological Applications* 10:1833-1848.
- 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.
- Hijmans, R., L. Guarino, M. Cruz, and E. Rojas. 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter* 127:15-19.
- Hirzel, A., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83:2027-2036.
- Hugall, A., C. Moritz, A. Moussalli, and J. Stanisic. 2002. Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosiphia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences USA* 99:6112-6117.
- Kadmon, R., O. Farber, and A. Danin. 2003. A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications* 13:853-867.
- Kelly, M., Q. Guo, D. Liu, and D. Shaari. 2007. Modeling the risk for a new invasive forest disease in the United States: An evaluation of five environmental niche models. *Computers, Environment and Urban Systems* 31:689-710.
- Leathwick, J. R., and D. Whitehead. 2001. Soil and atmospheric water deficits and the distribution of New Zealand's indigenous tree species. *Functional Ecology* 15:233-242.
- Lobo, J., A. Jimenez-Valverde, and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology & Biogeography* 17:145-151.
- Manel, S., J. Dias, and S. Ormerod. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling* 120:337-347.
- Manel, S., H. Williams, and S. Ormerod. 2001. Evaluating presence-absence models in ecology: the

- need to account for prevalence. *Journal of Applied Ecology*:921-931.
- Mills, J., and J. Childs. 1998. Ecologic studies of rodent reservoirs: their relevance for human health. *Emerging Infectious Diseases* 4:529-537.
- Mitchell, T., and P. Jones. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* 25:693–712.
- Murphy, J., D. Sexton, D. Barnett, G. Jones, M. Webb, M. Collins, and D. Stainforth. 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* 430:768-772.
- Ortega-Huerta, M., and A. Peterson. 2008. Modeling ecological niches and predicting geographic distributions: a test of six presence-only methods. *Revista Mexicana de Biodiversidad* 79:205-216.
- Parra, J., C. Graham, and J. Freile. 2004. Evaluating alternative data sets for ecological niche models of birds in the Andes. *Ecography* 27:350-360.
- Pearce, J., and S. Ferrier. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133:225-245.
- Pearson, R., C. Raxworthy, M. Nakamura, and A. Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34:102-117.
- Pearson, R., W. Thuiller, M. Araújo, E. Martinez-Meyer, L. Brotons, C. McClean, L. Miles, P. Segurado, T. Dawson, and D. Lees. 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography* 33:1704-1711.
- Peterson, A., and K. Cohoon. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling* 117:159-164.
- Peterson, A., M. Papes, and J. Soberón. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213:63-72.
- Peterson, A., V. Sanchez-Cordero, E. Martínez-Meyer, and A. Navarro-Sigüenza. 2006. Tracking population extirpations via melding ecological niche modeling with land-cover information. *Ecological Modelling* 195:229-236.
- Peterson, A., and J. Shaw. 2003. *Lutzomyia* vectors for cutaneous leishmaniasis in Southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *International Journal of Parasitology* 33:919–931.
- Peterson, A., and D. Vieglais. 2001. Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *Bioscience* 51:363-371.
- Peterson, T., M. Papes, and M. Eaton. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30:550-560.
- Phillips, S. 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al.(2007). *Ecography* 31:272-278.
- Phillips, S., M. Dudík, and R. Schapire. 2004. A maximum entropy approach to species distribution modeling. Pp. 83-84 in A. I. C. P. Series, ed. *Proceedings of the Twenty-first International conference on Machine Learning*. ACM Press New York, NY, USA, Banff, Alberta, Canada
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231-259.
- Pontius, R. 2000. Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing* 66:1011-1016.
- Proctor, E. 2004. Reducing variation in georeferenced locality descriptions. Pp. 191. *Geography & Human Environmental Studies*. San Francisco State University, San Francisco.
- Raxworthy, C., E. Martinez-Meyer, N. Horning, R. Nussbaum, G. Schneider, M. Ortega-Huerta, and A. Townsend Peterson. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426:837-841.
- Reddy, S., and L. Davalos. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography* 30:1719-1727.
- Reese, G., K. Wilson, J. Hoeting, and C. Flather. 2005. Factors affecting species distribution predictions: a simulation modeling experiment. *Ecological Applications* 15:554-564.
- Roura-Pascual, N., A. Suarez, C. Gómez, P. Pons, Y. Touyama, A. Wild, and A. Peterson. 2004. Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proceedings of the Royal Society B*: 271:2527-2535.
- Rowe, R. 2005. Elevational gradient analyses and the use of historical museum specimens: a cautionary tale. *Journal of Biogeography* 32:1883-1897.
- Soberon, J., and T. Peterson. 2004. Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359:689-698.
- Stockwell, D. 2006. *Niche Modeling: Predictions from Statistical Distributions*. Chapman & Hall/CRC, Boca Raton FL.
- Stockwell, D., and D. Peters. 1999. The GARP modelling system: problems and solutions to

- automated spatial prediction. *International Journal of Geographical Information Science* 13:143-158.
- Stockwell, D., and A. Peterson. 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling* 148:1-13.
- Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography*:234-240.
- Tsoar, A., O. Allouche, O. Steinitz, D. Rotem, and R. Kadmon. 2007. A comparative evaluation of presence-only methods for modelling species distribution. *Diversity & Distributions* 13:397-405.
- Underwood, E., R. Klinger, and P. Moore. 2004. Biodiversity Research Predicting patterns of non-native plant invasions in Yosemite National Park, California, USA. *Diversity & Distributions* 10:447.
- Visser, H., and T. De Nijs. 2006. The map comparison kit. *Environmental Modelling and Software* 21:346-358.
- Welk, E., K. Schubert, and M. Hoffmann. 2002. Present and potential distribution of invasive garlic mustard (*Alliaria petiolata*) in North America. *Diversity and Distributions* 8:219-233.
- Wieczorek, J., Q. Guo, and R. Hijmans. 2004. The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science* 18:745-767.
- Williams, S., and J. Hero. 2001. Multiple determinants of Australian tropical frog biodiversity. *Biological conservation* 98:1-10.
- Wisz, M., R. Hijmans, J. Li, A. Peterson, C. Graham, and A. Guisan. 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14:763-773.
- Zimmermann, N., T. Edwards, G. Moisen, T. Frescino, and J. Blackard. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology* 44:1057-1067.