



Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar

Richard G. Pearson^{1*}, Christopher J. Raxworthy², Miguel Nakamura³ and A. Townsend Peterson⁴

¹Department of Herpetology & Center for Biodiversity and Conservation, American Museum of Natural History, Central Park West at 79th Street, New York, NY, USA,

²Department of Herpetology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA, ³Centro de Investigacion en Matematicas, A.C. Apartado Postal 402, Guanajuato, Gto., 36000, Mexico and ⁴Natural History Museum & Biodiversity Research Center, the University of Kansas, Lawrence, KS 66045-2454, USA

ABSTRACT

Aim Techniques that predict species potential distributions by combining observed occurrence records with environmental variables show much potential for application across a range of biogeographical analyses. Some of the most promising applications relate to species for which occurrence records are scarce, due to cryptic habits, locally restricted distributions or low sampling effort. However, the minimum sample sizes required to yield useful predictions remain difficult to determine. Here we developed and tested a novel jackknife validation approach to assess the ability to predict species occurrence when fewer than 25 occurrence records are available.

Location Madagascar.

Methods Models were developed and evaluated for 13 species of secretive leaf-tailed geckos (*Uroplatus* spp.) that are endemic to Madagascar, for which available sample sizes range from 4 to 23 occurrence localities (at 1 km² grid resolution). Predictions were based on 20 environmental data layers and were generated using two modelling approaches: a method based on the principle of maximum entropy (Maxent) and a genetic algorithm (GARP).

Results We found high success rates and statistical significance in jackknife tests with sample sizes as low as five when the Maxent model was applied. Results for GARP at very low sample sizes (less than c. 10) were less good. When sample sizes were experimentally reduced for those species with the most records, variability among predictions using different combinations of localities demonstrated that models were greatly influenced by exactly which observations were included.

Main conclusions We emphasize that models developed using this approach with small sample sizes should be interpreted as identifying regions that have similar environmental conditions to where the species is known to occur, and not as predicting actual limits to the range of a species. The jackknife validation approach proposed here enables assessment of the predictive ability of models built using very small sample sizes, although use of this test with larger sample sizes may lead to overoptimistic estimates of predictive power. Our analyses demonstrate that geographical predictions developed from small numbers of occurrence records may be of great value, for example in targeting field surveys to accelerate the discovery of unknown populations and species.

Keywords

Environmental niche modelling, geckos, jackknife, Madagascar, model validation, small sample size, species distribution modelling.

*Correspondence: Richard G. Pearson, Department of Herpetology & Center for Biodiversity and Conservation, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA. E-mail: pearson@amnh.org

INTRODUCTION

Techniques that characterize the geographical distributions of species abiotic niches by relating observed occurrence localities to environmental data have been widely applied across a range of biogeographical analyses (Guisan & Thuiller, 2005). Applications have included guiding field surveys to accelerate detection of unknown distributional areas and undiscovered species (Raxworthy *et al.*, 2003; Bourg *et al.*, 2005), projecting potential impacts of climate change (e.g. Iverson & Prasad, 1998; Thomas *et al.*, 2004; Thuiller *et al.*, 2005a), testing evolutionary hypotheses (e.g. Peterson *et al.*, 1999; Graham *et al.*, 2004b), predicting species invasions (e.g. Peterson, 2003; Thuiller *et al.*, 2005b) and supporting conservation planning (e.g. Araújo & Williams, 2000; Ferrier *et al.*, 2002). Some of the most promising applications of these models relate to poorly known tropical landscapes where biogeographical data are scarce (Raxworthy *et al.*, 2003). In such regions, distributional data for some difficult-to-detect species are often limited to small samples of observed localities (e.g. < 25) due to limited recent survey effort and the lack of precise locality data associated with some museum specimens (Graham *et al.*, 2004a; Soberón & Peterson, 2004). The potential for very limited samples of locality records to yield useful predictions of species occurrences thus warrants detailed analysis.

The modelling approach that we use aims to define the environmental conditions within which a species can persist by associating known distributional information with suites of environmental variables. Geographical regions presenting similar environments to where the species has been observed can thus be identified. The central premise of this approach is that the observed distribution of a species provides useful information as to its environmental requirements (Pearson & Dawson, 2003). The degree to which these models describe fully the range of conditions within which a species can persist (the fundamental niche, *sensu* Hutchinson, 1957) depends on the degree to which the environmental dimensions examined actually define the species' distributional limits. In reality, additional factors not considered in the modelling (including biotic interactions, geographic barriers and history) mean that species rarely occupy all areas with suitable environments (Anderson *et al.*, 2002; Svenning & Skov, 2004; Araújo & Pearson, 2005). The output from niche-based distribution models must therefore be interpreted carefully (for discussion see Pearson & Dawson, 2003; Soberón & Peterson, 2005; Phillips *et al.*, 2006). Despite these caveats, distribution models have been shown to yield highly informative biogeographical information (e.g. Fleishman *et al.*, 2002; Bourg *et al.*, 2005).

Herein, we test our ability to model species potential distributions when fewer than 25 occurrence records are available. A number of previous studies have used distribution models with low numbers of records (e.g. Loiselle *et al.*, 2003, minimum sample size of four; Anderson & Martínez-Meyer, 2004, sample sizes of seven and 12; Ortega-Huerta & Peterson, 2004, minimum sample size of two) yet assessment of predictive performance has been dependent on the availability

of observed absence data (e.g. Loiselle *et al.*, 2003), which as negative data can be problematic (see later), or on the partitioning of data into training and test data sets, which can become very small (e.g. Anderson *et al.*, 2002; Anderson & Martínez-Meyer, 2004). In some cases, the lack of occurrence records has meant that no independent test of model quality was carried out (e.g. Ortega-Huerta & Peterson, 2004). Whilst some studies have demonstrated deterioration in predictive performance as sample sizes are decreased (Stockwell & Peterson, 2002; Reese *et al.*, 2005), we are not aware of any studies that have developed and applied a suitable test statistic for investigating predictive performance with very low sample sizes.

We describe a novel jackknife approach for testing distribution models that enables an assessment of predictive ability to be made when few observed locality records are available. Two modelling techniques are employed: the Maximum Entropy method (Maxent; Phillips *et al.*, 2006) and the Genetic Algorithm for Rule-Set Prediction (GARP; Stockwell & Peters, 1999). Since reliable absence data are rarely available for species that are difficult to detect in surveys, or from regions that have been poorly surveyed, the modelling techniques and validation approach that we apply use presence data only. We evaluate models based on the localities of 13 species of leaf-tailed *Uroplatus* geckos of Madagascar, a generally difficult-to-detect group of nocturnal lizards endemic to one of the most species-rich and threatened regions of the world (Myers *et al.*, 2000; Ricketts *et al.*, 2005).

MODEL TESTING WITH SMALL SAMPLE SIZES

Assessing the predictive performance of niche-based distribution models requires careful selection of (1) an approach for generating independent test data and (2) an appropriate test statistic. Problems associated with each of these choices are exacerbated when few observed locality points are available.

A legitimate measure of predictive success should make use of independent data not used to develop the model. We refer to this as 'test' data (sometimes termed 'evaluation' data), whilst data used to build the model can be called 'training' data (sometimes termed 'calibration' data; Fielding & Bell, 1997; Hastie *et al.*, 2001). Ideally, test data would be collected independently after the model has been developed. In practice, however, available data are often split into test and training sets using a partitioning method such as bootstrapping, randomization or *k*-fold partitioning (Fielding & Bell, 1997; Araújo *et al.*, 2005). Perhaps the simplest and most common approach is to split the available data into training and test sets, using either a random (e.g. Pearson *et al.*, 2002) or spatially stratified (e.g. Peterson & Shaw, 2003) partition. However, such approaches are not appropriate when available data are limited, since both training and test data sets become very small.

Several test statistics (or 'discrimination indexes') have been proposed and applied to test model performance (Fielding & Bell, 1997; Pearce & Ferrier, 2000). Statistics are often derived

from a 2×2 confusion matrix, which describes the frequency with which known presences and absences are correctly and incorrectly predicted. When only presence data are available, tests are limited to those that do not require absence data (use of 'observed' absences can only be justified when a site has been surveyed extensively, or when species can be confidently detected by a single survey; Anderson *et al.*, 2003). Commonly applied indexes, including kappa and the area under the receiver operating characteristic curve (AUC), are thus unsuitable for evaluation in poorly sampled regions for which absence data are not available (Boyce *et al.*, 2002; but see Phillips *et al.*, 2006). We also argue that observed absence data should be excluded from validations on theoretical grounds: when a model is used to identify *potential* suitable habitat, it is not correct to judge false-positive predictions as failures since absence at a given locality may be caused by factors not included in the model (e.g. historical contingency, dispersal limitation and biotic interactions; Anderson *et al.*, 2003; Pearson & Dawson, 2003; Soberón & Peterson, 2005).

To provide informative predictions, it is necessary for a model to successfully predict a high proportion of test localities (i.e. have a low omission rate) whilst not predicting as suitable such a large proportion of the study area as to make the model statistically indistinguishable from a random prediction (Anderson *et al.*, 2002). When test data have been partitioned from the full data set, statistical significance can be assessed using a chi-square test or, for small sample sizes, an exact upper-tailed binomial probability (e.g. ten test localities; Anderson *et al.*, 2002). This approach tests whether test points fall into areas predicted present more often than expected at random, given the overall proportion of the study area predicted to be present (Anderson *et al.*, 2002). However, an alternative significance test is required for the jackknife validation approach described below, because multiple models are generated, each with a different proportion of the study region predicted present.

Model testing using a statistic based on a jackknife technique

Because of the importance of model assessment, and the reality of small numbers of observed localities for many species, we implemented a jackknife (or 'leave-one-out') procedure. Each observed locality was removed once from the set of data and a model built using the remaining $n - 1$ localities. Hence, for a species with n observed localities, n separate models were built for testing. Predictive performance was then assessed based on the ability of each model to predict the single locality excluded from the training data set.

The test of significance that we implemented is as follows. We denote by n the number of observations. As already stated, a given prediction method (e.g. Maxent, GARP) is applied successively to sets of $n - 1$ localities, by leaving out one observation at a time. Let p_i be the proportion of the study area predicted present when having deleted the i th point. Furthermore, let X_i be a success-failure variable to indicate if the i th

point is included in the i th predicted area or not (X_i takes on the value 1 if such point is included in the prediction and 0 otherwise).

Let H denote the assumption of a completely random assignment. Under H , X_i is essentially a random trial with probability of success P_i . It is thus of interest to examine if observed successes constitute evidence against H , in the direction of 'better than at random'. This is accomplished by a P value, based on a suitable test criterion and a probability distribution under H (see Sprott, 2000, Chapter 6). The test criterion (or test statistic) adopted here, D , is simply

$$D = \sum X_i(1 - P_i).$$

Note that this amounts to weighting the successes (and not the failures); a success ($X_i = 1$) carrying greater weight ($1 - P_i$) if it has occurred under a small assumed probability. This D effectively ranks possible values of X s according to the evidence they provide against H . If d denotes the observed experimental value of D , the corresponding P value is computed as the probability under H that $D \geq d$. Since D is dependent on the proportional areas predicted in the n runs, a universal table of P values cannot be created and the probability calculation must be achieved by exhaustive case-by-case summation. Examine all possible ways in which the array of n values of X s can occur, and sum probabilities under H for all of these arrays that satisfy $D \geq d$. The individual probability of a given array is the product of corresponding P_i s and $(1 - P_i)$ s. For example, if $n = 4$ and the array under consideration is (1, 0, 1, 0), its individual probability is $P_1(1 - P_2)P_3(1 - P_4)$. We have made available a program for calculating this P value (Supplementary Appendix S1).

We note that there is an implicit technical assumption behind the previous calculation: that the jackknife trials are independent. In a strict sense, this assumption is not true because each jackknife trial is sharing some of the data. However, when sample sizes are small each locality is expected to have a large influence on the model projection, giving varying results from different combinations of available localities. We further address this issue in the Discussion, and contend that for small sample sizes the P value thus described is at least approximately correct, and therefore provides a useful measure of predictive ability.

Spatially autocorrelated localities

Species observations and collecting events tend, inevitably, to be clustered around field camps and settlements. This gives rise to the possibility of inflating validation statistics by including localities that are not spatially independent (Hampe, 2004; Luoto *et al.*, 2005). To lessen this effect, for each jackknife model run we conservatively removed from the training data set all localities situated within 10 km of the test locality. Based on the substantial local variation in topography and climatic conditions that exists in Madagascar (as shown in our environmental layers), we consider localities separated by at

least 10 km to exhibit sufficient potential variation as to be considered spatially independent.

DATA AND MODEL BUILDING

Study region and species sampling

The island of Madagascar (the fourth largest island in the world) has a surface area of 587,000 km², and exhibits substantial environmental gradients, with diverse climates and complex topography. The interior of the island forms a high plateau above 1000-m elevation, with four complex massif systems exceeding 1900 m, and summits up to 2876 m. The steep eastern escarpment of the high plateau results in much of the rainfall from the Indian Ocean trade winds falling on the eastern side of the island, and consequently creates a western rain-shadow. Recorded mean annual rainfall ranges from 350 to 3753 mm, and the island is heavily influenced by cyclones during the rainy season between January and April (Donque, 1972). Recorded mean annual temperatures range from 13.9 to 27.4°C, with increasing seasonality in temperatures in the south, resulting from the latitudinal range of the island (12–25°) that straddles the Tropic of Capricorn.

The biota of Madagascar has long been recognized as one of the world's most threatened, due to both the high levels of diversity and endemism on the island, and the decline of natural habitats (e.g. Myers *et al.*, 2000; Ricketts *et al.*, 2005). The Gondwanan origin and long isolation (88 Myr BP) of Madagascar (Storey *et al.*, 1995; Raxworthy *et al.*, 2002) has resulted, for many biological groups, in the evolution of species-rich clades that exhibit species-level regional endemism. One such group is the *Uroplatus* leaf-tailed geckos (Gekkonidae, Reptilia). These are endemic to Madagascar, and are distributed throughout all regions of the island (except, so far as we know, the extreme south). No species is cosmopolitan across the island; each is associated with a specific primary vegetation type and elevational range.

Current *Uroplatus* taxonomy recognizes 11 species (Raxworthy, 2003), but we here also consider three other species that are in the process of being formally described (C. J. Raxworthy & R. A. Nussbaum, in preparation), and we treat the two subspecies of *Uroplatus sikorae* (*Uroplatus sikorae sikorae* and *Uroplatus sikorae sameiti*) as good species based on the diagnostic features of each taxon (Böhme & Ibish, 1990; C. J. Raxworthy & R. A. Nussbaum, in preparation). Until recently, *Uroplatus* geckos had been rarely collected, and there were relatively few specimens held in museum collections (see Bauer & Russell, 1989). This situation presumably developed because all *Uroplatus* species are arboreal and only active at night, and diurnal collecting is extremely difficult due to their cryptic habits of either retreating to arboreal refugia (e.g. under bark or within palm axils) or else blending onto the surface of tree trunks using skin fringes on the head, body, limbs and tail (Raxworthy, personal observation). However, since 1985, extensive new collections of *Uroplatus* have been made as a result of herpetological surveys led by R. A.

Nussbaum and one of us (CJR) throughout most regions of Madagascar (see Andriamialisoa & Langrand, 2003).

For this study, we selected *Uroplatus* to serve as a test case for distribution modelling for the following reasons: (1) the availability of accurate georeferenced locality data associated with specimen vouchers, (2) the unambiguous taxonomy (in reference to our revision of this group), (3) the species locality samples sizes available to us were typically small, ranging from 2 to 23 localities, (4) the group exhibits regional endemism in most areas of Madagascar, thus making this group a useful test case for the island, and (5) the pressing conservation and evolutionary applications for validated distribution models for this genus. The locality data sets were compiled based on voucher specimens held at the American Museum of Natural History (AMNH) and the University of Michigan Museum of Zoology (UMMZ), and supplemented with other museum records that could be georeferenced (Bauer & Russell, 1989; Böhme & Ibish, 1990; Raselimanana, 1998, 1999; Andreone *et al.*, 2001; Rakotomalala *et al.*, 2001; Rakotomalala, 2002; Böhme & Schönecker, 2003). Because of the resolution of the environmental layers (1 km²), where specimens had been collected in close proximity to each other, only one occurrence record per grid cell was included in this analysis.

Environmental coverages

We assembled 20 environmental coverages relating to three principal traits: temperature, precipitation and topography (Supplementary Appendix S2). All environmental variables were resampled to an oblique Mercator projection at 1 km² resolution. Eleven temperature-derived variables were extracted from the WorldClim data base (Hijmans *et al.*, 2005; <http://www.worldclim.org/>), which is a set of global climate layers generated through interpolation of climate data from weather stations on a 30'' grid (c. 1 km² resolution). Weather station data for Madagascar covered the period 1930–90, and included temperature records from approximately 117 stations (R. Hijmans, personal communication).

Four precipitation variables were derived from NOAA's Famine Early Warning System (FEWS) data archive (<http://www.cpc.ncep.noaa.gov/products/fews/data.html>). FEWS precipitation estimates are generated at 0.1° resolution using a method that incorporates multiple data sources: numerical model analyses, satellite imaging and recorded surface data (Herman *et al.*, 1997; updated technical description at http://www.cpc.ncep.noaa.gov/products/fews/RFE2.0_tech.pdf). Precipitation variables were calculated from 10-day estimates for the period January 1996–December 2004. Mean monthly values for February and August were included as individual variables so as to represent wet and dry periods, respectively. FEWS precipitation data were considered advantageous over estimates derived by interpolation from weather station records (e.g. Hijmans *et al.*, 2005), since merging data from multiple sources has been shown to reduce bias and random error significantly compared to individual precipitation data sources (Xie & Arkin, 1996). We expect that the advantages of

incorporating satellite and modelled data in precipitation estimates will be of particular value in Madagascar, where weather stations are at low density and interpolation of precipitation between localities is thus particularly uncertain.

Topographical variables were taken from the US Geological Survey's Hydro1k data base (<http://edcdaac.usgs.gov/gtopo30/hydro/index.asp>). Hydro1k provides data layers derived from a 30'' global digital elevation model. To avoid problems with continuity of radially distributed variables, aspect was converted into two linear variables describing 'northness' and 'eastness' (Supplementary Appendix S1).

Modelling algorithms

Several alternative methods have been used to model ecological niches and predict the geographical distributions of species. In many cases, methods requiring both presence and absence records are used (e.g. classification and regression trees; Bourg *et al.*, 2005), but reliable absence data are rarely available in poorly sampled regions, or for species that are easily missed during surveys. To model based on presence records only, methods requiring both presence and absence records have been applied by sampling 'pseudo-absences' from the study area in place of real absence data (Ferrier *et al.*, 2002; Engler *et al.*, 2004). Another approach is to use background environmental data for the entire study area (Hirzel *et al.*, 2002). Other methods rely solely on presence records by assessing similarity with observed occurrences in environmental space (e.g. Nix, 1986; Carpenter *et al.*, 1993).

Studies have identified considerable differences between predictions obtained from different modelling algorithms, emphasizing the importance of careful selection of appropriate methods and the need to assess results from more than one approach (Thuiller *et al.*, 2004; Pearson *et al.*, 2006). Here, we apply the Maxent (Phillips *et al.*, 2006) and GARP (Stockwell & Peters, 1999) approaches, both of which have characteristics that make them appropriate for our application, and for our presence-only occurrence data.

Maxent

Maxent is a general-purpose method for characterizing probability distributions from incomplete information, and has recently been applied to modelling species distributions (Phillips *et al.*, 2006). For a concise mathematical definition of Maxent, discussion of its application to species distribution modelling and initial testing of the approach see Phillips *et al.* (2006). In estimating the unknown probability distribution defining a species' distribution across a study area, Maxent formalizes the principle that the estimated distribution must agree with everything that is known (or inferred from the environmental conditions at the occurrence localities) but should avoid placing any unfounded constraints. The approach is thus to find the probability distribution of maximum entropy – that which is closest to uniform – subject to constraints imposed by the information available regarding

the observed distribution of the species and environmental conditions across the study area. Maxent computes a probability distribution based on environmental variables spread over the entire study area; an approach to dealing with lack of absence data most similar to that of Hirzel *et al.* (2002). However, since our study region contains a very large number of pixels (approaching 600,000) the implementation that we used took a random sample of 100,000 'background' pixels to represent the environmental conditions present in the region.

We implemented Maxent using version 1.8.2 of the software developed by S. Phillips and colleagues (for free download see: <http://www.cs.princeton.edu/~schapire/maxent/>). Recommended default values were used for the convergence threshold (10^{-5}) and maximum number of iterations (500). Suitable regularization values, included to reduce overfitting, were selected automatically by the program. Selection of 'features' (environmental variables or functions thereof) was also carried out automatically, following default rules dependent on the number of presence records. Maxent assigns a probability of occurrence to each cell in the study area. Because these probabilities must sum to 1, each cell's probability is usually extremely small, making model output difficult to interpret. We therefore present model predictions as cumulative probabilities, wherein the value of a given grid cell is the sum of that cell and all other cells with equal or lower probability, multiplied by 100 to give a percentage (Phillips *et al.*, 2006). The Maxent output is thus a continuous variable ranging from 0 to 100, indicating relative suitability (not probability of occurrence). Maxent is a promising new method for modelling species potential distributions and has been shown to perform well in comparison with alternative approaches (Elith *et al.*, 2006).

GARP

GARP has been tested and applied more widely (e.g. Peterson *et al.*, 1999; Martínez-Meyer *et al.*, 2004; Roura-Pascual *et al.*, 2004). GARP is a machine-learning approach that develops a set of conditional rules to relate observed occurrences to environmental variables (Stockwell & Peters, 1999). Predicted presences are defined by envelope (i.e. upper and lower bounds for each environmental variable), atomic (i.e. specific values or categories for each variable), and logistic regression (logit) rules. The set of rules is developed using a genetic algorithm, which refines the solution in an evolutionary manner by testing and selecting rules on random subsets of available data. The algorithm requires absence samples for rule development, but is implemented to deal with presence-only data by selecting 1250 pseudo-absence localities at random from the study area (Stockwell & Peters, 1999).

We used the Desktop GARP application (version 1.1.6; <http://www.lifemapper.org/desktopgarp/>) and followed common procedure for implementation. For each model run, we used 50% of the localities for training, with a convergence limit of 0.01 and maximum number of iterations of 1000. Given the heuristic nature of GARP processing each run provides a

different binary solution, so we applied a modification of the 'best subsets' procedure described by Anderson *et al.* (2003). For each species, we ran 100 GARP models (each time randomly selecting a new 50% of localities for training, to ensure all data were used across the set of models) and selected the 20 models that omitted fewest of the remaining 50% of localities (i.e. an extrinsic test; Anderson *et al.*, 2003); of these models, we selected the 10 closest to the median predicted area. This approach favours predictions with low omission error whilst removing models that are overfit or that present unduly large predicted areas. The final GARP prediction was produced by summing the 10 selected models, giving a prediction that ranges from 0 to 10, and increases in increments of 1.

Decision thresholds

To aid model validation and interpretation, it is usually desirable to distinguish 'suitable' from 'unsuitable' areas by setting a decision threshold above which model output is considered to be a prediction of presence (Pearson *et al.*, 2004). Although validation statistics are available that are independent of a threshold (e.g. AUC), these are unsuitable when only presence data are available (Boyce *et al.*, 2002; but see Phillips *et al.*, 2006). The jackknife validation approach that we use here requires application of a threshold.

Many different approaches have been employed for setting thresholds (Liu *et al.*, 2005), yet most techniques depend on balancing false-positive and false-negative predictions, making them applicable only to cases in which absence data are available. For the presence-only case, we considered here two alternative thresholds. First, we chose the lowest predicted value associated with any one of the observed presence records; we term this the 'lowest presence threshold' (LPT). This approach can be interpreted ecologically as identifying pixels predicted as being at least as suitable as those where a species' presence has been recorded; it is thus conservative, identifying the minimum predicted area possible whilst maintaining zero omission error in the training data set.

The second approach is intended to be more liberal by incorporating a larger predicted area. We thus applied fixed thresholds that rejected only the lowest 10% of possible predicted values: for Maxent we used a threshold of 10 (T10), and for GARP a threshold of 1 (T1). These thresholds were chosen following initial tests that gave LPT values higher than the selected fixed thresholds. Whilst somewhat arbitrary, the fixed thresholds provide an alternative against which the LPT can be compared. It is notable that fixing the GARP threshold at 1 is equivalent to the 'any model predicts' criterion used for the best subsets models by Raxworthy *et al.* (2003).

Jackknife model testing

The jackknife approach described earlier was implemented to test predictive performance. Thus, multiple predictions were

made per species, with one of the observed localities excluded in each case. For each prediction, a decision threshold was applied (based on the training localities) and the ability to predict the excluded locality was tested. A *P* value was then calculated for each species across the set of jackknife predictions using the program made available as Supplementary Appendix S2.

Testing the impacts of varying sample size

To explore the effects of low sampling effort, we studied changes in model performance as sample size was reduced artificially. Here we used Maxent and analysed the two species with the largest numbers of localities: *Uroplatus sikorae* (23 localities) and *Uroplatus sameiti* (16 localities). For each species, localities were removed in random order, one at a time, from the training data set and a model developed at each sample size. Changes in predictive performance were examined by calculating the proportion of cells predicted as present both by the model trained on all available localities and the model trained on a reduced set of localities. Five different random sequences of locality removal were run for each species. This experimental approach also enabled the relationship between sample size and LPT to be investigated.

RESULTS

Models were developed for 13 *Uroplatus* species. Two described species, *Uroplatus alluaudi* and *Uroplatus pietschmanni*, which could not be modelled here, have just three and two localities, respectively. These were too few spatially independent localities to enable model testing, even using our jackknife approach, and thus these species were not considered suitable for modelling.

Sample size influences on model projections and test performance

Projected potential distributions for six example species using Maxent (T10) are presented in Fig. 1. In each case, the model was trained using all available localities. Figure 1(a)–(c) shows models for species with sample sizes ≥ 15 ; in each instance, the principal distributional areas are identified and high success rates (i.e. low omission rates) and statistical significance were obtained during testing (Table 1). The projections presented in Fig. 1(d) and (e) were trained using six localities and show high, and significant, success rates in jackknife tests. In contrast, the model for *Uroplatus malahelo* (Fig. 1f) was trained using only four localities and, although the projection appears to identify a distinct area that could be interpreted as being the main area of distribution, inspection of the test statistics reveals that its predictive power was poor (two of four excluded localities predicted).

Jackknife test results show high success rates and statistical significance using Maxent (T10) for sample sizes ≥ 5 (Table 1).

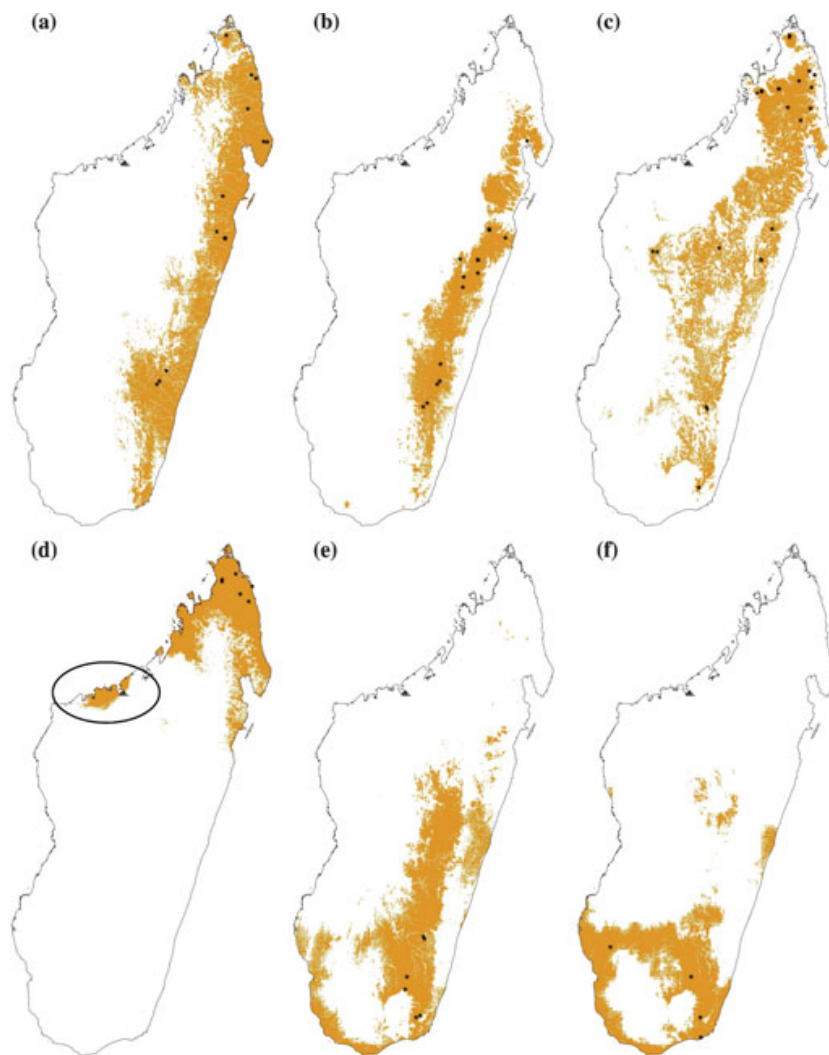


Figure 1 Modelled distributions for six example species, using Maxent and applying a threshold of 10: (a) *U. fimbriatus*, (b) *U. phantasticus*, (c) *U. sikorae*, (d) *U. sp. A*, (e) *U. malama*, (f) *U. malahelo*. Predicted areas are shaded, and observed localities are shown as black stars. The circled area in (d) is an example of a disjunct area of 'overprediction'.

Table 1 Jackknife tests of distribution models for 13 species of leaf-tailed gecko (*Uroplatus* spp.)

Species	Locality sample size	Maxent, LPT		Maxent, T10		GARP, LPT		GARP, T1	
		Successes	P value	Successes	P value	Successes	P value	Successes	P value
<i>U. malahelo</i>	4	0	1.0	2	0.0481	0	1.0	0	1.0
<i>U. sp. C</i>	5	2	5.27e-05	5	4.72e-09	1	0.0092	2	1.00e-04
<i>U. sp. A</i>	6	2	0.005	5	2.53e-05	2	1.0	2	0.0092
<i>U. malama</i>	6	2	0.0067	6	2.71e-05	1	1.0	1	0.0227
<i>U. guentheri</i>	7	5	2.26e-04	7	2.04e-05	2	0.0803	2	0.0099
<i>U. lineatus</i>	10	8	1.45e-04	10	1.32e-06	1	0.1273	6	9.79e-07
<i>U. ebonaui</i>	11	7	1.06e-06	10	1.57e-08	9	1.33e-09	11	1.02e-07
<i>U. henkeli</i>	13	10	3.33e-04	10*	1.76e-04	4	0.0361	9	1.50e-04
<i>U. sp. B</i>	13	10	1.68e-07	10*	1.97e-06	5	1.24e-05	9*	8.25e-09
<i>U. fimbriatus</i>	15	13	2.75e-08	13	2.09e-07	7	4.16e-05	11	2.11e-07
<i>U. phantasticus</i>	15	12	4.59e-10	13	5.31e-09	11	8.26e-11	12	1.28e-08
<i>U. sameiti</i>	16	15	3.84e-10	15*	8.57e-10	10	3.03e-07	13	4.28e-08
<i>U. sikorae</i>	23	20	2.73e-10	21	1.45e-10	11	3.45e-04	18	1.05e-06

LPT, lowest presence threshold; T10, fixed threshold of 10; T1, fixed threshold of 1.

*Denotes cases where the fixed threshold was greater than the LPT, thus leading to the omission of one or more localities in the projection based on all localities.

Test results showed that a statistically significant model was built with as few as four localities (Maxent, T10; $P < 0.05$), but the success rate at such a low sample size was poor (50%), suggesting that this was not a good model. Results for Maxent (LPT) and GARP (LPT and T1) were less good, particularly at sample sizes of less than about 10, although jackknife success rates generally increased with sample size.

A closer look at jackknife models

Example results from jackknife tests are presented for *Uroplatus eburnei* in Fig. 2. The example demonstrates the exclusion from model building of one of two localities owing to close proximity to the test point (Fig. 2a) and the unsuccessful prediction of an excluded locality despite it being geographically close to the predicted area (Fig. 2b). The disjunct locality excluded in Fig. 2(c) and (d) was not successfully predicted in either case, yet it is notable that when the decision threshold was reduced to a fixed value of 10, the model identified a

suitable area that is disjunct from other predicted regions and very close to the excluded locality (Fig. 2d).

Influence of decision threshold on model projections and test performance

Using a fixed decision threshold below the LPT generally increased the success rate in jackknife tests (Table 1). Although using a lower threshold results in a higher proportion of the study area being predicted as present, models that were statistically significant using the LPT remained significant with the fixed threshold. For example, Fig. 3 presents models using each threshold approach with both modelling algorithms for *Uroplatus lineatus*. For each modelling method, lowering the threshold identified additional regions of suitable conditions, notably extending south of the observed localities on the eastern escarpment. Reducing the decision threshold can thus uncover potentially informative distributional areas, as noted above in Fig. 2(c) and (d).

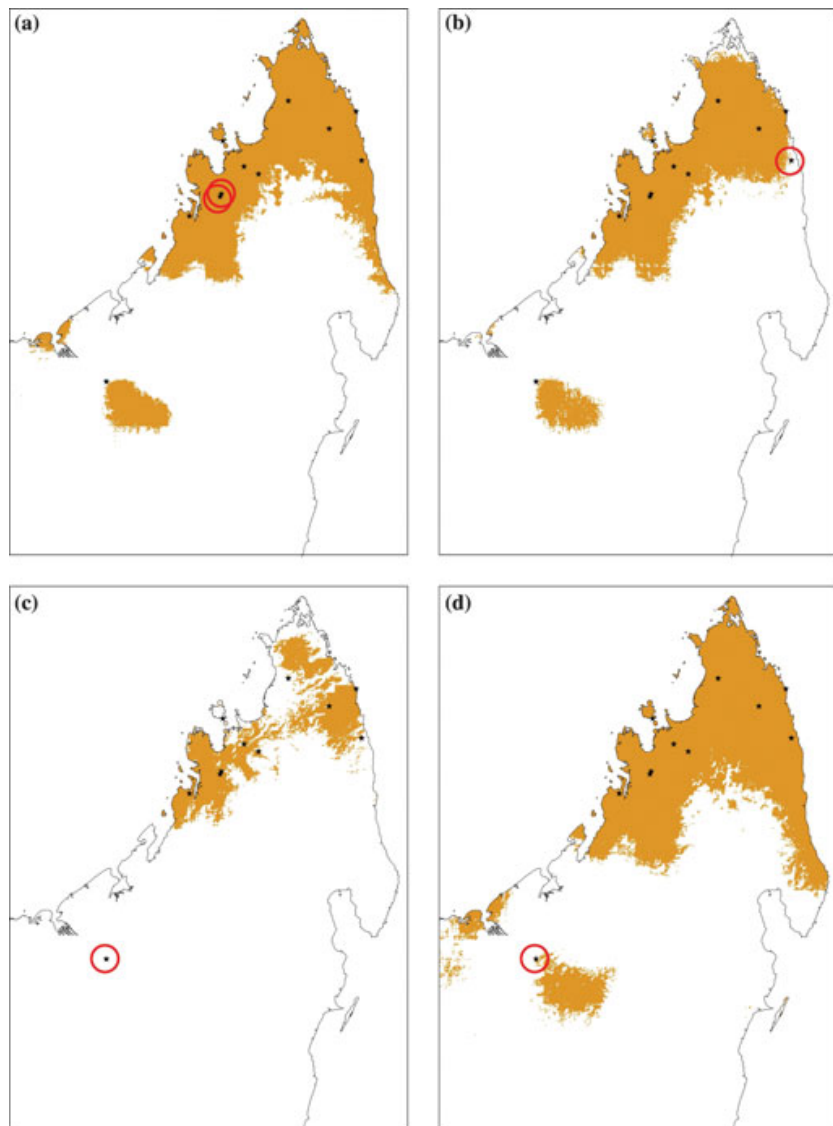


Figure 2 Example jackknife tests for *U. eburnei* using Maxent. In each case the circled localities were excluded from model training. Lowest presence thresholds based on the training localities were applied in (a), (b) and (c), and a fixed threshold of 10 was applied in (d). Two points were excluded in (a) since a second point was located < 10 km from the test point. Both points were successfully predicted by the model. The excluded point in (b) was not successfully predicted, yet is located in close proximity to the predicted area. The excluded point in (c) and (d) was not successfully predicted in either case, but if the threshold is reduced from a lowest presence value of 46 (c) to a fixed value of 10 (d), it is apparent that the model predicts suitable area in close proximity to the excluded point.

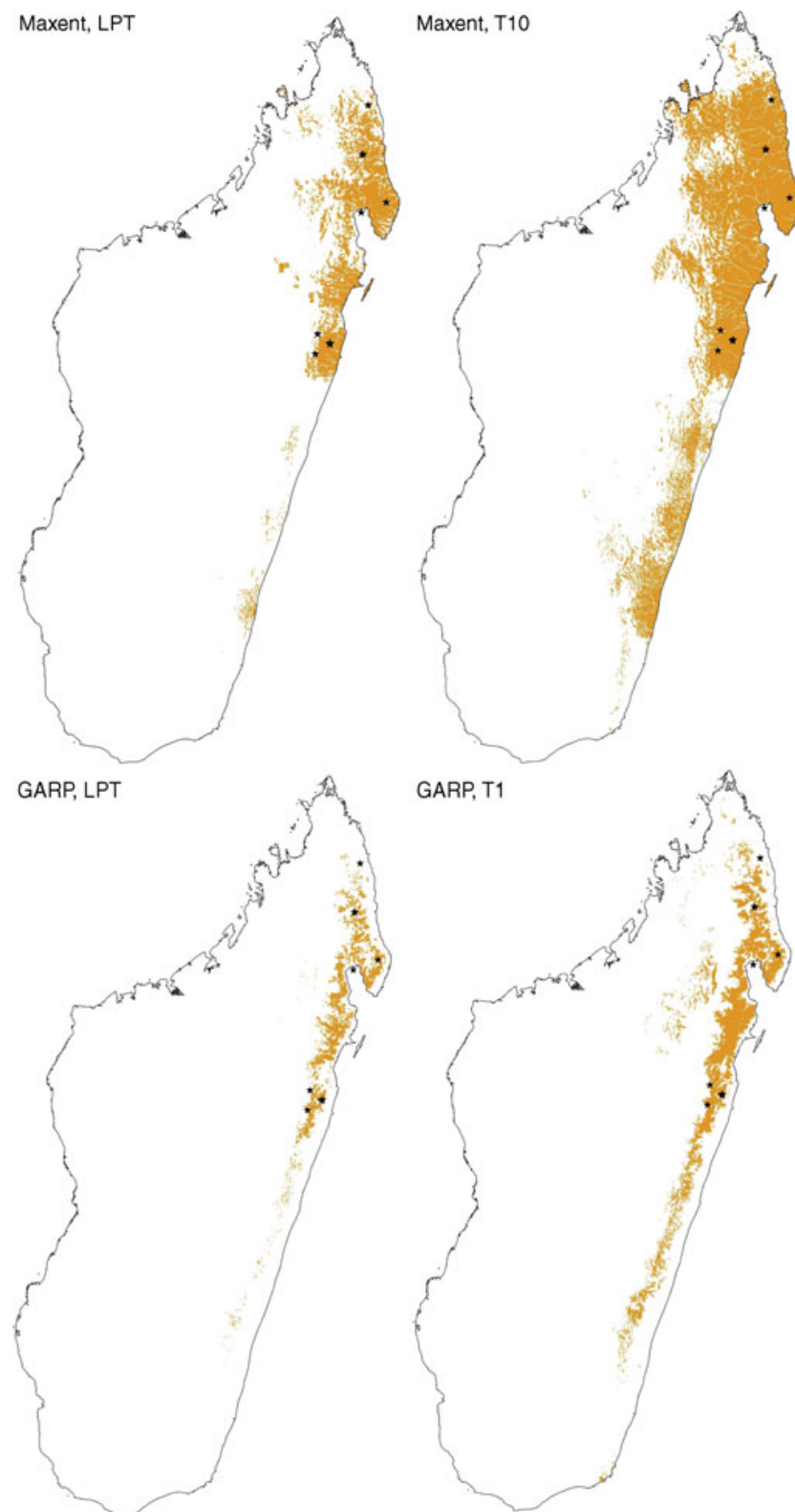


Figure 3 Modelled distributions for *U. lineatus* using two algorithms (Maxent and GARP) and alternative decision thresholds: lowest presence threshold (LPT) and fixed thresholds of 10 for Maxent (T10) and 1 for GARP (T1). Predicted areas are shaded, and observed locality points are shown as black stars.

In some jackknife tests, application of the LPT with GARP resulted in successes but *P* values equal to 1 since the entire region was predicted to be suitable. This was the case with two excluded localities for *Uroplatus* sp. A and for one excluded locality for *Uroplatus malama* (see Table 1). In these instances,

all best-subset GARP models excluded one of the training localities, resulting in a LPT of zero. Application of a fixed threshold of 1 resulted in non-zero omission in the training data set, but yielded statistically significant models. Similarly, for one species (*Uroplatus* sp. B) the GARP prediction based

on all observed localities had a LPT of zero, thus requiring application of a fixed threshold of 1 in order to give a meaningful prediction.

LPT values varied broadly with both Maxent (Fig. 4a) and GARP (Fig. 4b). In the case of Maxent, LPT generally increased as sample size decreased. A similar trend with Maxent was apparent when we experimentally reduced the sample size of the training data set for the two species with the largest sample sizes (Fig. 4c,d).

Comparison of modelling algorithms

Performance of the Maxent and GARP algorithms in the jackknife tests was different, with Maxent achieving higher success rates for almost all species (Table 1). Differences were especially marked at very low sample sizes (≤ 10), with Maxent (T10) achieving higher, and more significant, success rates.

Whilst identifying broadly similar patterns of distribution, the Maxent and GARP projections differed in a number of respects. Taking the example of *Uroplatus lineatus* (Fig. 3), each algorithm predicted suitable areas restricted to the north-east of the country, and extending south of the observed localities when the decision threshold was reduced. However, the Maxent models predicted a generally broader area of suitable conditions, extending further west in the north of the country and into more coastal regions in the east. Although the GARP models visually appeared to fit the observed localities more closely, the ability to predict excluded localities was considerably higher in Maxent projections (Table 1).

Across all species modelled, the proportion of the study area predicted as present was significantly different between algorithms when the LPT was applied (paired *t*-test: for LPT

$P = 0.034$; for fixed thresholds $P = 0.061$). In general, Maxent predicted a larger proportion of the study area as present compared to GARP (Supplementary Appendix S3).

Effects of reducing sample size on model performance

As expected, artificially reducing sample sizes by removing localities from the training data set had a negative impact on model performance (Fig. 5). Model performance reduced especially dramatically at very small sample sizes ($< c. 5$). Of note was a great deal of variability between results obtained when using different random sequences for the removal of localities, demonstrating the sensitivity of model output to individual localities when sample sizes are very low. It is also notable that models with a fixed threshold of 10 tended to maintain a higher proportion of presences correctly predicted as sample size was reduced (i.e. the proportion of cells predicted as present both by the model trained on all localities and by models trained on a reduced set of localities tended to be higher for T10 rather than LPT models).

DISCUSSION

Modelling with small numbers of localities

To provide useful biogeographical information, niche-based distribution models use environmental data to yield results that cannot be derived solely from the original occurrence data. For example, the area circled in Fig. 1(d) does not include observed localities, yet the model indicated that environmental conditions there are similar to where the species is known to occur. We have demonstrated elsewhere

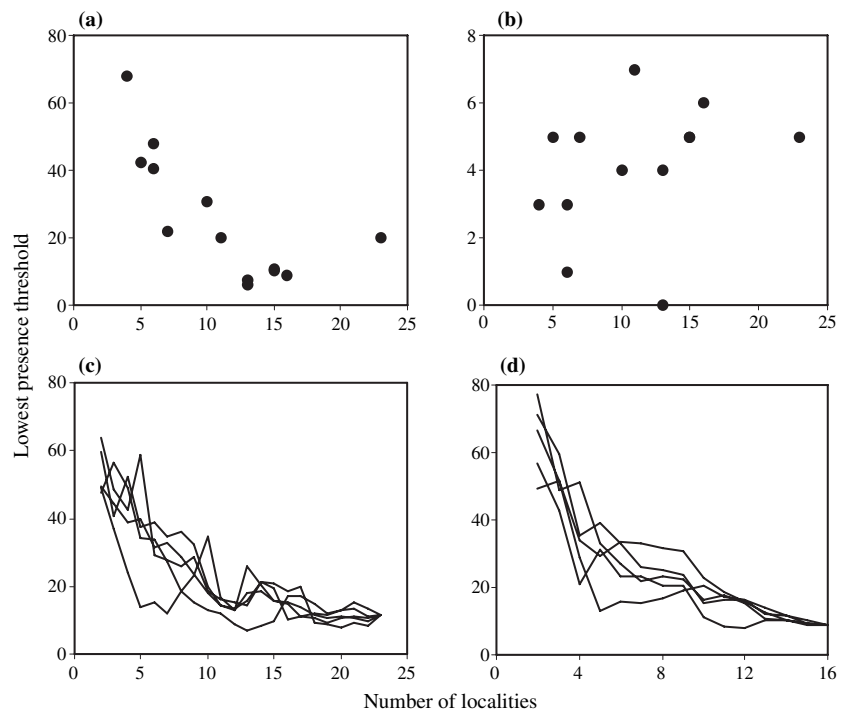


Figure 4 Relationship between lowest presence threshold (LPT) and number of training localities. (a) and (b) show results for all species using Maxent (a) and GARP (b) models trained on all observed localities. (c) and (d) show changes to LPT with Maxent for *U. sikorae* and *U. sameiti*, respectively, as localities are sequentially removed in random order (five different random sequences presented).

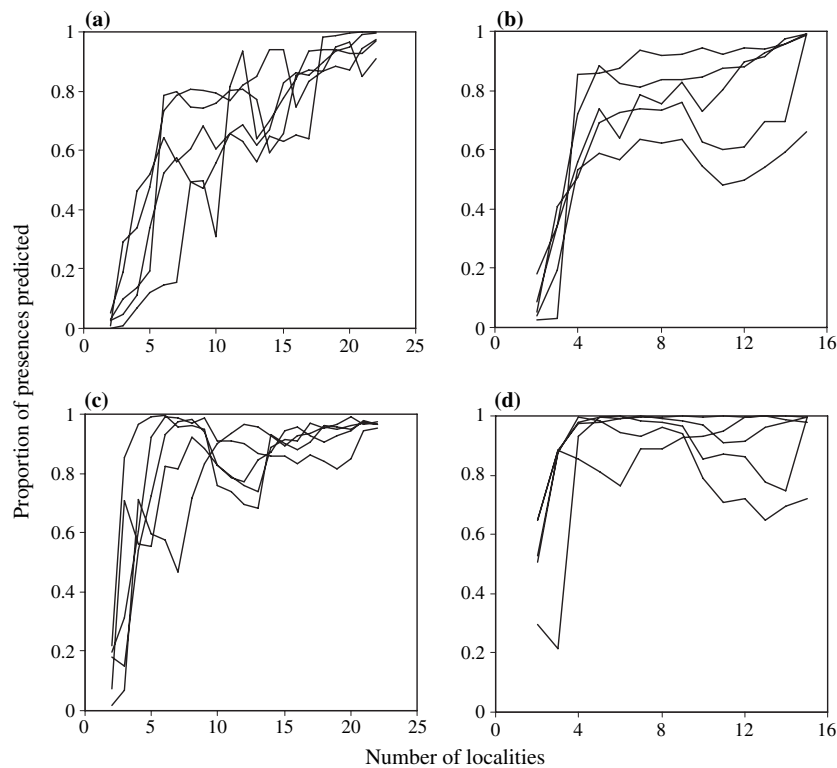


Figure 5 The impact of reducing sample size on model performance for *U. sikorae* (left) and *U. sameiti* (right). Localities were removed sequentially (one by one) in random order and a model trained with the remaining data. The proportion of presences correctly predicted was calculated for each model, where 'correct' was defined by the model trained on all available localities. Each plot shows five different random sequences. (a) and (b) Maxent models using the lowest presence threshold (LPT). (c) and (d) Maxent models using a fixed threshold of 10 (T10).

(Raxworthy *et al.*, 2003) that such regions of model 'overprediction' have the potential to identify unknown distributional areas and unknown species. The ability to provide biogeographical information of this kind is a strength of this modelling approach, and predictive ability at low sample sizes should be judged in this context.

Based on our jackknife tests, we have demonstrated significant predictive ability for Maxent (T10) models with as few as five observed localities. In an application of GARP with bird data for Mexico, Stockwell & Peterson (2002) found good predictive performance with as few as 10 localities [the success rate at predicting presence and (assumed) absence of a species based on 10 records was commonly 90% of that achievable with over 200 records]. Our application of the jackknife test with Madagascar geckos suggests that sample sizes may be reduced below 10 if Maxent is used. Reducing the minimum number of sample localities required to build useful distribution models can greatly increase the proportion of species that can be studied using these techniques. For example, Stockwell & Peterson (2002) noted that ≥ 50 localities are available for 20% of bird species in Mexico, ≥ 20 localities for 39% of species, and ≥ 10 localities for 65% of species. Reducing the requirement to ≥ 5 localities, we find that 88% of species in the Mexican bird atlas (Peterson *et al.*, 1998) could be studied using these techniques.

However, variability between predictions obtained when artificially removing localities using different random sequences demonstrated that models were greatly influenced by exactly which observations are included when sample sizes are small. This variability illustrates the inherent uncertainty in

model predictions constructed from few occurrence records. Addition of new records from future field surveys has the potential to greatly impact model predictions. The importance of each record will depend on whether that locality represents a unique environment not represented by the other sample points. For some species, a few sample localities may be sufficient to characterize the environmental niche, whereas in other cases the same number of samples may be inadequate to represent the range of conditions under which the species exists. Careful analysis of model output using the methods we have described can help to identify cases wherein useful information can be obtained from a few localities.

Choice of a decision threshold

Applications of GARP that apply the best-subsets approach have tended to apply two alternative decision thresholds: 'any model predicts' and/or 'all models predict' (e.g. Rice *et al.*, 2003; Roura-Pascual *et al.*, 2004). For Maxent, Phillips *et al.* (2006) have emphasized the need to investigate rules for setting thresholds further, since no such precedents exist. Our analyses support application of a threshold below that of the LPT, since jackknife success rates were higher in these cases. Furthermore, Maxent models with a fixed threshold of 10 tended to maintain a higher proportion of presences correctly predicted as sample size was artificially reduced. Also, LPT values were found to decrease as sample sizes became larger. These results support relaxation of the decision threshold below the LPT when small numbers of presence-only data are available.

Adjusting the decision threshold changes the proportion of the study area predicted to be present and affects proportions of observed records that are successfully predicted. For model applications in well-sampled regions with large numbers of locality records, it is common to set a balance between false-positive and false-negative predictions (Thuiller *et al.*, 2003; Pearson *et al.*, 2004). However, in cases where locality records are few, and there is high confidence that species identifications are accurate and localities are correctly georeferenced, it is appropriate to ensure there are no false-negative predictions (i.e. enforce a zero omission rate since failure to predict any of the observed localities is a clear error). This is achieved with the LPT, which provides the lowest predicted proportional area whilst maintaining zero training omission, but our jackknife tests demonstrate improved predictive ability if the LPT is reduced.

Choice of a decision threshold should be influenced by the proposed application of the model. We have demonstrated the possibility of uncovering potentially important distributional areas (areas of 'overprediction') by setting thresholds relatively low. A threshold below the LPT is thus appropriate for applications that aim to guide fieldwork toward identification of unknown distributional areas and undiscovered species. However, Loiselle *et al.* (2003) have argued that reducing false-positive predictions (i.e. minimizing the proportion of the study area predicted as present) is of principal concern in conservation applications because overestimating species occurrences may misdirect conservation action.

Use of the LPT may therefore be appropriate in cases where a more conservative prediction is desired. A notable advantage of the LPT is that it has a straightforward ecological interpretation, in identifying sites that are at least as suitable as those where a species' presence has been recorded. In contrast, applying a more liberal threshold requires that a subjective decision be made as to how much the LPT should be lowered.

Choice between alternative algorithms

Choice of an appropriate modelling algorithm is important since it is well known that different methods can give very different predictions (Segurado & Araújo, 2004; Thuiller *et al.*, 2004; Elith *et al.*, 2006; Pearson *et al.*, 2006). Our analyses support the use of Maxent when sample sizes are very small since success rates in jackknife tests were higher than with GARP. Maxent also generally predicted a larger proportion of the study area as being present, thus making the approach suited to the identification of new distributional areas in poorly known regions.

Whilst GARP has the advantage of having already undergone extensive testing and application, the concise mathematical definition of Maxent means that the approach and assumptions are transparent; furthermore, the algorithm is deterministic. Maxent thus offers excellent potential for extracting useful biogeographical information from small samples of locality records.

Statistical assumptions

Analytical study of the degree of independence of jackknife trials is intractable, because of the very complex nature of most algorithmic modelling techniques used for niche prediction. (A technical report, by E. Díaz-Francis and M. Nakamura addressing the potential effect of varying degrees of correlation on the efficacy of the test statistic, is forthcoming.) We here argue that possible strong correlations tend to appear in this problem when the samples are large, and therefore that when the sample sizes are small, the proposed *P* value is approximately correct. Examining the leave-one-out predictions for a species with a small sample size typically shows that each prediction is very different. In contrast, when the sample size is larger, many of the jackknifed predictions are alike. This is because groups or clusters of observations in sufficient numbers tend to jointly determine a predicted niche, so that strong dependence is induced in jackknife trials.

The striking (and cautionary) effect of this lack of independence can be illustrated by taking an example of a larger sample size and testing predictive performance using the alternative approach of randomly splitting data into two disjunct sets (test and training) and calculating a binomial probability. For *U. sikorae*, this approach gives a *P* value different by several orders of magnitude (five random splits, 30% training, Maxent LPT, results in *P* values from 0.0001 to 0.0129) compared with the *P* value reported in Table 1 (2.73×10^{-10}). The dependence between trials is evident in this example when viewing the individual jackknife models, as many of the predictions are very similar. If this caveat for larger samples is not properly understood, our *P* value can wrongly be interpreted as providing more optimistic evidence of predictive power than is warranted.

CONCLUSIONS

The approach that we have presented for testing distribution models when locality records are scarce has demonstrated the predictive ability of these models, yet the predictions should be interpreted with care. It would be inappropriate to interpret modelled distributions as representing actual limits to the range of a species. Rather, the models identify regions that have similar environmental conditions to where the species currently maintains populations. These data can provide valuable biogeographical information, for example in targeting surveys to accelerate the discovery of unknown populations and species (Raxworthy *et al.*, 2003; Bourg *et al.*, 2005).

We have presented a methodological approach that shows much potential for extracting useful biogeographical information from small samples of observed presence localities. The approach can form part of an iterative procedure for generating new biogeographical data: (1) models that show significant predictive ability in jackknife tests are used to guide field surveys, (2) field surveys provide new locality records that provide an independent test of the models, and can be used to

generate improved model projections for guiding future surveys.

Further research is required to investigate the impacts of factors including sampling bias (Zaniewski *et al.*, 2002; Reddy & Davalos, 2003), spatial autocorrelation (Luoto *et al.*, 2005) and data resolution (Karl *et al.*, 2000). However, our analyses support the careful application of these techniques in poorly known regions, such as Madagascar, where modelled biogeographical data can provide valuable information for evolutionary and ecological study, and for informing conservation priorities.

ACKNOWLEDGEMENTS

Ronald A. Nussbaum (University of Michigan Museum of Zoology) kindly provided additional *Uroplatus* locality data and access to museum specimens. We thank Robert Hijmans for advice regarding the WorldClim data set, and Tim Love for providing access to the FEWS precipitation estimates. We are grateful to Ned Horning and Kevin Koy for GIS advice, to Eloísa Díaz-Francés and David Sprott for discussion, to Robert Anderson and Steven Phillips for discussion and comments on the manuscript, and to two anonymous referees. Funding for this project was provided by the National Science Foundation (NSF) under grants DEB 0423286 and DEB 9984496, and the National Aeronautic and Space Administration (NASA) for the support of distribution modelling work at the American Museum of Natural History (NASA grant NAG5-12333). Support for fieldwork in Madagascar was provided by NSF, the National Geographic Society and Earthwatch.

REFERENCES

- Anderson, R.P. & Martínez-Meyer, E. (2004) Modelling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biological Conservation*, **116**, 167–179.
- Anderson, R.P., Gómez-Laverde, M. & Peterson, A.T. (2002) Geographical distributions of spiny pocket mice in South America: insights from predictive models. *Global Ecology and Biogeography*, **11**, 131–141.
- Anderson, R.P., Lew, D. & Peterson, A.T. (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, **162**, 211–232.
- Andreone, F., Vences, M. & Randrianirina, J.E. (2001) Patterns of amphibian and reptile diversity at Berera Forest (Sahamalaza Peninsula), NW Madagascar. *Italian Journal of Zoology*, **68**, 235–241.
- Andriamialisoa, F. & Langrand, O. (2003) The history of zoological exploration in Madagascar. *The natural history of Madagascar* (ed. by S.M. Goodman and J.P. Benstead), pp. 1–15. University of Chicago Press, Chicago.
- Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, **28**, 693–695.
- Araújo, M.B. & Williams, P.H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**, 331–345.
- Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species-climate envelope models under climate change. *Global Change Biology*, **11**, 1504–1513.
- Bauer, A.M. & Russell, A.P. (1989) A systematic review of the genus *Uroplatus* (Reptilia: Gekkonidae), with comments on its biology. *Journal of Natural History*, **23**, 169–203.
- Böhme, A. & Ibish, P. (1990) Studien an *Uroplatus*. I Der *Uroplatus-fimbriatus* -komplex. *Salamandra*, **26**, 246–259.
- Böhme, A. & Schönecker, P. (2003) Eine neue art der gattung *Uroplatus* Duméril, 1805 aus ost-Madagaskar (Reptilia: Squamata: Gekkonidae). *Salamandra*, **31**, 129–138.
- Bourg, N.A., McShea, W.J. & Gill, D.E. (2005) Putting a CART before the search: successful habitat prediction for a rare forest herb. *Ecology*, **86**, 2793–2804.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E. & Schmiegelow, F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling*, **157**, 281–300.
- Carpenter, G., Gillson, A.N. & Winter, J. (1993) DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation*, **2**, 667–680.
- Donque, G. (1972) The climatology of Madagascar. *Biogeography and ecology of Madagascar* (ed. by R. Battistini and G. Richard-Vindard), pp. 87–144. W. Junk, The Hague.
- Elith, J., Graham, C. & the NCEAS Species Distribution Modelling Group (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Engler, R., Guisan, A. & Rechsteiner, L. (2004) An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology*, **41**, 263–274.
- Ferrier, S., Watson, G., Pearce, J. & Drielsma, M. (2002) Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation*, **11**, 2275–2307.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Fleishman, E., MacNally, R. & Fay, J.P. (2002) Validation tests of predictive models of butterfly occurrence based on environmental variables. *Conservation Biology*, **17**, 806–817.
- Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004a) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, **19**, 497–503.
- Graham, C.H., Ron, S.R., Santos, J.C., Schneider, C.J. & Moritz, C. (2004b) Integrating phylogenies and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**, 1781–1793.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

- Hampe, A. (2004) Bioclimatic models: what they detect and what they hide. *Global Ecology and Biogeography*, **11**, 469–471.
- Hastie, T., Tibshirani, R. & Friedman, J. (2001) *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York.
- Herman, A., Kumar, V.B., Arkin, P.A. & Kousky, J.V. (1997) Objectively determined 10-day African rainfall estimates created for famine early warning systems. *International Journal of Remote Sensing*, **18**, 2147–2159.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.
- Hirzel, A.H., Hausser, J., Chessel, D. & Perrin, N. (2002) Ecological-niche factor analysis: how to compute habitat-suitability map without absence data. *Ecology*, **83**, 2027–2036.
- Hutchinson, G.E. (1957) Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, **22**, 415–457.
- Iverson, L.R. & Prasad, A.M. (1998) Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs*, **68**, 465–485.
- Karl, J.W., Heglund, P.J., Garton, E.O., Scott, J.M., Wright, M.N. & Hutto, R.L. (2000) Sensitivity of species habitat-relationship model performance to factors of scale. *Ecological Applications*, **10**, 1690–1705.
- Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.
- Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G. & Williams, P.H. (2003) Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, **17**, 1591–1600.
- Luoto, M., Poyry, J., Heikkinen, R.K. & Saarinen, K. (2005) Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecology and Biogeography*, **14**, 575–584.
- Martínez-Meyer, E., Peterson, A.T. & Navarro-Sigüenza, A.G. (2004) Evolution of seasonal ecological niches in the *Passerina* buntings (Aves: Cardinalidae). *Proceedings of the Royal Society of London B, Biological Sciences*, **271**, 1151–1157.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.
- Nix, H.A. (1986) A biogeographic analysis of Australian Elapid Snakes. *Atlas of elapid snakes* (ed. by R. Longmore), pp. 4–15. Australian Government Publishing Service, Canberra.
- Ortega-Huerta, M.A. & Peterson, A.T. (2004) Modelling spatial patterns of biodiversity for conservation prioritization in North-eastern Mexico. *Diversity and Distributions*, **10**, 39–54.
- Pearce, J. & Ferrier, S. (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.
- Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.
- Pearson, R.G., Dawson, T.P., Berry, P.M. & Harrison, P.A. (2002) SPECIES: a spatial evaluation of climate impact on the envelope of species. *Ecological Modelling*, **154**, 289–300.
- Pearson, R.G., Dawson, T.P. & Liu, C. (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, 285–298.
- Pearson, R.G., Thuiller, W., Araújo, M.B., Martínez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P. & Lees, D.C. (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography*, **33**, 1704–1711.
- Peterson, A.T. (2003) Predicting the geography of species' invasions via ecological niche modelling. *Quarterly Review of Biology*, **78**, 419–433.
- Peterson, A.T. & Shaw, J.J. (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.T., Navarro-Sigüenza, A.G. & Benítez-Díaz, H. (1998) The need for continued collecting: a geographic analysis of Mexican bird specimens. *Ibis*, **140**, 288–294.
- Peterson, A.T., Soberón, J. & Sánchez-Cordero, V. (1999) Conservatism of ecological niches in evolutionary time. *Science*, **285**, 1265–1267.
- Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.
- Rakotomalala, D. (2002) Diversité des reptiles et amphibiens de la Réserve Spéciale de Manongarivo, Madagascar. *Boissiera*, **59**, 339–358.
- Rakotomalala, D., Raholimavo, E., Talata, P. & Rajeriarison, E. (2001) Les amphibiens et reptiles du Parc National de Ranomafana et de la zone forestière le reliant au Parc National d'Andringitra. *Recherche pour le Développement, Série Sciences Biologiques*, **17**, 133–163.
- Raselimanana, A.P. (1998) Inventaire biologique, Forêt d'Andranomay, Anjozorobe: La diversité de la faune de reptiles et d'amphibiens. *Recherche pour le Développement, Série Sciences Biologiques*, **13**, 43–59.
- Raselimanana, A.P. (1999) Inventaire biologique de la Réserve Spéciale de Pic d'Ivohibe et du couloir forestier qui la relie au Parc National d'Andringitra: L'herpétofaune. *Recherche pour le Développement, Série Sciences Biologiques*, **15**, 81–97.
- Raxworthy, C.J. (2003) Introduction to the reptiles. *The natural history of Madagascar* (ed. by S.M. Goodman and J.P. Benstead), pp. 934–949. University of Chicago Press, Chicago.
- Raxworthy, C.J., Forstner, M.R.J. & Nussbaum, R.A. (2002) Chameleon radiation by oceanic dispersal. *Nature*, **415**, 784–787.
- Raxworthy, C.J., Martínez-Meyer, E., Horning, N., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A. & Peterson, A.T.

- (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**, 837–841.
- Reddy, S. & Davalos, L.M. (2003) Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, **30**, 1719–1727.
- Reese, G.C., Wilson, K.R., Hoeting, J.A. & Flather, C.H. (2005) Factors affecting species distribution predictions: a simulation modelling experiment. *Ecological Applications*, **15**, 554–564.
- Rice, N., Martínez-Meyer, E. & Peterson, A.T. (2003) Ecological niche differentiation in the *Aphelocoma* jays: a phylogenetic perspective. *Biological Journal of the Linnean Society*, **80**, 369–383.
- Ricketts, T.H., Dinerstein, E., Boucher, T., Brooks, T.M., Butchart, S.H.M., Hoffmann, M., Lamoreux, J.F., Morrison, J., Parr, M., Pilgrim, J.D., Rodrigues, A.S.L., Sechrest, W., Wallace, G.E., Berlin, K., Bielby, J., Burgess, N.D., Church, D.R., Cox, N., Knox, D., Loucks, C., Luck, G.W., Master, L.L., Moore, R., Naidoo, R., Ridgely, R., Schatz, G.E., Shire, G., Strand, H., Wettengel, W. & Wikramanayake, E. (2005) Pinpointing and preventing imminent extinctions. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18497–18501.
- Roura-Pascual, N., Suarez, A.V., Gomez, C., Pons, P., Touyama, Y., Wild, A.L. & Peterson, A.T. (2004) Geographical potential of Argentine ants (*Linepithema humile* Mayr) in the face of global climate change. *Proceedings of the Royal Society of London Series B, Biological Sciences*, **271**, 2527–2534.
- Segurado, P. & Araújo, M.B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**, 1555–1568.
- Soberón, J. & Peterson, A.T. (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **359**, 689–698.
- Soberón, J. & Peterson, A.T. (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, **2**, 1–10.
- Spratt, D.A. (2000) *Statistical inference in science*. Springer-Verlag, New York.
- Stockwell, D.R.B. & Peters, D.P. (1999) The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Systems*, **13**, 143–158.
- Stockwell, D.R.B. & Peterson, A.T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148**, 1–13.
- Storey, M., Mahoney, J.J., Saunders, A.D., Duncan, R.A., Kelly, S.P. & Coffin, M.F. (1995) Timing of hot spot-related volcanism and the breakup of Madagascar and India. *Science*, **267**, 852–855.
- Svenning, J.-C. & Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.
- Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., Ferreira de Siquira, M., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L. & Williams, S.E. (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.
- Thuiller, W., Vaydera, J., Pino, J., Sabate, S., Lavorel, S. & Garcia, C. (2003) Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology and Biogeography*, **12**, 313–325.
- Thuiller, W., Araújo, M.B., Pearson, R.G., Whittaker, R.J., Brotons, L. & Lavorel, S. (2004) Biodiversity conservation: uncertainty in predictions of extinction risk. *Nature*, **430**, 33.
- Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T. & Prentice, I.C. (2005a) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences USA*, **102**, 8245–8250.
- Thuiller, W., Richardson, D.M., Pysek, P., Midgley, G.F., Hughes, G.O. & Rouget, M. (2005b) Niche-based modelling as a tool for predicting the global risk of alien plant invasions at a global scale. *Global Change Biology*, **11**, 2234–2250.
- Xie, P. & Arkin, P.A. (1996) Analysis of global monthly precipitation using gauge observations, satellite estimates, and numerical model prediction. *Journal of Climate*, **9**, 840–858.
- Zaniewski, A.E., Lahmann, A. & Overton, J.M. (2002) Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling*, **157**, 261–280.

SUPPLEMENTARY MATERIAL

The following supplementary material is available for this article:

Appendix S1. Executable program (pValueCompute.exe), with Help file (help.txt), for calculating the *P* value described in this paper.

Appendix S2. A list of environmental variables used in the modelling.

Appendix S3. Table showing the proportion of the study area predicted as present by each modelling approach for all species.

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/abs/> (This link will take you to the article abstract.)

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

BIOSKETCHES

Richard Pearson is a researcher at the American Museum of Natural History (AMNH) where he is associated with both the Center for Biodiversity and Conservation and the Department of Herpetology. Richard's research falls principally within the fields of biogeography and spatial ecology.

Chris Raxworthy is associate curator in the Department of Herpetology at the AMNH and has studied the amphibians and reptiles of Madagascar since 1985.

Miguel Nakamura is based at the Centro de Investigacion en Matematicas (CIMAT), Mexico, where his research focuses on the application of statistics in biodiversity studies.

Town Peterson is professor and curator of ornithology at the University of Kansas and has worked extensively on the geography and ecology of species distributions.

Editor: Jon Sadler