Species distribution models do not account for abundance: the case of arthropods on Terceira Island

Alberto Jiménez-Valverde[1,2,]*, Francisco Diniz[1], Eduardo B. de Azevedo[3] & Paulo A. V. Borges[1]

[1] *Azorean Biodiversity Group (CITA-A), Universidade dos Açores, Dep. de Ciências Agrárias, Terra-Chã, 9700-851 Angra do Heroísmo, Terceira, Açores, Portugal (*corresponding author's e-mail: ajvalv@ku.edu)*

[2] *Present address: Natural History Museum and Biodiversity Research Center, The University of Kansas, Lawrence, Kansas 66045, USA*

[3] *CMMG (Centro de Estudos do Clima, Meteorologia e Mudanças Globais), Universidade dos Açores, Dep. de Ciências Agrárias, Terra-Chã, 9700-851 Angra do Heroísmo, Terceira, Açores, Portugal*

The objective of this study is to investigate whether presence/absence models can be used as surrogates of arthropod abundance, and eventually under which circumstances such surrogacy is guaranteed. Presence/absence data for 48 arthropod species from Terceira Island were modelled using artificial neural networks. Probabilities of occurrence were correlated with abundance data from a standardized arthropod survey programme. Although a tendency was found for vagile species to show relationships, only nine species showed significant positive correlations between probability of presence and abundance. Five of these were exotic spider species with high abundances and wide distributions in several human-modified habitats. The patchy distribution of pristine habitats, the capacity to reach them and the probable low dependence on limiting resources, other than food, enhance the relationship. A lack of significant correlations for the majority of the species may be due to historical events, inappropriate scale, demographic controls of density, or the incapacity of presence/absence models to account for environmental suitability. The difficulty to identify *a priori* the species for which the relationship will hold advises against the use of species distribution models as surrogates of arthropod abundance.

Introduction

Species occurrence and abundances are key variables in modern ecological and conservation sciences (*see* Gaston & Blackburn 2000 and references therein). Moreover, understanding spatial and temporal patterns of distribution and abundance is critical to assess the threat status of species (Regan *et al.* 2000, Conrad *et al.* 2006). The abundance of a species is ultimately a function of its birth and death rates, which depend on its fitness in different environments (Boyce

& McDonald 1999, Pearce & Ferrier 2001, Tyre *et al*. 2001, but *see* Van Horne 1983). Besides, abundance is closely related to persistence, a property of the population that is desirable to maximize in reserve selection schemes (Mangel & Tier 1994, Araújo & Williams 2000). It follows that predicting population densities across its distributional range is of great value for conservation and management purposes, in order to assess the impact of environmental changes on organisms (Boyce & McDonald 1999, Joseph *et al*. 2006, Smith *et al*. 2006).

However, the logistical difficulties of obtaining abundance data in many locations are particularly relevant for arthropods. Many projects are currently compiling information on species distributions (e.g. GBIF at www.gbif.org, ATLANTIS www.azoresbioportal.angra.uac.pt; Borges 2005, Hortal *et al*. 2007), but spatial data from these databases have no associated abundance data, which makes them at least incomplete. In fact, presence data are much easier to survey, and are available in numerous biological databases (Soberón & Peterson 2004, Hortal *et al*. 2007). This is one of the reasons why predictive distribution models have developed rapidly in the last decade (Scott *et al*. 2002). By quantifying species–environment relationships and prediction of species' geographic distributions from confirmed occurrences, they assume a direct link between the probability of presence and environmental suitability (Pearce & Ferrier 2001, Peterson 2006). Such models can be used to develop management decisions and conservation strategies (Schadt *et al*. 2002, Barbosa *et al*. 2003, Hirzel *et al*. 2004, Russell *et al*. 2004), among other practical and theoretical applications (*see* Peterson 2006 for a review). Their usefulness would be greatly strengthened if a relationship between local probability of occurrence and abundance were to exist. Moreover, several authors have recently pointed out that relating probability of occurrence to abundance could be a powerful way of validating presence/absence models (Albert & Thuiller 2008, Lobo *et al*. 2008).

A positive relationship between the probability of occurrence and local abundance is expected when there is a strong relation between population densities and environmental variables (Pearce & Ferrier 2001). Moreover, the relationship will be enhanced if resources are limiting and patchy and if the densities of the populations are at their upper limit, as at low densities, other factors apart from environment may be determining the size of the populations (Mitchell 2005, Nielsen *et al*. 2005). However, factors other than environment will make species reach high densities in places with low probability of occupancy (Van Horne 1983, Tyre *et al*. 2001, Nielsen *et al*. 2005). For example, biotic interactions such as aggregation, intraspecific competition or predator pressure can force individuals to concentrate in unsuitable sink marginal habitats. Thus, the relationship between probability of occurrence and local abundance is a complex phenomenon that depends on many different factors acting in isolation or in synergy. However, little empirical work has been undertaken to study the capacity of occurrence models to account for local abundance. To our knowledge, only two studies specifically address the question: (i) Pearce and Ferrier (2001) tested the capacity of logistic presence/absence models to predict population densities of several species of vertebrates and vascular plants, but found little correlation between the probability of presence and abundance among occupied sites; (ii) likewise, Nielsen *et al*. (2005), studying a fern and a moose, obtained disappointing results regarding probability of presence as a surrogate of abundance, attributing their negative results to unmeasured factors, such as site history, that ultimately determine abundance. Recently, in a study not focused on this question, Real *et al*. (2009) found significant relationships between environmental favourability and abundance in two vertebrates in Spain, the Iberian lynx and its prey, the European rabbit. All these results are inconclusive and no inference about the circumstances, if any, under which probability of occurrence acts as a surrogate for local abundance can be deduced.

In this study, we use information derived from a replicated, intensive arthropod sampling scheme on Terceira Island (Azores) in several habitats (*see* Borges *et al*. 2005a, 2006). We evaluated the degree to which environmental suitability assessed with presence/absence models account for abundance estimates. By comparing numerous arthropod species with diverse biological characteristics, we tried to assess the con-

ditions under which the relationship holds. We hypothesize that the wider the distributions and the higher the global abundance, the stronger the relation will be. Thus, we expect the relationship to be enhanced in the case of vagile species able to colonize all suitable habitats. As occurrence models can have two different interpretations (potential versus actual distributions), depending on the kind of distribution data and predictors used (Soberón & Peterson 2005, Jiménez-Valverde *et al.* 2008), we developed two types of models to assess if there are any differences in their relation with abundance.

Data and methods

Biological data

For this study we used a large dataset of epigean arthropod species sampled with pitfall traps in a standardised, large-scale biodiversity study in the Azores (BALA Project — **B**iodiversity of **A**rthropods from the **L**aurisilva of the **A**zores; for more details *see* Borges *et al.* 2005a, 2006, Gaston *et al.* 2006, Hortal *et al.* 2006). Pitfall traps measure activity densities rather than true abundances of species (Luff 1975). Thus, for the same species, they can be used to compare between-sites relative number of individuals and assume that they are a direct estimate of between-sites relative true abundances (Curtis 1980, Downie *et al.* 2000). As in Borges *et al.* (2006), only data from the most intensively surveyed island, Terceira, were included in the analysis. Fieldwork was conducted between 1999 and 2005, during the summer months following a gradient of anthropogenic modification from pristine, high-elevation native forest to low-elevation anthropogenic intensive pasture (Borges *et al.* 2008, Cardoso *et al.* 2009). At each location ($N = 81$), one 150×5 m transect was used to record species occurrences (presence or absence data) and estimate their local abundances as the number of recorded individuals, assuming that, as pointed out before, these intraspecific between-sites relative numbers are direct indicators of the intraspecific between-sites relative size of local densities. A total of 30 pitfall traps were used in each transect (*see* also Borges *et al.*

2005a, 2006). Data were referred to 500×500 m UTM grids. In cases where two or more transects fell in the same grid (8 cases, 67 final UTM grids with abundance data), the mean abundance for each species was used. BALA transects provided abundance data (Table 1) as well as presence/absence information. In addition, we collated presences of each species in 500×500 m grids from the GIS-based ATLANTIS database (*see* Borges 2005 and www.azoresbioportal.angra.uac.pt) to increase their number for occurrence models.

We chose a 500×500 m spatial resolution as we consider it to represent an appropriate balance between the incidences of the fauna on BALA transects (both presence/absence as well as relative abundances for each species among cells) and minimization of the effects of other variables (competition, stochasticity, etc.) on population abundances and species presences. A lower resolution (larger cell size) for such a small territory as Terceira Island would suppose the loss of a large amount of variability in the predictors and a drastic reduction in sample size. A 500×500 m spatial resolution has been previously used in other works with arthropods in the Azores (e.g., Hortal *et al.* 2005).

We discarded species from the BALA (Borges *et al.* 2005a, 2006) and ATLANTIS (Borges 2005, www.azoresbioportal.angra.uac.pt) databases if they were: (i) recorded in < 10 localities with non-zero abundances; (ii) recorded in < 10 localities with abundances = 0 (real absences in real distribution models, *see* below); or (iii) with < 10 probable absences (absences in potential distribution models, *see* below). From an initial pool of 207 species, 48 were selected for further analysis (*see* Appendix).

Predictor variables

Landscape heterogeneity is a key factor in determining species distribution, abundance and diversity. We used volcanic, topographic, climatic and land-use variables to model the real and potential distribution of species (*see* below). As in our GIS database the original spatial resolution of raster maps was at 100×100 m, all maps in the database were resampled to a 500×500 m resolution.

Table 1. Characteristics of the training data for the potential distribution models (sample size and prevalence), accuracy statistics (sensitivity, specificity and AUC) and results of the correlation analysis between probability of occurrence and abundance (r_s and p values).

Species	Sample size	Prevalence	Sensitivity	Specificity	AUC	r_s	p
Araneae							
Agyneta decora	380	0.10	1.00	0.99	1.00	−0.45	0.1058
Dysdera crocata	195	0.63	0.98	0.97	1.00	**0.37**	**0.0141**
Eperigone bryantae	138	0.49	0.97	0.97	1.00	0.42	0.1083
Eperigone fradeorum	82	0.40	0.97	0.98	1.00	0.60	0.0515
Erigone atra	386	0.10	0.97	0.97	0.99	−0.01	0.9827
Erigone autumnalis	135	0.43	0.97	0.96	1.00	**0.45**	**0.0162**
Erigone dentipalpis	143	0.41	0.98	0.99	1.00	**0.58**	**0.0020**
Ero furcata	402	0.11	0.98	0.98	0.99	0.36	0.1284
Lepthyphantes acoreensis	450	0.10	0.98	0.98	1.00	–	–
Meioneta fuscipalpis	260	0.10	0.96	0.97	0.99	−0.31	0.3217
Oedothorax fuscus	105	0.70	0.99	1.00	1.00	**0.41**	**0.0258**
Ostearius melanopygius	153	0.40	0.98	0.98	1.00	0.27	0.3215
Pachygnatha degeeri	244	0.13	0.97	0.97	1.00	0.31	0.2613
Palliduphantes schmitzi	410	0.10	0.98	0.98	0.99	–	–
Pardosa acoreensis	456	0.27	0.98	0.99	1.00	0.26	0.0682
Porrhomma borgesi	230	0.10	1.00	0.99	0.99	−0.37	0.2122
Prinerigone vagans	91	0.60	0.98	0.97	1.00	**0.46**	**0.0455**
Rugathodes acoreensis	380	0.10	1.00	1.00	1.00	0.08	0.7509
Blattaria							
Zetha vestita	102	0.44	0.98	0.98	0.98	−0.15	0.5636
Chilopoda							
Geophilus truncorum	150	0.10	0.93	0.96	0.99	0.34	0.3650
Coleoptera							
Aleochara bipustulata	118	0.76	0.76	0.77	0.82	−0.08	0.7781
Amischa analis	165	0.50	0.97	0.98	1.00	−0.01	0.9545
Anisodactylus binotatus	154	0.69	0.99	1.00	1.00	−0.03	0.9271
Calosoma olivieri	280	0.52	0.94	0.95	0.96	−0.03	0.9233
Chaetocnema hortensis	300	0.10	0.97	0.97	1.00	0.06	0.8362
Cordalia obscura	257	0.51	0.99	0.99	1.00	0.33	0.1792
Epuraea biguttata	111	0.32	0.94	0.95	0.99	−0.46	0.1834
Heteroderes azoricus	590	0.11	0.98	0.98	1.00	−0.35	0.2703
Hirticomus quadriguttatus	240	0.10	1.00	1.00	1.00	0.40	0.2224
Ocypus (Pseudocypus) aethiops	128	0.60	0.90	0.90	0.94	−0.07	0.6769
Pseudoophonus rufipes	226	0.45	0.99	0.99	1.00	−0.28	0.4024
Pterostichus (Argutor) vernalis	69	0.43	0.97	0.97	0.99	−0.21	0.4909
Rugilus orbiculatus orbiculatus	233	0.18	0.98	0.98	1.00	0.20	0.4493
Sericoderus lateralis	126	0.33	0.98	0.98	1.00	**0.63**	**0.0485**
Sitona discoideus	98	0.27	0.96	0.96	0.99	0.02	0.9518
Sphenophorus abbreviatus	153	0.18	0.93	0.94	1.00	0.54	0.0862
Stelidota geminata	170	0.10	0.94	0.94	0.99	0.25	0.3620
Typhaea stercorea	460	0.10	0.98	0.98	1.00	−0.02	0.9390
Xantholinus longiventris	549	0.15	0.99	0.99	1.00	0.40	0.1402
Dermaptera							
Forficula auricularia	307	0.38	0.99	0.99	1.00	0.00	0.9968
Diplopoda							
Cylindroiulus propinquus	547	0.15	0.97	0.97	1.00	**0.38**	**0.0236**
Polydesmus coriaceus	115	0.37	0.98	0.97	1.00	0.18	0.3402
Hemiptera							
Anoscopus albifrons	385	0.12	0.98	0.97	0.99	0.20	0.3377
Aphrodes hamiltoni	380	0.10	0.97	0.98	1.00	0.07	0.6726
Nabis pseudoferus ibericus	73	0.71	0.96	0.96	0.99	0.11	0.7136
Pseudoescropiones							
Chthonius ischnocheles	180	0.53	0.94	0.94	0.99	−0.12	0.5039
Chthonius tetrachelatus	127	0.14	0.72	0.73	0.80	0.37	0.1782
Opiliones							
Homalenotus coriaceus	124	0.31	0.97	0.98	1.00	**0.51**	**0.0172**

Five categories of maps were included, and were used as predictors:

i.   A digital elevation model (DEM) and derived maps. A raster DEM was developed by interpolating the elevation curves of the Digital Chart developed by the Cartographic Service of the Portuguese Army. We extracted mean and minimum elevation. The standard deviation of elevation in each 500×500 m grid was calculated as a measure of topographic ruggedness.

ii. Climate data obtained from the CIELO Model (Azevedo 1996, Azevedo *et al*. 1999), developed to simulate local climate in island environments. CIELO is a physically based model that simulates the climatic variables on an island using data from the synoptic reference of a meteorological station. The domain of computation is a raster GIS parameterized with a DEM (*see* also Borges *et al*. 2006 for an application of such data). The CIELO model has been calibrated and validated for Terceira Island, and is now available through the CLIMAAT project (CLIMAAT Interreg_IIIB, MAC 2.3/A3, Azevedo 2003). The variables extracted were annual potential evapotranspiration, mean annual temperature and mean annual precipitation.

iii. Volcanic island map from the work of Victor Hugo Forjaz and colleagues (Forjaz 2004) that describes in detail all the volcanic systems of Terceira and all other Azorean islands. A categorical map was created with a total of 25 volcanic systems.

iv. Concerning land-use cover, we used the recently available habitat classification for Terceira Island (F. Dinis unpubl. data). The main land-uses are native forest (mainly composed of Laurisilva forest), semi-natural grassland, intensive pasture, exotic forest, urban areas, vines/orchards, abandoned fields and unproductive areas. The percentage of each land use in each 500×500 grid cell was computed.

v.   Two additional layers were obtained from the Digital Chart developed by the Cartographic Service of the Portuguese Army: presence of water sources (rivers and lakes) and distance to urban areas.

The strong environmental gradient of Terceira Island, from high (wet and cold) to low (dry and hot) elevations, makes topographic and climatic variables highly correlated. To simplify these data, a principal components analysis (PCA) was constructed using the standardized normal scores of mean and minimum altitude and the three climate variables. PCA analysis selected one component (herein called climatic factor) with an eigenvalue higher than 1 (4.8) which explained 96.5% of total variance in the climatic and topographic data. While the climatic factor describes the broad-scale environmental gradient of the island, the other variables refer to characteristics of the environment that do not show such broad spatial patterns (topographic ruggedness, presence of water sources), to historic events (volcanic systems) or to land use/habitat type (land-use cover, presence of water sources). Thus, the climatic factor was used to characterise the Grinnellian niche of the species (*sensu* Soberón 2007) and to produce estimations of their potential distributions; the other variables were used to estimate the real distributions.

Modelling presence/absence data

Presence/absence data were modelled using artificial neural networks (ANN), a machine-learning method able to deal with non-linear relationships between the dependent variable and the predictors and able to account for complex data (Olden *et al*. 2008, Özesmi *et al*. 2006a, Recknagel 2001, Venables & Ripley 2002). ANNs have proved to outperform other conventional techniques in species distribution modelling studies (e.g. Olden & Jackson 2002, Segurado &Araújo 2004).

In this study, a Multi-Layer Perceptron (feed-forward neural network with back propagation for training, Isasi Viñuela & Galván León 2004) with one hidden layer was used (Fig. 1). Neurons are organized in different layers: an input layer, a hidden layer and an output layer. The input layer has a number of neurons equal to the number of independent variables, while the output layer has only one neuron. The hidden layer contains 15 neurons to achieve a trade off between predictive accuracy and overfitting (A. Jiménez-Val-
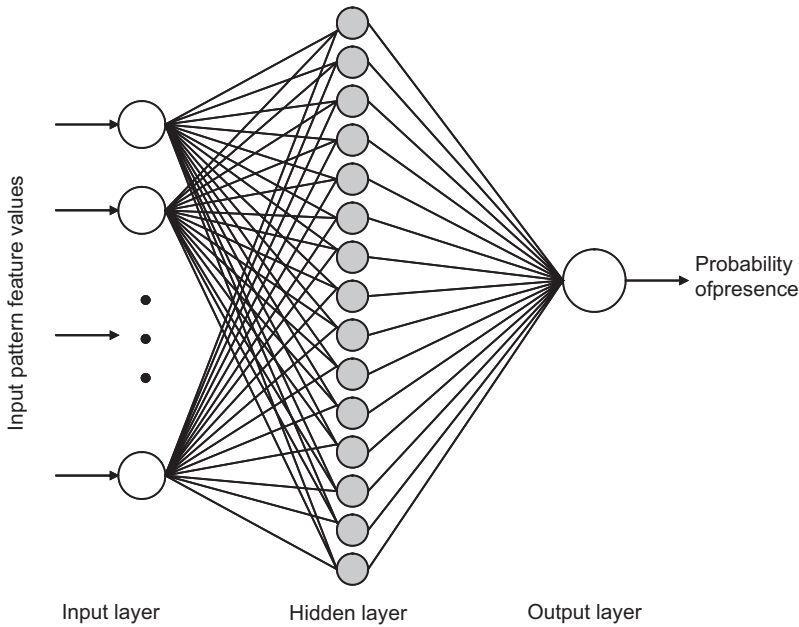
Fig. 1. Architecture of the Multi-Layer Perceptron used in this study. In the case of the potential distribution, one neuron forms the input layer; for real distributions 13 neurons formed the input layer.

verde unpubl. data). Each neuron is connected by axons to all neurons of the next layer in a unidirectional way, and the intensity of the signal is defined by the weight assigned to each connexion. Thus, the state of the hidden neurons is defined by the weighted sum of the inputs and an activation non-linear function (logistic function in this study). Initial weights were set to 0.1. In the training process, weights are changed iteratively (maximum number of iterations = 2000) to minimize the least-square error function. Thus, a Multi-Layer Perceptron with one hidden layer basically performs a linear combination of sigmoid function of the inputs. For further details *see*, for example, Ripley (1996), Isasi Viñuela and Galván León (2004) and Özesmi *et al*. (2006b). Neural networks were fitted in R (R Development Core Team 2004) using the nnet package (Venables & Ripley 2002).

Two kinds of models were developed for every species, each using a different type of absence information and different predictor variables to account for both extremes of the potential-actual distribution gradient (*see* Soberon & Peterson 2005 and Jiménez-Valverde *et al*. 2008 for an extensive conceptual framework):

1. In order to account for the potential distribution (climate-related potentiality), we tabu-

lated maximum and minimum values of the climatic factor at presence points (presences from the BALA survey and ATLANTIS database), so defining the climate envelope for each of the 48 species (Busby 1991, Lobo *et al*. 2006). From the area outside the envelope, probable absences were randomly selected to generate training datasets with prevalence as low as 0.1 (i.e., number of presences/[number of presences + number of absences] = 0.1) for each species if possible (Jiménez-Valverde & Lobo 2006b, 2007a). In the modelling process, only the climatic factor was used as predictor, as including other variables would imply accounting for factors that restrict potential distributions to realized ones (*see* above).

2. To account for the real distribution, presences from the BALA survey and ATLANTIS database and absences obtained in the BALA transects were used (real absences). In this case, all 13 (including the climatic factor) variables were used in the modelling process to introduce historical and anthropogenic factors in the models (*see* above).

Model evaluation

Models were evaluated using a 'leave-one-out'

jackknife approach: that is, one observation was excluded, the model parameterized with the remaining $n - 1$ observations, a predicted probability is obtained for the excluded observation, and the procedure is repeated n times (Olden *et al*. 2002). With these new jackknife probabilities, the area under the ROC curve (AUC), a measure of overall discriminatory power (Swets 1988, Fielding & Bell 1997), was computed (but *see* recent criticisms of this measure in Lobo *et al*. 2008 and Peterson *et al*. 2008). Also, sensitivity (presences correctly predicted) and specificity (absences correctly predicted) were calculated using the threshold which minimizes their difference (Jiménez-Valverde & Lobo 2006c, 2007b). Notice that potential distribution models will unavoidably show high values of discrimination capacity (*see* Jiménez-Valverde & Lobo 2007a). All evaluation computations were run in R (R Development Core Team 2004) using the ROCR package (Sing *et al*. 2005) and custom scripts.

Relationship between probability of presence and abundance

Spearman rank correlations between predicted probabilities of presence (from potential and real distribution models) and observed abundance data from locations with values > 0 were calculated. Positive correlations including 0 data points indicating absence would mean good discrimination capacity of presence/absence models, but not necessarily good capacity to explain variation in abundance. In total, 48×2 independent correlation analyses were performed. No α-adjustment method for multiple tests was applied, to avoid increasing type II errors (the probability of accepting the null hypothesis when the alternative is true; Perneger 1998, Moran 2003, Gotelli & Ellison 2004). Then, we studied if there was a relationship between Spearman correlation coefficients (r_s) and p values with mean abundances (*see* Appendix) and the number of occupied sites by each species, as well as if there was any difference in r_s scores and p values between Araneae (high vagile taxa) and Coleoptera (low vagile taxa; note that the concepts of high and low vagility are relative).

Results

All models of the potential distribution showed high predictive power, with small variation in accuracy measures (Table 1). The mean discrimination AUC score was 0.99, being higher than 0.8 in all cases, which indicates that the models are highly discriminatory (Swets 1988). This means that, in the worst model, a presence will be predicted as having a higher probability than an absence 80% of the time. Sensitivity and specificity were quite high, with mean values of 0.96, higher than 0.72 in all cases. Only eight models showed probabilities of presence significantly correlated with abundance (*Dysdera crocota*, *Erigone autumnalis*, *Erigone dentipalpis*, *Oedothorax fuscus*, *Prinerigone vagans*, *Sericoderus lateralis*, *Homalenotus coriaceus*, and *Cylindroiulus propinquus*) at $p < 0.05$, only one of them significant at $p < 0.01$ (*Erigone dentipalpis*). The discrimination capacity of the models does not affect the relationship as r_s scores were not correlated with AUC values ($r = 0.08^{ns}$). No significant correlation was found between neither mean abundance nor the number of occupied sites and r_s scores ($r = 0.14^{ns}$ and $r = 0.19^{ns}$, respectively). P values were not correlated with mean abundance or the number of occupied sites either ($r = -0.20^{ns}$ and $r = -0.22^{ns}$, respectively). Mean r_s scores were higher in Araneae (mean ± SD = 0.214 ± 0.333) than in Coleoptera (0.068 ± 0.299) but the difference was not statistically significant ($t = 1.371$, $p = 0.180$). P values were significantly lower in Araneae than in Coleoptera (Mann-Whitney *U*-test: $Z = -2.924$, $p = 0.004$).

On the contrary, real distribution predictions were not as accurate as the potential ones and showed higher variation in accuracy measures (Table 2). The mean AUC value was 0.74, with 33 species showing AUC scores ≥ 0.70, indicating useful models (Swets 1988). Mean values of sensitivity and specificity were both 0.70. Only *Erigone autumnalis* and *Geophilus truncorum* showed significant correlations between probability of presence and abundance at $p < 0.05$, only one of them significant at $p < 0.01$ (*Erigone autumnalis*). Again, the discrimination capacity of the models does not affect the relationship as r_s scores were not correlated with AUC values ($r = -0.25^{ns}$). No significant correlations

Table 2. Characteristics of the training data for the real distribution models (sample size and prevalence), accuracy statistics (sensitivity, specificity and AUC) and results of the correlation analysis between probability of occurrence and abundance (r_s and p values).

Species	Sample size	Prevalence	Sensitivity	Specificity	AUC	r_s	p
Araneae							
Agyneta decora	85	0.45	0.63	0.64	0.68	−0.27	0.3408
Dysdera crocata	145	0.85	0.62	0.64	0.67	−0.17	0.2686
Eperigone bryantae	112	0.60	0.61	0.62	0.65	0.35	0.1856
Eperigone fradeorum	90	0.37	0.61	0.61	0.67	−0.06	0.8537
Erigone atra	81	0.48	0.49	0.48	0.50	0.40	0.1584
Erigone autumnalis	96	0.60	0.67	0.68	0.76	**0.63**	**0.0003**
Erigone dentipalpis	97	0.60	0.72	0.72	0.78	0.12	0.5650
Ero furcata	86	0.53	0.61	0.60	0.59	−0.08	0.7356
Lepthyphantes acoreensis	82	0.55	0.76	0.76	0.81	0.15	0.4707
Meioneta fuscipalpis	80	0.33	0.77	0.78	0.81	0.54	0.0696
Oedothorax fuscus	100	0.74	0.65	0.65	0.72	0.19	0.3135
Ostearius melanopygius	111	0.54	0.80	0.80	0.82	0.03	0.9183
Pachygnatha degeeri	84	0.37	0.84	0.83	0.87	−0.32	0.2424
Palliduphantes schmitzi	83	0.49	0.66	0.67	0.70	0.02	0.9420
Pardosa acoreensis	135	0.90	0.57	0.57	0.62	0.20	0.1670
Porrhomma borgesi	74	0.31	0.65	0.67	0.75	0.05	0.8799
Prinerigone vagans	103	0.53	0.73	0.73	0.78	0.34	0.1535
Rugathodes acoreensis	75	0.51	0.84	0.84	0.89	−0.01	0.9565
Blattaria							
Zetha vestita	73	0.62	0.71	0.71	0.78	−0.33	0.1935
Chilopoda							
Geophilus truncorum	66	0.23	0.53	0.51	0.56	**0.70**	**0.0254**
Coleoptera							
Aleochara bipustulata	129	0.68	0.81	0.80	0.85	−0.05	0.8562
Amischa analis	116	0.69	0.75	0.75	0.85	−0.09	0.7015
Anisodactylus binotatus	152	0.69	0.73	0.72	0.80	0.02	0.9562
Calosoma olivieri	186	0.77	0.76	0.76	0.85	−0.51	0.0918
Chaetocnema hortensis	73	0.41	0.70	0.70	0.75	0.14	0.6386
Cordalia obscura	162	0.78	0.80	0.81	0.90	0.19	0.4480
Epuraea biguttata	85	0.42	0.69	0.69	0.73	0.16	0.4731
Heteroderes azoricus	109	0.55	0.95	0.96	0.99	−0.30	0.3458
Hirticomus quadriguttatus	74	0.32	0.92	0.92	0.97	0.24	0.4789
Ocypus (Pseudocypus) aethiops	91	0.85	0.44	0.43	0.41	0.19	0.2349
Pseudoophonus rufipes	150	0.67	0.78	0.78	0.84	−0.12	0.7269
Pterostichus (Argutor) vernalis	74	0.41	0.70	0.70	0.76	−0.16	0.6069
Rugilus orbiculatus orbiculatus	85	0.51	0.70	0.69	0.75	0.00	0.9962
Sericoderus lateralis	95	0.43	0.80	0.81	0.82	0.29	0.4082
Sitona discoideus	70	0.37	0.65	0.66	0.69	0.19	0.5450
Sphenophorus abbreviatus	79	0.35	0.71	0.73	0.76	0.60	0.0509
Stelidota geminata	61	0.28	0.53	0.52	0.50	−0.21	0.4408
Typhaea stercorea	93	0.49	0.91	0.91	0.97	−0.19	0.5257
Xantholinus longiventris	120	0.68	0.71	0.71	0.74	0.20	0.4630
Dermaptera							
Forficula auricularia	157	0.74	0.80	0.80	0.88	0.23	0.4303
Diplopoda							
Cylindroiulus propinquus	96	0.83	0.75	0.75	0.78	0.03	0.8744
Polydesmus coriaceus	67	0.64	0.49	0.50	0.48	0.12	0.5081
Hemiptera							
Anoscopus albifrons	76	0.61	0.43	0.43	0.46	0.29	0.1563
Aphrodes hamiltoni	56	0.68	0.84	0.83	0.94	0.06	0.7301
Nabis pseudoferus ibericus	94	0.52	0.67	0.67	0.73	−0.03	0.9140
Pseudoescropiones							
Chthonius ischnocheles	114	0.82	0.63	0.65	0.68	0.00	0.9770
Chthonius tetrachelatus	62	0.29	0.67	0.66	0.66	−0.01	0.9742
Opiliones							
Homalenotus coriaceus	73	0.53	0.64	0.65	0.71	0.39	0.0775

were found between neither mean abundance nor the number of occupied sites and r_s scores ($r = 0.07^{ns}$ and $r = 0.04^{ns}$, respectively). P values were not correlated with mean abundance or the number of occupied sites either ($r = -0.13^{ns}$ and $r = -0.08^{ns}$, respectively). Mean r_s scores were higher in Araneae (mean ± SD = 0.117 ± 0.262) than in Coleoptera (0.031 ± 0.254), although the difference was not statistically significant ($t = 1.016$, $p = 0.317$). P values were not significantly lower in Araneae than in Coleoptera (Mann-Whitney U-test: $Z = -0.881$, $p = 0.378$).

Discussion

General trends

As in previous analyses (Pearce & Ferrier 2001, Nielsen *et al*. 2005), for most species analysed, we did not find a significant relationship between local probability of occurrence and local abundance. However, for eight exotic and one native species (*see* Appendix), we found probabilities of presence significantly correlated with density. No relation with mean abundance or number of occupied sites was found, but there is a tendency for vagile species (spiders) to show more apparent relationships between probability of occurrence and abundance. In fact, five out of nine significant cases were spiders. Spiders show a great dispersal capacity, as suggested by the fact that due to passive air dispersal using silk threads (ballooning), they can be found in the air at high altitudes (Freeman 1946) and are among the first colonizers of islands (Edwards & Thornton 2001). This characteristic, together with being generalist predators (Wise 1993), may result in spiders being in a higher degree of equilibrium with the environment than other taxa, something that, as explained in more detail in the next section, would enhance the relationship between probability of occurrence and local abundance.

Concrete cases

The five spider species for which significant correlations were found are widely distributed in anthropogenic Azorean habitats, namely inten-

sive pasture (*see* Borges & Brown 2001, 2004). The five spider species are well adapted pasture-dwelling spiders, at least four of which (*Prinerigone vagans*, *Erigone dentipalpis*, *Erigone autumnalis* and *Oedothorax fuscus*) normally occur in heavily grazed pastures (De Keer & Maelfait 1988, Borges & Brown 2001). Borges and Brown (2001) showed that in Azorean intensive and semi-natural pastureland, diverse and structurally complex pasture sites supported more individual web-building spiders per unit area than less complex ones. In fact, physical complexity of the grassland and sub-shrub vegetation layers has been pointed out as one of the most important deterministic factors of spider distribution, especially for web-building species (Grill *et al*. 2005, Jiménez-Valverde & Lobo 2007c and references therein). Probably the most direct effect of habitat complexity is the availability of a greater number of microniches (Jiménez-Valverde & Lobo 2007c). Azorean pastures are usually organized into small (around 100 × 100 m) stone-fenced fields grazed regularly (every three weeks in the summer months) by milk cattle. Therefore, the landscape is organized in a mosaic of recently grazed fields either with or without cows and with high structural heterogeneity of vegetation, thus providing suitable conditions in terms of habitat structure throughout most of Terceira. Moreover, Erigonidae is the family most common in aeroplankton samples (Dean & Sterling 1985, Bishop 1990), indicating a huge capacity for long-distance dispersal and, thus, for reaching and establishing in pastures with high habitat complexity. This, together with the fact that all the spiders were sampled between grazing periods when the grass reaches its highest density, suggests that habitat structure is not a limiting factor in spider populations. Also, pastures are highly productive ecosystems (Curry 1994), so these spider species live in habitats with high food availability, probably indirectly related to the climatic gradient of the island. These factors suggest that the patchy distribution of resources in the Azorean mosaic pastureland favours the strength of the relationship between occurrence and abundance for these habitat specialist species (*see* also Nielsen *et al*. 2005), together with the fact that no other factor except food availability appears to be a major determinant of their abun-

dance at this scale (*see* Wise 1993: chapter 2). In fact, significant relationships are mostly found when modelling the potential distribution, i.e., when only the climatic factor is used as explanatory variable and so the probability of occurrence may reflect climatic favourability.

Besides the tendency observed for spiders, four other arthropod species pertaining to different taxa and with different characteristics showed positive relationships between the probability of occurrence and abundance. *Cylindroiulus propinquus* (Diplopoda) and *Homalenotus coriaceus* (Opiliones) are common arthropod species in several Azorean native and altered habitats, but not particularly common in intensive pastures. *Geophilus truncorum* (Chilopoda) is mainly a forest species and *Sericoderus lateralis* (Coleoptera) is one of the most common beetles in most habitats in the Azores.

Why is the probability of occurrence not predicting abundance?

Despite the generally observed trend for spiders to show a positive relationship between predicted occurrence and abundance, the probability of occurrence was a poor predictor of local abundance for most species. In fact, for other species with the same characteristics as the ones for which significant relations have been found, we observed no capacity to account for abundance in their occurrence models. Instead, other variables, such as site history, scale or species life history may have influenced the spatial variation in local abundance of Azorean epigean arthropod species. That many other arthropod species have wide distributions on the island suggests that their local abundance arises from non-environmental controls of density (Van Horne 1983), rather than on their ability to reach and occupy potential habitats.

The adequate sampling of arthropods in space and time is a very hard task (Jones and Eggleton 2000, Jiménez-Valverde & Lobo 2006a, Cardoso *et al.* 2007). Additionally, usually only one-year abundance estimations are available for a number of sites. However, arthropod populations are known to fluctuate markedly over time (Stewart-Oaten *et al.* 1995, Heller *et al.* 2006).

These fluctuations may not depend on environmental factors but rather on demographic ones, including demographic stochasticity, finally altering occupancy patterns (Hanski 1999) and may dilute the probability of a significant occurrence-density relationship in many species (Nielsen *et al.* 2005). Thus, it could be argued that, although the relationship between presence/abundance exists for many other species under ideal conditions, difficulties in the estimation of density results in the lack of recognition of the pattern. However, we do not believe this methodological drawback to be determining our results as the model of abundance–variance–occupancy of He and Gaston (2003) fits very well to the same abundance used in our study (Gaston *et al.* 2006), supporting the robustness of the data.

All these considerations are based on the assumption that occurrence models reflect habitat suitability. First, source–sink dynamics (*sensu* Shmida & Wilson 1985) explains how a species can occur in unsuitable places, maintaining populations due to immigration (Shmida & Wilson 1985, Pulliam 2000). Relocation to suboptimal habitats can be observed if the suitable environment is scarce (Braunisch *et al.* 2008). Additionally, dispersal limitation can explain how absences can be found in environmentally suitable places (Pulliam 2000). Second, and more importantly from a conceptual point of view, the modelling objective changes depending on the nature of the absence and the predictors used (Soberón & Peterson 2005, Jiménez-Valverde *et al.* 2008). When climatic variables and probable absences located outside the envelope defined by those predictors are used, the models are probably forecasting the potential niche concept in a Grinnellian sense (our potential distribution approach; Soberón 2007). However, if real absences and/or variables that account for historical or human disturbance events are used, then the models may approach the real geographic distribution (our real distribution approach). If, as in the case of the Azorean arthropod fauna, species are highly influenced by land-use variables, then these variables will force the algorithm in such a way that the climatic factors lose relevance and climatic suitability may be diluted. That is, we would be generating a geographical range map rather than an environmental suitability map. As

a consequence, the relationship between probability of occurrence and abundance will be lost, as our results for potential vs. real distribution models indicate. Besides, even potential distribution models based on non-causal correlative relationships are hardly able to describe the fundamental niche of the species (Kearney & Porter 2009). Hence, their interpretation in terms of environmental suitability may be questioned.

Conclusions

Generally, the spatial distribution of individuals is fundamental to understanding macroecological patterns (Gaston 1996). Our analyses have shown that, although there is a tendency for spiders (vagile species and generalist predators) to show relationships between probability of occurrence and local abundance, this is weak and there is no general applicability of the pattern. Species for which the relation holds are difficult to identify *a priori*, so using environmental suitability (i.e., probability of presence) as a surrogate for arthropod relative local abundance estimates should be ruled out; abundance is a quite variable parameter which depends on many factors apart from environment. Consequently, despite the relatively easy access of occurrence data and the straightforward application of distribution modelling techniques, the use of occurrence models as density indicators of the relative population size of arthropod species cannot be generalized. Finally, although the way occurrence models are built is critical as it will determine the concept being modelled, i.e., environmental potential *vs.* realized geographical distribution, they may not reflect environmental suitability.

References

Albert, C. & Thuiller, W. 2008: Favourability functions versus probability of presence: advantages and misuses. — *Ecography* 31: 417–422.

Araújo, M. B. & Williams, P. H. 2000: Selecting areas for species persistence using occurrence data. — *Biological Conservation* 96: 331–345.

Azevedo, E. B. 1996: *Modelação do Clima Insular à Escala Local. Modelo CIELO aplicado à ilha Terceira.* — Ph.D. thesis, Universidade dos Açores, Angra do Heroísmo.

Azevedo, E. B. 2003: *Projecto CLIMAAT — Clima e Meteorologia dos Arquipélagos Atlanticos.* — PIC Interreg_IIIB-Mac 2.3/A3.

Azevedo, E. B., Pereira, L. S. & Itier, B. 1999: Modelling the local climate in island environments: water balance applications. — *Agricultural Water Management* 40: 393–403.

Barbosa, A. M., Real, R., Olivero, J. & Vargas, J. M. 2003: Otter (*Lutra lutra*) distribution modeling at two resolution scales suited to conservation planning in the Iberian Peninsula. — *Biological Conservation* 114: 377–387.

Bishop, L. 1990: Meteorological aspects of spiders ballooning. — *Environmental Entomology* 19: 1381–1387.

Borges, P. A. V. 2005: Introduction. — In: Borges, P. A. V., Cunha, R., Gabriel, R., Martins, A. M. F., Silva, L. & Vieira, V. (eds.), *A list of the terrestrial fauna (Mollusca and Arthropoda) and flora (Bryophyta, Pteridophyta and Spermatophyta) from the Azores*: 11–20. Direcção Regional de Ambiente e Universidade dos Açores, Horta, Angra do Heroísmo and Ponta Delgada.

Borges, P. A. V. & Brown, V. K. 2001: Phytophagous insects and web-building spiders in relation to pasture vegetation complexity. — *Ecography* 24: 68–82.

Borges, P. A. V. & Brown, V. K. 2004: Arthropod community structure in pastures of an island archipelago (Azores): looking for local-regional species richness patterns at small-scales. — *Bulletin of Entomological Research* 94: 111–121.

Borges, P. A. V., Aguiar, C., Amaral, J., Amorim, I. R., André, G., Arraiol, A., Baz, A., Dinis, F., Enghoff, H., Gaspar, C., Ilharco, F., Mahnert, V., Melo, C., Pereira, F., Quartau, J. A., Ribeiro, S., Ribes, J., Serrano, A. R. M., Sousa, A. B., Strassen, R. Z., Vieira, L., Vieira, V., Vitorino, A. & Wunderlich, J. 2005a: Ranking protected areas in the Azores using standardized sampling of soil epigean arthropods. — *Biodiversity and Conservation* 14: 2029–2060.

Borges, P. A. V., Vieira, V., Dinis, F. & Jarroca, S. 2005b: List of Arthropods (Arthropoda). — In: Borges, P. A.

V., Cunha, R., Gabriel, R., Martins, A. M. F., Silva, L. & Vieira, V. (eds.), *A list of the terrestrial fauna (Mollusca and Arthropoda) and flora (Bryophyta, Pteridophyta and Spermatophyta) from the Azores*: 163–221. Direcção Regional de Ambiente and Universidade dos Açores, Horta, Angra do Heroísmo and Ponta Delgada.

Borges, P. A. V., Lobo, J. M., Azevedo, E. B., Gaspar, C., Melo, C. & Nunes, L. V. 2006: Invasibility and species richness of island endemic arthropods: a general model of endemic vs. exotic species. — *Journal of Biogeography* 33: 169–187.

Borges, P. A. V., Ugland, K. I., Dinis, F. O. & Gaspar, C. 2008: Insect and spider rarity in an oceanic island (Terceira, Azores): true rare and pseudo-rare species. — In: Fattorini, S. (ed.), *Insect ecology and conservation*: 47–70. Research Signpost.

Boyce, M. S. & McDonald, L. L. 1999: Relating populations to habitats using resource selection functions. — *Trends in Ecology and Evolution* 14: 268–272.

Braunisch, V., Bollmann, K., Graf, R. F. & Hirzel, A. H. 2008: Living on the edge — modelling habitat suitability for species at the edge of their fundamental niche. — *Ecological Modelling* 214: 153–167.

Busby, J. R. 1991: BIOCLIM: A bioclimate analysis and prediction system. — In: Margules, C. R. & Austin, M. P. (eds.), *Nature conservation: cost effective biological surveys and data analysis*: 64–68. CSIRO, Australia.

Cardoso, P., Lobo, J. M., Calvo, S., Dinis, F., Gaspar, C. & Borges, P. A. V. 2009: A spatial scale assessment of habitat effects on arthropod communities of an oceanic island. — *Acta Oecologica* 35: 590–597.

Cardoso, P., Silva, I., Oliveira, N. G. & Serrano, A. M. 2007: Seasonality of spiders (Araneae) in Mediterranean ecosystems and its implications in the optimum sampling period. — *Ecological Entomology* 32: 516–526.

Conrad, K. F., Warren, M. S., Fox, M., Parsons, M. S. & Woiwod, I. P. 2006: Rapid declines of common, widespread British moths provide evidence of an insect biodiversity crisis. — *Biological Conservation* 132: 279–291.

Curry, J. P. 1994: *Grassland invertebrates — ecology, influence on soil fertility and effects on plant growth*. — Chapman & Hall, London.

Curtis, D. J. 1980: Pitfalls in spider community studies (Arachnida: Araneae). — *Journal of Arachnology* 8: 271–280.

Dean, D. A. & Sterling, W. L. 1985: Size and phenology of ballooning spiders at two locations in eastern Texas. — *Journal of Arachnology* 13: 111–120.

De Keer, R. & Maelfait, J.-P. 1988: Observations on the life cycle of *Erigone atra* Araneae, Erigoninae in a heavily grazed pasture. — *Pedobiologia* 32: 201–212.

Downie, I. S., Ribera, I., McCracken, D. I., Wilson, W. L., Foster, G. N., Waterhouse, A., Abernethy, V. J. & Murphy, K. J. 2000: Modelling populations of *Erigone atra* and *E. dentipalpis* (Araneae: Linyphiidae) across an agricultural gradient in Scotland. — *Agriculture, Ecosystems and Environment* 80: 15–28.

Edwards, J. S. & Thornton, W. B. 2001: Colonization of an island volcano, Long Island, Papua New Guinea, and an emergent island, Motmot, in its caldera lake. VI. The pioneer arthropod community of Motmot. — *Journal of*

Biogeography 28: 1379–1388.

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.

Forjaz, V. H. (ed.) 2004: *Atlas Básico dos Açores*. — OVGA — Observatório Vulcanológico dos Açores, Ponta Delgada.

Freeman, J. A. 1946: The distribution of spiders and mites up to 300 feet in the air. — *Journal of Animal Ecology* 15: 69–74.

Gaston, K. J. 1996: The multiple forms of the interspecific abundance–distribution relationship. — *Oikos* 75: 211–220.

Gaston, K. J. & Blackburn, T. M. 2000: *Pattern and process in macroecology*. — Blackwell Science, London.

Gaston, K. J., Borges, P. A. V., He, F. & Gaspar, C. 2006: Abundance, spatial variance and occupancy: arthropod species distribution in the Azores. — *Journal of Animal Ecology* 75: 646–656.

Gotelli, N. J. & Ellison, A. M. 2004: *A primer of ecological statistics*. — Sinauer Associates, Inc., MA, USA.

Grill, A., Knoflach, B., Cleary, D. F. R. & Kati, V. 2005: Butterfly, spider, and plant communities in different land-use types in Sardina, Italy. — *Biodiversity and Conservation* 14: 1281–1300.

Hanski, I. 1999: *Metapopulation ecology*. — Oxford University Press, Oxford.

He, F. & Gaston, K. J. 2003: Occupancy, spatial variance, and the abundance of species. — *American Naturalist* 162: 366–375.

Heller, N. E., Sanders, N. J. & Gordon, D. M. 2006: Linking spatial and temporal scales in the study of an Argentine ant invasion. — *Biological Invasions* 8: 501–507.

Hirzel, A. H., Posse, B., Oggier, P.-A., Crettenand, Y., Glenz, C. & Arlettaz, R. 2004: Ecological requirements of a reintroduced species, with implications for release policy: the bearded vulture recolonizing the Alps. — *Journal of Applied Ecology* 41: 1103–1116.

Hortal, J., Borges, P. A. V., Dinis, F., Jiménez-Valverde, A., Chefaoui, R. M., Lobo, J. M., Jarroca, S., Azevedo, E. B., Rodrigues, C., Madruga, J., Pinheiro, J., Gabriel, R., Cota Rodrigues, F. & Pereira, A. R. 2005: Using ATLANTIS — Tierra 2.0 and GIS environmental information to predict the spatial distribution and habitat suitability of endemic species. — In: Borges, P. A. V., Cunha, R., Gabriel, R., Martins, A. M. F., Silva, L. & Vieira, V. (eds.), *A list of the terrestrial fauna (Mollusca and Arthropoda) and flora (Bryophyta, Pteridophyta and Spermatophyta) from the Azores*: 69–113. Direcção Regional de Ambiente and Universidade dos Açores, Horta, Angra do Heroísmo and Ponta Delgada.

Hortal, J., Borges, P. A. V. & Gaspar, C. 2006: Evaluating the performance of species richness estimators: Sensitivity to sample grain size. — *Journal of Animal Ecology* 75: 274–287.

Hortal, J., Lobo, J. M. & Jiménez-Valverde, A. 2007: Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife (Canary Islands). — *Conservation Biology* 21: 853–863.

Isasi Viñuela, P. & Galván León, I. M. 2004: *Redes de Neuronas Artificiales. Un enfoque práctico*. — Pearson

Educación, Madrid.

Jiménez-Valverde, A. & Lobo, J. M. 2006a: Establishing reliable spider (Araneae, Araneidae & Thomisidae) assemblage sampling protocols: estimation of species richness, seasonal coverage and contribution of juvenile data to species richness and composition. — *Acta Oecologica* 30: 21–32.

Jiménez-Valverde, A. & Lobo, J. M. 2006b: Distribution determinants of endangered Iberian spider *Macrothele calpeiana* (Araneae, Hexathelidae). — *Environmental Entomology* 35: 1491–1499.

Jiménez-Valverde, A. & Lobo, J. M. 2006c: The ghost of unbalanced species distribution data in geographic model predictions. — *Diversity and Distributions* 12: 521–524.

Jiménez-Valverde, A. & Lobo, J. M. 2007a: Potential distribution of the endangered spider *Macrothele calpeiana* (Walckenaer, 1805) (Araneae, Hexathelidae) and the impact of climate warming. — *Acta Zoologica Sinica* 53: 865–876.

Jiménez-Valverde, A. & Lobo, J. M. 2007b: Threshold criteria for conversion of probability of species presence to either-or presence–absence. — *Acta Oecologica* 31: 361–369.

Jiménez-Valverde, A. & Lobo, J. M. 2007c: Determinants of local spider (*Araneidae & Thomisidae*) species richness on a regional scale: climate and altitude *vs*. habitat structure. — *Ecological Entomology* 32: 113–122.

Jiménez-Valverde, Lobo, J. M. & Hortal, J. 2008: Not as good as they seem: the importance of concepts in species distribution modelling. — *Diversity and Distributions* 14: 885–890.

Jones, D. T. & Eggleton, P. 2000: Termite species assemblages in tropical forests: testing a rapid biodiversity assessment protocol. — *Journal of Applied Ecology* 37: 191–203.

Joseph, L. N., Field, S. A., Wilcox, C. & Possingham, H. P. 2006: Presence–absence *versus* abundance data for monitoring threatened species. — *Conservation Biology* 20: 1679–1687.

Kearney, M. & Porter, W. 2009: Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. — *Ecology Letters* 12: 334–350.

Lobo, J. M., Verdú, J. R. & Numa, C. 2006: Environmental and geographical factors affecting the Iberian distribution of flightless *Jekelius* species (Coleoptera: Geotrupidae). — *Diversity and Distributions* 12: 179–188.

Lobo, J. M., Jiménez-Valverde, A. & Real, R. 2008: AUC: A misleading measure of the performance of predictive distribution models. — *Global Ecology and Biogeography* 17: 145–151.

Luff, M. L. 1975: Some features influencing the efficiency of pitfall traps. — *Oecologia* 19: 345–357.

Mangel, M. & Tier, C. 1994: Four facts every conservation biologist should know about persistence. — *Ecology* 75: 607–614.

Mitchell, S. C. 2005: How useful is the concept of habitat? A critique. — *Oikos* 110: 634–638.

Moran, M. D. 2003: Arguments for rejecting the sequential Bonferroni in ecological studies. — *Oikos* 100: 403–405.

Nielsen, S. E., Johnson, C. J., Heard, D. C. & Boyce, M. S.

2005: Can models of presence-absence be used to scale abundance? Two case studies considering extremes in life history. — *Ecography* 28: 197–208.

Olden, J. D. & Jackson, D. A. 2002: A comparison of statistical approaches for modelling fish species distributions. — *Freshwater Biology* 47: 1976–1995.

Olden, J. D., Jackson, D. A. & Peres-Neto, P. 2002: Predictive models of fish species distributions: a note on proper validation and chance predictions. — *Transactions of the American Fisheries Society* 131: 329–336.

Olden, J. D., Lawler, J. J. & Poff, N. L. 2008: Machine learning methods without tears: a primer for ecologists. — *The Quarterly Review in Biology* 83: 171–193.

Özesmi, U., Tan, C. O. & Özesmi, S. L. 2006b: Methodological issues in building, training, and testing artificial neural networks in ecological applications. — *Ecological Modelling* 195: 83–93.

Özesmi, U., Tan, C. O., Özesmi, S. L. & Robertson, R. J. 2006a: Generalizability of artificial neural network models in ecological applications: predicting nest occurrence and breeding success of the red-winged blackbird *Agelaius phoeniceus*. — *Ecological Modelling* 195: 94–104.

Pearce, J. & Ferrier, S. 2001: The practical value of modelling relative abundance of species for regional conservation planning: a case study. — *Biological Conservation* 98: 33–43.

Perneger, T. V. 1998: What's wrong with Bonferroni adjustments. — *British Medical Journal* 316: 1236–1238.

Peterson, A. T. 2006: Uses and requirements of ecological niche models and related distributional models. — *Biodiversity Informatics* 3: 59–72.

Peterson, A. T., Papeş, M. & Soberón, J. 2008: Rethinking receiver operating characteristic analysis applications in ecological niche modelling. — *Ecological Modelling* 213: 63–72.

Pulliam, H. R. 2000: On the relationship between niche and distribution. — *Ecology Letters* 3: 349–361.

R Development Core Team 2004: *R: A language and environment for statistical computing*. — R Foundation for Statistical Computing, Vienna, Austria, available at www.R-project.org.

Real, R., Barbosa, A. M., Rodríguez, A., García, F. J., Vargas, J. M., Palomo, J. & Delibes, M. 2009: Conservation biogeography of ecologically interacting species: the case of the Iberian lynx and the European rabbit. — *Diversity and Distributions* 15: 390–400.

Recknagel, F. 2001: Applications of machine learning to ecological modelling. — *Ecological Modelling* 146: 303–310.

Regan, H. M., Colyvan, M. & Burgman, M. A. 2000: A proposal for fuzzy International Union for the conservation of Nature (IUCN) categories and criteria. — *Biological Conservation* 92: 101–108.

Ripley, B. D. 1996: *Pattern recognition and neural networks*. — Cambridge University Press, Cambridge.

Russell, K. R., Mabee, T. J. & Cole, M. B. 2004: Distribution and habitat of Columbia torrent salamanders at multiple spatial scales in managed forests of northwestern Oregon. — *Journal of Wildlife Management* 68: 403–415.

Schadt, S., Revilla, E., Wiegand, T., Knauer, F., Kaczensky,

P., Breitenmoser, U., Bufka, L., Červený, J., Koubek, P., Huber, T., Staniša, C. & Trepl, L. 2002: Assessing the suitability of central European landscapes for the reintroduction of Eurasian lynx. — *Journal of Applied Ecology* 39: 189–203.

Scott, J. M., Heglund, P. J., Haufler, J. B., Morrison, M., Raphael, M. G., Wall, W. B. & Samson, F. (eds.) 2002: *Predicting species occurrences*. *Issues of accuracy and scale*. — Island Press, Covelo, CA.

Segurado, P. & Araújo, M. B. 2004: An evaluation of methods for modelling species distributions. — *Journal of Biogeography* 31: 1555–1568.

Shmida, A. & Wilson, M. V. 1985: Biological determinants of species diversity. — *Journal of Biogeography* 12: 1–20.

Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. 2005: *ROCR: Visualizing the performance of scoring classifiers*. — R package ver. 1.0-1, available at rocr.bioinf.mpi-sb.mpg.de.

Smith, R. S., Gaston, K. J., Warren, P. H. & Thompson K. 2006: Urban domestic gardens (VIII): environmental correlates of invertebrate abundance. — *Biodiversity and Conservation* 15: 2515–2545.

Soberón, J. 2007: Grinnellian and Eltonian niches and geographic distribution of species. — *Ecology Letters* 10: 1115–1123.

Soberón, J. & Peterson, A. T. 2004: Biodiversity informatics: managing and applying primary biodiversity data. — *Philosophical Transactions of the Royal Society of London B* 359: 689–698.

Soberón, J. & Peterson, A. T. 2005: Interpretation of models of fundamental ecological niches and species' distributional areas. — *Biodiversity Informatics* 2: 1–10.

Stewart-Oaten, A., Murdoch, W. W. & Walde, S. J. 1995: Estimation of temporal variability in populations. — *American Naturalist* 146: 519–535.

Swets, J. A. 1988: Measuring the accuracy of diagnostic systems. — *Science* 240: 1285–1293.

Tyre, A. J., Possingham, H. P. & Lindenmayer, D. B. 2001: Inferring process from pattern: can territory occupancy provide information about life history parameters? — *Ecological Applications* 11: 1722–1737.

Van Horne, B. 1983: Density as a misleading indicator of habitat quality. — *Journal of Wildlife Management* 47: 893–901.

Venables, W. N. & Ripley, B. D. 2002: *Modern applied statistics with S*. — Springer, New York.

Wise, D. H. 1993: *Spiders in ecological webs*. — Cambridge University Press, New York.

Appendix. Colonization status (END = Azorean endemic; MAC = Macaronesian endemic; N = native; I = introduced; – = no information available; following Borges *et al.* 2005b) and mean abundance (and range) of each species.

Species	Colonization status	Mean abundance (range)	Species	Colonization status	Mean abundance (range)
Araneae			*Cordalia obscura*	I	4.06 (1–24)
Agyneta decora	I	3.21 (1–9)	*Epuraea biguttata*	I	4.00 (1–11)
Dysdera crocata	I	4.14 (1–21)	*Heteroderes azoricus*	END	74.50 (3–204)
Eperigone bryantae	I	4.06 (1–19)	*Hirticomus quadriguttatus*	N	94.64 (3–589)
Eperigone fradeorum	I	5.00 (1–33)	*Ocypus (Pseudocypus) aethiops*	N	10.10 (1–60)
Erigone atra	I	9.29 (1–41)	*Pseudoophonus rufipes*	I	5.73 (1–18)
Erigone autumnalis	I	23.32 (1–69)	*Pterostichus (Argutor) vernalis*	I	14.77 (1–81)
Erigone dentipalpis	I	47.23 (1–121)	*Rugilus orbiculatus orbiculatus*	N	6.41 (1–27)
Ero furcata	I	1.74 (1–9)	*Sericoderus lateralis*	I	2.70 (1–7)
Lepthyphantes acoreensis	END	5.42 (1–18)	*Sitona discoideus*	I	5.00 (1–14)
Meioneta fuscipalpis	I	18.83 (1–68)	*Sphenophorus abbreviatus*	I	2.00 (1–6)
Oedothorax fuscus	I	442.07 (10–1081)	*Stelidota geminata*	I	9.33 (1–58)
Ostearius melanopygius	I	6.00 (1–32)	*Typhaea stercorea*	I	18.85 (1–120)
Pachygnatha degeeri	I	7.93 (1–32)	*Xantholinus longiventris*	I	1.93 (1–6)
Palliduphantes schmitzi	MAC	2.65 (1–16)	**Dermaptera**		
Pardosa acoreensis	END	22.22 (1–188)	*Forficula auricularia*	–	13.29 (1–42)
Porrhomma borgesi	END	1.54 (1–3)	**Diplopoda**		
Prinerigone vagans	I	8.37 (1–37)	*Cylindroiulus propinquus*	I	3.60 (1–10)
Rugathodes acoreensis	END	4.94 (1–17)	*Polydesmus coriaceus*	I	4.83 (1–26)
Blattaria			**Hemiptera**		
Zetha vestita	–	1.47 (1–3)	*Anoscopus albifrons*	–	6.36 (1–40)
Chilopoda			*Aphrodes hamiltoni*	END	5.43 (1–19)
Geophilus truncorum	N	1.11 (1–2)	*Nabis pseudoferus ibericus*	N	4.62 (1–16)
Coleoptera			**Pseudoescropiones**		
Aleochara bipustulata	I	3.00 (1–15)	*Chthonius ischnocheles*	–	5.74 (1–14)
Amischa analis	I	4.80 (1–15)	*Chthonius tetrachelatus*	–	2.73 (1–8)
Anisodactylus binotatus	I	3.00 (1–9)	**Opiliones**		
Calosoma olivieri	N	2.33 (1–8)	*Homalenotus coriaceus*	–	22.00 (1–147)
Chaetocnema hortensis	I	12.00 (1–71)			