2014

Methodological advancements for improving performance and generating ensemble ecological niche models

Robert Boria
CUNY City College of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: http://academicworks.cuny.edu/cc_etds_theses

Part of the Biology Commons

Recommended Citation
http://academicworks.cuny.edu/cc_etds_theses/537

This Thesis is brought to you for free and open access by the City College of New York at CUNY Academic Works. It has been accepted for inclusion in Master's Theses by an authorized administrator of CUNY Academic Works. For more information, please contact AcademicWorks@cuny.edu.
METHODOLOGICAL ADVANCEMENTS FOR IMPROVING PERFORMANCE AND
GENERATING ENSEMBLE ECOLOGICAL NICHE MODELS

ROBERT A. BORIA

Department of Biology, City College of the City University of New York, New York, NY
10031, USA

Masters Thesis Committee:
Dr. Robert P. Anderson (advisor)
Dr. Amy C. Berkov
Dr. Ana Carnaval

08 August 2014
Table of Contents

Chapter 1: Spatial filtering to reduce sampling bias can improve the performance of ecological niche models.......................................................... 3

Chapter 2: A single-algorithm approach to ensemble niche modeling provides critical estimates of uncertainty for cross-time projections......................................................... 12

Acknowledgments................................................................................................................. 29

References.............................................................................................................................. 30

Tables................................................................................................................................ 36

Figures................................................................................................................................ 40

Supplementary Material........................................................................................................... 48
Chapter 1: Spatial filtering to reduce sampling bias can improve the performance of ecological niche models.

The results in this chapter have been published in *Ecological Modelling, 275*, 73–77.

**Abstract**

This study employs spatial filtering of occurrence data with the aim of reducing overfitting to sampling bias in ecological niche models (ENMs). Sampling bias in geographic space leads to localities that may also be biased in environmental space. If so, the model can overfit to those biases. As a preliminary test addressing this issue, we used Maxent, bioclimatic variables, and occurrence localities of a broadly distributed Malagasy tenrec, *Microgale cowani* (Family Tenrecidae: Subfamily Oryzorictinae). We modeled the abiotically suitable area of this species using three distinct datasets: unfiltered, spatially filtered, and rarefied unfiltered localities. To quantify overfitting and model performance, we calculated evaluation AUC, the difference between calibration and evaluation AUC (= AUC\(_{\text{diff}}\)), and omission rates. Models made with the filtered dataset showed lower overfitting and better performance than the other two suites of models, having lower omission rates and AUC\(_{\text{diff}}\), and a higher AUC\(_{\text{evaluation}}\). Additionally, the rarefied unfiltered dataset performed better than the unfiltered one for three evaluation metrics, likely because the larger datasets reinforced the biases. These results indicate that spatial filtering of occurrence localities may allow biogeographers to produce better models.

**Introduction**

Ecological niche models (ENMs) are a correlative approach aiming to approximate the abiotically suitable area of a species by comparing environmental conditions at localities where the species occurs with the overall conditions available in the study region (see Peterson *et al.*, [2006]).
The increased prevalence of online databases of occurrence localities and climatic variables has resulted in an increase in the production of ENMs (Hijmans et al., 2005; Kozak et al., 2008). Although correlative ENMs are utilized widely in the fields of ecology, evolution, and conservation biology, their mainstream acceptance has outpaced methodological research and refinement.

Here, I study one area needing methodological improvement: the effect of sampling bias. Frequently, researchers sample easily accessible areas (i.e., near major roads or towns), leading to geographic clusters of localities (Hijmans et al., 2000; Kadmon et al., 2004; Reddy & Dávalos, 2003). These sampling biases artificially increase spatial auto-correlation of the localities. Such a situation can cause the model to overfit to environmental biases that correspond to these influences in geographic space. Overfitting occurs when a model fits too tightly to calibration data, limiting its ability to predict independent evaluation data. Eliminating artifactual clusters of localities is also important for model evaluation, as calibration localities that are next to evaluation localities lead to inflated values of performance (Hijmans, 2012; Veloz, 2009).

In this study, I aim to reduce the effect of sampling bias by spatially filtering the occurrence dataset, which should reduce the degree of overfitting in the model. Ideally, when information quantifying sampling effort exists (e.g., via a target group), it can be used in model calibration to correct for sampling bias (Anderson, 2012; Phillips et al., 2009). However, researchers frequently do not have access to such information. In contrast, the method applied here can be employed generally. Several studies have used filtering (= thinning) techniques (Anderson & Raza, 2010; Carroll, 2010; Pearson et al., 2007; Veloz, 2009) to reduce the effects of sampling biases, but I know of none that has explicitly tested whether this method improves
the performance of ENMs. If it does, an ENM made with the filtered dataset should show lower overfitting and higher performance in predicting independent evaluation data.

**Methods and Materials**

*Occurrence and environmental data*

Madagascar is home to four endemic radiations of extant terrestrial mammals, including nesomyine rodents, lemurs, euplerid carnivorans, and tenrecs. The latter show considerable morphological variation, forming an extraordinary adaptive radiation (Olson & Goodman, 2003), with the most taxonomically diverse genus being the shrew tenrecs (*Microgale* spp.; 22 currently recognized extant species; Goodman *et al*., 2006; Olson, 2013; Olson *et al*., 2009; Soarimalala & Goodman, 2011). Perhaps the most common, widespread, and well-documented species, Cowan’s shrew tenrec (*M. cowani*) is found throughout what remains of Madagascar’s humid forests at elevations ranging from 530-2,500 m (Soarimalala & Goodman, 2011). This swath spans several different vegetational zones, including forests ranging from lowland to upper montane, as well as above forest-line in ericoid alpine formations. This species appears to be a generalist among shrew tenrecs and accounts for over one-fifth of *Microgale* specimens in European and North American museums (Olson, unpub.). Because its range and habitat requirements are relatively well known, *M. cowani* represents a suitable species for the current study.

Occurrence localities were compiled from field collections and associated notes, examination of museum specimens, and literature (Fig. 1.1). The environmental data were obtained from WorldClim.org (Hijmans *et al*., 2005; at 30” resolution). These 19 bioclimatic variables employed reflect aspects of temperature and precipitation and have been used
successfully for niche models of small non-volant montane mammals (e.g., Jezkova et al., 2009; Davis et al., 2007). I delimited a custom study region for each model, specifically by drawing a rectangle around localities and adding a 0.5° buffer (Anderson and Raza, 2010; Barve et al., 2011; see Fig. 1.1).

Experimental design

As a first exploration, I built models using Maxent version 3.3.2k. Maxent is a presence-background algorithm that compares occurrence localities with a sample of background pixels to create a prediction of suitability (Phillips et al., 2006; Phillips and Dudik, 2008). Maxent has performed well in comparison with other techniques and is commonly used (Elith et al., 2006; Wisz et al., 2008) but sensitive to sampling biases (Anderson and Gonzalez, 2011; Phillips et al., 2009). In addition to sampling bias, two other issues can affect overfitting in niche models: correlations among environmental variables and the level of model complexity. To simplify the current experiment, I held those factors constant. Specifically, I used all 19 bioclimatic variables and employed default Maxent settings for the given sample size: feature class (linear, quadratic, and hinge) and regularization multiplier value (1). I note, however, that Maxent employs regularization to reduce complexity, not all variables are necessarily included in the final model (Phillips and Dudik, 2008).

For filtering, I randomly removed localities that were within 10 km of one another, keeping the most localities possible. The 10 km distance was chosen based on the high spatial heterogeneity of the mountains in Madagascar, and the same distance has been used in previous studies in mountainous areas with high geographical heterogeneity (Pearson et al., 2007; Anderson and Raza, 2010). This distance was not chosen to approximate the species’ dispersal
capabilities, but rather to reduce the inherent geographic biases associated with collection data. There were 57 unique localities before filtering and 31 unique localities after filtering (see Fig. 1.1). I used the Geographic Distance Matrix Generator version 1.2.3 to calculate the geographic distance between each pair of localities (Ersts, 2012). For each cluster of localities less than 10 km apart, I determined the maximum number of localities that could be retained. When more than one co-optimal solution existed for a given cluster, I selected one randomly. To test for the expected effect of reducing sampling bias vs. simply the effect of sample size, I also randomly rarefied the unfiltered dataset to match the number of localities of the filtered dataset. Hence, I used three different datasets for modeling: unfiltered, filtered, and rarefied unfiltered. To explore the possibility that the spatial filter used here removed localities with novel environmental conditions, I plotted the values of annual mean temperature and annual mean precipitation at each locality (see Supplementary Fig. S1.2).

An overfit model has an overly complex relationship between the occurrence localities of a species and associated environmental variables (Peterson et al., 2011). To quantify overfitting as well as general model performance, I implemented a variation of $k$-fold cross-validation. To provide strong tests, I divided the localities geographically into $k = 3$ bins (see Fig. 1.1). Each bin was constructed to contain approximately the same number of localities but occupy different portions of geography (Radosavljevic and Anderson, in press). This allowed the models to be evaluated on spatially segregated (spatially independent) evaluation data, avoiding the inflation of evaluation metrics due to spatial autocorrelation between calibration and evaluation datasets (Hijmans, 2012; Veloz, 2009). Such evaluations also are necessary for evaluating model transfer across space or time (e.g., for climate change studies; Anderson 2013). In each iteration, the models were calibrated using $k - 1$ bins and evaluated on the withheld bin (Fielding and Bell,
1997; Peterson et al., 2011). This was done until all bins were used once for evaluation (i.e., three iterations in total). Maxent sampled background data for the environmental variables from only the regions corresponding to the bins used during calibration (following Phillips, 2008; Radosavljevic and Anderson, in press). These methods allowed quantification of overfitting and performance after transfer (Peterson et al., 2011; Araújo and Rahbek, 2006; Bahn and McGill, 2013). The model from each iteration was then projected to the full study region to allow for evaluation and visualization.

I evaluated overall model performance via threshold-independent and threshold-dependent measures that assess various aspects of performance and overfitting. The threshold-independent metrics derive from the Area Under the Curve (AUC) of the Receiver Operating Characteristic plot, a rank-based measure of overall discriminatory ability of the model. Accordingly, the AUC calculated on evaluation localities \((\text{AUC}_{\text{evaluation}})\) constituted our measure of overall model performance. The other threshold-independent measure was \(\text{AUC}_{\text{diff}}\) = \(\text{AUC}_{\text{calibration}}\) minus \(\text{AUC}_{\text{evaluation}}\). The smaller the difference between the two, the less overfitting in the model (Warren and Seifert, 2011). Because comparisons between AUCs calculated using presence-background data are only valid when study regions are identical, I calculated AUCs over the entire study region. For each iteration of each treatment, I obtained \(\text{AUC}_{\text{diff}}\) and \(\text{AUC}_{\text{evaluation}}\), and then I averaged the values across the three geographic bins.

Complementarily, I employed two threshold-dependent measures: omission rates based on two threshold rules (10% calibration omission rate and lowest presence threshold, LPT = 0% calibration omission rate; Pearson et al., 2007; = minimum training presence threshold of Maxent software). Omission rates are the proportion of evaluation localities that fall outside of the model once converted to a binary prediction. Omission rates provide information regarding
both discriminatory ability and overfitting, evaluated at particular thresholds. In general, lower omission indicates higher performance (better discrimination between suitable versus unsuitable areas). In addition, overfit models have omission rates higher than theoretical expectations (Radosavljevic and Anderson, in press; Shcheglovitova and Anderson, 2013). The lowest presence threshold sets the threshold at the lowest value of the prediction for any pixel that contains a calibration locality (Anderson and Gonzalez, 2011) and has an expected omission rate of zero for evaluation localities. Similarly, the 10% calibration omission rate rule sets the threshold at a value that excludes the 10% of calibration localities with lowest prediction (Anderson and Gonzalez, 2011) and has an expected omission rate of 0.10. I obtained the two threshold-dependent measures and averaged their values as for AUC.

Results

Evaluations

The models made with the filtered dataset were superior to those from the other datasets for all metrics. Regarding the threshold-independent measures, the filtered datasets led to the highest average $AUC_{\text{evaluation}}$ (highest overall performance) and the lowest average $AUC_{\text{diff}}$ (lowest overfitting; table 1.1; Supplementary table 1.1). For the threshold-dependent measures, models made with the filtered datasets showed the lowest average omission rate for each of the threshold rules (lowest overfitting and highest discriminatory ability at those thresholds; table 1.1; Supplementary table 1.2). Despite the observed pattern for $AUC_{\text{evaluation}}$, average differences among the three suites of models were small in magnitude. In contrast, the magnitude of differences in the other three measures was substantial. The models made with the filtered dataset showed the lowest divergence from the expected values for both omission rates,
indicating the lowest overfitting. Importantly, the models made with the rarefied unfiltered dataset outperformed the unfiltered dataset for three of the four evaluation metrics (AUC\textsubscript{diff}, AUC\textsubscript{evaluation}, and LTP omission rate), but for all measures they still performed more poorly than the filtered datasets. Of note, spatial filtering did not remove any localities holding novel environmental conditions for the examined variables (Supplementary Fig. 1.2).

Qualitative evaluations in geographic space illustrate spatial patterns that help interpret the quantitative results (Fig. 1.1). Inspection of the three respective projection regions (each corresponding to the respective bin that was withheld during calibration), using the LPT threshold indicated the extent of predicted areas and the localities that were omitted for the unfiltered and the filtered datasets. All analyses predicted moderate to high suitability for the species throughout several extensive areas, in accordance with known natural history information for this widespread species occurring across a broad elevational range and in several mesic vegetational zones. However, the filtered dataset showed a generally broader prediction. The map for the rarefied unfiltered dataset (Supplementary Fig. 1.1) was similar to that for the unfiltered dataset. The models made with the filtered dataset predicted more evaluation localities than those for the other two treatments (Fig. 1.1; table 1.1).

**Discussion**

**Comparisons among datasets**

I show that for *Microgale cowani*, spatial filtering of occurrence datasets led to a reduction of overfitting and an increase in performance. The effects of filtering were not merely due to sample size, because the filtered dataset also outperformed the rarefied unfiltered dataset. The fact that the rarefied dataset outperformed the unfiltered dataset could be due to the larger
dataset reinforcing the biases (as in Anderson and Gonzalez, 2011). Although the three datasets led to similar levels of overall performance based on AUC evaluation, the filtered dataset did yield the highest value for that measure. Hence, all measures indicated that the filtered dataset provides the best prediction of spatially independent occurrence localities of this species.

Whereas sampling bias generally goes unaccounted for in ecological niche modeling, the current results support the value of spatial filtering and point to several directions for future research. When available, data quantifying sampling biases should be used to correct for those biases (e.g., via a target group; Phillips et al., 2009). However, for the many cases when such data are not available, spatial filtering of datasets may improve model performance. The spatial distance used here for filtering was chosen based on the heterogeneity of the study system. In future research, I recommend that researchers conduct filtering experiments to determine the optimal geographic distance for the organism(s) and system(s), while retaining the species-specific niche signal (Anderson, 2012). I predict that mountainous regions will require a spatial filter that is smaller compared with regions having more homogenous environments. Furthermore, in addition to filtering in geographic space, researchers should consider the possibility of filtering localities in environmental space (Varela et al., in press; see Supplementary Fig. 1.2). Finally, future comprehensive filtering experiments should be undertaken while varying the environmental datasets employed (e.g., different numbers of variables, and considering the correlations among them) and model complexity (e.g., in Maxent by changing feature classes and regularization), ideally with both simulated and real species.
Chapter 2: A single-algorithm approach to ensemble niche modeling provides critical estimates of uncertainty for cross-time projections

Abstract

Ecological niche models are utilized widely in ecology, evolution, and conservation biology, but model variability (and uncertainty) remain an underappreciated issue. Generally, either a single model from one algorithm or an ensemble of single models from different algorithms is used to provide a prediction. Additionally, several recent studies have shown the need to tune model settings for a single algorithm. Nevertheless, uncertainty cannot be measured with a single model, and multiple settings may lead to essentially co-optimal models. To address this issue, I present a consensus ecological niche model (ENM) using a single-algorithm approach, while adjusting model settings to maximize performance. As an example, I used MaxEnt, 19 bioclimatic variables, and occurrence records of four Malagasy tenrecs. I calibrated and evaluated preliminary models using a jackknifing approach, tuning two model settings (feature classes and the regularization multiplier) to estimate optimal model complexity. Based on omission rates and AUC values, I chose a suite of top-performing preliminary models, and then generated a consensus prediction by averaging the values for each grid cell. I qualitatively compared the ensemble prediction with those made using the default settings and the optimal settings, respectively. Furthermore, I calculated the standard deviation to obtain a map showing variation in geography among the co-optimal models (i.e. uncertainty). I then did the same after retrojecting the predictions to Last Glacial Maximum and projecting them to 2070. I calculated Schoner’s $D$ between co-optimal settings, and similarly performing top models showed high variation in their geographic predictions for three of the four species. Additionally, models
created with different settings generated different reconstructions of past suitable areas or future projections. In conclusion, the ensemble model allows researchers to use several high-performing models for a single algorithm, which may vary in geography, to generate a robust prediction and detect areas of discrepancy.

**Introduction**

Correlative Ecological niche models (ENMs), often also termed Species Distribution Models (SDMs), aim to approximate the abiotically suitable areas for a species by comparing environmental conditions at localities where the species occurs with the overall range of conditions available in the study region (see Peterson et al. 2011; Anderson, 2012 for relevant terminology). Such models hold important uses when transferred across time and space, applications that require model generality. Perhaps the most frequent implementations of ENMs are presence-only techniques: models built with only occurrence localities and either pseudo-absences or background localities (Elith et al., 2006). Although correlative models are utilized widely in ecology, evolution, and conservation biology, several outstanding issues remain (Anderson, 2012, 2013). One critical issue is model uncertainty, which can have many different sources. Two important sources are algorithm choice and algorithm parameterization. Whereas uncertainty due to algorithm choice has been explored in previous studies (i.e., Diniz-Filho et al., 2009; Garcia et al., 2012), uncertainty pertaining to their parameterization has seldom been addressed, despite the call for such studies (Araújo & Guisan, 2006). This aspect of uncertainty seems crucial, given that its effects can propagate through to later applications of the models (e.g., future climatic estimates based on global circulation models, and different future emissions scenarios).
Here, I explore the effects of algorithm parameterization (i.e., model tuning or smoothing) and its associated maps of uncertainty, with a special emphasis on transfers across different time periods (past and future). A primary challenge in niche modeling is to balance goodness-of-fit with complexity, especially if the desired outcome is to transfer across time or space (Peterson et al., 2011; Anderson, 2013). Recently, several studies have demonstrated the need to tune model settings for a single algorithm (e.g., for the presence-background technique Maxent, Anderson & Gonzalez, 2011; Elith et al., 2011; Merow et al., 2013). However, most studies rely on default settings during the model building stage (reviewed by Merow et al., 2013; Yackulic et al., 2012), although those settings have been shown to produce overfit models in several instances (Radosavljevic & Anderson, 2014; Shcheglovitova & Anderson, 2013; Warren & Seifert 2011). Additionally, most studies generally develop a single final model, either by default settings or by optimizing settings, and interpret only that prediction. In summary, measures of uncertainty are generally not taken into account when using presence-only models, even for the model-calibration stage.

Another approach to correlative modeling is to generate an ensemble model, where several model outputs are combined to generate a single prediction. These model outputs can be a combination of the results of different algorithms, different input data (occurrence localities and environmental variables), and/or different model parameters (Araújo & New, 2007). Generally, ‘ensemble modeling’ has been used for forecasting suitable areas in reference to climate change, because it has been shown that different algorithms and different future climate estimates can give vastly different predictions (i.e. Araújo et al., 2005; Garcia et al., 2012; Pearson et al., 2005). Because several different model outputs are combined, measures of uncertainty can be assessed (i.e. Garcia et al., 2012; Marmion et al., 2009). However, a main
drawback when using several different algorithms is that typically the default settings are used for each algorithm, especially because of the difficulty of tuning model settings in current software implementations (Diniz-Filho et al., 2009, Thuiller et al., 2009), despite the fact that, conceptually, varying model parameters has already been put forward for the ensemble approach (Araújo & New, 2007).

Here, I combine the two approaches, by tuning model settings for a single algorithm and combining the best performing models with an ensemble approach to generate a consensus prediction and maps of uncertainty. As an example, I study four closely related Malagasy tenrecs and use the common presence-background algorithm MaxEnt (Phillips et al., 2006; Phillips & Dudík, 2008). I perform model-tuning exercises based on sequential criteria following the recommendations of Shcheglovitova & Anderson (2013), but because multiple settings may lead to essentially the same level of performance, I choose the top 10% of models (= co-optimal solutions) (see below; Muscarella et al., in review; Shcheglovitova & Anderson 2013).

Subsequently, I measure the geographic concordance among the co-optimal models using Schoener's $D$ statistic. Then, I visually compare the default, optimal, and ensemble predictions for current climatic conditions. Additionally, I generate maps of uncertainty for the current climate ensemble prediction. Furthermore, to examine the effects of these various approaches for climate change studies, I project the three models to the Last Glacial Maximum (21 kBP) and to 2070. I visually evaluate those predictions to determine if each of the three models agreed regarding the areas suitable and the connectivity among them. I know of no study that has evaluated the differences among co-optimal solutions for Maxent or of any that used a single algorithm to generate ensemble predictions and maps of uncertainty. I predict that the co-optimal solutions will vary in geography, but the extent of variation is unknown and likely will differ
among species. Furthermore, theory suggests that such differences in geography will be exacerbated when the models are transferred across time and space (Peterson et al., 2011).

**Methods**

*Input data*

Madagascar is home to four endemic radiations of terrestrial mammals: lemurs, euplerid carnivorans, nesomyine rodents, and tenrecs. The latter have evolved multiple ecomorphologies, but the most taxonomically diverse are the shrew tenrecs (*Microgale* spp.), with 22 currently recognized extant species (Goodman et al., 2006; Olson, 2013; Olson et al., 2009). Here, I modeled the abiotically suitable areas for four tenrecs: *Microgale cowani* (*n* = 56 localities), *M. gracilis* (*n* = 20), *M. gymnorrhyncha* (*n* = 31), and *M. thomasi* (*n* = 28). The occurrence localities and their associated georeferences for each species were compiled from field collections and associated notes, examination of museum specimens, and the literature (and scrutinized by taxonomic specialists, L.E. Olson and S.M. Goodman).

Generally, researchers tend to sample more easily accessible areas, leading to localities that are geographically clustered (Hijmans et al., 2000; Kadmon et al., 2004; Reddy and Dávalos, 2003). This can artificially increase spatial auto-correlation of the resulting localities for a given species and interfere with characterization of the niche signal. Eliminating artifactual clusters of localities is also important for model evaluation, as evaluation localities that lie close to calibration localities lead to inflated values of performance (Hijmans, 2012; Veloz, 2009). To reduce the effects of sampling bias on model performance, I spatially filtered the occurrence dataset to ensure that no two localities were within 10 km of one another (while retaining the most localities possible) using the recently developed *spThin* package in R (Aiello-Lammens
et al., in review; v. 3.03, R Development Core Team, 2014). This distance was chosen because a recent study showed that spatially filtering the *Microgale cowani* dataset by a 10 km distance reduced overfitting and improved model performance (Chapter 1; Boria et al., 2014). Because all four species share similar distributions, and result from the same sampling efforts for the target group of small non-volant mammals, I used the same 10 km distance for each species. Each dataset was reduced after applying the spatial filter: *M. cowani* (*n* = 32), *M. gracilis* (*n* = 15), *M. gymnorrhyncha* (*n* = 20), and *M. thomasi* (*n* = 17).

The present-day environmental data were obtained from WorldClim.org (Hijmans et al., 2005; at 30′ resolution). These 19 bioclimatic variables reflect aspects of temperature and precipitation and have been used successfully for niche models of many small mammals, including a montane tenrec in Madagascar (Boria et al., 2014). Although these variables can be correlated, Maxent is a machining learning algorithm that employs regularization to reduce complexity, and not all variables are necessarily included in the final model (see below for more details; Phillips & Dudík, 2008; Elith et al., 2011). To approximate modeling assumptions more closely regarding biotic interactions and dispersal ability, I delimited a custom study region for each species, specifically by drawing a minimum convex polygon around the localities and adding a 0.5° buffer (Anderson & Raza, 2010; Barve et al., 2011). Background localities for calibration were taken from only within the delimited study region.

The past climatic conditions (Last Glacial Maximum; 21 kBP; 2.5′ resolution) and the future projection (2070; 30′ resolution; emission scenario rcp26) were generated by the Community Climate System Model (CCSM) (Kiehl & Gent 2004; Gent et al., 2011). These layers consisted of the same 19 environmental variables as the current climatic conditions and were downloaded from WorldClim (http://worldclim.org/). As an example for transfers across
time, I only used these two climate scenarios, but note that great variability also can exist among climate modeling groups and (for future projections) among emissions scenarios.

Ecological niche modeling

I calibrated models using the presence-background technique MaxEnt version 3.3.2k (Phillips et al., 2006; Phillips & Dudík, 2008), which has performed well in comparison with other techniques (Elith et al., 2006; Wisz et al., 2008; Merow & Silander 2014; but see Royle et al., 2012; Fitzgerald et al., 2013 for criticisms of some aspects of MaxEnt). For MaxEnt, the default settings (= Auto features) were determined based on a broad study for a large number of species (with many different sample sizes) and in many different parts of the world (Phillips & Dudík, 2008). However, for various reasons, the default settings are expected to and have been shown to produce overfit models (for more details see Radosavljevic & Anderson, 2014).

Users are easily able to adjust critical model settings for Maxent in two ways: feature classes (FC) and regularization multiplier (RM). Feature classes determine the flexibility of the allowed modeled responses to the environmental predictor variables (e.g., linear, quadratic, product, … j; see below). The RM (a single coefficient applied to each of the $\beta_j$ regularization values specific to each feature class $j$) penalizes models for including additional parameters that do not introduce new information; hence, regularization tends to avoid overly complex models (Phillips & Dudík, 2008; Merow et al., 2013; Radosavljevic & Anderson, 2014; Shcheglovitova & Anderson 2013). When using default settings, sample size determines what FCs are allowed and the level of regularization (Phillips & Dudík, 2008).

To determine species-specific model settings, I tuned by varying different combinations of FC and RM and applying a jackknife approach of occurrence localities for evaluating model
performance (Shcheglovitova & Anderson, 2013). I used the recently developed R package ENMeval (Muscarella et al., in review; v. 3.03, R Development Core Team, 2014), and all tuning analyses were conducted with the raw output format of MaxEnt (although cumulative or logistic scalings would not affect the omission rates and AUC values calculated here). The jackknife procedure is a special case of $k$-fold cross validation, where $k = n$. Here, a model is calibrated using $n - 1$ localities and evaluated on the withheld locality; this is done until all localities have been used as an evaluation record (Pearson et al., 2007). This technique has been shown to identify better-performing and more realistic models built with small sample sizes (Shcheglovitova & Anderson 2013). For consistency of methods and limitations of the jackknife procedure for large sample sizes (which increases the likelihood that calibration localities and evaluation localities will be close to each other in space), I randomly rarefied the dataset of Microgale cowani from 32 localities to 25. Because of the small number of localities for each species, I limited the FCs allowed to the simplest combinations: Linear; Hinge; Linear and Quadratic; and Linear, Hinge, and Quadratic. Furthermore, I employed a range of regularization multiplier values, 0.5 – 6.0 with 0.5 intervals. This led to a total of 48 unique combinations of FCs and RMs for each species.

**Model tuning and optimality criteria**

To identify the optimal combination of model settings, and subsequently the co-optimal settings, I evaluated model performance using sequential criteria (Muscarella et al., in review; Shcheglovitova & Anderson 2013). To avoid models that are overfit to the calibration data, I first employed a threshold-dependent measure, specifically an omission rate (the proportion of evaluation localities that fall outside of the prediction once converted to a binary prediction). The
omission rate used here was based on the lowest presence threshold (LPT = 0% calibration omission rate; Pearson et al., 2007; = minimum training presence threshold of Maxent software). The lowest presence threshold rule sets the threshold at the lowest value of the prediction for any pixel that contains a calibration locality (Pearson et al., 2007; Anderson & Gonzalez, 2011) and has an expected omission rate of zero for evaluation localities. The higher the omission rate (above that expected value), the more overfit a model. This thresholding rule has been shown to be effective, and perhaps even conservative, for small sample sizes (Shcheglovitova & Anderson 2013). For each of the different combinations (FC/RM), I obtained an omission rate, and then averaged the values across the jackknife iterations (Shcheglovitova & Anderson 2013).

There can be several setting combinations that obtain the lowest omission rate, I apply another measure of evaluation, one that maximizes discriminatory ability. Specifically, as a secondary sequential criterion, I ranked the models with lowest omission rate by a rank-based threshold-independent metric: the Area Under the Curve (AUC) of the Receiver Operating Characteristic plot. Parallel to the earlier mentioned omission rates, for each iteration I obtained an AUC_{evaluation} score, and then averaged the values across the jackknife iterations (per combination of FC and RM; Shcheglovitova & Anderson 2013). I determined the optimal solution for each species (of the 48 combinations assessed here), as the model with the lowest average omission rate and, subsequently, the highest average AUC_{evaluation}.

*Model projection*

For each species, using the optimal settings and all filtered localities, a final optimal model was generated in the species-specific study region; that model was then projected to all of Madagascar (and surrounding islands) for current climatic conditions, LGM conditions, and
2070 conditions. For visualization purposes and calculations of Schoener’s $D$ statistic (see below), I used the logistic output format for the optimal, default, and co-optimal solutions (see below; for more details about logistic format see Phillips & Dudík, 2008; Merow et al., 2013). For comparison purposes, I built a model using all filtered localities and the default settings (FC=Linear, Quadratic, and Hinge; and RM=1.0). These models were also projected to all of Madagascar (and surrounding islands) for current climatic conditions, LGM conditions, and 2070 conditions. I visually inspected these two predictions in comparison with those resulting in the ensemble model (see below).

**Ensemble models and their associated uncertainty**

There can be several different solutions that have the same omission rate and very similar AUC$_{\text{evaluation}}$ scores. I chose the top 10% of models (out of the 48 FC and RM combinations) and defined those 5 as the co-optimal solutions. To determine variability in geography among the co-optimal solutions, I measured the pairwise similarity by calculating Schoener's $D$ values using the R package `ENMeval` (Muscarella et al., in review, Warren et al., 2009). Schoener's $D$ is a pixel-by-pixel comparison between two predictions with scores ranging from 0 (no overlap) to 1 (identical models) (Warren et al., 2009).

Following the projection methods provided above, I generated a prediction for each of the co-optimal solutions and all filtered localities in the custom study regions. Specifically, I projected each of the models to all of Madagascar (and surrounding islands) for current climatic conditions, LGM conditions, and 2070 conditions. I then created an ensemble niche model for each species by averaging the five raster grids from the co-optimal settings using the R package `Raster` (Hijmans et al., 2009, v. 3.03, R Development Core Team, 2014). For experimental
simplicity and because the co-optimal solutions are all high-performing models, I chose to average the rasters, but note that other methods exist as well for creating consensus models via ensemble modeling (see Araújo et al., 2005; Marmion et al., 2009). To obtain a map showing variation among co-optimal predictions in geography, I calculated the standard deviation of the predictions using the Raster package in R (=uncertainty map; Hijmans et al., 2009, v. 3.03, R Development Core Team, 2014). Additionally, I created average predictions for the LGM and the year 2070.

As another approach to measuring uncertainty (and because many studies employ thresholding rules), I converted each of the co-optimal models into binary predictions using the LPT threshold. I summed the binary maps for each species. The values of these maps range from 0 (where all of the models predict no suitable area) to 5 (where all of the predictions indicate suitable areas). To quantify the concordance among binary predictions for each species, I calculated the pairwise percentage of agreement. I did so by adding the total number of pixels that hold the same binary prediction (suitable or unsuitable) between the respective two models and dividing by the total number of pixels.

**Results**

*Model tuning*

Using the sequential criteria, the model tuning exercises yielded several clear trends regarding regularization multiplier and the evaluation statistics. For all FCs, omission rate generally decreased with increasing regularization multiplier (See Supplementary Fig. 2.1A, C, E, G for details). Additionally, for each FC, $\text{AUC}_{\text{evaluation}}$ also decreased with an increasing regularization multiplier (Supplementary Fig. 2.1B, D, F, H).
For the four species, the default settings were never identified as optimal. The optimal model settings (FC + RM) were the same for three species (*Microgale gracilis*, *M. gymnorrhyncha*, and *M. thomasi*): Linear and Quadratic with a RM of 0.5; furthermore, the optimal setting for *M. cowani* contained the same FC combination with a RM of 2.5 (Table 2.1). The Linear and Quadratic FCs allow less complex responses than the default settings, which also included the Hinge FC. However, perhaps in a compensatory fashion, the 0.5 RM was smaller than the default setting. In contrast, the 2.5 RM is larger than the default setting (penalizing complexity) (Table 2.1).

The five model settings considered as the co-optimal solutions essentially led to the same evaluation statistics (Table 2.1). For each species, several model setting combinations achieved the lowest average omission rate, and furthermore many of those combinations had AUC_{evaluation} scores that were very similar to the respective optimal model (Supplementary Fig. 2.1A–H; Table 2.1). Many of the co-optimal solutions consisted of Linear and Quadratic; or Linear, Quadratic, and Hinge FCs. Only two species had co-optimal model combinations that used only the Hinge feature class: *Microgale gracilis* (one of the co-optimal combinations) and *M. gymnorrhyncha* (two of the co-optimal combinations). The co-optimal RMs ranged from 0.5 to 3.5. Finally, the default settings were not part of the co-optimal solutions for *M. cowani* and *M. gracilis*, but they were identified as one of the five co-optimal settings for *M. gymnorrhyncha* and *M. thomasi*.

*Ensemble models and maps of uncertainty*

The co-optimal solutions varied widely in geography for three of the four species. For *Microgale gracilis, M. gymnorrhyncha*, and *M. thomasi*, Schoener’s *D* consistently showed high
variability among co-optimal solutions (Table 2.2). In contrast, the co-optimal solutions for *M. cowani* yielded very similar predictions in geography (from 0.92 to 0.98; Table 2.2). For illustration of each category, I present geographic predictions for one species that varied greatly (*M. gracilis*) and the one species that showed little variation (*M. cowani*; Fig. 2.1 & Fig. 2.2). For the binary predictions, the percentage of concordance between the co-optimal solutions for each species ranged from 93% to 99% (Supplementary Table 2.1). In contrast to the results for Schoener’s *D*, the co-optimal settings led to binary predictions that were relatively similar (for all species).

Qualitative evaluations among the predictions in geographic space (default, optimal, and ensemble) for the three time periods revealed several key differences, especially for *Microgale gracilis*. Inspection of the three types of predictions indicated differences that ranged from minimal (*M. cowani*) to very strong (*M. gracilis*) in the extent of predicted areas for each of the three time periods (Fig. 2.1 & Fig. 2.2). For each time period, the default settings for *M. cowani* led to a more restricted prediction than did either the optimal or ensemble predictions (Fig. 2.1). The optimal and ensemble predictions for *M. cowani* were usually very similar for each respective time slice; however, one important difference emerged for the current climate predictions. For that time period, the ensemble model predicted the suitable areas as contiguous, but the optimal model indicated clear intervening unsuitable areas (Fig. 2.1). Interestingly, for *Microgale gracilis*, the optimal solution was more similar to the default settings for the three time slices, than was the ensemble prediction. Again, for that species, the ensemble model predicted a larger geographic extent of suitable areas than the other two models. The most striking difference for *M. gracilis* appeared for the LGM predictions. The optimal settings
predicted suitable areas as disjunct, while the default settings and the ensemble models predicted a contiguous suitable area throughout Madagascar’s eastern mountains (Fig. 2.2).

The maps of uncertainty show the variation among co-optimal solutions in geographic space (Fig. 2.3). First, the standard deviation maps revealed the areas of the predictions that were highly variable among the model settings. *Microgale cowani* shows the most variation in the middle portion of Madagascar, and *M. gracilis* also shows a great deal of variation in the middle, and additionally in northern part of the island (Fig. 2.3). Additionally, the summed binary models for each species show that once converted into a binary prediction, many of the co-optimal solutions predict very similar suitable areas (Fig. 2.3; Supplementary table 2.1). The specific areas of variability according to the standard deviation map were also areas of discordance for each of the binary predictions.

**Discussion**

*Importance of model tuning and co-optimal solutions*

The default settings were not selected as optimal for any of the four species, which is consistent with results from other recent studies with Maxent (Radosavljevic & Anderson, 2014; Shcheglovitova & Anderson, 2013; Warren & Seifert 2011). The co-optimal solutions had essentially the same evaluation statistics (table 2.1), with less than a 0.01 difference in their AUC\textsubscript{evaluation} scores. Although the default settings were not part of the co-optimal solutions for *M. cowani* and *M. gracilis*, they were identified as a co-optimal setting for *M. gymnorrhyncha* and *M. thomasi*. Hence, for some species, the default settings may not lead to models that are overfit, but rather perform well as originally intended by Phillips & Dudík (2008).
Although the co-optimal solutions were very similar in their evaluation statistics, the settings for three of the four species led to models that varied considerably in geographic space (table 2.2; Fig. 2.3). For example, the co-optimal solutions for Microgale gracilis were only 81% similar on average according to Schoener’s $D$ (table 2.2). Additionally, among the five co-optimal solutions for $M. gracilis$, only one used Hinge FC. All of the other co-optimal settings were either Linear and Quadratic, or Linear, Quadratic, and Hinge. This could indicate that for the species used in this study, there were several different mathematical solutions to construct quantitatively similarly performing models that nevertheless differ in geographic space. When measuring similarity among the models constructed with the co-optimal settings for the current climatic conditions when the LPT threshold was applied, much of the difference dissipates (table s2.1). However, when the thresholding rule was applied to models projected to the LGM for $M. gracilis$, I identified key differences between the three different model types. Lastly, the maps of uncertainty indicate the areas in geographic space that vary with regard to relative suitability (Fig. 2.3). This becomes especially important when no thresholding rule is applied to the predictions (as shown in tables 2.2 and s2.1).

Conclusions and recommendations for transfers across time and space

This preliminary study using one algorithm, Maxent, illustrates the utility of making an ensemble prediction by tuning model parameters. Using this single-algorithm ensemble approach, I was able to identify that co-optimal solutions varied in geographic space. Although the predictions were visually similar in current climatic conditions when projected onto geography, these differences become exacerbated as models were projected through time. Generally, the variability within a single algorithm (most notably with different model settings)
within studies that utilize ENMs is rarely assessed, but the current results indicate that it can substantially affect the generated hypotheses and conclusions.

There are several caveats in the current study that should be addressed in future studies. The jackknife approach does not test the performance of a model after transfers across space, which should be used when the desired use is to transfer across time periods (Araújo & Guisan, 2006). Preferably, researchers will have a sufficient number of occurrence records to perform tests of spatially independent evaluations; unfortunately, however, most species are poorly documented and a lack of information limits these tests. Additionally, I did not fully explore parameter space with MaxEnt; therefore, other studies can determine if using a greater variety of settings brings more similarity among co-optimal predictions. Furthermore, studies can employ other criterion methods to generate the co-optimal solutions, such as using AIC (Warren & Seifert, 2011). Lastly, the number of co-optimal solutions can be varied and concordance among the settings can be observed. Despite these caveats, the current results demonstrate that different model settings, even if they show similar quantitative performance, can generate vastly different reconstructions of past suitable areas or future projections (Fig. 2.2).

This approach is scalable, and future studies should be undertaken to improve the transfer of ENMs across space and time. As mentioned above, several different algorithms can be tuned independently and their outputs combined, in a hierarchal manner, to generate an higher-level ensemble of algorithm-specific co-optimal models, as originally envisioned by Araújo & New (2007). Additionally, different climate retro/projections for a given time period can also be used. Furthermore, there are different methods available to combine the model outputs, and several of those methods can be used to determine if there is a difference in projection scenarios (see Marmion et al., 2009). The ultimate goal for correlative approaches should be to generate an
appropriate number of tuned models from different algorithms and applied to different climate projections, leading to a probability density function reflecting suitability (Araújo & New, 2007).
Acknowledgments

I thank Robert Anderson, Ana Carnaval, and Amy Berkov for their support and feedback throughout my masters. Additionally, I recognize Link Olson and Steven Goodman for collaborative efforts and providing the locality data used in both projects. I thank Jason Brown for providing the environmental layers for the last glacial maximum. I acknowledge the Anderson lab for all of their help and support throughout my tenure in the lab. Past and present members have helped me become a better scientist, and more importantly a better person. Robert Anderson has been an amazing advisor and taught me how to be a professional. He has never stopped forcing me to challenge myself beyond my own perceived limits. My time in his lab has changed my career aspirations entirely, and for that I am truly grateful. I thank Matthew Aiello-Lammens, Babatunde Alade, Amy Berkov, Ana C. Carnaval, Peter J. Galante, Maria Gavrutenko, Beth Gerstner, Jamie Kass, Mariano Soley-Guardia, Sara Varela, and an anonymous reviewer for their helpful comments and/or discussion on one or more versions of both chapters. I thank Jason Brown and Diego Alvarado-Serrano for their thoughts and comments throughout my second chapter.

This research was made possible by funding from the U. S. National Science Foundation (NSF DEB-1119915 and DEB-1120904) and the Luis Stokes Alliance for Minority Participation (Bridge to Doctorate Fellowship). Additional funds to present results were provided by the International Biogeography Society and by the City College Professional Development Institute.
References


Erratum: Evolution 2865: 1215.


Tables

*Table 1.1*

Summary statistics for performance and overfitting in experiments assessing the effects of geographic filtering in ecological niche models of the shrew tenrec *Microgale cowani* in Madagascar. The three datasets of localities (unfiltered, filtered, and rarefied unfiltered) were each subjected to spatially segregated (independent) evaluations using geographically structured $k$-fold cross-validation. Results are averaged for one measure of performance (overall discriminatory ability; $\text{AUC}_{\text{evaluation}}$) and three measures that include assessment of overfitting: $\text{AUC}_{\text{diff}}$ ($\text{AUC}_{\text{calibration}} - \text{AUC}_{\text{evaluation}}$), and average omission rates (at the lowest presence threshold and 10% calibration omission threshold). Note that the filtered dataset showed the lowest value for each measure related to overfitting, as well as the highest value for the measure of overall discriminatory ability.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$\text{AUC}_{\text{diff}}$</th>
<th>$\text{AUC}_{\text{evaluation}}$</th>
<th>Lowest presence threshold</th>
<th>10% calibration omission threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfiltered</td>
<td>0.057</td>
<td>0.903</td>
<td>0.193</td>
<td>0.263</td>
</tr>
<tr>
<td>Rarefied unfiltered</td>
<td>0.043</td>
<td>0.912</td>
<td>0.164</td>
<td>0.382</td>
</tr>
<tr>
<td>Filtered</td>
<td>0.028</td>
<td>0.914</td>
<td>0.100</td>
<td>0.194</td>
</tr>
</tbody>
</table>
Table 2.1

Summary statistics for the 5 co-optimal solutions in model tuning experiments for ecological niche models of *Microgale cowani, M. gracilis, M. gymnorrhyncha, and M. thomasi* in Madagascar. I calibrated and evaluated preliminary models using a jackknifing approach, tuning two model settings (feature classes and the regularization multiplier) to estimate optimal model complexity via a sequential procedure. Results are averaged for one measure of overfitting (average omission rates at the lowest presence threshold) and one measure of overall discriminatory ability; \( \text{AUC}_{\text{evaluation}} \). Note the overall similarity among the different model settings for each species. See supplementary appendix for comparisons with the 48 settings combinations that showed poorer performance for each species.
<table>
<thead>
<tr>
<th>Species</th>
<th>Model settings</th>
<th>Lowest presence threshold (omission rate)</th>
<th>AUC evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Mircogale cowani</em></td>
<td>LQ_2.5</td>
<td>0.040</td>
<td>0.883</td>
</tr>
<tr>
<td></td>
<td>LQ_3</td>
<td>0.040</td>
<td>0.882</td>
</tr>
<tr>
<td></td>
<td>LQH_3</td>
<td>0.040</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td>LQH_3.5</td>
<td>0.040</td>
<td>0.878</td>
</tr>
<tr>
<td></td>
<td>LQH_3.5</td>
<td>0.040</td>
<td>0.878</td>
</tr>
<tr>
<td><em>M. gracilis</em></td>
<td>LQ_0.5</td>
<td>0.067</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td>LQ_1</td>
<td>0.067</td>
<td>0.918</td>
</tr>
<tr>
<td></td>
<td>LQ_1.5</td>
<td>0.067</td>
<td>0.909</td>
</tr>
<tr>
<td></td>
<td>H_2.5</td>
<td>0.067</td>
<td>0.909</td>
</tr>
<tr>
<td></td>
<td>LQH_2.5</td>
<td>0.067</td>
<td>0.908</td>
</tr>
<tr>
<td><em>M. gymnorrhyncha</em></td>
<td>LQ_0.5</td>
<td>0.050</td>
<td>0.925</td>
</tr>
<tr>
<td></td>
<td>LQH_1</td>
<td>0.050</td>
<td>0.918</td>
</tr>
<tr>
<td></td>
<td>LQH_1.5</td>
<td>0.050</td>
<td>0.917</td>
</tr>
<tr>
<td></td>
<td>H_1.5</td>
<td>0.050</td>
<td>0.916</td>
</tr>
<tr>
<td></td>
<td>H_2</td>
<td>0.050</td>
<td>0.915</td>
</tr>
<tr>
<td><em>M. thomasi</em></td>
<td>LQ_0.5</td>
<td>0.063</td>
<td>0.888</td>
</tr>
<tr>
<td></td>
<td>LQH_3</td>
<td>0.063</td>
<td>0.885</td>
</tr>
<tr>
<td></td>
<td>LQH_3.5</td>
<td>0.063</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>LQH_2</td>
<td>0.063</td>
<td>0.883</td>
</tr>
<tr>
<td></td>
<td>LQH_1</td>
<td>0.063</td>
<td>0.883</td>
</tr>
</tbody>
</table>
Table 2.2

Pairwise Schoener’s D statistic measuring the amount of geographic similarity for the 5 co-optimal solutions in ecological niche models of *Microgale cowani*, *M. gracilis*, *M. gymnorhyncha*, and *M. thomasi* in Madagascar. Within each species, note the high level of variability between pairs of co-optimal models made with different settings, although those differences were much less for *M. cowani*.

<table>
<thead>
<tr>
<th>Model settings</th>
<th>LQ_2.5</th>
<th>LQ_3</th>
<th>LQH_3</th>
<th>LQH_3.5</th>
<th>LQ_3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M. cowani</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LQ_2.5</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQ_3</td>
<td>0.953</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_3</td>
<td>0.952</td>
<td>0.965</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_3.5</td>
<td>0.920</td>
<td>0.962</td>
<td>0.950</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>LQ_3.5</td>
<td>0.936</td>
<td>0.974</td>
<td>0.966</td>
<td>0.978</td>
<td>1</td>
</tr>
<tr>
<td>LQ_0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LQ_1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LQ_1.5</td>
<td>0.802</td>
<td>0.935</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>H_2.5</td>
<td>0.647</td>
<td>0.748</td>
<td>0.789</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>LQH_2.5</td>
<td>0.734</td>
<td>0.846</td>
<td>0.882</td>
<td>0.873</td>
<td>1</td>
</tr>
<tr>
<td>LQ_0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LQ_1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LQ_1.5</td>
<td>0.872</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_1.5</td>
<td>0.863</td>
<td>0.930</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>H_1.5</td>
<td>0.812</td>
<td>0.872</td>
<td>0.901</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>H_2</td>
<td>0.787</td>
<td>0.824</td>
<td>0.867</td>
<td>0.936</td>
<td>1</td>
</tr>
<tr>
<td>LQ_0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LQ_1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LQ_1.5</td>
<td>0.744</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_1.5</td>
<td>0.692</td>
<td>0.936</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_2</td>
<td>0.820</td>
<td>0.898</td>
<td>0.837</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>LQH_1</td>
<td>0.863</td>
<td>0.763</td>
<td>0.711</td>
<td>0.849</td>
<td>1</td>
</tr>
</tbody>
</table>
Maps showing the effects of geographic filtering on ecological niche models of the shrew tenrec *Microgale cowani* in Madagascar. Results are shown for unfiltered localities (A and C) and geographically filtered localities (B and D). Geographically structured *k*-fold cross-validation (see localities indicated within *k* = 3 rectangles) allowed evaluation of performance of spatially segregated (spatially independent) localities. Models calibrated with two of the three bins were then projected to the geographic region corresponding to the withheld bin (shown in C and D). The right panel provides details for a region in northern Madagascar for each dataset (E, unfiltered; F, filtered), with arrows indicating localities omitted by the respective prediction. Increasingly warm colors indicate successively stronger predictions. For each withheld region, areas in white indicate those pixels with values below the LTP threshold determined based on calibration data. Note that, overall, the filtered dataset led to a broader prediction, omitted fewer evaluation localities, and predicted the evaluation localities more strongly.
filtered

unfiltered

suitability

filtered

Kilometers

unfiltered

filtered
Fig. 2.1

Maps showing the three different ecological niche models for *Microgale cowani* in Madagascar: those based on default settings (left), optimal settings (middle), and an ensemble of co-optimal settings (right). Results are shown for three different time slices: current (top), the Last Glacial Maximum (21 kBP; middle) and 2070 (bottom). Increasingly warm colors indicate successively stronger predictions (higher suitability). Areas in white indicate pixels with values below the LTP threshold determined based on calibration localities. Note, overall, the high similarity between the optimal settings and the ensemble of co-optimal settings, but the ensemble prediction shows a broader extent for each time slice.
Maps showing the three different ecological niche models for *Microgale gracilis* in Madagascar: those based on default settings (left), optimal settings (middle), and an ensemble of co-optimal settings (right). Results are shown for three different time slices: current (top), the Last Glacial Maximum (21 kBP; middle) and 2070 (bottom). Increasingly warm colors indicate successively stronger predictions (higher suitability). Areas in white indicate pixels with values below the LTP threshold determined based on calibration localities. Note, overall, the differences between the ensemble model and the other two prediction methods, especially the differences among the LGM predictions.
Fig. 2.3

Maps showing two different methods of displaying uncertainty for *Microgale cowani* (top) and *M. gracilis* (bottom) in Madagascar: the standard deviation of the continuous output (left) and the binary ensemble method (right). In the former, increasingly darker red colors indicate more variation between different model settings. For the latter, increasingly darker blue colors indicated areas where the more of the binary co-optimal models predicted suitable conditions.
Microgale cowani

Microgale gracilis

Standard deviation

Binary ensemble
Supplementary documents

*Supplementary Table s1.1*

Summary statistics for threshold-independent measures of performance and overfitting in experiments assessing the effects of geographic filtering in ecological niche models of the shrew tenrec *Microgale cowani* in Madagascar. The three datasets of localities (unfiltered, filtered, and rarefied unfiltered) were each subjected to spatially segregated (independent) evaluations using geographically structured $k$-fold cross-validation. Results are for each iteration, as named by the individual bin withheld for evaluation. Raw results and the associated averages are provided for one measure of performance (overall discriminatory ability; $\text{AUC}_{\text{evaluation}}$) and one measure of overfitting, $\text{AUC}_{\text{diff}} (\text{AUC}_{\text{calibration}} - \text{AUC}_{\text{evaluation}})$.

<table>
<thead>
<tr>
<th></th>
<th>Unfiltered</th>
<th>Filtered</th>
<th>Rarefied unfiltered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calibration</td>
<td>Evaluation</td>
<td>Calibration</td>
</tr>
<tr>
<td>Bin 1</td>
<td>0.967</td>
<td>0.853</td>
<td>0.942</td>
</tr>
<tr>
<td>Bin 2</td>
<td>0.944</td>
<td>0.929</td>
<td>0.931</td>
</tr>
<tr>
<td>Bin 3</td>
<td>0.967</td>
<td>0.926</td>
<td>0.954</td>
</tr>
<tr>
<td>Average</td>
<td>0.960</td>
<td>0.903</td>
<td>0.942</td>
</tr>
<tr>
<td>$\text{AUC}_{\text{diff}}$</td>
<td>0.057</td>
<td></td>
<td>0.028</td>
</tr>
</tbody>
</table>
**Supplementary Table s1.2**

Summary statistics for threshold-dependent measures of performance and overfitting in experiments assessing the effects of geographic filtering in ecological niche models of the shrew tenrec *Microgale cowani* in Madagascar. The three datasets of localities (unfiltered, filtered, and rarefied unfiltered) were each subjected to spatially segregated (independent) evaluations using geographically structured *k*-fold cross-validation. Results are for each iteration, as named by the individual bin withheld for evaluation. Raw results and the associated averages are provided for two measures of performance and overfitting: omission rate at the lowest presence threshold and 10% calibration omission threshold.

<table>
<thead>
<tr>
<th>Omission rates</th>
<th>Unfiltered</th>
<th>Filtered</th>
<th>Rarefied unfiltered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LPT 10%</td>
<td>LPT 10%</td>
<td>10% LPT</td>
</tr>
<tr>
<td>Bin 1</td>
<td>0.421</td>
<td>0.474</td>
<td>0.200 0.300 0.100</td>
</tr>
<tr>
<td>Bin 2</td>
<td>0.000</td>
<td>0.105</td>
<td>0.000 0.182 0.091</td>
</tr>
<tr>
<td>Bin 3</td>
<td>0.158</td>
<td>0.211</td>
<td>0.100 0.100 0.300</td>
</tr>
<tr>
<td>Average</td>
<td>0.193</td>
<td>0.263</td>
<td>0.100 0.194 0.164</td>
</tr>
</tbody>
</table>
**Supplementary Table s2.1**

Pairwise binary concordance values among binary predictions for the 5 co-optimal solutions in ecological niche models of *Microgale cowani*, *M. gracilis*, *M. gymnorhyncha*, and *M. thomasi* in Madagascar. Note the high concordance between all pairwise comparisons when the lowest presence training threshold rule was applied to each co-optimal setting.

<table>
<thead>
<tr>
<th>Model settings</th>
<th>LQ_2.5</th>
<th>LQ_3</th>
<th>LQH_3</th>
<th>LQH_3.5</th>
<th>LQ_3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LQ_2.5</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQ_3</td>
<td>0.968</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_3</td>
<td>0.989</td>
<td>0.973</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_3.5</td>
<td>0.945</td>
<td>0.978</td>
<td>0.9518</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>LQ_3.5</td>
<td>0.9715</td>
<td>0.993</td>
<td>0.978</td>
<td>0.973</td>
<td>1</td>
</tr>
<tr>
<td>LQ_0.5</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_1</td>
<td>0.971</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_1.5</td>
<td>0.954</td>
<td>0.983</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>H_2.5</td>
<td>0.944</td>
<td>0.957</td>
<td>0.956</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>LQH_2.5</td>
<td>0.979</td>
<td>0.980</td>
<td>0.965</td>
<td>0.961</td>
<td>1</td>
</tr>
<tr>
<td>LQ_0.5</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQ_1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_1.5</td>
<td>0.970</td>
<td>0.988</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>H_1.5</td>
<td>0.963</td>
<td>0.975</td>
<td>0.9758</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>H_2</td>
<td>0.964</td>
<td>0.974</td>
<td>0.976</td>
<td>0.961</td>
<td>1</td>
</tr>
<tr>
<td>LQ_0.5</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_1</td>
<td>0.967</td>
<td>1</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_2</td>
<td>0.961</td>
<td>0.993</td>
<td>1</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LQH_2.5</td>
<td>0.968</td>
<td>0.992</td>
<td>0.986</td>
<td>1</td>
<td>—</td>
</tr>
<tr>
<td>LQH_1</td>
<td>0.967</td>
<td>0.976</td>
<td>0.979</td>
<td>0.976</td>
<td>1</td>
</tr>
</tbody>
</table>
Supplementary Figure s1.1

Maps showing the results of the rarefied unfiltered dataset (A) on ecological niche models of the tenrec *Microgale cowani* in Madagascar. Geographically structured *k*-fold cross-validation (see localities indicated within *k* = 3 rectangles) allowed evaluation of model performance of spatially segregated (spatially independent) localities. Models calibrated with two of the three bins were then projected to the geographic region corresponding to the withheld bin (shown in B). Increasingly warm colors indicate successively stronger predictions. For each withheld region, areas in white indicate those pixels with values below the LTP threshold determined based on calibration data.
unfiltered rarefied

B

0  suitability  1.0
Supplementary Figure s1.2

Bivariate plot of values for occurrence localities of two environmental variables (mean annual precipitation and mean annual temperature) for each dataset of the tenrec *Microgale cowani* in Madagascar. These variables were chosen because they provide an intuitive general overview of environmental space inhabited by the species. Note that when applying the spatial filter, no localities with unique environmental conditions were removed for these two variables.
Supplementary Figure s2.1

Results of threshold-dependent and threshold-independent evaluations in tuning experiments of *Microgale cowani* (A, B), *M. gracilis* (C, D), *M. gymnorrhyncha* (E, F), and *M. thomasi* (G, H) in Madagascar. The omission rate for each model combination was calculated by averaging each jackknife iteration using the lowest presence threshold (A, C, E, and G). The AUC<sub>evaluation</sub> for each model combination also was calculated by averaging each n-jackknife iteration (B, D, F, and H).
C.

![Graph](image)

D.

![Graph](image)
E. OR vs RM

F. AUC evaluation vs RM
G. OR vs RM

H. AUC vs RM