Modelling the partially unknown distribution of wall lizards (*Podarcis*) in North Africa: ecological affinities, potential areas of occurrence, and methodological constraints

A. Kaliontzopoulou, J.C. Brito, M.A. Carretero, S. Larbes, and D.J. Harris

Abstract: Species distribution modelling (SDM) is a powerful tool to investigate various biological questions with a spatial component, but is also sensitive to presence-data characteristics, particularly data precision and clustering. Here, we investigate the effect of these two factors on SDM using Maxent as the modelling technique and wall lizards (genus *Podarcis* Wagler, 1830) from North Africa as a model system. *Podarcis* are not ubiquitous in Africa as they are in Europe, but their ecological and distributional characteristics in this area are poorly known. Our results show that the most important environmental factors related to the distribution of this genus in North Africa are humidity, habitat type, and temperature. The areas of potential distribution predicted by models based on data sets with different precision and clustering characteristics show high relatedness to coastal areas and mountain ranges and extend to areas were presence records for these lizards are lacking. Our comparison of models based on different data sets indicates that finer scale models, even if based on fewer presence locations, outperform coarser scale ones. Data clustering does not have a negative effect on model performance, but is rather overcome by sample-size effects. Similar approaches may be of general application to other stenoic species for which available locations are scarce in comparison with the extension of the study area.

Résumé : La modélisation de la répartition des espèces (SDM) est un outil puissant pour étudier diverses questions biologiques à composante spatiale, mais elle est sensible aux caractéristiques des données, particulièrement leur précision et leur regroupement. Nous examinons ici les effets de ces deux facteurs sur la SDM en utilisant Maxent comme méthode de modélisation et des lézards des murailles (le genre *Podarcis* Wagler, 1830) d'Afrique du Nord comme système modèle. Les *Podarcis* ne sont pas ubiquistes en Afrique comme ils le sont en Europe, mais les caractéristiques de leur répartition et de leur écologie dans cette région restent mal connues. Nos résultats montrent que les facteurs du milieu les plus importants pour expliquer la répartition de ce genre en Afrique du Nord sont l'humidité, le type d'habitat et la température. Les zones de répartition potentielle prédites par les modèles d'après des bases de données présentant des précisions et des caractéristiques de regroupement diverses montrent une forte affinité avec les régions côtières et les chaînes montagneuses et comprennent des zones où il n'existe pas actuellement de confirmation de la présence de ces lézards. Nos comparaisons des modèles élaborés à partir de différentes bases de données indiquent que les modèles à échelle plus fine, même s'ils se basent sur moins de localités de présence, sont plus performants que les modèles à échelle plus grossière. Le regroupement des données n'a pas d'effet négatif sur la performance du modèle, car cet effet est oblitéré par les effets de la taille de l'échantillon. Des approches semblables pourraient être d'application générale pour l'étude d'autres espèces sténoïques pour lesquelles il existe peu de données de répartition, compte tenu de l'étendue de la zone d'étude.

[Traduit par la Rédaction]

Introduction

Species distribution modelling (SDM) is rapidly being incorporated to an increasing number of biological applications (Araújo and Guisan 2006) as an innovative tool to investigate ecological and evolutionary questions and to enhance conservation (Elith et al. 2006). SDM tries to evaluate how a set of environmental factors influence a species' distribution. Therefore, it provides an estimate of the species' fundamental niche (Hutchinson 1957), which can be used to predict its potential geographical distribution. In practice, information on a number of environmental parameters is combined to known locations of presence and a model is fitted

A. Kaliontzopoulou.[1] Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Campus Agrário de Vairão, 4485-661 Vairão, Portugal; Departament de Biologia Animal (Vertebrats), Facultat de Biologia, Universitat de Barcelona, Avinguda Diagonal, 645, 08028 Barcelona, Spain.
J.C. Brito, M.A. Carretero, and D.J. Harris. Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Campus Agrário de Vairão, 4485-661 Vairão, Portugal.
S. Larbes. Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO), Campus Agrário de Vairão, 4485-661 Vairão, Portugal; Département de Biologie, Faculte des Sciences Biologiques et Agronomiques, Université M. Mammeri, Tizi-Ouzou, Algeria.

[1]Corresponding author (e-mail: antigoni@mail.icav.up.pt).

doi:10.1139/Z08-078

to describe the requirements of the target species (Anderson et al. 2003). Recently, numerous statistical approaches have been developed for this purpose, including envelope models, regression analysis, factor analysis, genetic algorithms, and maximum entropy (reviewed by Elith et al. 2006).

However, all these techniques present inherent problems. Sample size (Stockwell and Peterson 2002; Elith et al. 2006; Hernandez et al. 2006), bias in data distribution (Guisan et al. 2006; Segurado et al. 2006), and geographic accuracy of the records (Hernandez et al. 2006) may affect modelling performance. Also, the question of scale is inherent to all modelling techniques (Huettmann and Diamond 2006; Guisan et al. 2007). In practical terms, all the above factors are inter-related and involve decision-making by the investigator. For example, obvious trade-offs exist between sample size and clustering of records (Segurado et al. 2006) and one has to choose between a complete but clustered or a random but reduced data set, since clustered field observations will usually result in spatially autocorrelated data. Another dilemma concerns the use of published data and species records from collections. Although field observations can now be precisely referenced using Global Positioning Systems (GPS), species records published in atlases, scientific and naturalistic journals, or obtained from museum collections often fail to have the desired geographic precision (Graham et al. 2004; Rowe 2005). This constrains the geographic scale of the analysis and a decision has to be made between many records of low precision or fewer records of high precision. Obviously, these questions are especially relevant for species with fragmented or partially unknown distributions, as well as for species occurring in conflictive areas of difficult access.

Because of the geographically asymmetrical knowledge on distribution and their complex evolutionary history, *Podarcis* from North Africa constitute an interesting model system to investigate both methodological and biological questions related to SDM. Wall lizards of the genus *Podarcis* Wagler, 1830 are members of the family Lacertidae, distributed around the Mediterranean basin. They are known to be ecological generalists, occupying a wide variety of habitats (Arnold 1987). Although detailed distributional and comprehensive ecological data exist for most European species of the genus, this is not the case for North Africa. In this region the genus attains the southern limit of its distribution, ranging from Morocco through northern Algeria, and reaching west to northwestern Tunisia (Fig. 1A). Although detailed distribution data are available for Morocco (Bons and Geniez 1996), Algeria is practically unexplored (but see Chirio and Blanc 1997) and few records exist for Tunisia (Blanc 1979).

This situation is further obscured by the taxonomic complexity that characterizes Iberian and North African *Podarcis* lizards (i.e., the species complex of *Podarcis hispanica* Steindachner, 1870), owing to the lack of concordance between current taxonomy and phylogenetic relationships (Harris and Sá-Sousa 2002; Harris et al. 2002; Pinho et al. 2007). Members of the genus from Morocco and southern Spain were recently re-elevated to specific status, as *Podarcis vaucheri* (Boulenger, 1905) (Busack et al. 2005). However, since no populations from Algeria and Tunisia were analysed, the remaining North African *Podarcis* are still

maintained under *P. hispanica* sensu lato. Another study has indicated a more complex situation and suggested that taxonomic re-evaluation is needed, since at least two lineages are present in the area (Pinho et al. 2007). Together with the difficulties related to fieldwork in Algeria, this has hindered a clear definition of the southern limits of the distribution of this group. SDM is an indirect approach to resolving such questions and can be efficiently applied to infer the ecological affinities and distribution limits of this group in North Africa. The potential existence of more than one phylogenetic entity in the area does not invalidate the use of SDM. When, as in the case of North African *Podarcis*, the physiology of such entities is phylogenetically conserved (Amaral and Carretero 2005; Carretero et al. 2006) and their ranges are parapatric (Pinho et al. 2007), their joint distributions can be efficiently modelled.

In this study, we use maximum entropy modelling to investigate the distribution of *Podarcis* lizards in North Africa based on geographic distributional data and environmental predictor variables, with the following objectives: (*i*) to determine which environmental factors are correlated with the distribution of these lizards in the southern limit of the distribution of the genus, (*ii*) to identify potential areas of occurrence in the study area by applying models based on those factors, and (*iii*) to evaluate the effect of common methodological restrictions, namely data precision and clustering of presence records, on the quality of such models.

Materials and methods

Study area and presence records

The study area encompasses the overall range of *Podarcis* in North Africa (Fig. 1A). Records for the species were collected during fieldwork and exact coordinates were marked with GPS. Moreover, published localities of occurrence of *Podarcis* in the study area (Blanc 1979; Bons and Geniez 1996; Chirio and Blanc 1997) were located at a 10 km × 10 km scale. Because of the uncertainty of precision of published data, these localities could not be considered at a finer geographic scale. We used zone 31N of the World universal transverse mercator (UTM) grid, which is the central zone of the study area, and projected data from all published localities onto this grid. Data sets were reduced to one record per cell for each geographic scale.

Data sets examined

To examine the effect of data precision and data clustering on the performance of Maxent modelling, we used four alternative data sets (Figs. 1B–1D). As mentioned above, we considered two initial data sets: one with exact records at an approximately 1 km^2 (30 s × 30 s) geographic scale (1ALL) and another including all available records at a 10 km^2 scale (10ALL). Both data sets were examined for clustering using the nearest neighbour index (NNI; spatial statistics toolbox in ArcMap version 9.2; Environmental Systems Research Institute, Inc. 2006). NNI assesses the degree of clustering of the data and is the ratio of the observed distance between points divided by the expected distance for a random distribution. For NNI = 1, the distribution is random; for NNI > 1, the distribution is dispersed; for NNI < 1, the distribution is clustered (Cliff and Ord 1973). Since the initial data sets

Fig. 1. (A) Location of the study area and (B–E) presence data sets used. (B) 1 km^2 scale, all data; (C) 1 km^2 scale, non-autocorrelated data; (D) 10 km^2 scale, all data; (E) 10 km^2 scale, non-autocorrelated data. The shaded area represents the mask used to model the potential distribution of wall lizards (genus *Podarcis*).
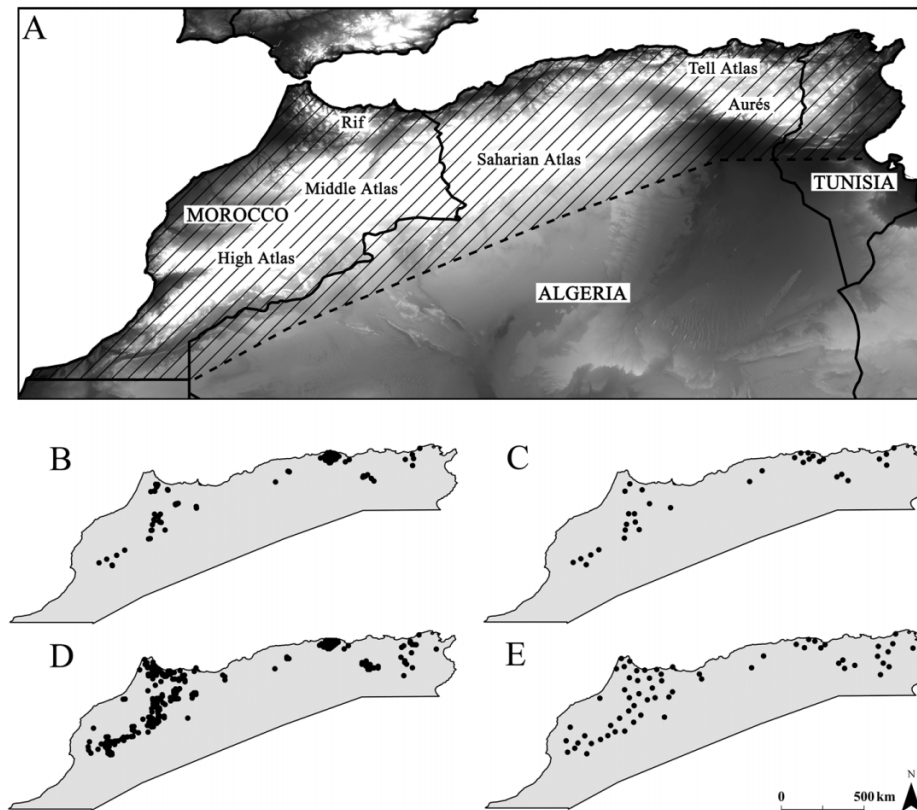

Table 1. Results of the nearest neighbour index (NNI) analysis.

Data set	n	Observed mean distance	Expected mean distance	NNI	Z score
1ALL	185	0.061	0.239	0.255	−19.39 × SD
10ALL	243	14.478	30.154	0.480	−15.50 × SD
1RED	34	0.473	0.548	0.863	−1.53 × SD
10RED	64	51.100	58.559	0.873	−1.95 × SD

Note: n is the number of sampling points and the Z score is the statistic value for testing the null hypothesis of a random distribution of points, expressed as a multiple of standard deviations (SD). The Z-score critical values for significance of $\alpha = 0.05$ are $Z = \pm1.96$ SD. See Material and methods for data set codes. Distances are in decimal-degrees for data sets at the 1 km scale, while in kilometres for data sets at the 10 km scale.

showed clustering, we eliminated clustered points using a random interactive process until the distribution did not significantly differ from random, i.e., the NNI was not statistically different from 1. This procedure resulted in two reduced data sets (1RED and 10RED, respectively, without autocorrelated points; Table 1).

Modelling techniques

We used maximum entropy modeling of species geographic distributions (Maxent version 3.0; available from http://www.cs.princeton.edu/~schapire/maxent [accessed 20 July 2007]) with default parameters and partitioned the data to training and test samples (80% and 20% of presence points, respectively, i.e., Anderson et al. 2003). We chose Maxent because it has been shown to perform better than other established methods, among both presence-only (Bioclim, Domain) and presence–absence (GAM, GLM, GARP) techniques, particularly with small sample sizes (Elith et al. 2006; Hernandez et al. 2006; Phillips et al. 2006). Moreover, Maxent presents several advantages (see Phillips et al. 2006; Phillips and Dudík 2008), the most important of which (at least in terms of experimental design and data availability) is that it admits both continuous and categorical predictor variables and it can be very effective with a relatively small number of presence records (Pearson et al. 2007).

Maxent modelling estimates the potential distribution of a species by finding the distribution of maximum entropy (i.e., closest to uniform), with the constraint that the expected value of each predictor variable under this estimated distribution has to match its empirical average, i.e., the mean

Table 2. Environmental variables used to model the distribution of *Podarcis* lizards in North Africa.

Variable (unit)	Ranges and classes	Code	Source
Altitude (m above sea level)	0–4016	ALT	USGS 2004
Slope (%)	0–36.22	SLOPE	USGS 2004
Land cover	1. Lowland evergreen forest	LANDCOV	GLC 2003
	2. Submontane and montane forest		
	3. Humid areas		
	4. Closed deciduous forest		
	5. Open deciduous shrubland		
	6. Sparse grassland		
	7. Croplans		
	8. Sandy desert and dunes		
	9. Stony desert		
	10. Bare rock		
	11. Salines		
	12. Water bodies		
	13. Cities		
Annual precipitation (mm)	27–1432	ANN_PR	Hijmans et al. 2005
Precipitation in the driest month (mm)	0–29	PR_DRY	Hijmans et al. 2005
Precipitation seasonality (mm)	19–110	PR_SEAS	Hijmans et al. 2005
Annual mean temperature (°C)	1.5–24.3	ANN_X_T	Hijmans et al. 2005
Temperature annual range (°C)	10.9–42.3	T_ANN_RNG	Hijmans et al. 2005
Minimum temperature of coldest month (°C)	–15 to 13.3	TMN_COLD	Hijmans et al. 2005
Maximum temperature of warmest month (°C)	22.2–46.7	TMX_WARM	Hijmans et al. 2005

value of a random set of points sampled from the distribution in question (Phillips et al. 2004, 2006; Phillips and Dudík 2008). This is achieved through an iterative algorithm, which begins with a uniform probability distribution with zero gain. The gain is a measure of likelihood of the samples and it expresses how much higher is the average sample likelihood compared with that of a random background pixel. The algorithm sequentially updates the weights of predictor variables until it converges to the optimum potential distribution and the gain of the final model can be interpreted as a measure of how much better the predicted distribution fits the sample points compared with a theoretical uniform distribution (Phillips et al. 2004, 2006; Phillips and Dudík 2008).

Different models were tested with receiver operating characteristics (ROC) plots. ROC curves plot true-positive rate against false-positive rate (Phillips et al. 2004) and the area under the curve (AUC) was used as a measure of the overall fit of the model.

Environmental predictor variables

We considered a set of 10 uncorrelated ($r < 0.8$) environmental factors (i.e., ecogeographical variables, EGVs) that were selected to describe habitat variability. Three types of EGVs were considered (Table 2): (1) topographical, i.e., altitude and slope derived from a digital elevation model (USGS 2004); (2) ecological, i.e., a land-cover EGV derived from satellite data (GLC 2003); (3) climatic, i.e., seven climate grids at approximately 1 km^2 precision (30 s × 30 s) describing precipitation and temperature means, extremes, and variability (Worldclim version 1.4; Hijmans et al. 2005). EGVs were all continuous, except for land cover, which was categorical (Table 2). For 10 km^2 models, the resolution of all EGVs was decreased to a grid cell size of 10 km^2 to match the precision of published species records.

The importance of each EGV in explaining the observed distribution was evaluated by jacknife analysis of the AUC with training and test data. For this purpose, Maxent sequentially excludes one EGV and creates a model with the remaining. This way it determines the AUC that is lost by exclusion of that particular EGV and estimates how much information that EGV has which is not contained in others. Additionally, to evaluate how much information that EGV contains on its own, Maxent creates models considering each EGV individually by fixing the rest to their mean value. To simultaneously assess both measures of variable importance and incorporate information from different models, we examined the scatter plot of "mean individual AUC" (the mean AUC of models with a single EGV along alternative data sets) vs. "mean AUC loss" (the mean AUC lost when that EGV was excluded from the model, along alternative data sets).

Response to predictor variables

To investigate the precise dependence of the distribution observed on individual environmental factors, we produced univariate models. This is necessary because the correlation between predictor variables might produce spurious results when examining all the variables together (Phillips et al. 2006). Consequently, we produced Maxent models with each of the five most important variables (individual AUC > 0.75) separately, using the four alternative data sets. This allowed us to obtain response curves that describe the relation between the probability of occurrence of *Podarcis* in the study area and the most important environmental factors.

**Predicting potential areas of occurrence**

To integrate information provided by both models at each geographic scale, we combined predictions to generate an

Table 3. Maxent modelling results for the four models built.

Model	*n* train	Iterations	Training AUC	*n* test	Test AUC	AUC SD
1ALL	148	500	0.986	37	0.982	0.007
1RED	28	440	0.963	6	0.833	0.077
10ALL	195	500	0.931	48	0.883	0.021
10RED	52	500	0.866	12	0.854	0.039

Note: *n* train is the training sample size, *n* test is the test sample size, AUC is the area under the curve, and SD is standard deviation.

Fig. 2. Scatter plots of individual area under the curve (AUC) vs. AUC loss for the ecogeographical variables (EGVs) evaluated at 1 km^2 (A) and 10 km^2 (B) geographic scales. Values for each EGV represent the mean of the models developed at each scale. For variable codes see Table 2.
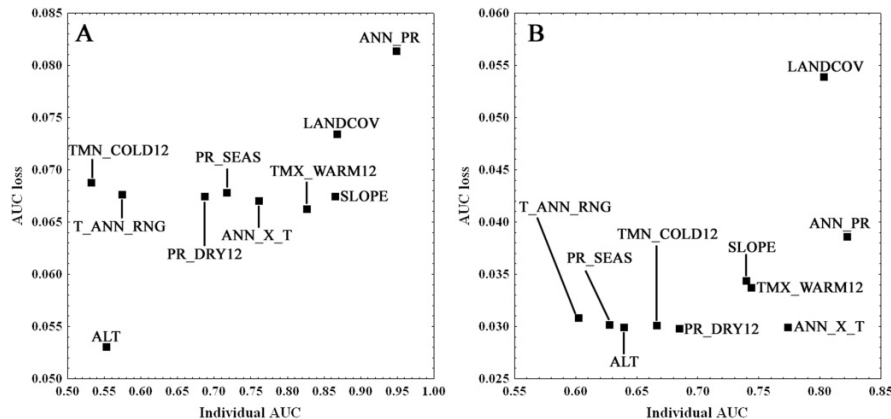
ensemble prediction (Araújo and New 2007). For this purpose, we attributed to each cell of the study area the mean value of the probability calculated for this cell by both models. Predictions are given in the logistic format of Maxent, the value of each cell representing the exponential of the entropy of its raw distribution. This logistic probability can be interpreted as predicted probability of presence, ranging from 0 to 1.

Results

Model performance

All models performed very well and closely fitted the presence points of *Podarcis* in the study area, as suggested by both training and test AUC values (Table 3), which were above 0.9 for all the training data sets except for 10RED with a training AUC of 0.8656. An examination of both training and test AUCs for the four models built (Table 3) revealed the existence of three sets of models: with both training and test AUCs below 0.9 (10RED), with training AUC above 0.9 but test AUC below 0.9 (1RED, 10ALL), and with both training and test AUCs above 0.9 (1ALL). Although differences between the models in terms of AUC are minor, the model that best predicts occurrence of the species in presence areas is 1ALL.

Explanatory variables

Examination of the scatter plots of mean individual AUC vs. mean AUC loss for each EGV (Fig. 2) revealed that the five most important variables determining the distribution of *Podarcis* lizards in North Africa are annual precipitation, land cover, annual mean temperature, slope, and maximum temperature in the warmest month of the year. The response

curves produced by univariate models (Fig. 3) give more insights into the precise effect of each variable on the distribution of *Podarcis* in the study area. A positive relationship is observed with annual precipitation; the contribution to the prediction increases in a nearly linear fashion, reaching its maximum at around 1000–1100 mm under all models considered. In contrast, annual mean temperature seems to be restrictive, since a relatively high contribution is observed for temperatures from 3 to 18 °C, with a sharp drop above this temperature. A similar pattern is observed for the maximum temperature in the warmest month of the year, the contribution to the prediction reaching its highest value at 29–30 °C. Considering land cover, the lowland, submontane, montane, and deciduous forests, humid areas, and water bodies have a high contribution, while grasslands and croplands exhibit notably lower contributions. Sandy desert, dunes, and stony desert make no contributions to the prediction. Finally, a positive relationship is observed with slope, with quite high contribution above 5%, which stabilizes at 15%–20%.

Potential distribution of *Podarcis* in North Africa

As expected, the area predicted at the 10 km^2 scale was much larger than that predicted at the 1 km^2 scale, but both models provided high probabilities of occurrence for similar areas (Fig. 4). Areas encompassed by all models are closely related to coastal areas and mountain ranges, including the Rif, Middle, and High Atlas in Morocco; the Aurés, High Plateaux, isolated areas of the Saharian Atlas in Algeria; and the Tell Atlas in Algeria and Tunisia (for toponyms see Fig. 1A). At both geographic scales, the predictions include areas where *Podarcis* records are lacking, especially for Algeria.

Fig. 3. Response curves produced by univariate models of each data set on the five most important predictor variables. Solid black line: 1ALL; broken black line: 1RED; solid gray line: 10ALL; broken gray line: 10RED. The *y* axis (logarithmic contribution to raw prediction) represents the contribution of the examined variable to the exponent that is applied to calculate the maximum entropy probability of each pixel of the study area.
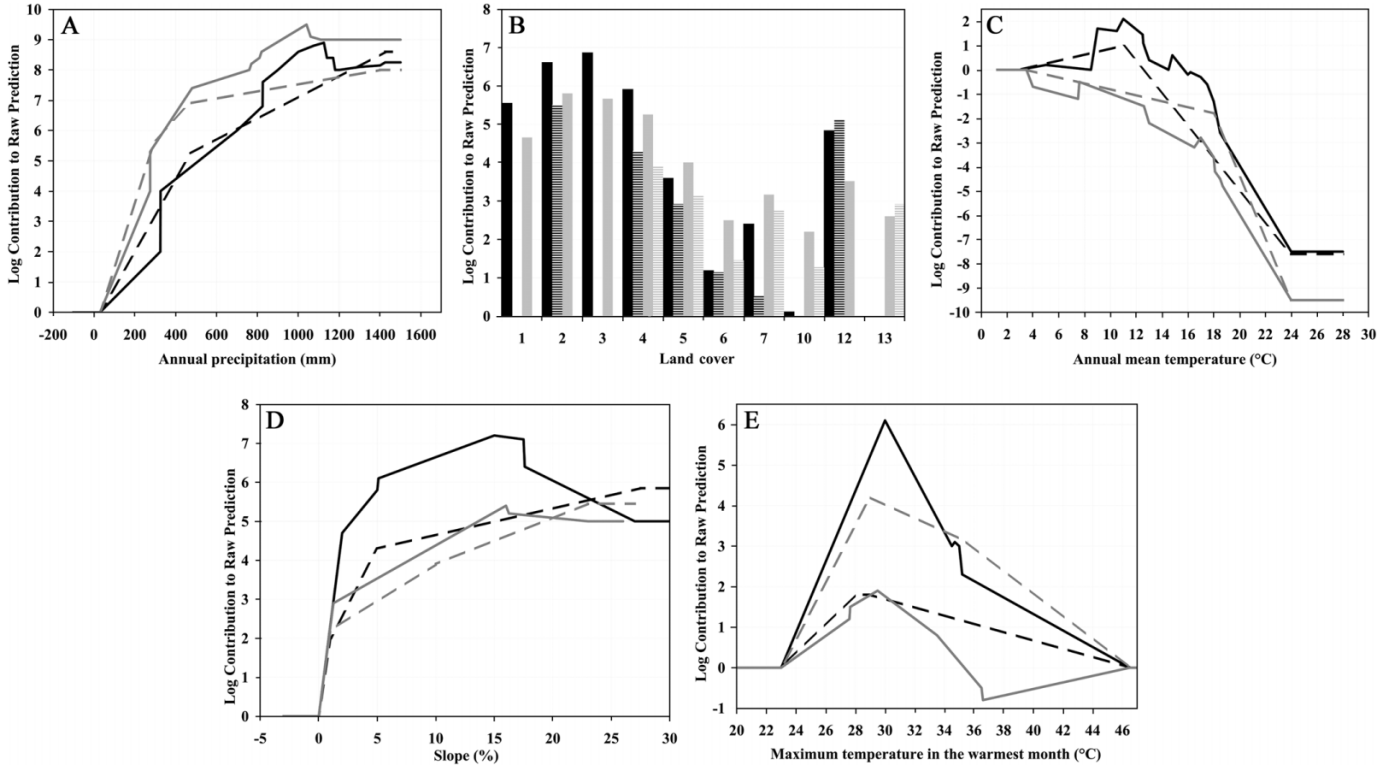
Discussion

SDM has been shown to be a tool that can be used to infer ecological requirements of organisms, study niche segregation and competition of co-existing species, facilitate fieldwork by the prediction of potential occurrence areas for rare species and improve conservation, and numerous other applications with a spatial component (Araújo and Guisan 2006; Elith et al. 2006; Guisan et al. 2006). Our results reinforce this view, offering insights into the ecological factors related to the distribution of *Podarcis* in North Africa and providing a map of potential occurrence. Additionally, a number of interesting methodological conclusions with relevance for broader applications can be drawn.

Ecological affinities of *Podarcis* in North Africa

As revealed by Maxent modelling, precipitation, habitat type, and temperature are important environmental factors in explaining the geographical distribution of wall lizards in North Africa (Fig. 3). This is not surprising given the Mediterranean origin of the genus and the environmental characteristics of the study area. Interestingly, the only other study using Maxent to explain the distribution of *Podarcis* species also highlighted the importance of humidity and habitat type for these lizards (Herkt 2007). Additionally, other modelling approaches applied to two forms of *P. hispanica* also revealed that temperature and climate type may be highly relevant (Sá-Sousa 2000).

In our case, a positive relationship existed between annual precipitation and the probability of occurrence of *Podarcis*

in the study area (Fig. 3A), emphasising the preference of these lizards for relatively humid environments and being additionally supported by the importance of humid areas and water bodies as a positive contributor related to land cover (Fig. 3B). This confirms previous observations that *Podarcis* inhabiting the north of Africa are dependent on permanent water courses (Schleich et al. 1996) and also agrees with the preference for forested habitats (see Fig. 3B, Table 2). In contrast, more desert-like habitats which favour arid steppe-land vegetation are completely avoided. This pattern is possibly related to the Mediterranean origin of these lizards, which usually prefer Mediterranean-type shrubby vegetation and temperate forests and avoid intensively cultivated areas (Miras et al. 2005).

However, the high values of both mean individual AUC and mean AUC loss indicate that land cover is an EGV relevant for the distribution of *Podarcis* on its own, which is not included in other EGVs. Therefore, the contribution of habitat type (as represented by the EGV of land cover) is not uniquely related to the humidity characteristics of corresponding classes. Probably, the habitat structure related to these vegetation types is also important, since these lizards prefer rocky habitats (Arnold 1987; Miras et al. 2005). The positive contribution of bare rock (Fig. 3B), as well as the response curve of the EGV "slope" (Fig. 3D), which suggests that plains are excluded (zero contribution to the prediction) and slopes above 5% are favoured, may be an additional indication of this fact.

Finally, the high importance of two temperature-related variables and their corresponding response curves also fa-

Fig. 4. Individual and ensemble predictions of *Podarcis* distribution obtained by the models at 1 km^2 ((A) all data; (B) reduced data set; (I) ensemble of A and B in the WGS 1984 projection) and 10 km^2 ((C) all data; (D) reduced data set; (II) ensemble of C and D projected at the UTM 31N zone). Dark tones represent low probabilities of occurrence and light tones represent high probabilities of occurrence. Total records of *Podarcis* presence used at each scale can be seen in the reduced shaded maps.
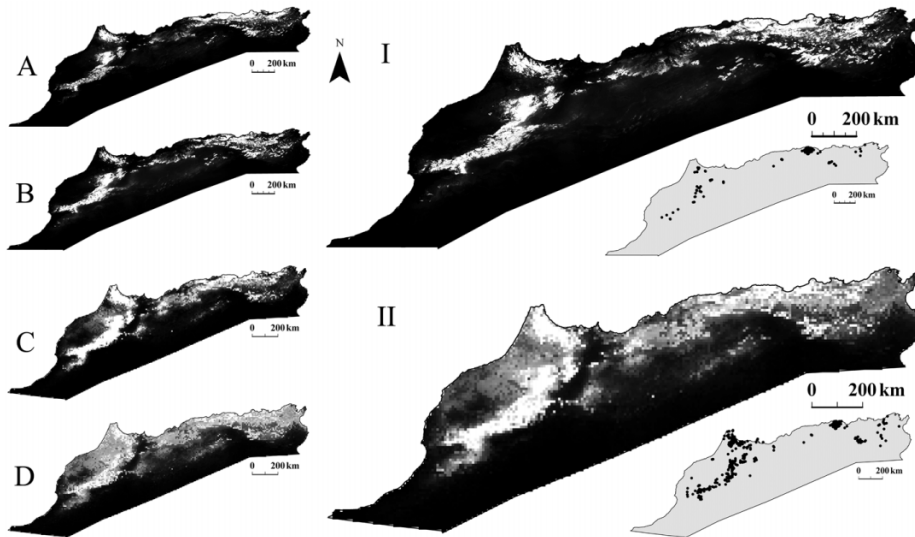
vour the view that *Podarcis* behave as specialists in North Africa. The response curves of annual mean temperature (Fig. 3C) and maximum temperature at the warmest month (Fig. 3E) indicate that these lizards are dependent on temperate areas, but avoid exceedingly warm conditions. *Podarcis* are known to select lower body temperatures than other North African lacertids like the *Acanthodactylus* Wiegmann, 1834 (Bauwens et al. 1995).

Potential distribution

The preference of *Podarcis* for humid conditions and forested habitats, but not extremely high temperatures, is reflected in their potential distribution model, which principally includes areas that lie either close to the Mediterranean coast or along mountain ranges. It is interesting that presence of *Podarcis* is predicted in areas where no records of occurrence exist. This could be a result of modelling weakness, but is probably rather due to undersampling of the areas in question. A cross-examination of the 1 km^2 scale prediction map (Fig. 4I) and the actual distribution of *Podarcis* in North Africa (Fig. 1D) reveals that 1 km^2 scale models accurately assign high probabilities of occurrence to areas where *Podarcis* exist, although these locations were not used for modelling. Reinforcing this observation, in the time between the elaboration of the models and the production of this manuscript, one of us (S.L.) confirmed the presence of *Podarcis* in localities where it was predicted but not previously recorded, confirming the good performance of the models in the field, as expected because of the relatively constant habitat preferences of *Podarcis* along the study area. Interestingly, *Podarcis* lizards were recorded in the Saharian Atlas, at locations corresponding to extremely isolated pixels with high probability of occurrence in the south of the Algerian part of the prediction map (Fig. 5). These isolated populations were clearly marginal, with very low densities, something that could result from *Podarcis* being at the southern limit of their distribution.

The environmental features related to the distribution of *Podarcis* in North Africa give indirect insights into the factors limiting the distribution of the genus southwards, specifically high humidity and moderate temperatures. The reproductive biology of these lizards is influenced by small fluctuations of both humidity (Galán 1996, 1999) and temperature (Van Damme et al. 1992; Castilla and Swallow 1996; Braña and Ji 2000), probably restricting the spread of the genus farther south. In the southern part of the study area, high probabilities of occurrence were only predicted for high mountain ranges, where the effect of altitude counterbalances the effect of latitude. An examination of the distribution of lacertid lizards in Morocco (Bons and Geniez 1996) indicates that while Mediterranean species, such as *Scelarcis perspicillata* (Duméril and Bibron, 1839), show a distribution that mostly coincides with the predicted distribution of *Podarcis*, more arid areas are occupied by other lacertids of African origin (i.e., members of the lacertid genera *Acanthodactylus* and *Mesalina* Gray, 1845).

Methodological considerations

Several methodological studies have examined the effect of sample size (Pearce and Ferrier 2000; Stockwell and Peterson 2002; Hernandez et al. 2006), data characteristics (Lennon 2000; Diniz-Filho et al. 2003; Segurado et al. 2006; Dormann 2007; Hawkins et al. 2007), and scale (Ferrier and Watson 1997; Tobalske 2002; Guisan et al. 2007) on distribution modelling performance. However, all of them were based on simulations or (and) different data partitioning of a single data set and did not include comparisons between field observations and data from publications and collections, which are more relevant for biologists. In concordance with previous studies (Elith et al. 2006; Hernandez et al. 2006; Phillips et al. 2006; Guisan et al. 2007), our results show that Maxent SDM performs very well over a wide range of initial data sets, at least for the case study explored. Other factors being equal, models inferred from

Fig. 5. New observations of *Podarcis* in Algeria, superimposed on the predictive map at 1 km² scale. (1) Tlemcen – Mafrouche National Park (34°51.00′N, 1°17.53′W); (2) Ain Guettaya, 7 km south of Charef (34°33.03′N, 2°47.77′E); (3) Ain Sidi Chikh, Djebel Ksel (33°43.94′N, 1°10.15′E).
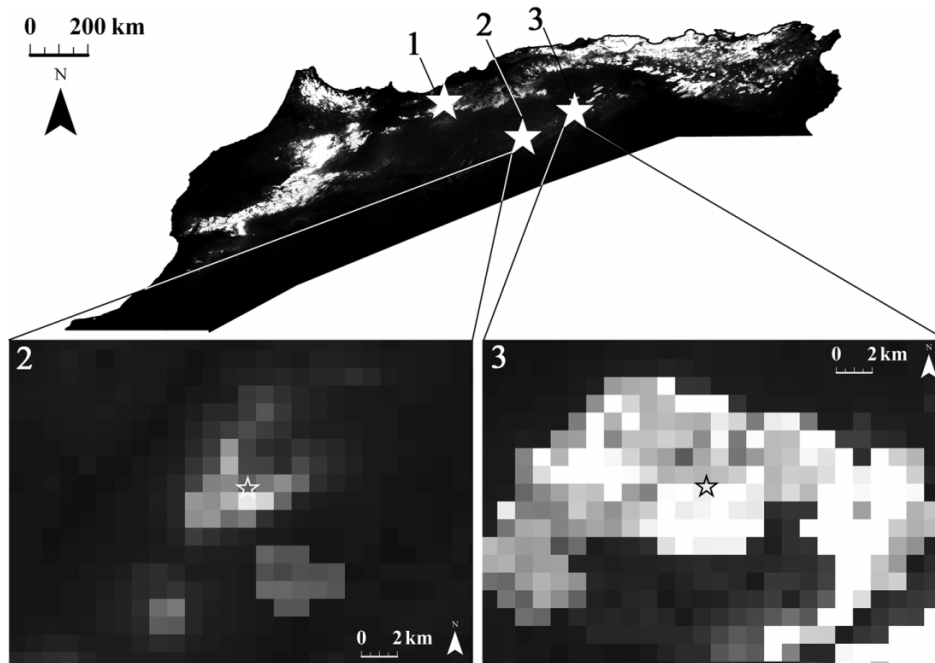
precise presence records performed better than those inferred from published data of uncertain precision. Similarly, models inferred using all the data available, although clustered, performed better than those using reduced, randomly arranged data sets.

In the context of the increasing availability of species records from natural-history collections (Graham et al. 2004) and published resources, data quality (Rowe 2005) and its effects on SDM are of relevance. Our comparison of Maxent models of different geographical precision shows a benefit in using fine-resolution models over more coarser resolution ones. Several studies report different degrees of sensitivity of SDM to change in grain size (Ferrier and Watson 1997; Tobalske 2002; Guisan et al. 2007). Regarding model structure and the importance of EGVs, examination of mean individual AUC and mean AUC loss (Fig. 2) reveals that grain size is relevant. Coarser grain models seem to have a lower capacity for detecting which variables are more important for species distribution. In our data, both mean individual AUC and mean AUC loss of EGVs showed lower values at 10 km² pixel size. This was reflected in the resulting probability predictions (Fig. 4A), with those for coarser grain size (Fig. 4C, 4D, 4II) being "fuzzy" compared with those for finer grain size (Figs. 4A, 4B, 4I). This result is probably related to the ecological characteristics of *Podarcis* (small body size and home range, aggregate distribution, and low dispersal) and therefore a higher resolution of EGVs is expected to better capture their requirements (Suárez-Seoane et al. 2002; Guisan and Hofer 2003).

Clustering of presence records is related to both sampling bias and spatial autocorrelation. Sampling bias may occur because investigators are likely to collect species records in a nonrandom, spatially clustered manner. The effect of spatial autocorrelation (lack of independence of neighbouring

points; Legendre 1993; Dormann 2007) on SDM has been extensively reviewed and several suggestions have been made to deal with this issue in modelling and ecological studies (Lennon 2000; Diniz-Filho et al. 2003; Segurado et al. 2006; Dormann 2007; Hawkins et al. 2007). Here we used an empirical approach to examine clustering effects on the performance of Maxent modelling. Our results indicate that complete data sets work better than those reduced to eliminate clustering, although the confounding effect of sample size does not permit direct comparison of the models (but see further below). Higher AUC values and lower AUC standard deviations (Table 3) characterize Maxent models produced with all the observations available. Maxent modelling, not being a regression-based technique and incorporating complicated interactions between predictor variables (Phillips et al. 2006; Phillips and Dudík 2008), should be less liable to autocorrelation effects than other methods, but this is a question that should be investigated further.

Although it has been repeatedly suggested that sample size strongly affects model accuracy (Pearce and Ferrier 2000; Stockwell and Peterson 2002), our results indicate that Maxent is quite robust to variations in sample size. All the models developed for *Podarcis* lizards in North Africa were at least useful (i.e., 0.75 < AUC < 0.9) and at least one of them was highly accurate (i.e., AUC > 0.9) (Swets 1988; Araújo et al. 2005). An evaluation of the effect of sample size on model accuracy along techniques (Hernandez et al. 2006) concluded that Maxent outperforms other methods, being accurate and stable across all sample-size categories tested. Our results reinforce this conclusion, since Maxent modelling seems to be very accurate along a wide range of sample sizes (34–243 in this study). However, two important conclusions can be drawn concerning its trade-offs with data precision and data clustering effects. First, data

precision seems to prevail over sample size, something that has also been indicated by other authors (i.e., Engler et al. 2004); models based on precise data performed better than those based on lower precision data, although sample sizes were higher for the latter. In contrast, sample size overcomes the importance of data clustering, both at 1 and 10 km² scales. Models based on complete data sets are more effective than those based on reduced data sets.

The narrow ecological requirements and restricted distribution of wall lizards in North Africa, examined with Maxent modelling, allowed us to determine putative environmental constraints and successfully predict species presence in unsampled areas orienting future samplings. Our analysis of methodological constraints on Maxent modelling shows that precise presence records are to be preferred over less accurate records, even if this reduces sample size. In contrast, clustering does not have a negative effect on Maxent performance, as sample size is more important. Similar approaches may be of general application to other stenoic species for which available locations are scarce in comparison with the study area.

Acknowledgements

References

Amaral, M.J., and Carretero, M.A. 2005. Preferred body temperatures of *Podarcis* lizards in Portugal: inter- and intraspecific variation. *In* Programme and Abstracts of the 13th Ordinary General Meeting of Societas Europaea Herpetologica, Bonn, Germany, 27 September – 2 October 1995. *Edited by* Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany. p. 24.

Anderson, R.P., Lew, D., and Peterson, A.T. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol. Model. **162**: 211–232. doi:10.1016/S0304-3800(02)00349-6.

Araújo, M.B., and Guisan, A. 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. **33**: 1677–1688. doi:10. 1111/j.1365-2699.2006.01584.x.

Araújo, M.B., and New, M. 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. **22**: 42–47. doi:10.1016/j.tree. 2006.09.010. PMID:17011070.

Araújo, M.B., Pearson, R.G., Thuiller, W., and Erhard, M. 2005. Validation of species–climate impact models under climate change. Glob. Change Biol. **11**: 1504–1513. doi:10.1111/j.1365-2486.2005.01000.x.

Arnold, E.N. 1987. Resource partition among lacertid lizards in southern Europe. J. Zool. Ser. B, **1**: 739–782.

Bauwens, D., Garland, T., Castilla, A.M., and Van Damme, R. 1995. Evolution of sprint speed in lacertid lizards: morphological, physiological and behavioral covariation. Evolution, **49**: 848–863. doi:10.2307/2410408.

Blanc, C.P. 1979. Observations sur *Lacerta hispanica* et *L. lepida* en Tunisie. Bull. Soc. Herpetol. Fr. **103**: 504–506.

Bons, J., and Geniez, P. 1996. Amphibians and reptiles of Morocco. Asociación Herpetológica Española, Barcelona.

Braña, F., and Ji, X. 2000. Influence of incubation temperature on morphology, locomotor performance, and early growth of hatchling wall lizards (*Podarcis muralis*). J. Exp. Zool. **286**: 422–433. doi:10.1002/(SICI)1097-010X(20000301) 286:4<422::AID-JEZ10>3.0.CO;2-D. PMID:10684565.

Busack, S.D., Lawson, R., and Arjo, W.M. 2005. Mitochondrial DNA, allozymes, morphology and historical biogeography in the *Podarcis vaucheri* (Lacertidae) species complex. Amphib.-Reptilia, **26**: 239–256. doi:10.1163/1568538054253438.

Carretero, M.A., Marcos, E., and de Prado, P. 2006. Intraspecific variation of preferred temperatures in the NE form of *Podarcis hispanica*. *In* Mainland and insular lacertid lizards: a Mediterranean perspective. *Edited by* C. Corti, P. Lo Cascio, and M. Biaggini. Firenze University Press, Florence, Italy. pp. 55–64.

Castilla, A.M., and Swallow, J.G. 1996. Thermal dependence of incubation duration under a cycling temperature regime in the lizard, *Podarcis hispanica atrata*. J. Herpetol. **30**: 247–253. doi:10.2307/1565516.

Chirio, L., and Blanc, C.P. 1997. Statut et distribution des reptiles dans le massif de l'Aures (Algérie). J. Afr. Zool. **111**: 205–232.

Cliff, A.D., and Ord, J.K. 1973. Spatial autocorrelation. Pion Limited, London.

Diniz-Filho, J.A.F., Bini, L.M., and Hawkins, B.A. 2003. Spatial autocorrelation and red herrings in geographical ecology. Glob. Ecol. Biogeogr. **12**: 53–64. doi:10.1046/j.1466-822X.2003. 00322.x.

Dormann, C.F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Glob. Ecol. Biogeogr. **16**: 129–138. doi:10.1111/j.1466-8238.2006.00279.x.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.C.C., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., and Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography, **29**: 129–151. doi:10.1111/j.2006.0906-7590.04596.x.

Engler, R., Guisan, A., and Rechsteiner, L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. J. Appl. Ecol. **41**: 263–274. doi:10.1111/j.0021-8901.2004.00881.x.

Environmental Systems Research Institute, Inc. 2006. ArcMap. Version 9.2 [computer program]. Environmental Systems Research Institute, Inc., Redlands, Calif.

Ferrier, S., and Watson, G. 1997. An evaluation of the effectiveness of environmental surrogates and modelling techniques in predicting the distribution of biological diversity. Environment Australia, Canberra, A.C.T.

Galán, P. 1996. Selección de lugares de puesta en una población del lacértido *Podarcis bocagei*. Rev. Esp. Herpetol. **10**: 97–108.

Galán, P. 1999. Demography and population dynamics of the lacertid lizard *Podarcis bocagei* in northwest Spain. J. Zool. Ser. B, **249**: 203–218.

GLC. 2003. The Global Land Cover for the Year 2000. Available from http://www-gem.jrc.it/glc2000/defaultGLC2000.htm [accessed 10 June 2007].

Graham, C.H., Ferrier, S., Huettmann, F., Moritz, C., and Peterson, A.T. 2004. New developments in museum-based informatics and applications in biodiversity analysis. Trends Ecol. Evol. **19**: 497–503. doi:10.1016/j.tree.2004.07.006. PMID:16701313.

Guisan, A., and Hofer, U. 2003. Predicting reptile distributions at

the mesoscale: relation to climate and topography. J. Biogeogr. **30**: 1233–1243. doi:10.1046/j.1365-2699.2003.00914.x.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C.C., and Aspinall, R. 2006. Making better biogeographical predictions of species' distributions. J. Appl. Ecol. **43**: 386–392. doi:10.1111/j.1365-2664.2006.01164.x.

Guisan, A., Graham, C.H., Elith, J., and Huettmann, F., and the NCEAS Species Distribution Modelling Group. 2007. Sensitivity of predictive species distribution models to change in grain size. Divers. Distrib. **13**: 332–340.

Harris, D.J., and Sá-Sousa, P. 2002. Molecular phylogenetics of Iberian wall lizards (*Podarcis*): is *Podarcis hispanica* a species complex? Mol. Phylogenet. Evol. **23**: 75–81. doi:10.1006/mpev. 2001.1079. PMID:12182404.

Harris, D.J., Carranza, S., Arnold, E.N., Pinho, C., and Ferrand, N. 2002. Complex biogeographical distribution of genetic variation within *Podarcis* wall lizards across the Strait of Gibraltar. J. Biogeogr. **29**: 1257–1262. doi:10.1046/j.1365-2699.2002.00744. x.

Hawkins, B.A., Diniz-Filho, J.A.F., Bini, L.M., De Marco, P., and Blackburn, T.M. 2007. Red herrings revisted: spatial autocorrelation and parameter estimation in geographical ecology. Ecography, **30**: 375–384.

Herkt, M. 2007. Modelling habitat suitability to predict the potential distribution of Erhard's wall lizard *Podarcis erhardii* on Crete. M.Sc. thesis, International Institute for Geo-Information Science and Earth Observation, Enschede, the Netherlands.

Hernandez, P.A., Graham, C.H., Master, L.L., and Albert, D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, **29**: 773–785. doi:10.1111/j.0906-7590.2006.04700.x.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. Int. J. Climatol. **25**: 1965–1978. doi:10.1002/ joc.1276.

Huettmann, E., and Diamond, A.W. 2006. Large-scale effects on the spatial distribution of seabird in the Northwest Atlantic. Landsc. Ecol. **21**: 1089–1108. doi:10.1007/s10980-006-7246-8.

Hutchinson, G.E. 1957. Concluding remarks. Cold Spring Harb. Symp. Quant. Biol. **22**: 415–427.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology, **74**: 1659–1673. doi:10.2307/1939924.

Lennon, J.L. 2000. Red-shifts and red herrings in geographical ecology. Ecography, **23**: 101–113. doi:10.1034/j.1600-0587. 2000.230111.x.

Miras, J.A.M., Cheylan, M., Nouira, M.S., Joger, U., Sá-Sousa, P., and Pérez-Mellado, V. 2005. *Podarcis vaucheri*. *In* 2006 IUCN red list of threatened species. International Union for Conservation of Nature and Natural Resources (IUCN), Gland, Switzerland. Available from http://www.iucnredlist.org [accessed 13 August 2007].

Pearce, J., and Ferrier, S. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecol. Model. **128**: 127–147. doi:10.1016/S0304-3800(99)00227-6.

Pearson, R.G., Raxworthy, C.J., Nakamura, M., and Peterson, A.T. 2007. Predicting species distributions from small numbers of oc-

currence records: a test case using cryptic geckos in Madagascar. J. Biogeogr. **34**: 102–117. doi:10.1111/j.1365-2699.2006. 01594.x.

Phillips, S.J., and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography, **31**: 161–175. doi:10.1111/j.0906-7590.2008. 5203.x.

Phillips, S.J., Dudík, M., and Schapire, R.E. 2004. A maximum entropy approach to species distribution modeling. *In* Proceedings of the Twenty-first International Conference on Machine Learning, Banff, Alta., July 2004. *Edited by* R. Greiner and D. Schuurmans. ACM Press, New York. pp. 655–662.

Phillips, S.J., Anderson, R.P., and Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. **190**: 231–259. doi:10.1016/j.ecolmodel.2005.03.026.

Pinho, C., Harris, D.J., and Ferrand, N. 2007. Comparing patterns of nuclear and mitochondrial divergence in a cryptic species complex: the case of Iberian and North African wall lizards (*Podarcis*, Lacertidae). Biol. J. Linn. Soc. **91**: 121–133. doi:10. 1111/j.1095-8312.2007.00774.x.

Rowe, R.J. 2005. Elevational gradient analyses and the use of historical museum specimens: a cautionary tale. J. Biogeogr. **32**: 1883–1897. doi:10.1111/j.1365-2699.2005.01346.x.

Sá-Sousa, P. 2000. A predictive distribution model for the Iberian wall lizard (*Podarcis hispanica*) in Portugal. Herpetol. J. **10**: 1–11.

Schleich, H.H., Kästle, W., and Kabisch, K. 1996. Amphibians and reptiles of North Africa. Koeltz Scientific Books, Koenigstein, Germany.

Segurado, P., Araújo, M.B., and Kunin, W.E. 2006. Consequences of spatial autocorrelation for niche-based models. J. Appl. Ecol. **43**: 433–444. doi:10.1111/j.1365-2664.2006.01162.x.

Stockwell, D.R.B., and Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. **148**: 1–13. doi:10.1016/S0304-3800(01)00388-X.

Suárez-Seoane, S., Osborne, P.E., and Alonso, J.C. 2002. Large-scale habitat selection by agricultural steppe birds in Spain: identifying species–habitat responses using generalized additive models. J. Appl. Ecol. **39**: 755–771. doi:10.1046/j.1365-2664. 2002.00751.x.

Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. Science (Washington, D.C.), **240**: 1285–1293. doi:10.1126/ science.3287615. PMID:3287615.

Tobalske, C. 2002. Effects of spatial scale on the predictive ability of habitat models for the green woodpecker in Switzerland. *In* Predicting species occurrences: issues of accuracy and scale. *Edited by* J.M. Scott, P.J. Heglund, F. Samson, J. Haufler, M. Morrison, M. Raphael, and B. Wall. Island Press, Covelo, Calif. pp. 197–204.

USGS. 2004. Shuttle radar topography mission (SRTM): mapping the world in 3 dimensions. United States Geological Survey National Center, Reston, Va. Available from http://srtm.usgs.gov/ [accessed 8 June 2007].

Van Damme, R., Bauwens, D., Braña, F., and Verheyen, R.F. 1992. Incubation temperature differentially affects hatching time, egg survival, and hatching performance in the lizard *Podarcis muralis*. Herpetologica, **48**: 220–228.