Comparison of prediction accuracy between Climatic Envelope Models (CEM)

Dissertation for M.Sc. in Ecology
by Asaf Tsoar

Submitted to the
Department of Evolution, Systematics and
Ecology,
The Hebrew University of Jerusalem

Supervisors: Prof. Ronen Kadmon
Prof. Dr. Hans-Ulrich Schnitzler

Jerusalem, October 2003

Abstract:

Maps of species distribution are vital tools in conservation planning. The gathering of data on species distribution is time and money consuming. Ecological niche modeling may usefully serve as an alternative way of creating satisfactory maps of species distribution. Climatic envelope models (CEMs) define the potential geographic range of a species by the totality of sites with environmental conditions similar to those prevailing at the sites for which records of that species are available. Even though these models are widely used in the literature only few studies have been made on differences in their reliability. The present study compares three different CEMs (BIOCLIM, Habitat and Mahalanobis), using data for a total of 66 species representing three faunistic groups: bats, snails and birds. An independent dataset was created to serve as a validation set.

There was no significant difference between the models in their ability to predict species distribution. The 'c' (false absence) score was relatively very low, in all species. In general, validation scores were similar to those found in previous studies. No significant difference was found between the Fauna groups.

The Incidence of species showed a significant negative quadratic regression with Kappa, reflecting the inherent bias of Kappa. The number of observations did not affect the ability of the models to predict species distribution. This is probably because only species with over 25 observations in the calibrating dataset were used in our tests.

There is a difference between the BIOCLIM and Mahalanobis models in over predicting false presence. BIOCLIM model over predicts false presence, its performance is enhanced by eliminating from the calibrating dataset 10% of the outliers. The BIOCLIM best predicted the bat dataset when all of the calibrating dataset was used. This is due to the low number of sites per species in that specific dataset.

Contents

Introduction

Maps of species distribution play an important role in applying conservation action plans (Primack 1998). However, distribution maps are only as good as the data used to generate them. As obtaining data on species distribution is both time and money consuming, many maps are based only on limited data and reflect the subjective intuition of naturalists and researchers (Shalmon 1993).

Models provide an alternative, objective and relatively inexpensive means of creating species distribution maps. GIS (Geographical Information System) and modern statistical methods enable us to predict distribution ranges using ecological niche models. These models rely on niche theory (Hutchinson 1957), and interpolate species distribution patterns from point records to their entire geographical range (Guisan & Zimmermann 2000).

Climatic Envelope Models (CEMs), define the potential geographic range of a species by the totality of sites with environmental conditions similar to those prevailing in the sites at which the species was recorded. Thus the potential range of a species is mapped into geographic space by the set of climatic conditions at record-sites. Such models outline a *climatic envelope* (a climatic niche), which covers the space created by mapping distribution in relation to selected climatic variables. This approach assumes that species can persist where their climatic requirements are satisfied. Over the past 20 years there has been a growing interest in such CEMs. Several such models have been based on different definitions of the multi-dimensional space (see summary by Guisan & Zimmerman 2000).

CEMs involve three conceptual steps (Hirzel et al. 2001). (1) The niche profile of a species is computed by assigning each cell to a combination of the chosen climatic parameters. (2) Next, the climatic data are used to compute the climatic envelope of the species. Models vary in the manner in which the climatic envelope of the species is defined. The geometry of the border represented by the climatic envelope, affects the manner in which the model interprets the relations between climatic variables. (3) Identification of cells that satisfy the climatic requirements of the species (i.e., the climatic envelope) on the map.

CEMs have been extensively used for conservation research (Honig et al. 1992; Sindel & Michael 1992; Martin 1996; Skidmore et al. 1996; Pearce & Lindenmayer 1998). They have proved particularly popular in attempts to predict responses of species distribution to climatic changes (Brereton et al. 1995; Eeley et al. 1999; Crumpacker et al. 2001). Yet, little is known

about the predictive power of such models, or the degree to which model predictions depend on the technique used to determine the climatic envelope (Guisan et al. 1998). This study compares the performance of three climatic envelope models: 1) BIOCLIM (Busby 1991); 2) Habitat (Walker & Cocks 1991) and 3) Mahalanobis (Farber & Kadmon 2003). Each model was applied to data for three widely different groups of animals: (i) insectivorous bats, (ii) nesting land birds, (iii) land snails. Predictions generated by the different models were evaluated using independent validation data, obtained by extensive sampling. Four complementary measures of accuracy were used to quantify accuracy of the predictive maps: Kappa, Sensitivity, Specificity and Overall Accuracy (Fielding & Bell 1997).

The models

1) BIOCLIM

Originally developed by Busby (1986), the BIOCLIM algorithm (Nix 1986; Busby 1991) derives the climatic range for each climatic variable. The model prediction is of Boolean (presence/absence) nature. BIOCLIM computes the climatic envelope on the basis of the extreme values of each climatic variable. The contour of the climatic envelope is of a rectilinear nature, and so ignores correlations between climatic parameters (Figure 1) and usually predicts a large potential niche. Removal of the outer ten percent of data often improves predictions (Honig et al. 1992; Law 1994). BIOCLIM is the most widely used CEM.

2) Habitat

The Habitat model Proposed by Walker and Cocks (1991), defines the potential niche as the convex hull of the climatic data, assuming that the envelope borderline matches linear interaction between climatic variables (Figure 1). Habitat differs from BIOCLIM in its ability to adjust the border of the envelope to correspond to the distribution of the data points in the climatic space. Consequently it predicts a smaller potential niche than BIOCLIM. As with BIOCLIM, the model prediction is Boolean.

3) <u>Mahalanobis</u>

Introduced to bioclimatic modeling by Farber and Kadmon (2003), the Mahalanobis model defines the climatic envelope of the species based on the Mahalanobis distance (Figure 1). It differs from the two other models in two most important ways: (1) By explicitly coping with correlations between the different variables. (2) By the fact that its borderline envelope is calculated from all dataset points (Farber 2000).

In analyzing the performance of the various models, I asked the following questions:

- How does the removal of outliers affect predictions?
- To what extent does the number of observations available for calibrating the model influence the accuracy of predictions?
- Are there any differences in predictive accuracy between the three groups of organisms?
- Can niche-width of a species serve as a clue to the accuracy of the relevant prediction?
- Do rare species differ from common ones in the accuracy of their predictive maps?

Climatic data of species records from the
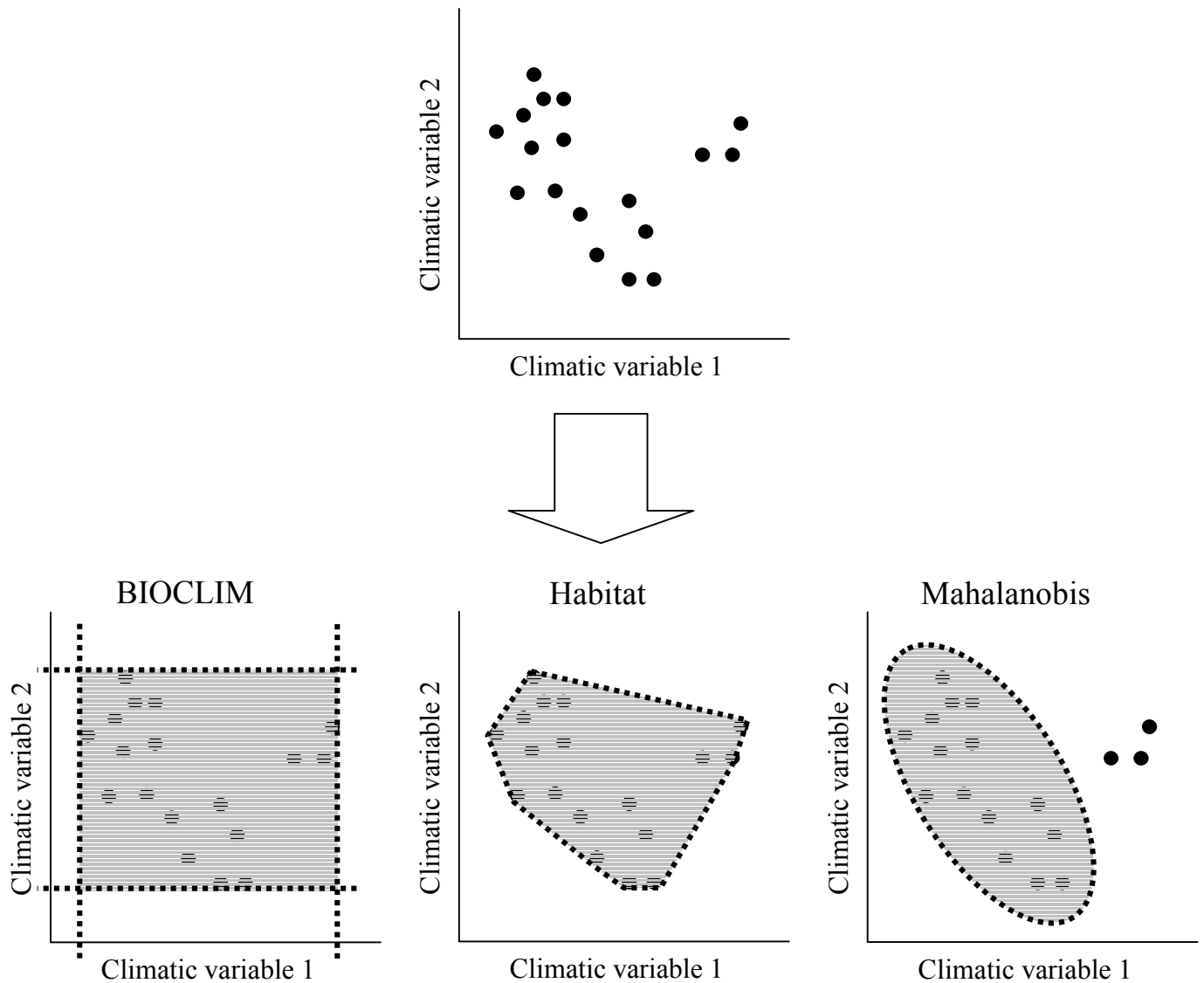calibration dataset

Figure 1: Schematic illustration of the climatic envelopes defined by the three models evaluated in this
study. The grey area is the potential niche predicted by each model.

Methods

Databases used for model calibration.

Data on bat distribution were obtained from several sources including observations by Menachem Dor (1930-1959); the bat database of the Israeli Mammalian Center of the Society for the Protection of Nature in Israel (SPNI) (1980-2002); David Harrison (1894-1963); David Harrison and Paul J. J. Bates (1866-1988); University of Tel Aviv Zoological collection (1931-1996); Yohay Barak, M.Sc. (1986-1988); Yohay Carmel, M.Sc. (1990-1993); Rachel Feldman, M.Sc.(1995-1997); and records of Carmi Korine from the Sde Boker area (1999-2001). Bird data were obtained from The University of Tel Aviv Zoological collection and the Israel Nature and Parks Authority. These data covered the years 1934-2002. The snail data were taken from the Mollusc Section, National Collections of Natural History, The Hebrew University of Jerusalem (collection period 1865-1998).

Previous studies have shown positive relationship between the amount of data and the accuracy of model predictions (Stockwell & Peterson 2002). Therefore our analysis is confined to species with at least 25 unique records and at least one observation in the validation dataset. These criteria yield a total of 66 species: 13 microchiropteran, 30 resident nesting land birds and 23 land snails (see appendix).

The climatic parameters used in the CEMs were: (i) mean annual rainfall, (ii) mean daily temperature of the hottest month (August), (iii) mean minimum temperature of the coldest month (January). The map of annual rainfall was digitized from 475 meteorological stations and corresponding isohyets (Kadmon & Heller 1998; Kadmon & Danin 1999). The mean temperature variables were calculated by interpolation from temperature data of 38 climatic stations (Kurtzman & Kadmon 1999).

The mahalanobis radius was tested in several sizes and was found that a mahalanobis radius size of 4 gave the highest significant Kappa for the dataset in this work. This radius size was found to best predict species distribution in Farber (2000) as well.

The bird and snail calibrating dataset used for the BIOCLIM model, excluded 10% outliers, for best prediction. Bat calibrating dataset for the BIOCLIM model and all other models used in this work, did not show a need to exclude outliers.

Validation data

I used an independent dataset for model validation, created by sampling the relevant fauna in 30 sampling sites of 1km^2 that were chosen to maximize representation of the climatic gradients of the research area. The 30 sampling sites were selected in a stratified random manner, using GIS technology (ESRI Inc. 1994). Each faunistic group was sampled in a matter that best fitted its specific nature. The sampling sites for bats, birds and snails were selected from the set of 30 sites. Some of the sampling sites for bats, were not chosen from the set of 30 sites, because of logistical and security considerations (e.g. the need to sample bats at night).

Bats were sampled in 19 sites; each site covering a 250m^2 was visited twice (two nights, once during spring time and the other in late summer), to increase the probability of obtaining a complete list of the species inhabiting that area. Bats were sampled during spring and summer from one hour before sunset, until half an hour after sunrise, using mist nets (total of 50 m x 2.5 m), frequency division bat detectors (ANABAT II, Titley Electronics) with data loggers (Delay switch, Titley Electronics), time expansion and heterodyne bat detectors (Petterson D-980 and Petterson D-200, Pettersson Elektronik). The combination of mist nets and echolocation analysis was intended to ensure the best possible representation of species at each site (O'Farrell & Gannon 1999).

Live bats were identified by relevant handbooks-identifying live bats. Methods for live bat identification were collected from the literature: (Koopman 1975; Makin 1977; Johnson 1980; Qumsiyeh 1985; Palmeirim 1990; Harrison & Bates 1991; Shalmon 1993; Wardhaugh 1999; Qumsiyeh 1996; Schober & Grimmberger 1997; Mendelssohn & Yom-Tov 1999; Taylor 2000). Visits to the collections of the British Museum of Natural History, the Harrison Museum and the Tel Aviv University Zoological Museum provided opportunities to check identifications. An echolocation reference library of almost all bat species of Israel was created for this work. Based on echolocation recordings collected in the course of my field work and on the available literature: (Waters & Jones 1995; Barataud 1996; Herr et al. 1997; Tupinier 1997; Vaughan et al. 1997; O'Farrell & Gannon 1999; Russ 1999; Schnitzler & Denzinger 1999 personal communication; Parsons & Jones 2000; Russo & Jones 2002; Limpens 2003) My own fieldwork involved capture, identification, and recording at time of release.

Birds were sampled in 20 sites by D. Rotem (Rotem 2003). The sampling procedure at each site was based on point counts (Bibby et al. 2000), gathered at five observation points

situated in a systematic array. Birds were sampled for 10 minutes at each point. Each site was sampled twice during the main breeding season, from March till July inclusive.

Snails were sampled at 27 sites by O. Steinitz (Steinitz 2003). Sampling in each site comprised collecting in nine quadrates of 100m^2, situated in a systematic array. Species were determined by their shells.

Accuracy assessment

In order to evaluate the accuracy of model predictions an error matrix for model assessment (Table 1) was constructed for each predictive map. This provided the basis for the calculation of four components: "a", "b", "c" and "d", which correspond to the frequencies of cases which represent the four possible outcomes of a comparison between a predictive map and the validation dataset (Table 1). The four measures of accuracy calculated (Table 2) are: Kappa (Cohen 1960), Sensitivity, Specificity, and Overall Accuracy (Legendre & Legendre 1998). The Kappa statistic evaluates the accuracy of prediction relative to the accuracy that might have resulted by chance alone (Cohen 1960; Monserud & Leemans 1992; Shao & Halpin 1995). Sensitivity represents the probability that the model correctly predicts a presence, whereas Specificity is the probability that the model correctly predicts an absence (Fielding & Bell 1997). The advantage of the parameters of accuracy is that they do not depend on the prevalence of a species. Overall Accuracy represents the correct prediction relative to all predictions made by the model. This parameter tends to ascribe relatively high accuracy for rare species (Fielding & Bell 1997; Manel et al. 1999).

Data analysis

Differences in predictive accuracy (expressed by error matrix and validation parameters) among the three models and the three fauna groups were tested using ANOVA, ANCOVA, and repeated measures ANOVA (Zar, 1999).

Regression analysis was used to evaluate the effect of three species parameters on the accuracy of model predictions: (1) Incidence (ACN); calculated as the relative validation sites, the species was found in, (2) number of observations (OBS) and (3) niche-width (NW); calculated as the relative range of all three climatic parameters.

Regressions were computed for each combination of models and fauna using three different functions, linear, logistic and quadratic (Sokal & Rohlf 1995). For each combination of model and fauna, the model showing the most significant fit was selected. If both linear and quadratic fits were significant, the fit with the best P value and higher statistic was selected.

Error matrix		Reality	
		Present	**Absent**
Prediction	**Present**	a	b
	Absent	c	d

Parameter	Name	Interpretation
a	Correct presences	Number of cells for which presence was correctly predicted by the model.
b	False presence	Number of cells for which the species was not found, but the model predicted presence.
c	False absence	Number of cells for which the species was found, but the model predicted absence.
d	Correct absences	Number of cells for which absence was correctly predicted by the model.

Table 1: Components of an error matrix. When a model map is compared to a map of actual distribution, the results can be summarized in a 2x2 matrix. The variables in this matrix (i.e. a,b,c and d) stand for the number of times each case occurred.

Kappa
$$\dfrac{\left(\dfrac{a+d}{n}\right) - \dfrac{(a+b)(a+c)+(c+d)(b+d)}{n^2}}{1 - \dfrac{(a+b)(a+c)+(c+d)(b+d)}{n^2}}$$

Overall Accuracy	Sensitivity	Specificity
$\dfrac{a+d}{n}$	$\dfrac{a}{a+c}$	$\dfrac{d}{b+d}$

Table 2: Measures of accuracy used in this study. Based on the 2x2 error matrix (Table 1), four measures of accuracy were used. 'Overall Accuracy' is the rate of correctly classified cells. The Kappa, normalizes this accuracy by the accuracy that might have occurred by chance alone. Sensitivity is the probability that the model will correctly classify a presence, and Specificity is the probability that the model will correctly classify an absence. In all formulas n=a+b+c+d.

Results

Distribution of the validation sites

The geographic and climatic distribution of the sites used for validating the models, developed for each faunistic group, are shown in figures 2-4. The distribution of the sites in the climatic space is defined by three variables: mean daily temperature of August, mean minimum temperature of January, and mean annual rainfall. The validation sites represent the entire range of climatic conditions prevailing in the study area.

Figure 2. Distribution of snail validation sites relative to climatic variables and the geographic distribution of the sites. In green is the climatic combination present in the study area.
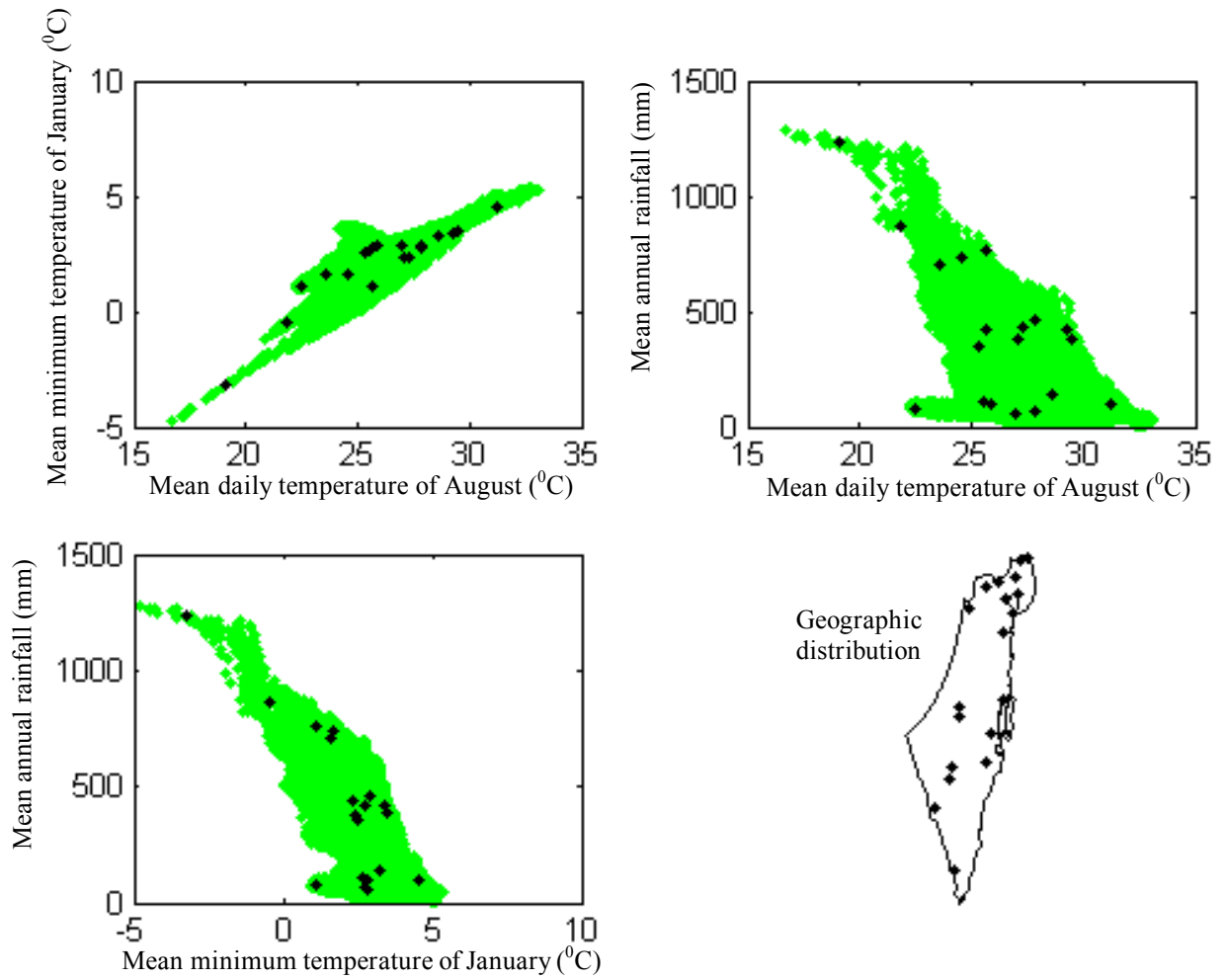
Figure 3. Distribution of the bat validation sites relative to climatic variables and the geographic distribution of the sites. In green is the climatic combination present in the study area.
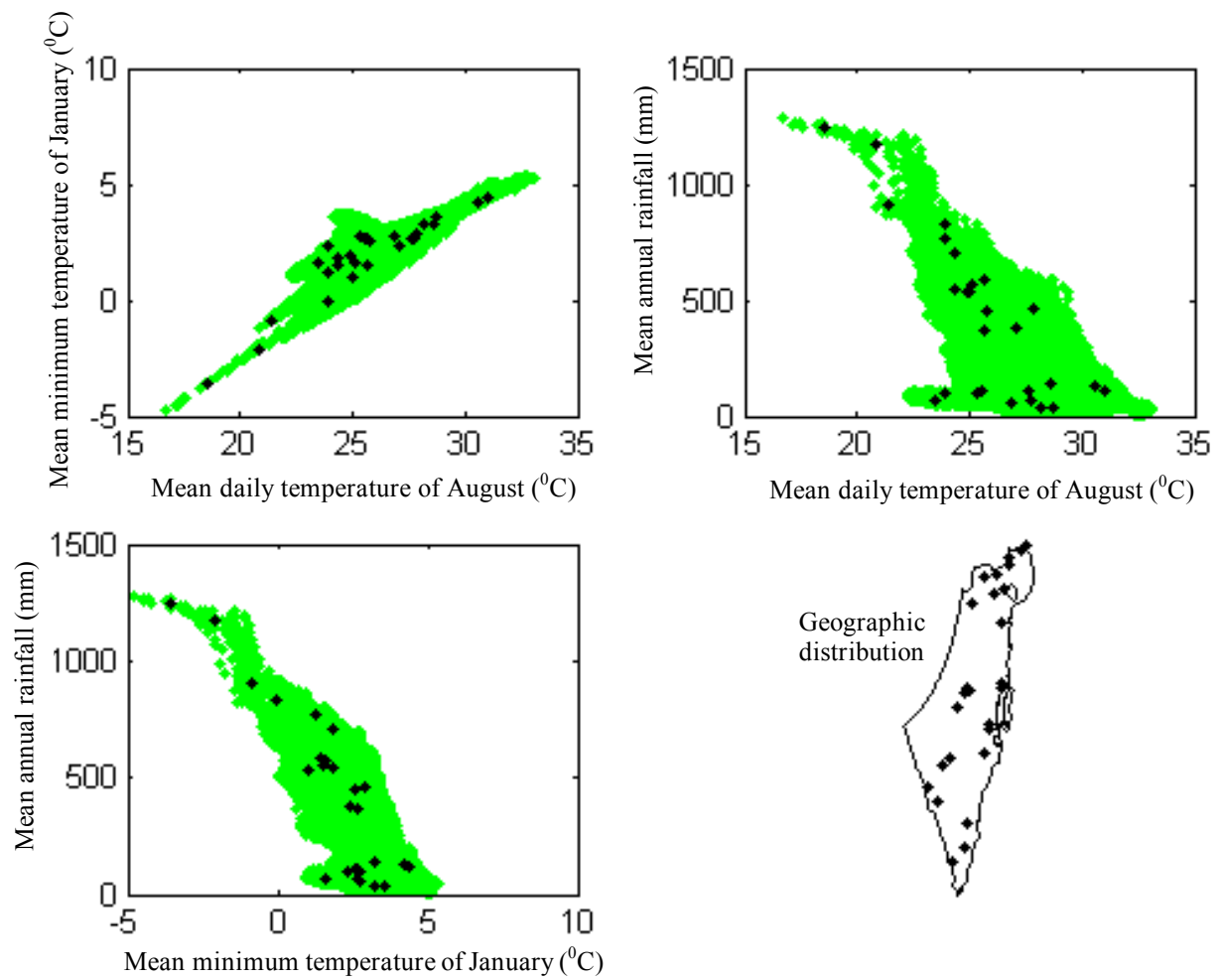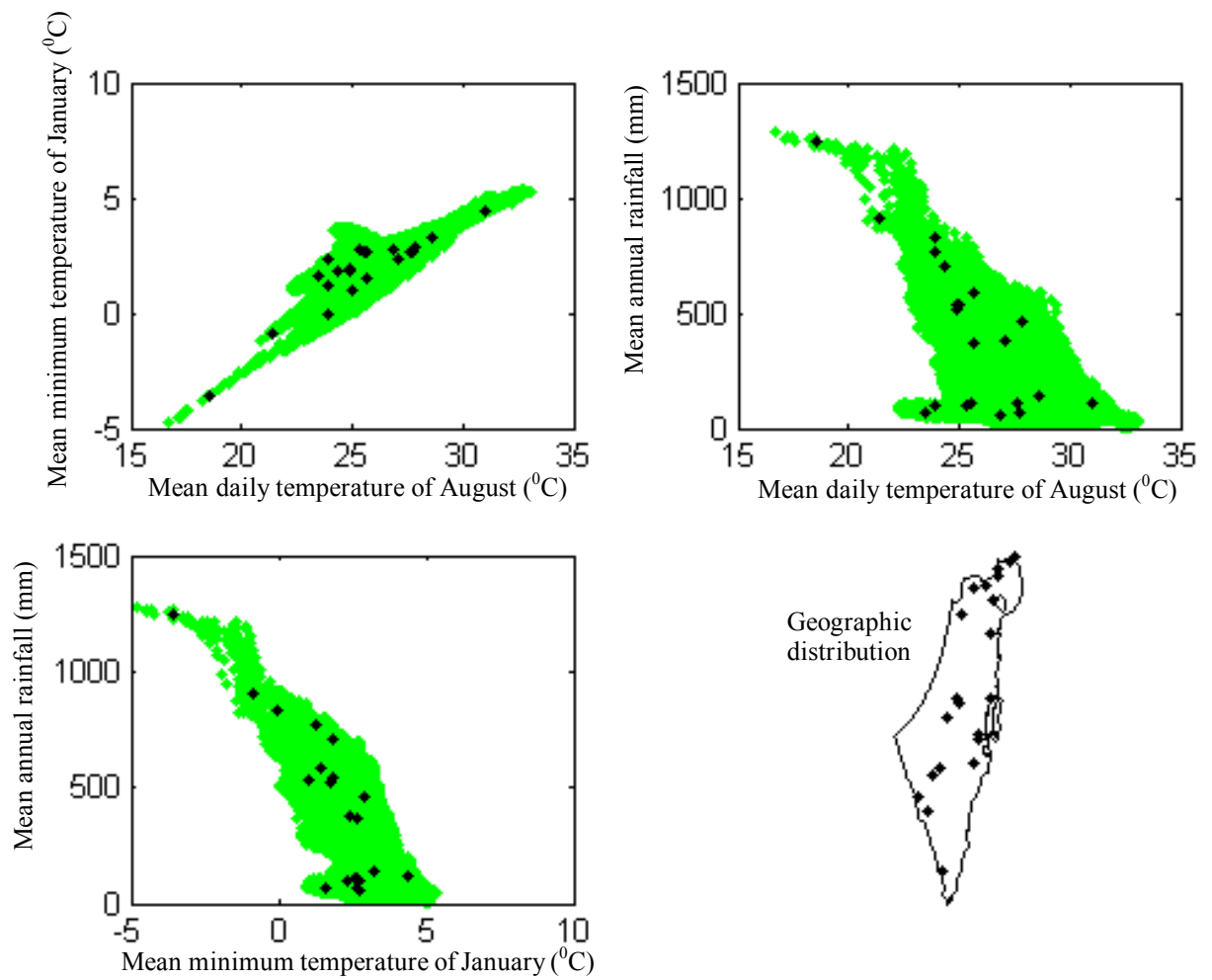
Figure 4. Distribution of the bird validation sites relative to climatic variables and the geographic distribution of the sites. In green is the climatic combination present in the study area.

Error matrix parameters

<u>Combined effects of models and fauna</u>

All error matrix parameters showed highly significant interactions between models and fauna (Figure 5, Tables 3, 4). Without covariates, all parameters showed a significant difference among the three models. The 'b' parameter showed a difference among the three fauna groups. The Mahalanobis model was significantly different from the other two models by its low scores for 'a' and 'b' and high scores for 'c' and 'd' (Figure 5). The 'c' parameter was very small compared to the other parameters. Bats had a significantly high 'b' and low 'd'. Addition of the covariates NW, ACN and OBS did not make a substantial difference in the results.

Figure 5: Error bars of error matrix parameters (Mean ± S.D.). Letters represent the different significant groups.

	a	b	c	d
Models	0.013	< 0.001	0.021	0.001
Fauna	0.694	0.025	0.321	0.660
Model*Fauna	< 0.001	< 0.001	< 0.001	< 0.001

Table 3. P values of repeated measures ANOVA testing the effects of models and fauna on error matrix parameters. Bold values are statistically significant.

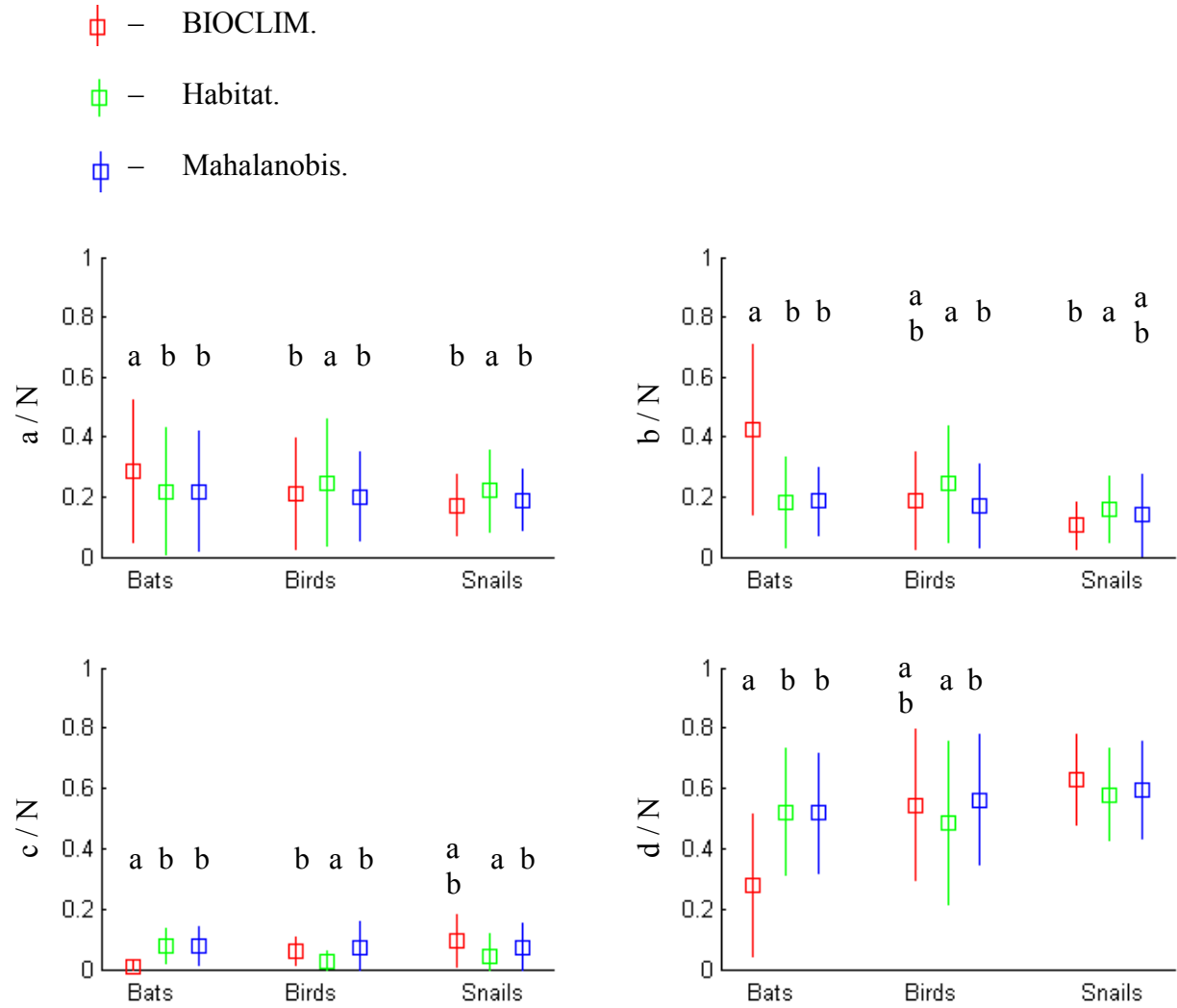|	a	b	c	d
Models	0.087	0.042	0.091	0.027
Fauna	0.403	0.001	0.077	0.001
Model*Fauna	< 0.001	< 0.001	0.003	< 0.001
Models*ACN	0.001	0.468	0.001	0.447
Models*NW	0.937	0.124	0.702	0.166
Models*OBS	0.008	0.022	0.009	0.047
ACN	< 0.001	< 0.001	0.960	0.002
NW	0.406	0.613	0.476	0.310
OBS	0.353	0.007	0.209	0.001

Table 4. P values of repeated measures ANOVA testing the effects of models and fauna on error matrix parameters, with ACN, NW, and OBS as covariates. Bold values are statistically significant.

Effect of fauna within models

In the BIOCLIM model, bats were significantly different from the other two fauna groups, by their relatively high score of 'b' and low scores of 'c' and 'd' (Tables 5, 6). In the Habitat model bats had a significantly higher 'c' than birds. The Mahalanobis model yielded no significant difference between fauna groups. Addition of the covariates to the ANOVA caused the difference between the fauna groups to be not significant.

	a	b	c	d
BIOCLIM	0.165	< 0.001	0.001	< 0.001
Habitat	0.838	0.139	0.013	0.336
Mahalanobis	0.852	0.575	0.959	0.518

Table 5. P values of one-way ANOVA testing for differences in error matrix parameters among fauna groups. A separate analysis was performed for each model. Bold values are statistically significant

		a	b	c	d
BIOCLIM	**Fauna**	0.121	0.611	0.121	0.611
	ACN	< 0.001	< 0.001	< 0.001	0.009
	NW	0.057	< 0.001	0.057	< 0.001
	OBS	0.312	0.008	0.312	0.008
Habitat	**Fauna**	0.337	0.654	0.337	0.654
	ACN	< 0.001	< 0.001	< 0.001	< 0.001
	NW	0.303	< 0.001	0.303	< 0.001
	OBS	0.045	0.018	0.045	0.018
Mahalanobis	**Fauna**	0.764	0.337	0.764	0.337
	ACN	< 0.001	< 0.001	< 0.001	< 0.001
	NW	0.486	< 0.001	0.486	< 0.001
	OBS	0.202	0.616	0.202	0.616

Table 6. P values of ANCOVA testing for differences in error matrix parameters among fauna groups, with ACN, NW, and OBS as covariates. A separate analysis was performed for each model. Bold values are statistically significant

<u>Effect of models within fauna</u>

In all three faunistic groups there were significant differences in error matrix parameters between the models (Tables 7, 8). The bats showed a pattern different from that of the birds or snails. In bats, the BIOCLIM model showed significantly higher scores, than the other two models, for 'a' and 'b' and lower scores for 'c' and 'd'. In birds, the Habitat model showed significantly higher scores for 'a', 'b' and significantly lower scores for 'c' and 'd'. The snails showed significant differences for 'a' and 'c'; where in 'a' the Habitat model was significantly higher then BIOCLIM and in 'c' the Mahalanobis was higher than Habitat. With covariates added, only the bat fauna in BIOCLIM showed significant differences between models resulting in a significantly high score of 'b' and a significantly low score of 'd'.

	a	b	c	d
Bats	0.001	< 0.001	0.001	0.002
Birds	0.004	0.028	0.001	0.017
Snails	0.005	0.007	0.023	0.107

Table 7. P values of repeated measures ANOVA testing for differences in error matrix parameters among models. A separate analysis was performed for each fauna. Bold values are statistically significant.

		a	b	c	d
Bats	**Models**	0.524	**< 0.001**	0.524	**< 0.001**
	Models*ACN	0.223	**0.012**	0.223	**0.012**
	Models*NW	0.176	**0.019**	0.176	**0.019**
	Models*OBS	0.430	**0.031**	0.434	**0.031**
	ACN	**< 0.001**	**0.004**	0.213	0.598
	NW	0.058	0.092	0.123	0.616
	OBS	0.071	**0.012**	0.349	0.126
Birds	**Models**	0.698	0.622	0.149	0.957
	Models*ACN	**< 0.001**	0.340	**< 0.001**	0.588
	Models*NW	0.777	0.232	**0.017**	0.594
	Models*OBS	0.260	0.328	0.115	0.285
	ACN	**< 0.001**	**0.011**	**0.026**	**0.020**
	NW	0.424	0.452	0.537	0.754
	OBS	0.749	**0.050**	0.260	**0.037**
Snails	**Models**	**0.034**	0.283	**0.023**	0.880
	Models*ACN	0.053	0.094	0.141	0.747
	Models*NW	0.525	0.636	0.504	0.695
	Models*OBS	0.178	0.780	0.204	0.786
	ACN	**< 0.001**	**0.047**	0.971	0.870
	NW	0.154	0.313	0.477	0.640
	OBS	**0.019**	0.826	0.684	0.308

Table 8. P values of repeated measures ANOVA testing for differences in error matrix parameters among models, with ACN, NW, and OBS as covariates. A separate analysis was performed for each fauna. Bold values are statistically significant.

Effect of species parameters

Incidence (ACN): Owing to the way ACN is calculated (i.e., (a+c)/N), there was a strong positive effect of ACN on the 'a' and 'c' parameters in all fauna groups (Table 9, Figures 6-9). There was a strong negative effect of ACN on the 'd' parameter in all fauna groups. The ACN did not show any consistent effect on the 'b' parameter. In bats it showed a negative effect, in snails a positive one, while in birds it did not show any significant affect.

Niche-width (NW): Birds and snails showed a significant positive effect of NW on 'a'. The effect of NW on 'b' and 'c' was not consistent (Table 10, Figures 6-9). All fauna showed a significant negative effect of NW on the 'd', this parameter had the highest coefficient of determination (R^2). In Birds all parameters showed a positive significant effect, except for the 'd' that had a negative one.

Number of Observations (OBS): For all fauna groups, there was a positive effect of OBS on the 'a' parameter and a negative one on 'd' (Table 11, Figures 6-9). The corresponding effects on 'b' and 'c' parameters were not consistent. In bats in the BIOCLIM model, only the 'a' parameter was significantly effected by OBS. The other two models showed significant effects on, 'a', 'c' and 'd'. In birds, most of the effects were statistically significant. In the BIOCLIM model all parameters were significant, in Habitat only the 'c' parameter was not significant. In the Mahalanobis model only the 'b' parameter was not significant. Snails showed a much clearer pattern: in all models, the 'b' and 'c' parameters were not significant whereas in the Mahalanobis model only the 'd' parameter was not significant.

		ACN		
		MODEL		
Fauna	**Validation parameters**	**BIOCLIM**	**Mahalanobis**	**Habitat**
Bats	**a**	0.994 lin (+)	0.943 lin (+)	0.951 lin (+)
	b	0.390 log (-)	0.369 log (-)	ns
	c	ns	0.435 lin (+)	0.392 log (+)
	d	ns	0.780 lin (-)	0.612 lin (-)
Birds	**a**	0.961 lin (+)	0.885 lin (+)	0.973 lin (+)
	b	ns	ns	ns
	c	0.405 log (+)	0.658 lin (+)	ns
	d	0.597 lin (-)	0.615 lin (-)	0.496 lin (-)
Snails	**a**	0.743 lin (+)	0.845 lin (+)	0.782 lin (+)
	b	ns	0.403 qua (+)	0.383 qua (+)
	c	0.621 lin (+)	0.734 lin (+)	0.421 qua (+)
	d	0.758 lin (-)	0.589 qua (-)	0.562 lin (-)

Table 9. R^2 values of regression models testing the effect of ACN on error matrix parameters. Regression models were constructed using linear models as well as quadratic and logarithmic transformations of ACN. The results given are those that showed the best fit. The sign indicates the direction (positive or negative) of the effect.

	NW			
		MODEL		
Fauna	**Validation parameters**	**BIOCLIM**	**Mahalanobis**	**Habitat**
Bats	a	ns	ns	ns
	b	ns	0.431 qua (-)	ns
	c	ns	ns	ns
	d	0.600 log (-)	0.401 lin (-)	0.368 lin (-)
Birds	a	0.472 lin (+)	0.282 log (+)	0.263 log (+)
	b	0.334 lin (+)	0.160 lin (+)	0.364 lin (+)
	c	ns	0.176 lin (+)	0.307 qua (+)
	d	0.798 lin (-)	0.596 lin (-)	0.706 lin (-)
Snails	a	0.263 log (+)	0.450 lin (+)	0.590 lin (+)
	b	0.231 lin (+)	ns	ns
	c	ns	ns	ns
	d	0.435 lin (-)	0.280 log (-)	0.428 log (-)

Table 10. R^2 values of regression models testing the effect of NW on error matrix parameters. Regression models were constructed using linear models as well as quadratic and logarithmic transformations of NW. The results given are those that showed the best fit. The sign indicates the direction (positive or negative) of the effect.

		OBS		
		MODEL		
Fauna	Validation parameters	BIOCLIM	Mahalanobis	Habitat
Bats	a	0.850 lin (+)	0.728 lin (+)	0.789 lin (+)
	b	ns	ns	ns
	c	ns	0.639 log (+)	0.656 qua (-)
	d	ns	0.832 lin (-)	0.820 lin (-)
Birds	a	0.554 lin (+)	0.428 log (+)	0.537 lin (+)
	b	0.182 log (+)	ns	0.175 log (+)
	c	0.138 lin (+)	0.493 lin (+)	ns
	d	0.692 log (-)	0.520 log (-)	0.676 log (-)
Snails	a	0.606 lin (+)	0.539 lin (+)	0.616 lin (+)
	b	ns	ns	ns
	c	ns	ns	ns
	d	0.232 lin (-)	ns	0.254 lin (-)

Table 11. R^2 values of regression models testing the effect of OBS on error matrix parameters. Regression models were constructed using linear models as well as quadratic and logarithmic transformations of OBS. The results given are those that showed the best fit. The sign indicates the direction (positive or negative) of the effect.
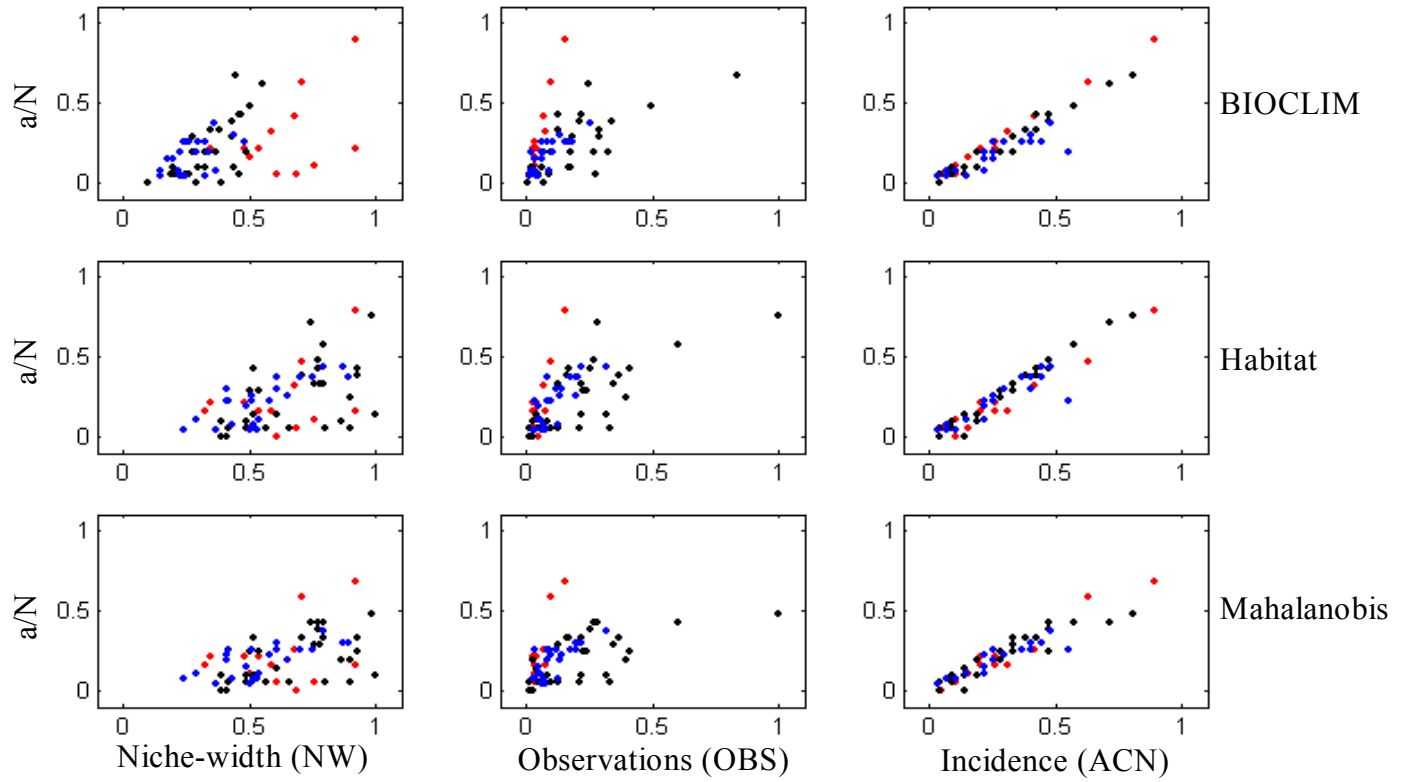
Figure 6. Plots of the effect of species parameters (NW, OBS, and ACN) on the normalized 'a' error matrix parameter (the 'a' was normalized by the number of validation cells: a/N). Bats in red, birds in black, snails in blue.
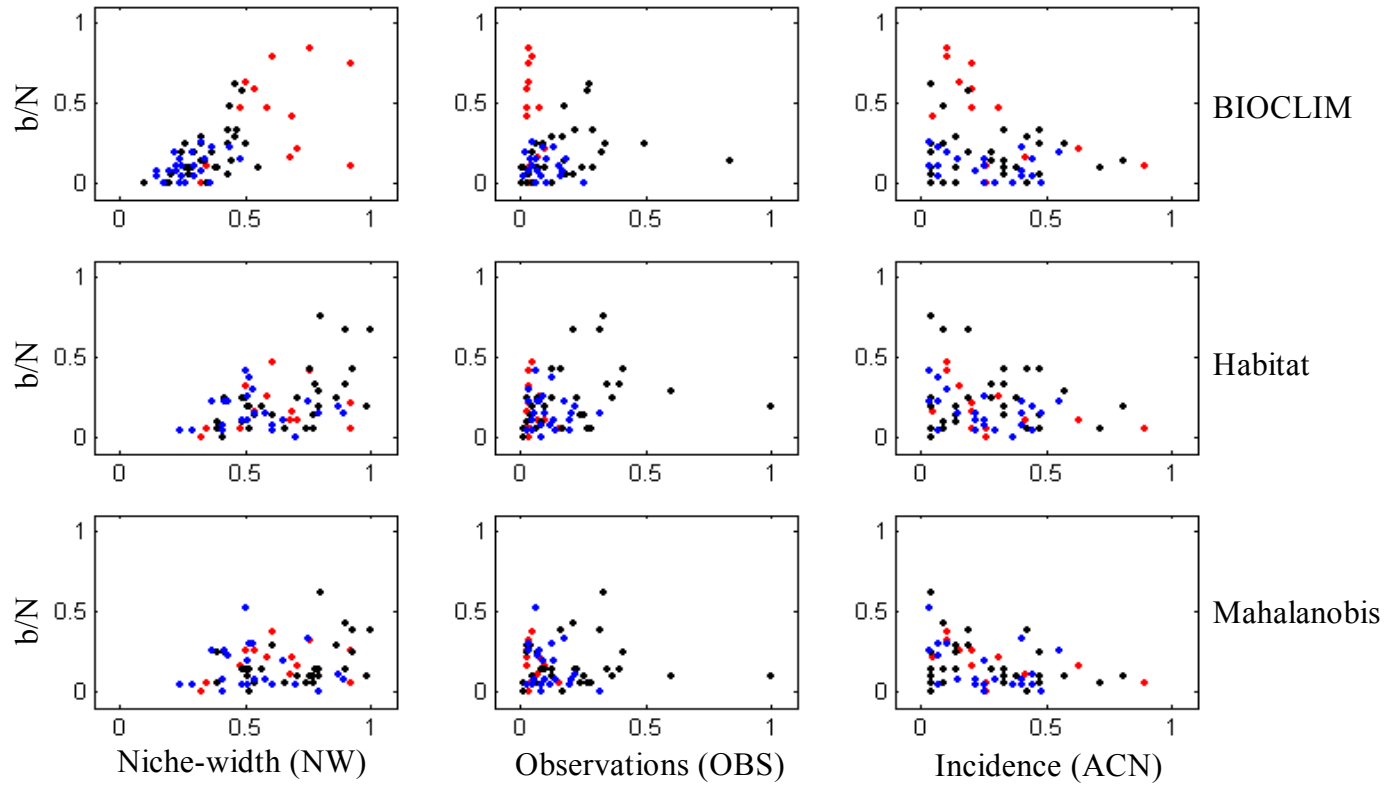
Figure 7: Plots of the effect of species parameters (NW, OBS, and ACN) on the normalized 'b' error matrix parameter (the 'b' was normalized by the number of validation cells: b/N). Bats in red, birds in black, snails in blue.
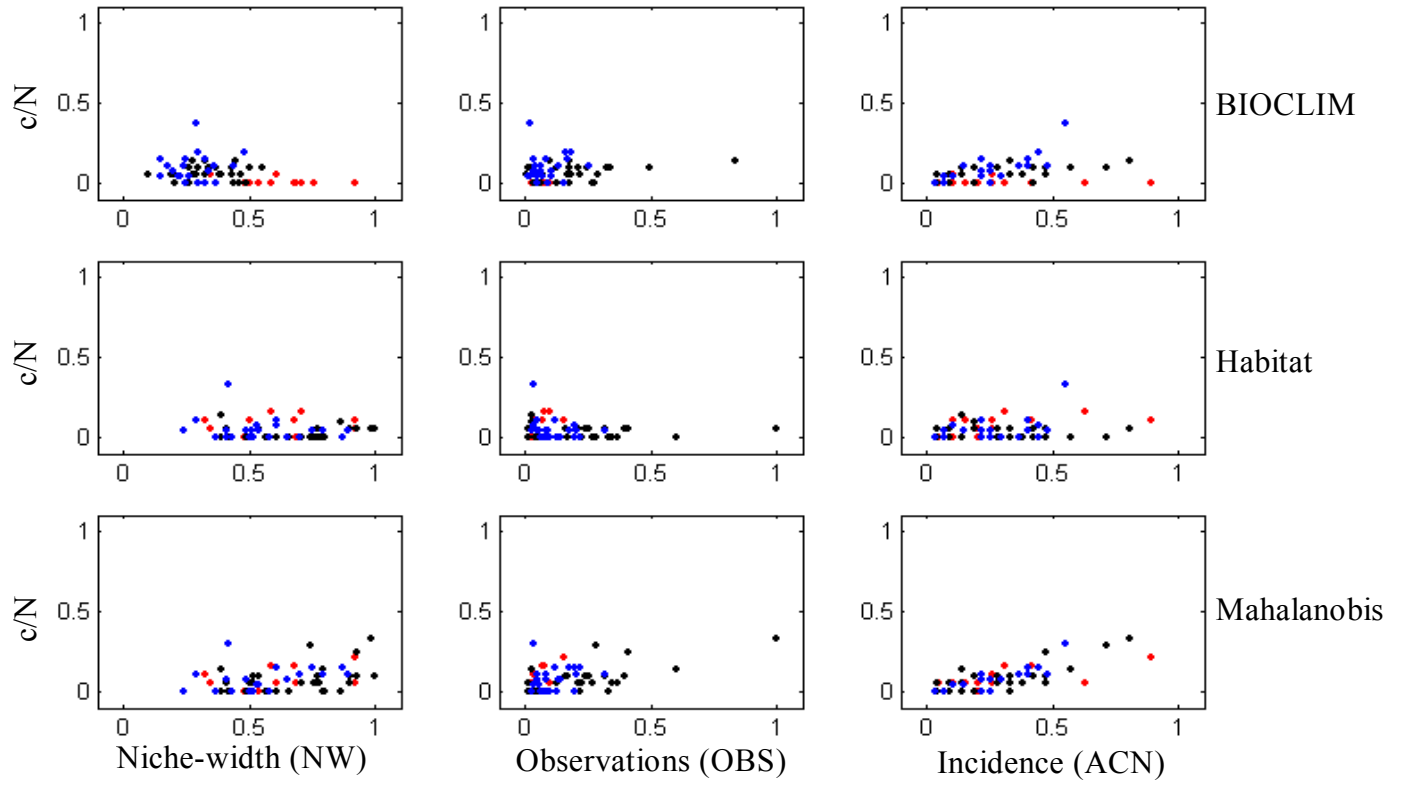
Figure 8: Plots of the effect of species parameters (NW, OBS, and ACN) on the normalized 'c' error matrix parameter (the 'c' was normalized by the number of validation cells: c/N). Bats in red, birds in black, snails in blue.
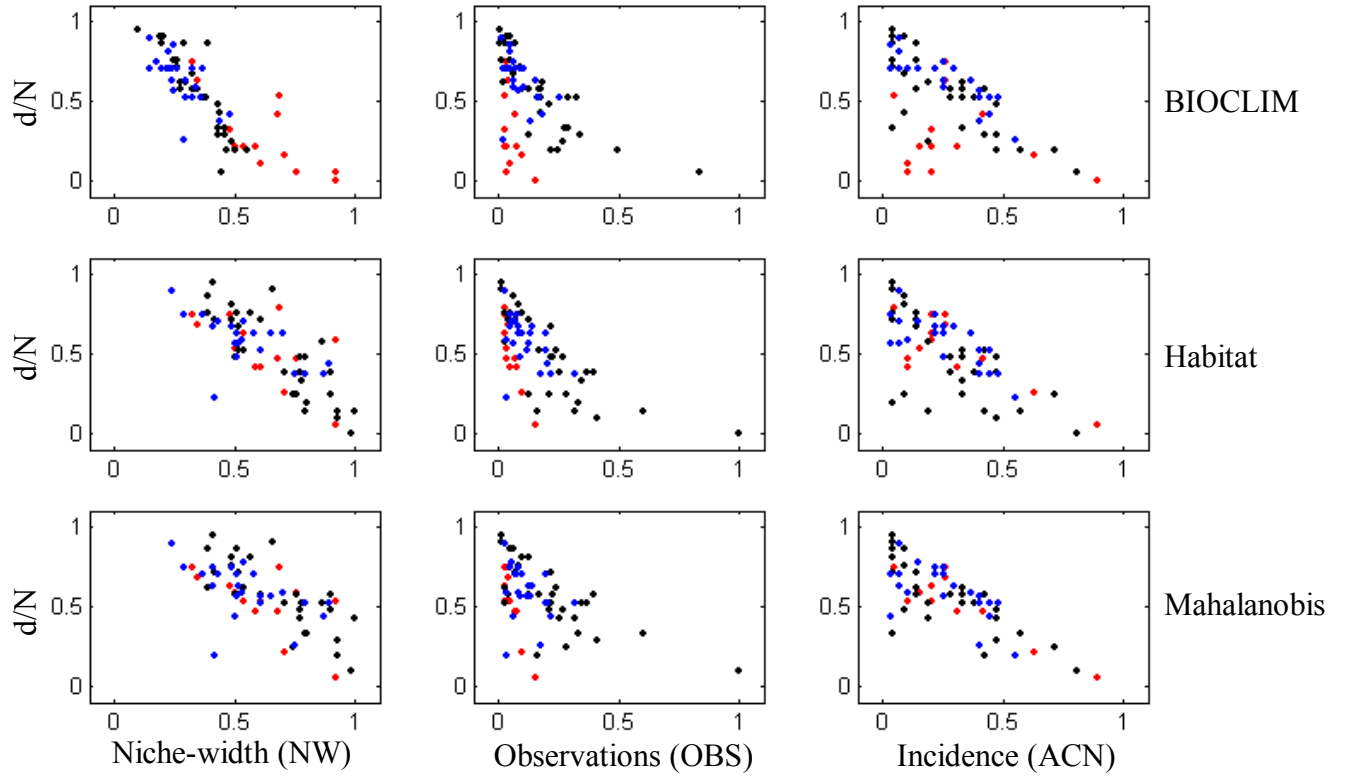
Figure 9: Plots of the effect of species parameters (NW, OBS, and ACN) on the normalized 'd' error matrix parameter (the 'd' was normalized by the number of validation cells: d/N). Bats in red, birds in black, snails in blue.

Measures of predictive accuracy

Combined effects of models and fauna

The Kappa statistic showed no significant difference between models or fauna. The interaction between the two effects was also not significant. Adding the covariates did not have a significant effect on these results (Figure 10, Tables 12, 13). Sensitivity showed a significant difference only for the interaction. Both Overall Accuracy and Specificity showed that BIOCLIM had significantly poorer performance than the other two models. Snails had significantly higher Specificity scores than the other groups. Bats in BIOCLIM had significantly lower validation scores than all other fauna and model groups. Adding covariates hardly affected the results.
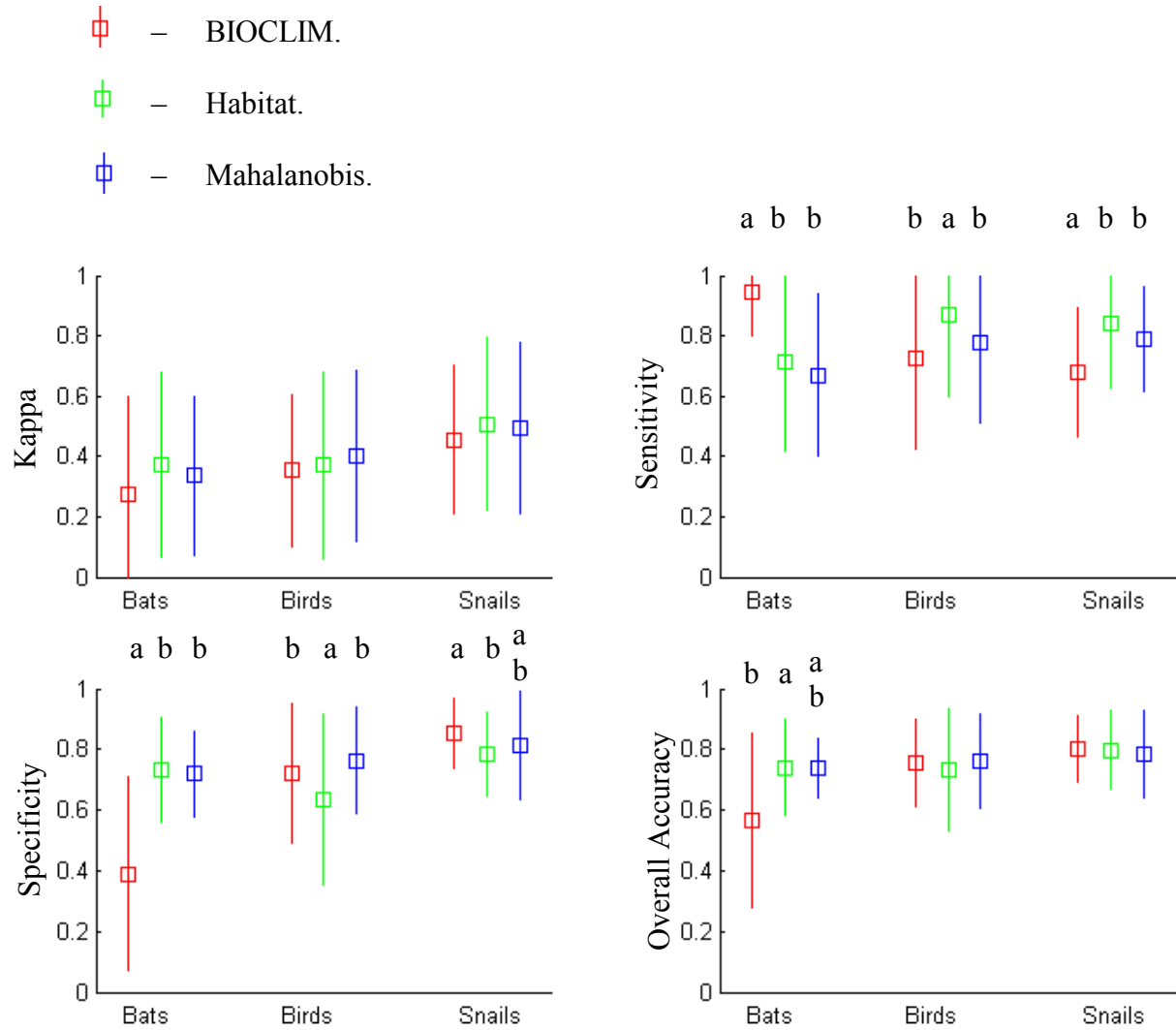
Figure 10. Error bars of model validation parameters (Mean ± S.D.). Letters represent the different significant groups.

	Kappa	Overall Accuracy	Sensitivity	Specificity
Models	0.092	0.002	0.180	< 0.001
Fauna	0.144	0.084	0.934	0.008
Model*Fauna	0.763	< 0.001	< 0.001	< 0.001

Table 12. P values of repeated measures ANOVA testing the effects of models and fauna on validation parameters. Bold values are statistically significant.

	Kappa	Overall Accuracy	Sensitivity	Specificity
Models	0.807	0.684	0.400	0.356
Fauna	0.161	0.014	0.848	0.001
Model * Fauna	0.993	<0.001	< 0.001	< 0.001
Models*ACN	0.290	0.009	0.784	0.272
Models*NW	0.890	0.209	0.562	0.266
Models*OBS	0.386	0.743	0.110	< 0.001
ACN	0.003	0.001	0.529	0.054
NW	0.041	< 0.001	0.478	< 0.001
OBS	0.292	0.124	0.240	0.011

Table 13. P values of the repeated measures ANOVA testing the effects of models and fauna on validation parameters, with ACN, OBS, and NW as covariates. Bold values are statistically significant.

Effect of fauna within models

Only the BIOCLIM model showed significant differences between the fauna groups (Tables 14, 15). Kappa did not show a significant difference between models, between fauna, or their interaction. It did not show any significant difference when the covariates were added. Bats in BIOCLIM had a significantly higher score in Overall Accuracy and Sensitivity while a relatively low score in Specificity. Adding covariates made the effect of fauna non significant in all models.

	Kappa	Overall Accuracy	Sensitivity	Specificity
BIOCLIM	0.126	0.001	0.008	< 0.001
Habitat	0.220	0.358	0.142	0.056
Mahalanobis	0.238	0.600	0.312	0.285

Table 14. P values of one-way ANOVA testing for differences in validation parameters, among fauna groups. A separate analysis was performed for each model. Bold values are statistically significant.

		Kappa	Overall Accuracy	Sensitivity	Specificity
BIOCLIM	Fauna	0.404	0.593	0.377	0.660
	ACN	0.001	< 0.001	0.222	0.010
	NW	< 0.001	< 0.001	0.059	< 0.001
	OBS	0.552	0.062	0.054	0.006
Habitat	Fauna	0.293	0.474	0.370	0.402
	ACN	0.002	0.001	0.868	0.016
	NW	0.114	< 0.001	0.400	< 0.001
	OBS	0.137	0.208	0.270	< 0.001
Mahalanobis	Fauna	0.325	0.328	0.396	0.352
	ACN	0.036	0.132	0.630	0.280
	NW	0.042	0.002	0.811	0.001
	OBS	0.684	0.358	0.903	0.719

Table 15. P values of ANCOVA testing for differences in validation parameters among fauna groups, with ACN, OBS and NW as covariates. A separate analysis was performed for each model. Bold values are statistically significant.

Effect of models within fauna

Kappa and Overall Accuracy showed the lowest differences between models in all faunistic groups (Tables 16, 17). The Kappa had no significant difference between fauna, nor did it show any significant difference when the covariates were added. In bats the BIOCLIM model showed a high score for Sensitivity and low scores for Specificity and Overall Accuracy. Birds showed a different pattern; the Habitat model had a significantly low Specificity score and a significantly high Sensitivity score. Snails showed a similar pattern as birds, except that the Mahalanobis model in Specificity was not significantly different from the other models. Sensitivity in the Mahalanobis was not significantly different from the Habitat model. Adding the covariates to the test resulted in all models becoming not significantly different from each other.

	Kappa	Overall Accuracy	Sensitivity	Specificity
Bats	0.276	0.011	0.011	< 0.001
Birds	0.499	0.376	0.039	< 0.001
Snails	0.371	0.664	0.001	0.003

Table 16. P values of repeated measures ANOVA testing for differences in validation parameters among models. A separate analysis was performed for each fauna. Bold values are statistically significant.

		Kappa	Overall Accuracy	Sensitivity	Specificity
Bats	**Models**	0.521	0.666	0.389	0.704
	Models*ACN	0.272	0.239	0.487	0.007
	Models*NW	0.122	0.017	0.394	0.049
	Models*OBS	0.159	0.216	0.589	0.009
	ACN	0.002	0.010	0.001	0.131
	NW	0.043	0.112	0.112	0.042
	OBS	0.006	0.071	0.002	0.142
Birds	**Models**	0.583	0.207	0.215	0.332
	Models*ACN	0.323	0.006	0.576	0.495
	Models*NW	0.574	0.013	0.547	0.165
	Models*OBS	0.580	0.758	0.387	0.015
	ACN	0.002	0.002	0.994	0.012
	NW	0.090	< 0.001	0.715	< 0.001
	OBS	0.047	0.81	0.622	0.002
Snails	**Models**	0.642	0.457	0.091	0.381
	Models*ACN	0.471	0.084	0.641	0.591
	Models*NW	0.968	0.996	0.821	0.726
	Models*OBS	0.693	0.828	0.449	0.577
	ACN	0.809	0.343	0.001	0.992
	NW	0.758	0.527	0.297	0.322
	OBS	0.188	0.195	0.034	0.345

Table 17. P values of repeated measures ANOVA testing for difference in validation parameters among models, with ACN, NW and OBS as covariates. A separate analysis was performed for each fauna. Bold values are statistically significant.

<u>Effect of species parameters</u>

   <u>Incidence (ACN)</u>: ACN had a clear quadratic effect on Kappa in all fauna groups (Table 18, figures 11-14). The responses of the other parameters to variation in ACN were not consistent.

   <u>Niche-width (NW)</u>: Birds showed a statistically significant effect of NW on all validation parameters. The effect on Specificity and Overall Accuracy was negative, while that on Sensitivity was not consistent (Table 19, figures 11-14). Snails hardly showed any significant effect. Bats had a negative effect of NW on Specificity.

   <u>Number of Observations (OBS)</u>: The bat fauna in all models did not show any significant effect of OBS on the validation parameters (Table 20, Figures 11-14). Snails showed no consistent effects of OBS on all parameters. Birds, showed a significant negative effect of the OBS on Overall Accuracy and Specificity.

Fauna	Validation parameters	MODEL		
		BIOCLIM	Mahalanobis	Habitat
Bats	Kappa	0.472 qua (-)	0.546 qua (-)	ns
	Sensitivity	ns	0.378 log (+)	ns
	Specificity	ns	ns	ns
	Overall Accuracy	0.402 log (+)	ns	ns
Birds	Kappa	0.267 qua (-)	ns	ns
	Sensitivity	ns	ns	ns
	Specificity	0.158 lin (-)	ns	0.142 lin (-)
	Overall Accuracy	ns	ns	ns
Snails	Kappa	0.384 qau (-)	0.434 qua (-)	0.633 qua (-)
	Sensitivity	ns	0.440 log (-)	ns
	Specificity	ns	0.314 qua (-)	0.347 qua (-)
	Overall Accuracy	0.385 qua (-)	0.385 qua (-)	0.443 qua (-)

Table 18. R^2 values of regression models testing the effect of ACN on validation parameters. Regression models were constructed using linear models as well as quadratic and logarithmic transformations of ACN. The results given are those that showed the best fit. The sign indicates the direction (positive or negative) of the effect.

Fauna	Validation parameters	MODEL		
		BIOCLIM	Mahalanobis	Habitat
Bats	Kappa	0.395 log (-)	0.414 log (-)	ns
	Sensitivity	ns	ns	ns
	Specificity	0.544 log (-)	0.620 log (-)	0.413 log (-)
	Overall Accuracy	ns	0.546 qua (+)	ns
Birds	Kappa	ns	0.412 qua (-)	0.422 qua (-)
	Sensitivity	0.236 log (+)	0.277 qua (-)	0.249 qua (-)
	Specificity	0.593 lin (-)	0.419 lin (-)	0.619 lin (-)
	Overall Accuracy	0.438 lin (-)	0.316 lin (-)	0.380 lin (-)
Snails	Kappa	ns	ns	ns
	Sensitivity	ns	ns	0.190 log (+)
	Specificity	0.315 lin (-)	ns	ns
	Overall Accuracy	0.198 lin (-)	ns	ns

Table 19. R^2 values of regression models testing the effect of NW on validation parameters. Regression models were constructed using linear models as well as quadratic and logarithmic transformations of NW. The results given are those that showed the best fit. The sign indicates the direction (positive or negative) of the effect.

| | | OBS | | |
| | | MODEL | | |
Fauna	Validation parameters	BIOCLIM	Mahalanobis	Habitat
Bats	**Kappa**	ns	ns	ns
	Sensitivity	ns	ns	ns
	Specificity	ns	ns	ns
	Overall Accuracy	ns	ns	ns
Birds	**Kappa**	ns	ns	ns
	Sensitivity	0.375 log (+)	ns	0.231 log (+)
	Specificity	0.477 lin (-)	0.133 lin (-)	0.474 lin (-)
	Overall Accuracy	0.288 log (-)	0.152 lin (-)	0.138 log (-)
Snails	**Kappa**	0.241 log (+)	ns	0.289 log (+)
	Sensitivity	ns	ns	0.432 log (+)
	Specificity	ns	ns	ns
	Overall Accuracy	ns	ns	ns

Table 20. R^2 values of regression models testing the effect of OBS on validation parameters. Regression models were constructed using linear models as well as quadratic and logarithmic transformations of OBS. The results given are those that showed the best fit. The sign indicates the direction (positive or negative) of the effect.
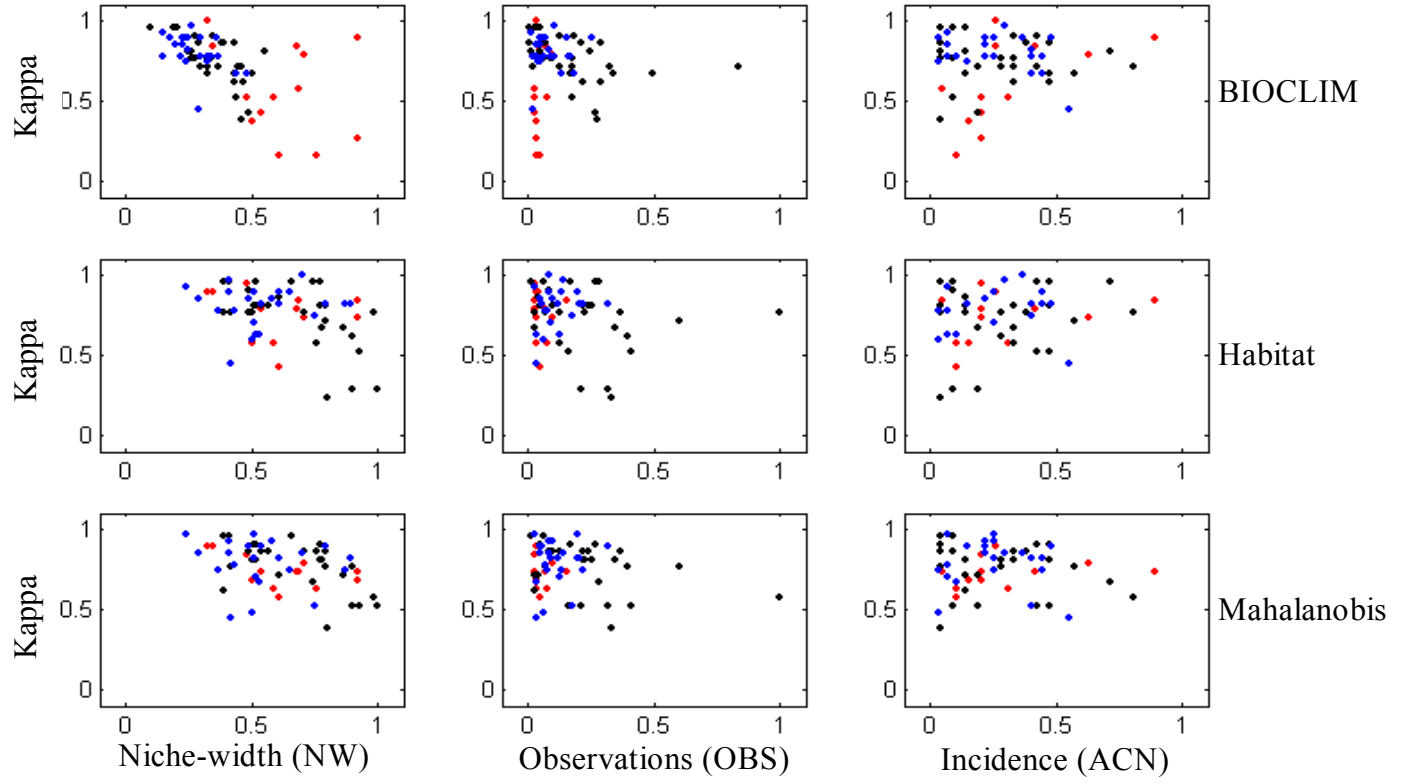
Figure 11. Plots of the effect of species parameters (NW, OBS, and ACN) on Kappa. Bats in red, birds in black, snails in blue.
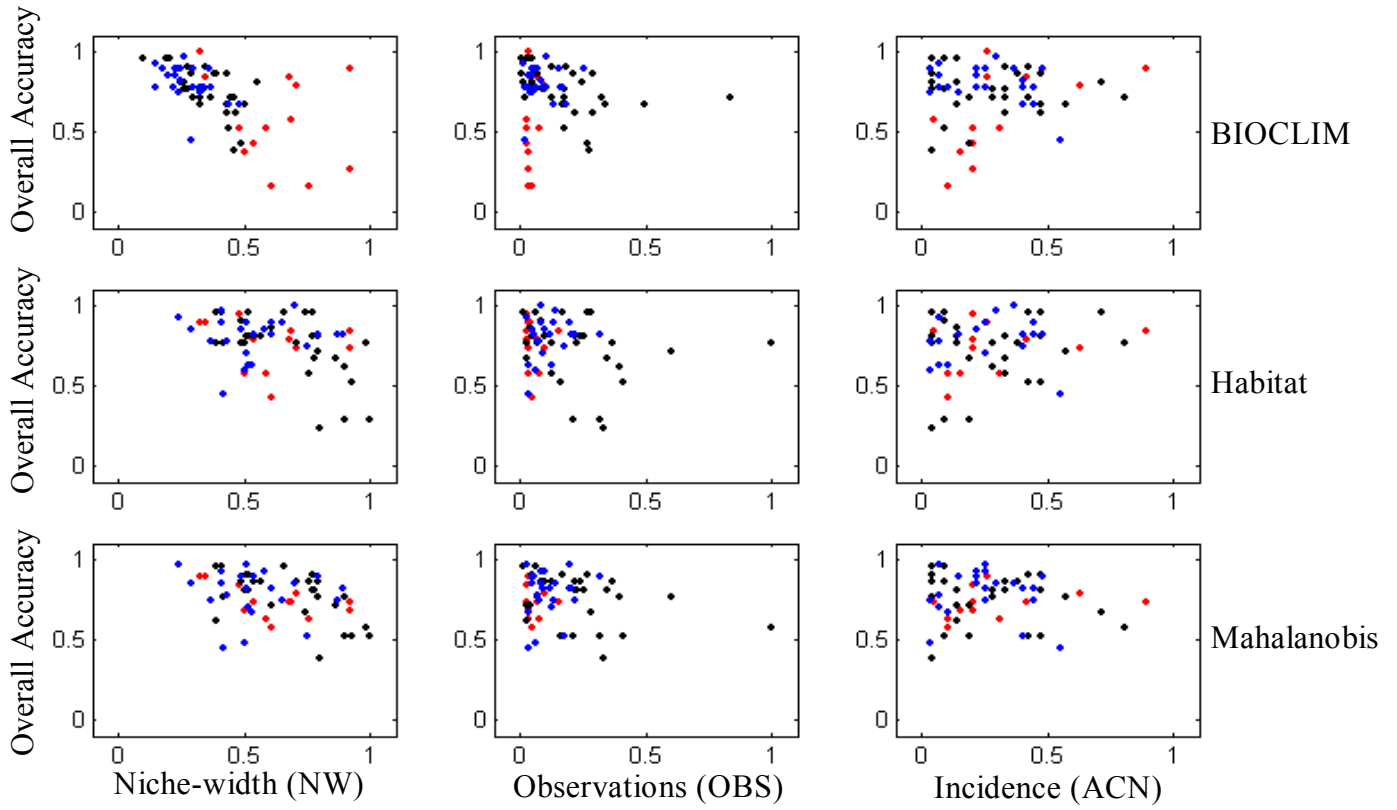
Figure 12. Plots of the effect of species parameters (NW, OBS, and ACN) on Overall Accuracy. Bats in red, birds in black, snails in blue.
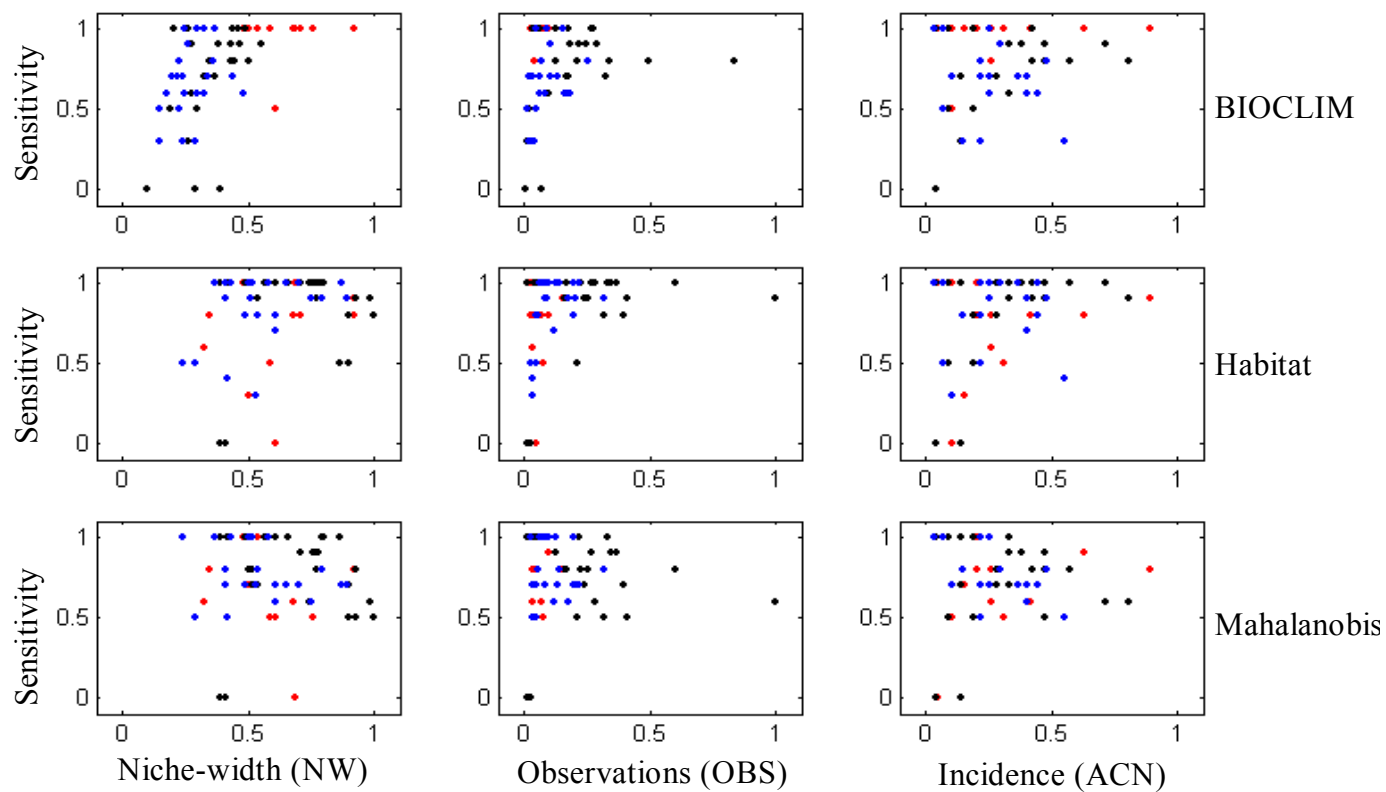
Figure 13. Plots of the effect of species parameters (NW, OBS, and ACN) on Sensitivity. Bats in red, birds in black, snails in blue.
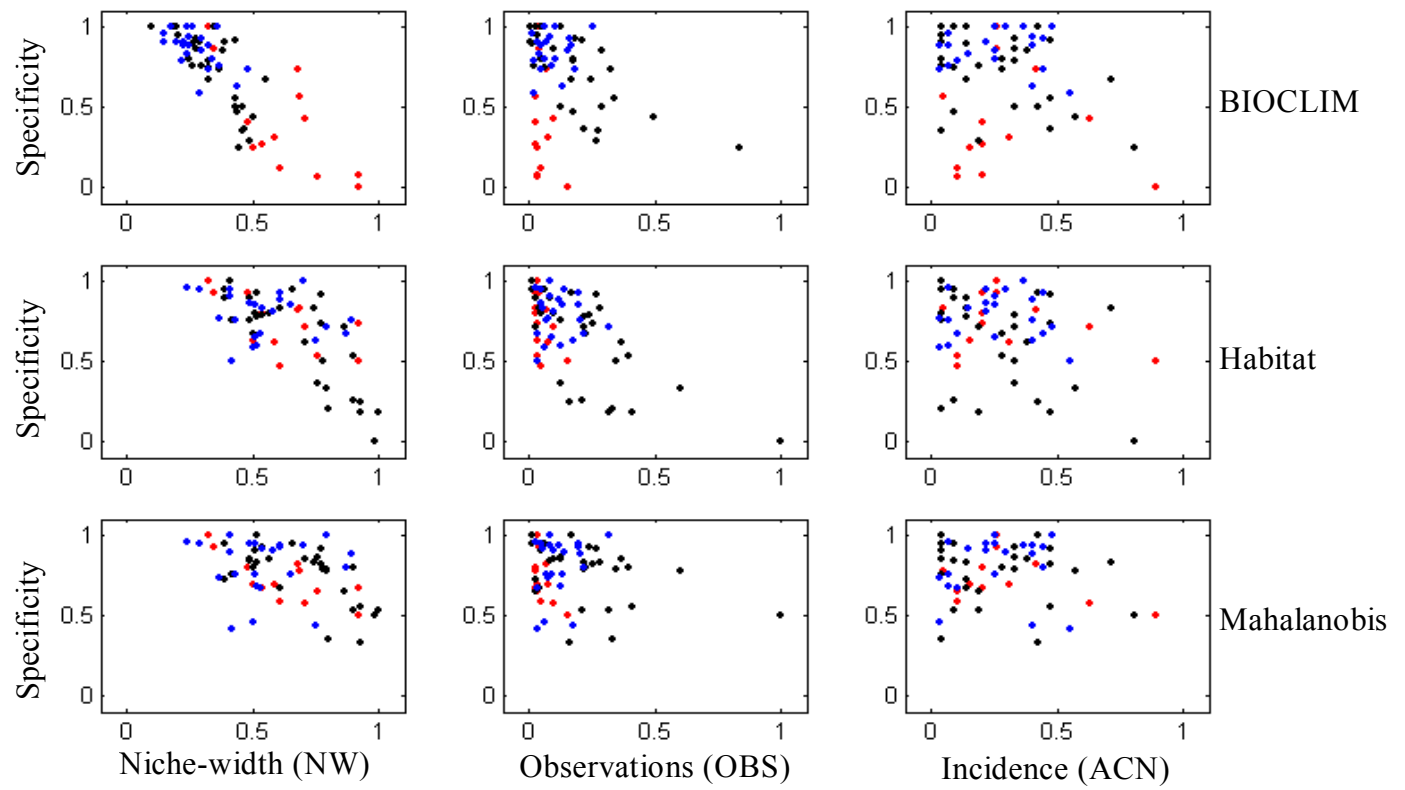
Figure 14. Plots of the effect of species parameters (NW, OBS, and ACN) on Specificity. Bats in red, birds in black, snails in blue.

Discussion

A variety of Climatic Envelope Models (CEMs) have been proposed in the literature (Guisan & Zimmerman 2000). Yet, very few studies have compared the accuracy of predictions derived from different CEMs. In principle, the accuracy of CEMs depends on three factors: (1) the accuracy of the dataset used for model calibration; (2) the accuracy of the climate dataset; and (3) the algorithm used to determine the climatic envelope. By constraining the first two factors, I was able to evaluate the effect of model algorithms on the accuracy of predictions generated by such models (Figure 16). In general, little difference was found between the models in their ability to correctly predict species distribution. Differences among fauna groups in predictive accuracy were also not statistically significant in most cases. Those differences that were found could be attributed to differences among species in properties such as incidence and niche-width.

Models

The absolute low value of the 'c' parameter in all models reflects little errors in predicting that a particular species will not be present where it really occurs (i.e., False present). By implication, CEMs are expected to make bigger errors in over predicting presence than in under predicting presence of a species. Previous works have found a similar pattern (Araujo & Williams 2000; Robertson et al. 2003). The BIOCLIM model differed from the other two models by the need to exclude 10% outliers from the dataset, in order for its prediction to be most accurate. This need is explained by the way the niche envelope is created in the BIOCLIM model which, by creating a rectangular niche envelope, over expects the presence of the species. This is reflected in relatively high 'a', 'b' and low 'c', 'd'. The Mahalanobis model, which creates elliptic niche envelopes (Farber & Kadmon 2003), predicts a more restricted distribution of species, and has a relatively high 'c', 'd' and low 'a', 'b'. Predictions of the habitat model resemble those of Mahalanobis, due to the convex hull shape of its envelope, which narrows the climatic envelope of the species.

The effect of excluding 10% of the outliers by the BIOCLIM model is evident in the bird and snail datasets. In bats the best predictions obtained with BIOCLIM were obtained without excluding outliers. The reason for this is the low number of observations per species, in the bat dataset (an average of 52.2 vs. 107.6 for snails and 175.8 for birds). Because of the low number

of observations, the elimination of outliers as done in BIOCLIM, decreased the number of observations tremendously and reduced the ability to accurately predict distribution ranges. Indeed, previous studies have found that a minimum of 50 sites in the calibrating dataset was needed for the BIOCLIM model to correctly predict distribution (McKenney et al. 1998; Stockwell & Peterson 2002; Kadmon et al. 2003). The present result corresponds to their findings.

Once the BIOCLIM-bats are excluded, there is little difference between models in their ability to predict species distribution. Interestingly, works comparing CEMs with other types of models, did not find a significant difference as well (Elith & Burgman 2001; Robertson et al. 2003). The Kappa parameter did not show any significant difference between the models. Once species properties is added to the test, most of the differences between the models become non significant, owing to their strong effect on the prediction of the model.

Earlier works on CEMs have resulted in similar validation results as obtained in this study (Mean Kappa for: BIOCLIM = 0.37, Habitat = 0.42, Mahalanobis = 0.42). Farber and Kadmon (2003) working on woody plants distribution in Israel, found the mean Kappa of the BIOCLIM model to be 0.41 while that predicted by the Mahalanobis was 0.48. GLM modeling made by Jaberg & Guisan (2001) on three species of bats found a Kappa of 0.27-0.63. Manel et al. (2001) tested three different models (logistic regression, discriminate analysis and neural networks) using data on river birds and got a mean Kappa of 0.15-0.36.

Fauna

Species properties, which depend on scale, patchiness and dispersal, constitute an important impact on the ability of a model to correctly predict distribution (Wiens 2001). The species analyzed in this work were chosen from three different systematic groups, differing in their manner of dispersal and their provenance in field surveys. The easier it is to find a species in the field, the more likely it would be included in the dataset, and so will have a bigger database, unless specific surveys were conducted. Snails have short-range dispersal and are easy to detect in the field, while, birds have long-range dispersal and are also easy to detect. Bats have long-range dispersal but are relatively hard to detect in the field. Yet the three fauna groups showed little difference in the ability of the models to correctly predict their distribution. Of the three models, only BIOCLIM showed a difference between the groups. This is explained by the fact

that in BIOCLIM-bats there was no exclusion of outliers and the dataset was incomplete probably due to the fact that bats are hard to detect. This significant difference disappeared once covariates were added to the test.

Unlike Pearce et al. (2001), there was no difference among fauna groups in the ability of the models to correctly predict the species distribution. It seems that the bat fauna is the odd one out of the three fauna groups; this is explained by the low number of observations per species in the dataset. Bat fauna in the other two models did not cause any significant difference on the performance of the models. BIOCLIM is sensitive to the number of sites because of its rectilinear envelope algorithm. The difference is not caused solely because of the dataset quality or the BIOCLIM algorithm, but the combination of the two; the low number of sites per species and the sensitivity of the BIOCLIM model to this sort of bias.

Species properties

The ACN showed a strong effect on the 'a', 'c' and 'd' parameters. The positive effect on 'a' and 'c' is expected due to the way the ACN is calculated i.e., (a+c)/N. The ACN was found to negatively effect the 'd' parameter as well. This finding was expected as a by product of the high correlation with the 'a' parameter. A negative quadratic effect of ACN was found on Kappa. This effect is due to the manner in which Kappa is calculated (Figure 15).

OBS was found to positively effect the 'a' and negatively effect the 'd'. This is due to the high correlation between number of observations in each dataset and species rarity. Rare species obviously have fewer observations in the dataset, and will have a high 'd' and low 'a'. Unlike earlier studies, in the present research Kappa did not show a significant affect, perhaps because we adapted the minimum of 25 observations as a limit for modeling. In previous studies, Kappa did not improve significantly when the number of observations in the calibrating dataset was increased above 50 (McKenney et al. 1998; Stockwell & Peterson 2002; Kadmon et al. 2003).

NW was found to negatively effect 'd' and positively effect 'a'. Because of the high correlation between climatic range and geographic range in the study area (Farber 2000), most species show a clear correlation between the geographic range and the width of their niche. NW had a negative effect on Specificity. This indicates that CEM is able to better predict true absence of a species, if that species has a narrow niche.

45

When evaluating individual species prediction maps, it is obvious that species with specific climatic restrictions are better predicted (Figure 17).

$$\frac{\left(\dfrac{a + d}{n}\right) - \dfrac{(a + b)(a + c) + (c + d)(b + d)}{n^2}}{1 - \dfrac{(a + b)(a + c) + (c + d)(b + d)}{n^2}} = \frac{(ad) - (bc)}{((ad) - (bc)) + (\dfrac{n}{2})(b + c)}$$

Figure 15: The Kappa bias is better explained by the simplified Kappa algorithm (the equation on the right). Notice that the more 'a' differs from 'd', the smaller the product of their multiplication. If 'b' and 'c' are constant the total score of Kappa will be lowest when 'a' or 'd' are smallest relative to the other and highest when a=d.

Considerations in creating distribution maps

The prediction ability of CEMs is similar between different organisms.

When making a distribution map from CEMs, one must decide whether the model should error towards an over predicting or under predicting species presence. In general, both types of models will be equally accurate, the difference between them being in the emphasis of the error. For example, if a researcher wishes to draw a risk assessment map of species distribution, he would be wise to prefer the BIOCLIM model. If, on the other hand, a map is needed to assess new potential areas to look for a certain species, Mahalanobis is appropriate (Figure 16). A different approach to this matter is to change the mahalanobis radius or the BIOCLIM, Habitat percentile range. An additional consideration is whether the prediction should include all the calibrating dataset points. Habitat and BIOCLIM include in their predictions all the observations used to calibrate them. Mahalanobis, on the other hand, does not. Its prediction size is relative to the mahalanobis radius selected (Farber 2000). A minimum of 50 record sites per species is essential for an accurate prediction.

PR = 50% **PR = 60%** **PR = 70%** **PR = 80%** **PR = 90%** **PR = 100%**
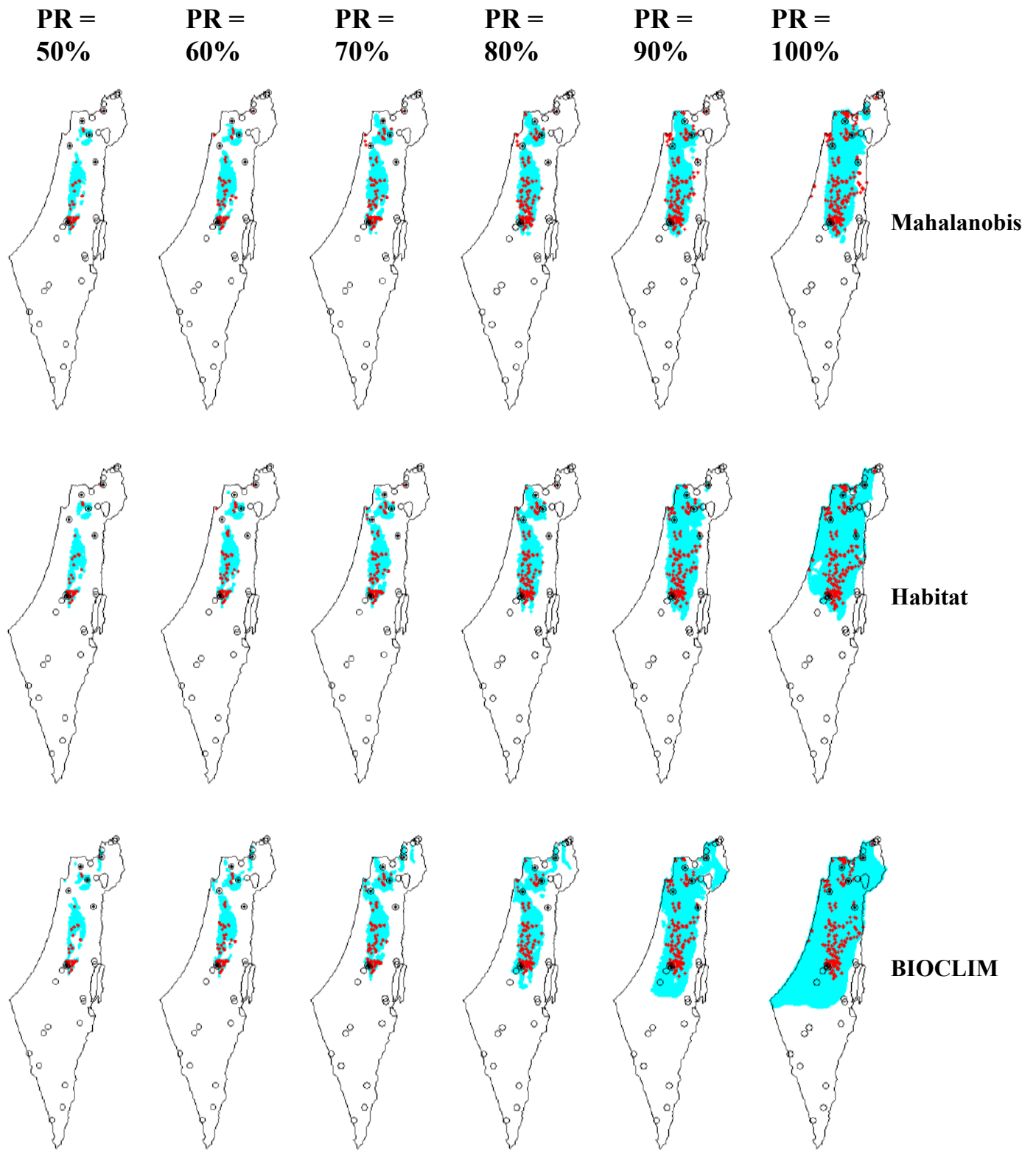
Mahalanobis

Habitat

BIOCLIM

Figure 16: Differences in all three models in their prediction, relative to outliers removed from the calibrating dataset. Red dots are the calibrating dataset, blue is the model prediction, and circles are the validation dataset (open circles: species not found, circles with black dots: species found). Data taken from the snail fauna (*Buliminus labrosus*).
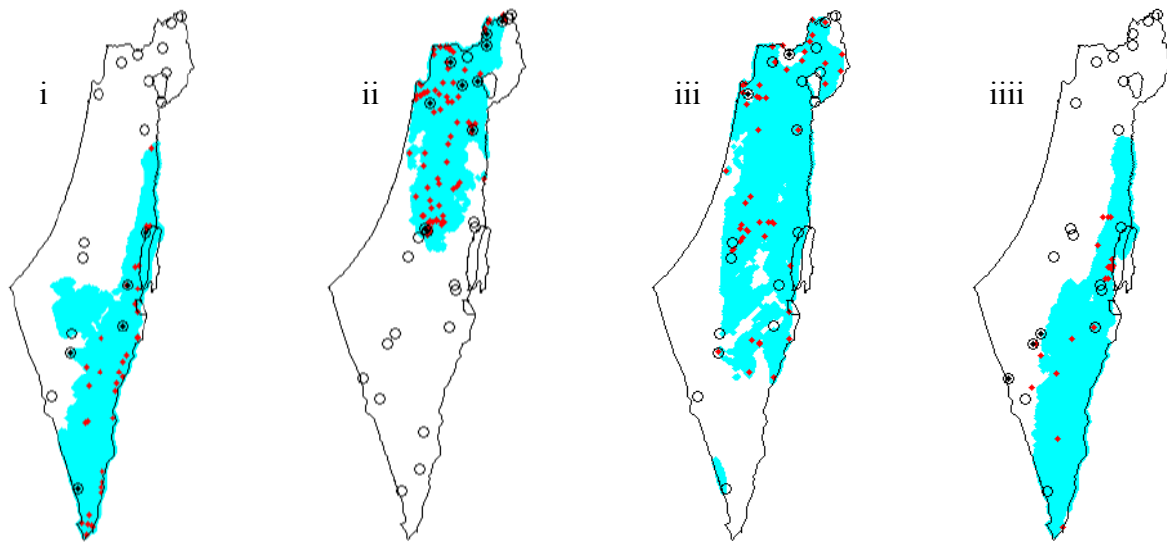
Figure 17. Models predicting species distribution.

Two maps with high prediction of species distribution (Kappa = 1), (i) a map of *Pipistrellus bodenheimeri* (bat fauna), a species inhabiting the desert climate zone. (ii) a map of *Sphincterochila cariosa* (snail fauna), a species inhabiting the Mediterranean climate zone.

Two maps with low prediction of species distribution (Kappa = -0.21), (iii) a map of *Rhinolophus hipposideros* (bat fauna), a species without a distinctive climatic distribution and rare. (iiii) a map of *Emberiza striolata* (bird fauna), a species found in micro-habitats of cliffs and rocky grounds in the desert climate zone.

Red dots are the calibrating dataset, blue is the model prediction, and circles are the validation dataset (open circles: species not found, circles with black dots: species found)

References

Araujo, M. B. & Williams, P. H. (2000) Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96,** 331-345.

Barataud, M. (1996) The world of bats. Sittelle, Mens.

Bibby, C., Jones, M. & Marsden, S. (1998) Bird Survey. Expedition Advisory Center, London.

Brereton, R., Bennett, S. & Mansergh, I. (1995) Enhanced greenhouse climate change and its potential effect of selected fauna of south-eastern australia: a trend analysis. *Biological Conservation*, **72,** 339-354.

Busby, J. R. (1991) BIOCLIM - A bioclimatic analysis and prediction system. *Nature Conservation* (eds. C. R. Margules & M. P. Austin), pp. 64-68. CSIRO, Australia.

Cohen, J. (1960) A coefficient of agreement of nominal scales. *Educational and Psychological Measurement* **20**, 37-46.

Crumpacker, D. W., Box, E. O. & Hardin, E. D. (2001) Implications of climatic warming for conservation of native trees and shrubs in Florida. *Conservation Biology*, **15,** 1008-1020.

Eeley, H. A. C., Lawes, M. J. & Piper, S. E. (1999) The influence of climate change on the distribution of indigenous forest in KwaZulu-Natal, South Africa. *Journal of Biogeography*, **26,** 595-617.

Elith, J. & Burgman, M. (2001) Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. in: *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds. Scott J.M., Heglund P.J., M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall & F. B. Samson), pp. 303-314. Island Press, Washington.

ESRI Inc. ARC/INFO. 7. 1994. Redlands, CA.

Farber, O. (2000) *Analysis, application and enhancement of Bioclimatic models for the prediction of species distribution*. M.Sc. The Hebrew University of Jerusalem.

Farber, O. & Kadmon, R. (2003) Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecological Modelling*, **160,** 115-130.

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.

Guisan, A., Theurillat, J. P. & Kienast, F. (1998) Predicting the potential distribution of plant species in an Alpine environment. *Journal of Vegetation Science*, **9,** 65-74.

Guisan, A. & Zimmermann, N. E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135,** 147-186.

Harrison, D. L. & Bates, P. J. J. (1991) The Mammals of Arabia. Harrison Zoological Museum, Kent.

Herr, A., Nicholas, I. K. & Atkinson, J. S. Identification of bat echolocation calls using a decision tree classification system. Complexity International 4. 1997.

 Hirzel, A., Helfer, V. & Mʾtral, F. (2001) Assessing habitat-suitability models with a virtual species. *Ecological Modelling*, **145,** 111-121.

Honig, M. A., Cowling, R. M. & Richardson, D. M. (1992) The invasive potential of Australian banksias in South African fynbos: A comparison of the reproductive potential of *Banksia ericifolia* and *Leucadendron laureolum*. *Australian Journal of Ecology*, **17,** 305-314.

Hutchinson, G. E. (1957) Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology*. **22**, 415-427.

Jaberg, C. & Guisan, A. (2001) Modelling the distribution of bats in relation to landscape structure in a temperate mountain environment. *Journal of Applied Ecology*, **38,** 1169-1181.

Kadmon, R. & Danin, A. (1999) Distribution of plant species in Israel in relation to spatial variation in rainfall. *Journal of Vegetation Science*, **10,** 421-432.

Kadmon, R., Farber, O. & Danin, A. (2003) A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications*, **13,** 853-867.

Kadmon, R. & Heller, J. (1998) Modelling faunal responses to climatic gradients with GIS: land snails as a case study. *Journal of Biogeography*, **25,** 527-539.

Koopman, K. F. (1975) Bats of the Sudan. *Bulletin of the American Museum of Natural History*, **154,** 353-444.

Kurtzman, D. & Kadmon, R. (1999) Mapping of temperature variables in Israel: a comparison of different interpolation methods. *Climate Research*, **13,** 33-43.

Law, B. S. (1994) Climatic limitation of the southern distribution of the common blossom bat *Syconycteris australis* in New South Wales. *Australian Journal of Ecology*, **19,** 366-374.

Legendre, L. & Legendre, P. (1998) Numerical ecology. Elsevier, Amsterdam.

Limpens, H. J. G. A. The use of bat detectors in identification and survey of bats. 2003 (in preparation).

Makin, D. (1977). The Distribution and Biology of Insectivorous Bats (Microchiroptera) of Israel. M.Sc. Thesis. The Hebrew University of Jerusalem (in Hebrew).

Manel, S., Dias, J. M., Buckton, S. T. & Ormerod, S. J. (1999) Alternative methods for

predicting species distribution: an illustration with Himalayan river birds. *Journal of*

Applied Ecology, **36,** 734-747.

Manel, S., Williams, H. C. & Ormerod, S. J. (2001) Evaluating presence-absence models in

ecology: the need to account for prevalence. *Journal of Applied Ecology*, **38,** 921-931.

Martin, W. K. (1996) The current and potential distribution of the common Myna *Acridotheres*

tristis in Australia. *Emu*, **96,** 166-173.

McKenney, D. W., Mackey, B. G., Bogart, J. P., McKee, J. E., Oldham, M. J. & Chek, A. (1998)

Bioclimatic and spatial analysis of Ontario reptiles and amphibians. *Ecoscience*, **5,** 18-30.

Mendelssohn, H. & Yom-Tov, Y. (1999) Mammalia of Israel. The Israel Academy of Sciences

and Humanities, Jerusalem.

Monserud, R. A. & Leemans, R. (1992) Comparing global vegetation maps with the Kappa

statistic. *Ecological Modelling*, **62,** 275-293.

Nix, H. A. (1986) A biogeographic analysis of Australian elapid snakes. *Atlas of Elapid snakes*

of Australia pp. 4-15. Australian Government Publications Service, Canberra.

O'Farrell, M. J. & Gannon, W. L. (1999) A comparison of acoustic versus capture techniques for

the inventory of bats. *Journal of Mammalogy*, **80,** 24-30.

Palmeirim, J. (1990) Bats of Portugal: zoogeography and systematics. The University of Kansas,

Lawrence.

Parsons, S. & Jones, G. (2000) Acoustic identification of twelve species of echolocating bat by

discriminant function analysis and artificial neural networks. *Journal of Experimental*

Biology, **203,** 2641-2656.

Pearce, J., Ferrier, S. & Scotts, D. (2001) An evaluation of the predictive performance of distributional models for flora and fauna in north-east New South Wales. *Journal of Environmental Management*, **62,** 171-184.

Pearce, J. & Lindenmayer, D. (1998) Bioclimatic analysis to enhance reintroduction biology of the endangered helmeted honeyeater (*Lichenostomus melanops cassidix*) in southeastern Australia. *Restoration Ecology*, **6,** 238-243.

Primack, R. B. (1998) Essentials of Conservation Biology. Sinauer Associates, Sunderland.

Qumsiyeh, M. B. (1985) The Bats of Egypt. Texas Tech University Press, Lubbock.

Qumsiyeh, M. B. (1996) Mammals of the Holy Land. Texas Tech University Press, Lubbock.

Robertson, M. P., Peter, C. I., Villet, M. H. & Ripley, B. S. (2003) Comparing models for predicting species' potential distributions: a case study using correlative and mechanistic predictive modelling techniques. *Ecological Modelling*, **164,** 153-167.

Rotem, D. (2003) *Predicting species distribution of land birds in Israel using logistic regression*. M.Sc. Thesis. The Hebrew University of Jerusalem (in preparation).

Russ, J. (1999) The Bats of Britain and Ireland. Alana Books, Shropshire.

Russo, D. & Jones, G. (2002) Identification of twenty-two bat species (Mammalia : Chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. *Journal of Zoology*, **258,** 91-103.

Schober, W. & Grimmberger, E. (1997) The Bats of Europe and North America. T.F.H. Publications.

Shalmon, B. (1993) A Field Guide to the Land Mammals of Israel: their tracks and signs. Keter, Jerusalem (in Hebrew).

Shao, G. F. & Halpin, P. N. (1995) Climatic controls of eastern north American coastal tree and shrub distributions. *Journal of Biogeography*, **22,** 1083-1089.

Sindel, B. M. & Michael, P. W. (1992) Spread and potential distribution of *Senecio madagascariensis* Poir. (fire weed) in Australia. *Australian Journal of Ecology*, 21-26.

Skidmore, A. K., Gauld, A. & Walker, P. A. (1996) Classification of kangaroo habitat distribution using three GIS models. *Int.Journal of GIS*, **10,** 441-454.

Sokal, R. R. & Rohlf, F. J. (1995) Biometry. W.H. Freeman, New York.

Steinitz, O. (2003) *Predicting patterns of species similarity using environmental and geographical distances*. M.Sc. Thesis. The Hebrew University of Jerusalem.

Stockwell, D. R. B. & Peterson, A. T. (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, **148,** 1-13.

Taylor, P. J. (2000) Bats of southern Africa. University of Natal Press, Pietermaritzburg.

Tupinier, Y. (1997) European Bats: their world of sound. Sittelle, Lyon.

Vaughan, N., Jones, G. & Harris, S. (1997) Habitat use by bats (Chiroptera) assessed by means of a broad- band acoustic method. *Journal of Applied Ecology*, **34,** 716-730.

Walker, P. A. & Cocks, K. D. (1991) HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecology and Biogeography Letters*, **1,** 108-118.

Wardhaugh, A. A. (1999) Bats of the British Isles. Shire natural history. Buckinghamshire.

Waters, D. A. & Jones, G. (1995) Echolocation call structure and intensity in 5 species of insectivorous bats. *Journal of Experimental Biology*, **198,** 475-489.

Wiens, J. A. (2001) Predicting species occurrences: progress, problems and prospects. in: *Predicting Species Occurrences: Issues of Accuracy and Scale* (eds. Scott J.M., Heglund

P.J., M. L. Morrison, J. B. Haufler, M. G. Raphael, W. A. Wall & F. B. Samson), Island Press, Washington.

Zar, J. H. (1999) Biostatistical Analysis. Prentice Hall, New Jersey.

Appendix – Species used in this work

Bat fauna

Species name	OBS	NW	ACN	Fauna
Asellia tridens	37	0.76	0.11	Bats
Eptesicus bottae	38	0.35	0.26	Bats
Myotis nattereri	26	0.69	0.05	Bats
Otonycteris hemprichi	25	0.48	0.21	Bats
Pipistrellus bodenheimeri	36	0.33	0.26	Bats
Pipistrellus kuhli	146	0.92	0.89	Bats
Plecotus austriacus	34	0.92	0.21	Bats
Rhinolophus ferrumequinum	68	0.68	0.42	Bats
Rhinolophus hipposideros	49	0.61	0.11	Bats
Rhinopoma hardwickei	71	0.59	0.32	Bats
Rhinopoma microphyllum	33	0.50	0.16	Bats
Tadarida teniotis	91	0.71	0.63	Bats
Taphozous nudiventris	25	0.54	0.21	Bats

Snail fauna

Species name	OBS	NW	ACN	Fauna
Buliminus glabratus	28	0.24	0.07	Snail
Buliminus labrosus	181	0.51	0.26	Snail
Buliminus therinus	72	0.37	0.04	Snail
Eopolita protensa	110	0.61	0.41	Snail
Euchondrus albulus	81	0.41	0.26	Snail
Euchondrus chondriformis	50	0.29	0.22	Snail
Euchondrus saulcyi	52	0.54	0.15	Snail
Euchondrus septemdentatus	189	0.89	0.41	Snail
Granopupa granum	36	0.42	0.56	Snail
Levantina caesareana	202	0.87	0.44	Snail
Levantina hierosolyma	164	0.75	0.41	Snail
Metafruticicola fourousi	47	0.49	0.22	Snail
Monacha haifaensis	185	0.61	0.44	Snail
Paramastus episomus	91	0.58	0.22	Snail
Pene sidoniensis	125	0.65	0.26	Snail
Rupestrella rhodia	34	0.53	0.11	Snail
Sphincterochila cariosa	78	0.70	0.37	Snail
Sphincterochila fimbriata	116	0.52	0.07	Snail
Sphincterochila prophetarum	85	0.51	0.26	Snail
Sphincterochila zonata	131	0.41	0.30	Snail
Trochoidea langloisiana	61	0.50	0.04	Snail
Trochoidea tuberculosa	66	0.43	0.07	Snail
Xeropicta vestalis	292	0.79	0.48	Snail

Bird fauna

Species name	OBS	NW	ACN	Fauna
Alectoris chukar	914	0.98	0.81	Birds
Ammomanes deserti	159	0.52	0.43	Birds
Ammoperdix heyi	222	0.54	0.33	Birds
Anthus similis	29	0.86	0.19	Birds
Athene noctua	307	0.80	0.05	Birds
Cercomela melanura	202	0.52	0.14	Birds
Cisticola juncidis	39	0.61	0.14	Birds
Columba livia	148	0.93	0.43	Birds
Corvus corone	338	0.71	0.38	Birds
Corvus rhipidurus	63	0.39	0.10	Birds
Corvus ruficollis	211	0.50	0.29	Birds
Dendrocopos syriacus	204	0.79	0.33	Birds
Emberiza striolata	25	0.39	0.14	Birds
Francolinus francolinus	90	0.57	0.05	Birds
Galerida cristata	374	0.93	0.48	Birds
Garrulus glandarius	366	0.90	0.29	Birds
Nectarinia osea	237	0.77	0.48	Birds
Oenanthe leucopyga	82	0.49	0.10	Birds
Parus lugubris	25	0.41	0.05	Birds
Parus major	248	0.77	0.48	Birds
Passer domesticus	295	1.00	0.19	Birds
Prinia gracilis	320	0.78	0.33	Birds
Pterocles coronatus	65	0.42	0.05	Birds
Pycnonotus xanthopygos	552	0.79	0.57	Birds
Scotocerca inquieta	120	0.76	0.33	Birds
Sitta neumayer	25	0.66	0.05	Birds
Streptopelia decaocto	263	0.74	0.71	Birds
Streptopelia senegalensis	198	0.90	0.10	Birds
Strix butleri	46	0.51	0.05	Birds
Turdoides squamiceps	118	0.49	0.05	Birds

השוואה של יכולת חיזוי בין מודלים ביו-אקלימיים המשמשים לניבוי דגמי תפוצה של מינים

מאת: אסף צוער

חיבור לקבלת תואר מוסמך באקולוגיה

מוגש לחוג לאבולוציה, סיסטמטיקה ואקולוגיה

האוניברסיטה העברית בירושלים

מנחים: פרופ' רונן קדמון
פרופ' הנס-אולריך שניצלר

תשרי, תשס"ד   ירושלים   נובמבר 2003

תודות:

לפרופ' רונן קדמון, שלימד אותי חשיבה מדעית מה היא, ושדלתו תמיד היתה פתוחה בפני.

ל Prof. Dr. Hans-Ulrich Schnitzler ול Dr. Annette Denzinger, שלימדו אותי את סוד שפת העטלפים.

ל Prof. Brock M. Fenton על כל הייעוץ ועזרה בשדה.

לאנשי העטלפים, שעזרו רבות בשדה ואיפשרו שימוש בנתוניהם: ד"ר דויד מייקין, ד"ר בני שלמון, פרופ' יורם יום-טוב, עמית דולב, עמרם צברי, ד"ר כרמי קורין, ד"ר יוחאי כרמל ורחלי פלדמן.

לאנשי המוזאונים, האוספים ומאגרי המידע: ד"ר לינדה וויטקאר, פרופ' יוסף הלר, הנק מיניס, פרופ' תמר דיין, צילה שריב, אריה לנדסמן, תרצה שטרן, David L. Harrison, Paul J. J. Bates, Richard Harbord, Daphne Hills, Paula Jenkins.

לכל אלו שעזרו לי בדיגומי השדה: יוסי יובל, עמיחי גוטר, צפנת פרוקצימר, הילה גיורא גיל ושרון תמיר, יהודה מירון, איתי שני, זיו יקוטיאלי, עירית שטייניץ וגלעד דנון.

לביולוגים המחוזיים ולפקחים של רשות הטבע והגנים, על הסיוע הרב בעבודת השדה.

לחברי למעבדה: עופר שטייניץ, דותן רותם, ד"ר דן מלקינסון, ד"ר שימרית להב-גינות, אנה טרכטנברוט, טל ומרב סיפן ומרים בלמקר.

לאורן פרבר, על שהכניסני לרזי ה MATLAB.

תודה חמה ומיוחדת לעדי בן-נון, על העזרה ב GIS והסבלנות להקשיב לקיטורים...

לפרופ' דני כהן וד"ר סלעית קרק, על הייעוץ לאורך הדרך. לפרופ' יעקב לורך ולד"ר נחמה בן-אליהו על העזרה בכתיבה.

תודה חמה ומיוחדת לפרופ' עוזי מוטרו, על הייעוץ הסטטיסטי ובעיקר על החיוך והמילה הטובה...

להורי שרה וחיים צוער ולאחותי הילה תמיר, על העזרה והעידוד לאורך כל הדרך.

לדנה, על הכל....

תקציר

מפות תפוצה של מינים מהוות כלי חשוב לממשק שמירת טבע. על מנת ליצור מפות תפוצה יש צורך בבסיס ידע נרחב בנוגע למין, בכל אזור תפוצתו, דבר הדורש משאבים רבים ובדרך כלל אינו אפשרי. אפשרות חלופית, מעשית יותר, היא להעריך (לנבא) את תפוצת המינים באמצעות מודלים מרחביים שנשענים על תיאוריית הנישה. לפי תיאוריה זו לכל מין דרישות סביבתיות המייחדות אותו, הדרושות לו לשם קיום. אם נזהה את הדרישות הייחודיות למין נוכל להעריך את תחום הנישה הפוטנציאלי שלו ולמפות אותה.

בספרות האקולוגית הוצעו מספר מודלים ביו-אקלימיים הנשענים על תורת הנישה. מודלים אלו מזהים בעזרת נתוני תפוצה של המין את דרישות הנישה האקלימית ומציגות אותה באמצעות מפה. למרות השימוש הרב של מודלים כאלו במחקר, מעט מאוד ניסיונות נעשו להשוות את יכולת הניבוי של מודלים שונים ולזהות את הקשר שבין יכולת הניבוי של המודל ותכונות המינים והנתונים הזמינים לפיתוחו. בעבודה זו בדקתי שלושה מודלים ביו-אקלימיים שונים (Mahalanobis ו BIOCLIM, Habitat) באמצעות נתוני תפוצה על 66 מינים שמייצגים שלוש קבוצות סיסטמטיות: עטלפי חרקים, ציפורים מקננות יציבות ושבלולי יבשה. בדקתי גם השפעה של פרמטרים הקשורים לביולוגיה של המין עצמו (שכיחות הופעה, רוחב נישה ומספר תצפיות במאגר המידע) על יכולת הניבוי של המודלים. לשם הערכת יכולת הניבוי של המודלים נדגמו אתרים ברחבי הארץ באופן בלתי תלוי למאגר המידע שבעזרתו כוילו המודלים.

טיב הניבוי של המודלים שנבדקו נמצא דומה לאלו שהוזכרו בספרות. למעט מודל ה-BIOCLIM בעטלפים, לא נמצא הבדל מובהק בין המודלים השונים. כמו כן לא נמצאו הבדלים בין הקבוצות הסיסטמתיות השונות, זאת בניגוד לתוצאות מעבודות קודמות. מודל ה BIOCLIM נתן תחזית טובה כאשר קוצצו 10% של התצפיות הקיצוניות מנתוני הכיול, אולם בשל מיעוט תצפיות במאגר העטלפים הורדה של 10% של התצפיות הקיצוניות פגעה בטיב הניבוי.

ניתוח הפרמטרים המתייחסים לתכונות המינים הראה שיש קשר הדוק בין נדירות המין לפרמטר Kappa. יחס זה מוסבר בשל היות Kappa תלוי ביחס תדירות המצאות/אי-המצאות של המין במאגר הנתונים. לא נמצא קשר בין מספר התצפיות ליכולת הניבוי של המודלים השונים, אך הסיבה לכך יכולה להיות העובדה שרק מינים בעלי מספר תצפיות גבוה מ-25 נכללו בניתוח. מחקרים קודמים הראו שמעבר ל 50 תצפיות אין שיפור משמעותי ביכולת הניבוי של המודל.

ההבדל העיקרי בין המודלים היה במידת טעות הניבוי של המצאות המין באתרים בהם הוא אינו קיים (False present). מודל ה-BIOCLIM חוזה תחום תפוצה גדול יחסית של המצאות המין, בשל יצירת מעטפת אקלים מלבנית המתעלמת מקורלציות בין משתני האקלים. מודל ה-Mahalanobis מראה טעות מסוג זה באופן מצומצם יותר, בשל אופייה האליפטי של המעטפת האקלימית ורגישותה לקורלציות בין המשתנים.

בחירת המודל ליצירת מפת תפוצה תלוי בצורך שלשמו מייצרים את המפה; אם יש צורך לוודא שרוב תחום התפוצה האמיתי של המין אכן מופיע במפה, וזאת על חשבון ניבוי יתר של המצאות המין, רצוי להשתמש במודל ה BIOCLIM. לעומת זאת, אם יש צורך בוודאות גבוהה יותר שבתחום הניבוי המין אכן קיים, מומלץ להשתמש במודל מסוג ה-Mahalanobis. שיקול נוסף הוא אם כלל האתרים בהם נצפה המין צריכים להיכלל בתחום מפת הניבוי. אם שיקול זה חשוב, יש להשתמש במודל BIOCLIM או Habitat.