Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Ecological Modelling 185 (2005) 299–308
ECOLOGICAL MODELLING
www.elsevier.com/locate/ecolmodel

Species-specific traits associated to prediction errors in bird habitat suitability modelling

Javier Seoane*, Luis M. Carrascal, César Luis Alonso, David Palomino

Department Biodiversidad y Biología Evolutiva, Museo Nacional de Ciencias Naturales, CSIC, C/José Gutiérrez Abascal 2, 28006 Madrid, Spain

Received 19 June 2004; received in revised form 15 December 2004; accepted 17 December 2004

Abstract

Although there is a wide range of empirical models applied to predict the distribution and abundance of organisms, we lack an understanding of which ecological characteristics of the species being predicted affect the accuracy of those models. However, if we knew the effect of specific traits on modelling results, we could both improve the sampling design for particular species and properly judge model performance. In this study, we first model spatial variation in winter bird density in a large region (Central Spain) applying regression trees to 64 species. Then we associate model accuracy to characteristics of species describing their habitat selection, environmental specialization, maximum densities in the study region, gregariousness, detectability and body size.

Predictive power of models covaried with model characteristics (i.e., sample size) and autoecological traits of species, with 48% of interspecific variability being explained by two partial least regression components. There are species-specific characteristics constraining abundance forecasting that are rooted in the natural history of organisms. Controlling for the positive effect of prevalence, the better predicted species had high environmental specialization and reached higher maximum densities. We also detected a measurable positive effect of species detectability. Thus, generalist species and those locally scarce and inconspicuous are unlikely to be modelled with great accuracy. Our results suggest that the limitations caused by those species-specific traits associated with survey work (e.g., conspicuousness, gregariousness or maximum ecological densities) will be difficult to circumvent by either statistical approaches or increasing sampling effort while recording biodiversity in extensive programs.

Keywords: Birds; Density; Habitat distribution models; Interespecific differences; Modelling errors; Regression trees

1. Introduction

The aim of modelling the distribution and abundance of organisms, which roots the modern concept of habitat suitability modelling, is far from new in ecology and conservation biology (MacArthur, 1972; Walter,

* Corresponding author. Present address: Departamento Interuniversitario de Ecología, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain. Tel.: +34 914973639; fax: +34 914978001.
E-mail address: javier.seoane@uam.es (J. Seoane).

0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2004.12.012

1973; Cody, 1985; Caughley and Gunn, 1996). However, practitioners of these disciplines have recently acquired a wide range of modelling methodologies, based on modern statistical approaches that benefit from continuous developments on both geographical information systems and remote sensing (Guisan and Zimmermann, 2000; Scott et al., 2002). Several empirical models to analyze distribution and abundance have been spreading during the last decade, such as generalized additive models (Leathwick, 1998; Lehmann, 1998; Seoane et al., 2004), classification and regression trees (De'Ath and Fabricius, 2000; Dzeroski and Drumm, 2003), neural networks (Lek et al., 1996; Özesmi and Özesmi, 1999; Dedecker et al., 2004) and distance-based models such as ecological niche factor analysis (ENFA) and climatic envelopes (e.g., Hirzel et al., 2002; Pearson and Dawson, 2003; Huntley et al., 2004; Remm, 2004). These techniques have been compared in terms of predictive accuracy and ease of use (Guisan et al., 1999; Manel et al., 1999; Olden and Jackson, 2002; Segurado and Araújo, 2004; Yen et al., 2004), with the conclusion that there is not a single best method. In fact, predictive accuracy varies more among species than among modelling techniques (Elith and Burgman, 2002; Thuiller et al., 2003).

Nevertheless, little is known about whether ecological traits of species may predict the errors in habitat suitability modelling (but see, Boone and Krohn, 1999; Kadmon et al., 2003). For example, among groups of animal species, the success of several modelling techniques relates inversely with spatial variability (mobility and nomadism) and niche width, but there are some effects which are not consistent across all biological groups (Pearce and Ferrier, 2000; Pearce et al., 2001). Similar effects have been found within particular groups of species, with negative effects of niche width and positive effects of commoness, abundance and detectability (Boone and Krohn, 1999; Kadmon et al., 2003). Nevertheless, among-species differences are often less clear-cut (see, for example, Elith and Burgman, 2002, who in a study of vascular plants did not find associations between specific traits and model discrimination ability). The analysis of the association between species biological traits and model accuracy is useful because if we knew the effect of specific traits on modelling results, we could improve the sampling design for some particular species (e.g., modifying survey intensity). We could also know the maximum accuracy attainable with the analytical approach, which would enable us to make informed judgements on model performances.

Birds are a suitable biological group to assess among-species differences in modelling accuracy because they show a wide range of ecological traits and they may be surveyed in large areas. Bird species differ greatly in stenotopy, abundance, geographical range, mobility and detectability, the main factors that could help to explain variation in modelling accuracy. These interspecific differences increase dramatically in winter, when birds are not constrained to a nest site or a fixed territory, and they may gather in nomadic flocks tracking feeding resources (Fretwell, 1972). In this paper, we study the effect of species' autoecological traits on predictability of habitat suitability models, working with wintering birds of Central Spain. First, we model spatial variation in bird density using regression trees. Second, the predictive power of these models are related to biological characteristics of species describing environmental preferences and specialization, maximum densities attained in the study region, gregariousness, detectability and body size.

2. Methods

2.1. Study area

The study area is located in the centre of the Iberian Peninsula, comprising Madrid province and surrounding areas of Segovia and Guadalajara (less than 50 km from the province border). Altitude ranges from 450 to 2450 m a.s.l. The climate is Mediterranean continental, with cold winters near the Guadarrama mountains and milder weather in the valleys of the Tajo basin. There is a wide variety of habitats in this region: autochthonous forests (pinewoods of *Pinus sylvestris* L., riparian woods, deciduous oakwoods of *Quercus pyrenaica* Willd. and evergreen holm-oakwoods of *Q. ilex* subsp. *ballota* [Desf.] Samp.), open wooded habitats (ash and holm-oaks parklands), scrublands, artificial and natural pasturelands, marshlands, rock outcrops, various agricultural formations (vineyards, olive plantations, extensive cereal croplands) and urban areas (from small villages to large cities) (Fig. 1).

Fig. 1. Location of the study area in the Iberian peninsula with the layout of surveys (dots). The outline represents Madrid province.

2.2. Survey data

Bird surveys were performed in wintertime (December, January and the first fortnight of February from 1981 to 2003) by the authors of this paper and were also obtained from published data in the literature (Santos et al., 1983; Potti, 1985a, 1985b; Santos et al., 1985; Monreal, 1986; Carrascal, 1988; Tellería et al., 1988; Carrascal et al., 2000). The survey method was the line transect with survey belts of 25 m at each side of the progression line (Bibby et al., 2000, a line transect of 1 km samples 5 ha). Transects were established throughout the study area covering nearly all habitats in the region. All surveys were carried out within homogeneous areas in windless days without precipitations, between 8:00–11:30 and 15:00–17:00 GMT, at a low speed (1–3 km/h). We gathered data for 77 transects, covering an area from 8 to 500 ha (median = 32 ha). Bird density was expressed in birds/10 ha. Due to sample limitations, statistical analyses were only performed with 64 species appearing in at least five surveys.

Each survey transect was characterized by its geographic location (latitude and longitude), altitude, and seven variables describing habitat structure and floristic composition. An index of structural complexity and vegetation volume (SCI) ranged from 0 to 5: 0—lacking or very sparse vegetation cover; (1) pasturelands; (2) shrublands with sparse vegetation cover made up of bushes lower than 0.5 m; (3) thick shrublands with bushes higher than 0.3 m in height; (4) parklands, narrow riparian woods, hedgerows; (5) dense forests with trees higher than 4 m (mainly > 8 m). Values of 0–1 were used to codify the absence (0) or the presence (1) of the following habitat attributes: agricultural use, urbanization, presence of water (pools, streams), rocky outcrops, coniferous trees, deciduous trees and evergreen trees.

Considering the density of species in the surveys, and the characteristics of the 77 line transects, the weighted means of each species in the variables describing survey plots were obtained. These weighted means will be used to obtain the environmental availability of selected habitats in the study region (see below) and the average complexity of selected habitats by each species.

Differences between species environmental preferences and the availability of those environments in the study region (END from environmental distance) were calculated by the Euclidean distance between the means of explanatory variables (excluding geographical coordinates) using the 77 survey plots (i.e., the availability sample) and the weighted means of each species in the variables describing these 77 survey plots (i.e., the preference sample). Before computing Euclidean distances, each variable was standardized to mean zero and S.D. = 1 (i.e., each variable weighed the same in distance estimations).

The sample of 77 surveys was grouped into 27 different habitat categories considering vegetation structure, floristic composition, altitudinal distribution and human impact. These 27 habitats are representative of the main environments for birds found in the study area, accounting for more than 99% of the area of the region. For each one of these 27 habitats, average density was obtained for the 64 studied species. Habitat breadth (HB) of species was calculated following the Levin's index divided by the number of habitat categories (Levins, 1968):

$$\mathrm{HB} = \frac{\left(\sum_{i=1}^{27} p_i^2\right)^{-1}}{27}$$

where p_i is the proportion of the density for each species measured in the habitat i (dividing density in habitat i

by the sum of all densities in the 27 habitats). This index ranges between 1 (evenly distributed across the 27 habitats) and 1/27 (only present in one habitat).

We also recorded the flock size of bird species in the study region during the winters of 1999–2003. When a species was contacted, we tried to count all individuals observed within a radius of 25 m from the flock centre. By means of this procedure we obtained enough data to make a coarse-grained description of the average group size of the 64 species included in data analyses (sample size for all species ranged between 6 and 108 groups; median = 34 groups).

The surveys used in this paper also counted the birds observed outside the transect belts. An index of lateral detectability was built (Järvinen and Väisänen, 1975), as the ratio between the birds belonging to each species observed inside the transect belt and the total amount of birds observed (i.e., the ratio of main belt to total belt observations). This index reflects important species characteristics related to the interaction with the observer, such as song or call intensity and audibility, conspicuousness, mobility, etc. It ranges between high values for inconspicuous species (e.g., >75% of individuals observed at less than 25 m at both sides of the observer) to low figures for more detectable species (e.g., <5% of observations at less than 25 m; see also Järvinen and Väisänen, 1976; Järvinen, 1978). The data on body mass for the 64 analyzed species were taken from Perrins (1998) (Fig. 1).

2.3. Statistical analyses

Species density (birds/10 ha) in survey samples were analyzed using regression trees (Clark and Pregibon, 1993; De'Ath and Fabricius, 2000). We built models with the geographic position of transects and the habitat characteristics as explanatory variables. Regression trees are of great interest in dealing with complex interactions between variables and can easily handle explanatory variables of different types (categorical, continuous with varying shapes). These models provide a set of dichotomous rules (splits) for dividing the data in homogeneous sets (leafs). To avoid overparameterization (i.e., growing a tree too large), we constrained tree complexity pruning by deviance on trees with a minimum of five samples per leaf. We used tree function of S-Plus 2000 (Clark and Pregibon, 1993).

The predictive power of models was evaluated with a data-splitting strategy (similar to that in Boyce et al., 2002). We obtained a random sample of 80% of original data (77 surveys × 80/100 = 62 surveys). We built a tree regression model with this sample. This model was used to predict densities in the other 'not-used' sample of 20% of the original data (77 surveys; 77 × 20/100 = 15 surveys). The densities actually measured in the test sample of 15 censuses were correlated with those predicted by the model built with the sample of 62 censuses. In each randomisation, we obtained a measurement of the agreement between predictions and observations in the census data, using the Spearman correlation (r_s) between the observed densities and those predicted by each tree regression model in the subsample of 15 testing cases. This process was repeated 50 times. Final estimates of agreement between predicted and observed densities were the average of the 50 Spearman rank correlations.

Interspecific variation in predictive power of regression trees was related to species-specific traits by means of partial least squares regression (hereafter PLSR), using species as the sample unit ($n = 64$). This is an extension of the multiple regression analysis where the effects of linear combinations of several predictors on a response variable can be analyzed. Associations are established with factors extracted from predictor variables that maximize the explained variance in the dependent variable (in our case, the predictive power of tree models). These factors are defined as a linear combination of independent variables, so the original multidimensionality is reduced to a lower number of factors to detect structure in the relationships between predictor variables, and between these factors and the response variable. The extracted factors are orthogonal (i.e., independent of each other) and they account for successive lower proportions of original variance. For more details on this statistical exploratory technique, see StatSoft (2001) and Tobias (2003).

The simplest model explaining the interspecific differences in predictive power of regression trees was obtained by means of stepwise regression analysis. The results were consistent in the forward and backward approaches with $P = 0.05$ as significance criterion to enter or to remove particular effects.

Bird species are evolutionarily related throughout a phylogenetic scheme, and therefore, they should not be treated as independent sample units (Felsenstein, 1985; Harvey and Purvis, 1991). This has been established as a common paradigm in evolutionary ecology research, although it is subjected to controversy and debate (Westoby et al., 1995; Ricklefs and Starck, 1996; Price, 1997). Several authors have pointed out that on many occasions similar results are obtained in phylogenetic and non-phylogenetic analyses (e.g., Price, 1997), and that in some instances ecologists are not interested in patterns of biological diversification across evolutionary time, but only in present day relationships comprising non-evolutionary associations under man-transformed environments. As in our case, we study the relationships among ecological traits of species and their present day distribution derived from inventories in a transformed landscape, we have simplified data analyses avoiding the complexities and drawbacks of comparative methods (i.e., uncertainty about models of evolutionary change, phylogeny topology or branch lengths).

3. Results

Table 1 shows the results of statistical models built with densities of 64 species in the sample of 77 survey transects using regression trees. All models significantly explained the data, in most of the occasions at $P < 0.001$. Original deviance explained varied between 11 and 79% (average = 40.6%). Predictive power of regression trees ranged between −0.02 and 0.74 (average $r_s = 0.363$).

Interspecific variability in predictive power of models covaried with ecological traits of species (Table 2). Forty-eight percent of interspecific variability in predictive power of regression trees models was explained by two partial least regression components. The first one (40.5% of variance) shows that better predicted species have higher prevalences, they have larger differences between their habitat selection and the availability of those environments in the study region (END), and reached higher maximum densities, specially in more structurally complex habitats (i.e., woodlands). The second component (7.5%) relates predictive power

Table 1
Values of specific traits and results of models built for predicting density (birds/10 ha; tree regression models were built using 77 census transects) of 64 bird species

	N	LV	p	D^2 (%)	r_s	D_{max}	END	HB	SCI	DET	FS	W
Aegithalos caudatus	28	6	***	75.0	0.60	10.6	2.4	0.22	4.5	67.1	5.9	7.5
Alauda arvensis	17	4	***	60.0	0.57	38.6	4.0	0.15	0.8	35.6	12.5	38.0
Alectoris rufa	18	6	***	34.0	0.15	3.5	2.3	0.18	2.1	50.3	2.4	525.0
Anthus pratensis	23	5	***	21.0	0.44	28.3	2.3	0.08	1.2	50.8	4.4	18.8
Buteo buteo	10	5	**	24.0	0.43	1.2	2.5	0.07	3.4	3.2	1.2	825.0
Carduelis cannabina	25	5	***	20.0	0.30	36.4	3.1	0.17	1.2	59.7	31.1	17.6
Carduelis carduelis	34	6	***	35.0	0.25	43.7	1.5	0.15	2.5	35.0	13.4	16.0
Carduelis chloris	12	5	***	33.0	0.26	8.1	3.2	0.14	1.9	43.8	9.4	26.5
Carduelis spinus	11	6	***	33.0	0.10	26.0	2.4	0.07	4.3	56.9	11.4	13.2
Certhia brachydactyla	34	6	***	49.0	0.71	11.8	2.3	0.24	4.5	43.3	1.2	8.3
Cettia cetti	13	4	***	44.0	0.69	16.1	3.3	0.07	2.9	68.7	1.1	13.3
Cisticola juncidis	11	4	***	38.0	0.46	2.3	3.1	0.10	2.4	49.4	1.3	8.9
Columba livia	14	6	***	67.0	0.53	17.8	4.6	0.06	1.7	41.9	3.0	270.0
Columba palumbus	24	6	***	29.0	0.43	27.8	3.9	0.10	3.2	13.1	15.5	485.0
Corvus corone	17	7	***	30.0	0.21	3.5	2.2	0.20	3.9	2.1	2.5	570.0
Corvus monedula	12	6	***	31.0	0.12	20.1	2.9	0.17	3.3	35.9	29.6	240.0
Cyanopica cyana	9	6	***	27.0	0.30	9.1	2.9	0.15	3.6	31.9	10.2	72.0
Dendrocopos major	15	7	***	29.0	0.37	1.9	2.4	0.18	5.1	39.6	1.1	80.0
Emberiza cia	23	9	***	36.0	0.04	5.3	1.5	0.30	3.0	56.4	3.5	23.5
Emberiza cirlus	8	5	***	26.0	0.07	2.8	2.2	0.10	2.9	51.4	4.5	25.5
Emberiza schoeniclus	5	3	***	41.0	0.47	6.1	3.3	0.08	1.3	53.7	2.4	18.0
Erithacus rubecula	45	8	***	79.0	0.45	10.4	1.7	0.42	3.3	44.9	1.0	16.7
Falco tinnunculus	6	5	NS	33.0	0.05	0.7	2.6	0.06	2.9	20.0	1.1	235.0
Fringilla coelebs	44	7	***	28.0	0.40	67.7	2.0	0.24	3.6	33.5	6.1	23.0
Galerida cristata	18	7	***	34.0	0.22	8.6	2.4	0.25	1.5	48.5	3.9	41.4

Table 1 (*Continued*)

	N	LV	p	D^2 (%)	r_s	D_{max}	END	HB	SCI	DET	FS	W
Galerida theklae	20	6	***	40.0	0.11	4.8	1.9	0.25	1.7	40.5	2.8	36.8
Garrulus glandarius	7	7	NS	11.0	0.04	0.3	2.1	0.07	3.8	4.8	1.5	174.0
Lanius excubitor	23	9	NS	33.0	0.04	1.0	1.1	0.29	2.5	35.4	1.1	63.5
Loxia curvirostra	9	3	***	45.0	0.70	8.3	4.6	0.06	5.8	20.0	8.8	39.2
Lullula arborea	13	6	***	28.0	0.22	4.4	2.4	0.17	3.6	37.7	3.7	26.1
Melanocorypha calandra	6	4	***	30.0	0.28	16.7	3.8	0.09	0.2	38.6	26.0	65.0
Miliaria calandra	16	5	***	30.0	0.23	11.7	2.1	0.26	2.1	43.5	6.2	43.0
Motacilla alba	35	7	***	37.0	0.48	7.9	2.0	0.18	2.8	41.5	2.2	21.0
Motacilla cinerea	9	3	***	35.0	0.25	3.1	3.6	0.07	3.5	75.6	1.2	18.0
Parus ater	21	6	***	70.0	0.71	19.4	3.6	0.16	5.0	54.9	5.0	9.9
Parus caeruleus	42	5	***	65.0	0.66	19.1	2.4	0.27	4.3	55.8	1.5	11.3
Parus cristatus	18	4	***	67.0	0.74	7.2	3.6	0.14	5.4	48.0	2.1	10.5
Parus major	49	8	***	58.0	0.57	15.6	1.7	0.39	3.8	45.7	1.6	16.8
Passer domesticus	25	5	***	78.0	0.66	148.4	4.4	0.11	2.1	63.9	16.6	28.0
Passer montanus	12	7	***	39.0	0.16	6.4	2.9	0.08	2.5	95.2	9.7	22.0
Petronia petronia	7	5	***	25.0	−0.02	3.4	2.1	0.14	2.0	31.1	9.9	31.0
Phoenicurus ochruros	21	7	***	37.0	0.24	4.4	1.3	0.16	2.4	66.8	1.1	16.5
Phylloscopus collybita	37	6	***	39.0	0.43	32.3	3.1	0.19	3.6	59.8	1.7	7.7
Pica pica	40	7	***	44.0	0.63	19.0	2.4	0.36	2.6	28.1	2.8	225.0
Picus viridis	23	8	***	47.0	0.39	1.2	1.7	0.17	3.5	11.1	1.0	200.0
Prunella modularis	19	9	***	35.0	0.17	2.8	1.1	0.18	2.9	40.5	1.1	19.3
Regulus ignicapillus	28	6	***	60.0	0.46	8.3	2.8	0.16	4.1	64.8	2.1	5.3
Regulus regulus	13	4	***	54.0	0.58	5.8	4.0	0.06	5.4	85.7	4.5	5.5
Remiz pendulinus	6	4	***	41.0	0.48	12.6	3.2	0.04	1.8	59.0	2.6	10.0
Saxicola torquata	28	8	***	57.0	0.38	3.6	2.4	0.26	1.5	48.5	1.4	15.2
Serinus citrinella	12	4	***	20.0	0.44	7.5	4.1	0.12	5.4	33.9	3.3	12.5
Serinus serinus	35	9	***	57.0	0.54	16.9	1.5	0.31	2.4	43.4	5.9	11.5
Sitta europaea	15	5	***	37.0	0.55	2.4	3.7	0.09	5.5	35.2	1.3	23.4
Sturnus unicolor	32	9	***	64.0	0.51	39.3	3.3	0.19	2.1	14.4	67.8	88.0
Sylvia atricaprilla	14	4	***	23.0	0.05	7.0	3.9	0.05	3.3	60.3	1.1	22.3
Sylvia melanocephala	13	4	***	46.0	0.50	2.8	4.2	0.13	2.9	49.2	1.0	11.2
Sylvia undata	22	9	***	48.0	0.24	3.0	2.1	0.24	2.1	45.4	1.2	10.5
Troglodytes troglodytes	28	7	***	58.0	0.38	3.4	2.4	0.27	3.8	54.0	1.1	8.8
Turdus iliacus	9	5	***	34.0	0.07	4.7	3.0	0.11	4.9	17.8	7.6	62.5
Turdus merula	52	8	***	55.0	0.31	14.9	2.0	0.38	3.0	44.6	1.3	86.1
Turdus philomelos	34	6	***	25.0	0.35	29.8	2.7	0.13	3.3	56.8	2.8	70.0
Turdus viscivorus	26	6	***	20.0	0.29	7.0	2.0	0.07	4.1	17.3	10.6	119.2
Upupa epops	5	4	NS	26.0	0.47	0.2	3.4	0.09	2.8	42.4	1.1	65.0
Vanellus vanellus	10	4	***	26.0	0.33	4.1	3.3	0.09	1.2	9.8	22.2	192.0

N: number of census transects where the species were present (prevalence); LV: number of final predicted values (leaves) in regression trees; p: significance of models ($^{***}P < 0.001$, N.S.: non-significant); D^2 (%): reduction of original deviance expressed in percentage; r_s: average Spearman correlation between predicted and observed densities in internal validation tests (50 repetitions); D_{max}: maximum abundance measured as the average of the three largest densities in the sample of 77 censuses; END: distance between environmental preferences and environments available in the study region; HB: habitat breadth in 27 different habitat categories considering vegetation structure, floristic composition, altitudinal distribution and human impact; SCI: structural complexity index of occupied habitats; DET: inconspicuousness measured as the percentage of birds detected in the proximity of the observer (less than 25 m); FS: average flock size; W: body mass in g.

of tree models with those birds that, having high environmental distance and large prevalences, are conspicuous and large-sized species, and they usually have small maximum abundances and do not gather in large flocks.

A stepwise multiple regression analysis explained 57.4% of the original interspecific variability in predictive power of models of the 64 studied species ($F_{2,61} = 28.29$, $P \ll 0.001$). The selected variables were the prevalence (standardized regression coeffi-

Table 2
Results of partial least square regression analyses relating predictive power of regression trees (see r_s in Table 1) to statistical model characteristics and ecological traits of species

	Regression trees (density: birds/10 ha)	
	COMP1	COMP2
Prevalence (N)	0.47***	0.31*
Environmental distance (END)	0.62***	0.63***
Habitat breadth (HB)	0.08	0.04
Structural complexity index (SCI)	0.37**	0.22
Maximum abundance (Dmax)	0.42***	−0.28*
Flock size (FS)	−0.02	−0.29*
Inconspicuousness (DET)	0.21	−0.38**
Body mass	−0.18	0.38**
R^2 (%)	40.5	7.5
P	≪0.001	0.002

For more details about variables see Table 1. R^2 (%): percentage of variance in interspecific differences in predictive power of tree models accounted for each component.
* $P < 0.05$.
** $P < 0.01$.
*** $P < 0.001$.

cient, $\beta = 0.61$, $P \ll 0.001$), and environmental distance (END; $\beta = 0.68$, $P \ll 0.001$).

4. Discussion

Line transects with survey belts (strip transects) have been used to sample avifaunas extensively (Hildén, 1986; Raven et al., 2003). Although it does not provide exact density estimations because detectabilities are far below 100% (usually 33–95%; Järvinen and Väisänen, 1975; Järvinen, 1978), relative abundance estimates are useful for comparisons between habitats or regions, for defining species-specific habitat selection, for analyzing macroecological patterns, or for building predictive maps of regional distribution or habitat suitability (Rotenberry and Wiens, 1980; Raven et al., 2003). Nevertheless, when forecasting bird distribution or abundance it is necessary to be aware of the accuracy of predictive models. In general, the proportion of explained variability in bird abundance derived from statistical models is used as a surrogate of model success. Interspecific discrepancies in modelling results are expected considering differences in the goodness-of-fit of species data to assumptions of statistical procedures or sample characteristics (e.g., sample size and prevalence). However, other ecological traits of species such as habitat use, habitat selection, foraging behaviour or abundance, could explain predictive power per se, considering the interaction between species and observers while sampling. This has been an unattended subject in modeling organism distributions (but see, Boone and Krohn, 1999; Kadmon et al., 2003).

Better predicted species are those most specialized in their habitat selection that also attain very high local densities. Stenotopic species, whose selected habitats are scarce and clearly identifiable in the study region, should be more likely to be accurately modelled and predicted because their sharply defined distributions allow the identification of unambiguous habitat selection patterns by splitting criteria in regression trees. On the other hand, the species showing higher maximum abundances are also those with larger variability in the modelled response (i.e., in density), because the density values in transects vary from 0 birds/10 ha (absence) to the maximum ecological density. Contrarily, less powerful models are expected in very rare or relatively scarce species, whose maximum abundances are low, and therefore the range in the response variable is narrow. Thus, a broad variability in the response variable helps to determine sharply defined distribution patterns explaining a large proportion of spatial variation in abundance. These two facts working together (stenotopic species that can reach high densities) allow researchers to obtain models that explain and predict a large proportion of spatial variation in abundance.

Habitat breadth does not contribute to explain interspecific variability in predictive power of models. Therefore, habitat width per se seems of little value because it is the interaction between habitat selection and environmental availability what really matters (END, see also, Boone and Krohn, 1999; Garrison and Lupo, 2002; Hepinstall et al., 2002).

Predictability of bird densities covaried with habitat selection of species according to the structural complexity of habitats: species inhabiting more wooded habitats were better predicted than open space birds. This result is probably an artefact derived from casual correlations between habitat structural complexity and other variables. Thus, species inhabiting forests during winter are specialists that occupy scarce environmental formations, and have large values of environmental distance (END; e.g., mature forests, montane pinewoods and riparian woodlands; *Regulus regulus*, *Dendrocopos major*, *Loxia curvirostra*, *Serinus citrinella*, *Sitta europaea*, etc.).

Inconspicuousness should limit the prediction of densities because less conspicuous species are prone to pass undetected in survey plots, more notably when they are rare. Linked with this phenomenon is the body size of species: small sized birds are prone to pass undetected because they are less visible, particularly in more vegetated habitats far away from the observer (Carrascal et al., 1989). As data noise increases, due to inaccuracy in density estimation, the probability of obtaining successful models decreases. Although this effect was significant, it played a limited role in explaining interspecific variations in the predictive power of models (they only entered the second component of PLS accounting for a low proportion of variance).

Grouping behaviour of species should constrain model accuracy and agreement between predicted and observed densities. For species gathering in large flocks, more variable estimations of relative abundance in quantitative inventory work are expected than in those living alone or in small flocks. Moreover, local abundance estimations should be very unstable in highly aggregated species, particularly under low sampling effort (Tellería, 1986). Thus, species living in large groups should preclude the attainment of good densities estimates (e.g., Anderson et al., 1998) or predictions in habitat suitability models. Our results point out a negative association between gregariousness and model predictive power (component 2 of PLSR in Table 2).

Finally, the prevalence of species in the sampling units is a variable that ought to be controlled for. Species prevalence has been found to affect several measures of model accuracy differently (Manel et al., 2001). In our case, the positive effect of prevalence on predictive power (Table 2) should be understood considering that more presences are necessary to build better models for the less registered species within the survey data base. Tree regression models with scarcely represented birds could be improved increasing the total number of transects performed under a well designed sampling protocol (e.g., stratified by habitats).

5. Conclusion

Predicting the distribution and abundance of species will still challenge the scientific community for a long time. New statistical approaches are incorporated to those previously available that circumvent statistical assumptions of classical models (for example, General Additive Models). Nevertheless, there are limits to model accuracy that could not be overcome by methodological refinements pursued by the researchers. This paper identifies three sources of restraints in the predictive power of habitat suitability modelling.

One of them is inherent to sampling effort in the rarest species. The obvious cure to this problem is to obtain large data bases incorporating a wealth of standardized surveys of the less represented species and habitats, which is costly in terms of time, money and human resources. On the other hand, there are species-specific characteristics constraining abundance forecasting that are rooted in the natural history of organisms. Among them, we have clearly identified maximum ecological density and conspicuousness. Finally, there is also a limitation considering the interaction between habitat selection of species and regional availability of those environments. A habitat specialist species (e.g., from coniferous forests or steppe environments) will not be accurately predicted in regions where those habitats are widely spread, while they will be properly forecasted in heterogeneous regions with a wide variety of environments. In our opinion, the limitations caused by those species-specific traits associated with survey work (e.g., conspicuousness, gregariousness or maximum ecological densities) will be difficult to circumvent by either statistical approaches or increasing sampling effort while recording biodiversity in extensive programs.

Acknowledgements

We would like to thank two anonymous referees for kindly improving a first version of the manuscript, and to Claire Jasinski for checking the English. Mark Boyce promptly assisted us with useful comments during the last revision.

References

Anderson, D.R., Moody, D.S., Smith, B.L., Lindzey, F.G., Lanka, R.P., 1998. Development and evaluation of sightability models for summer elk surveys. J. Wildl. Manage. 62, 1055–1066.

Bibby, C.J., Burgess, N.D., Hill, D.A., Mustoe, S.H., 2000. Bird Census Techniques, second ed. Academic Press, London.

Boone, R.B., Krohn, W.B., 1999. Modeling the occurrence of bird species: are the errors predictable? Ecol. Appl. 9, 835–848.

Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. Ecol. Model. 157, 281–300.

Carrascal, L.M., 1988. Influencia de las condiciones ambientales sobre la organización de la comunidad de aves invernante en un bosque subalpino mediterráneo. Doñana, Acta Vertebrata 15, 111–131.

Carrascal, L.M., Alonso, C.L., Palomino, D., 2000. Análisis ambiental de la influencia sobre la fauna silvestre del desdoblamiento y puesta en servicio del tramo 21,8–39,5 de la carretera M-501(Aves “no rapaces” en el entorno de la carretera M-501). Consejería de Transportes y Obras Públicas de la Comunidad de Madrid-CSIC, Madrid.

Carrascal, L.M., Díaz, J.A., Ruiz, M., 1989. Detectabilidad visual de aves en censos desde coche. Ardeola 36, 210–214.

Caughley, G., Gunn, A., 1996. Conservation Biology in Theory and Practice. Blackwell Science, Oxford.

Clark, L.A., Pregibon, D., 1993. In: Chambers, J.M., Hastie, T.J. (Eds.), Tree-based models. Statistical Models in S. Chapman & Hall, London, pp. 377–419.

Cody, M.L. (Ed.), 1985. Habitat Selection in Birds. Academic Press Inc., San Diego.

De’Ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81, 3178–3192.

Dedecker, A.P., Goethals, P.L.M., Gabriels, W., De Pauw, N., 2004. Optimization of Artificial Neural Network (ANN) model design for prediction of macroinvertebrates in the Zwalm river basin (Flanders, Belgium). Ecol. Model. 174, 161–173.

Dzeroski, S., Drumm, D., 2003. Using regression trees to identify the habitat preference of the sea cucumber (Holothuria leucospilota) on Rarotonga, Cook Islands. Ecol. Model. 170, 219–226.

Elith, J., Burgman, M., 2002. Predictions and their validation: rare plants in the Central Highlands, Victoria, Australia. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences. Issues of Scale and Accuracy. Island Press, Washington, pp. 303–313.

Felsenstein, J., 1985. Phylogenies and the comparative method. Am. Nat. 125, 1–15.

Fretwell, S.D., 1972. Populations in a Seasonal Environment. Princeton University Press, Princeton, New Jersey.

Garrison, B.A., Lupo, T., 2002. Accuracy of bird range maps based on habitat maps and habitat relationships models. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences. Issues of Scale and Accuracy. Island Press, Washington, pp. 367–375.

Guisan, A., Weiss, S.B., Weiss, A.D., 1999. GLM versus CCA spatial modelling of plant species distribution. Plant Ecol. 143, 107–122.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186.

Harvey, P.H., Purvis, A., 1991. Comparative methods for explaining adaptations. Nature 351, 619–624.

Hepinstall, J.A., Krohn, W.B., Sader, S.A., 2002. Effects of niche width on the performance and agreement of avian habitat models. In: Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), Predicting Species Occurrences. Issues of Scale and Accuracy. Island Press, Washington, pp. 593–606.

Hildén, O., 1986. Long-term trends in the Finnish bird fauna: methods of study and some results. Var Fagelvarld 11, 61–69.

Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology 83, 2027–2036.

Huntley, B., Green, R.E., Collingham, Y.C., Hill, J.K., Willis, S.G., Bartlein, P.J., Cramer, W., Hagemeijer, W.J.M., Thomas, C.J., 2004. The performance of models relating species geographical distributions to climate is independent of trophic level. Ecol. Lett. 7, 417–426.

Järvinen, O., 1978. Species-specific census efficiency in line transects. Ornis Scand. 9, 164–167.

Järvinen, O., Väisänen, R.A., 1975. Estimating relative densities of breeding birds by line transect method. Oikos 26, 316–322.

Järvinen, O., Väisänen, R.A., 1976. Estimating relative densities of breeding birds by the line transect method. IV. Geographical constancy of the proportion of main belt observations. Ornis Fenn. 53, 87–91.

Kadmon, R., Farber, O., Danin, A., 2003. A systematic analysis of factors affecting the performance of climatic envelope models. Ecol. Appl. 13, 853–867.

Leathwick, J.R., 1998. Are New-Zealand’s *Nothofagus* species in equilibrium with their environment? J. Veg. Sci. 9, 719–732.

Lehmann, A., 1998. GIS modeling of submerged macrophyte distribution using generalized additive models. Plant Ecol. 139, 113–124.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulanier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90, 39–52.

Levins, R., 1968. Evolutions in Changing Environments: Some Theoretical Explorations. Princeton University Press, Princeton, NJ, USA.

MacArthur, R.H., 1972. Geographical Ecology: Patterns in the Distribution of Species. Harper and Row, New York.

Manel, S., Dias, J.M., Ormerod, S.J., 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecol. Model. 120, 337–347.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38, 921–931.

Monreal, J., 1986. Evolución anual de los parámetros de la comunidad de aves de la vega del río Tajuña (Madrid). Tesis de licenciatura. Universidad Complutense de Madrid, Madrid.

Olden, J.D., Jackson, D.A., 2002. A comparison of statistical approaches for modelling fish species distributions. Freshwat. Biol. 47, 1976–1995.

Özesmi, S.L., Özesmi, U., 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. Ecol. Model. 116, 15–31.

Pearce, J., Ferrier, S., 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. Ecol. Model. 128, 127–147.

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. J. Environ. Manage. 62, 171–184.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecol. Biogeogr. 12, 361–371.

Perrins, C. (Ed.), 1998. The Complete Birds of the Western Paleartic on CD-ROM. Oxford University Press, Oxford.

Potti, J., 1985a. La sucesión de las comunidades de aves en los pinares repoblados de Pinus sylvestris del macizo de Ayllón (Sistema Central). Ardeola 32, 253–277.

Potti, J., 1985b. Las comunidades de aves del macizo de Ayllón. Universidad Complutense de Madrid, Madrid.

Price, T., 1997. Correlated evolution and independent contrasts. Philos. Trans. R. Soc. Lond. B 352, 519–529.

Raven, M.J., Noble, D.G., Baillie, S.R., 2003. The breeding bird survey 2002. BTO Research Report 334, British Trust for Ornithology, Thetford.

Remm, K., 2004. Case-based predictions for species and habitat mapping. Ecol. Model. 177, 259–281.

Ricklefs, R.E., Starck, J.M., 1996. Applications of phylogenetically independent contrasts: a mixed progress report. Oikos 77, 167–172.

Rotenberry, J.T., Wiens, J.A., 1980. Habitat structure, patchiness, and avian communities in North American steppe vegetation: a multivariate analysis. Ecology 61, 1228–1250.

Santos, T., Suárez, F., Tellería, J.L., 1983. The bird communities of the Spanish Juniper Woodland (*Juniperus thurifera* L.). In: Purroy, F.J. (Ed.), VII International Conference on Bird Census Work. Facultad de Biología, Universidad de León, León, pp. 79–88.

Santos, T., Tellería, J.L., Suárez, F., 1985. Guía de las aves invernantes en la Sierra de Madrid. Consejería de Agricultura y Ganadería de la Comunidad de Madri, Madrid.

Scott, J.M., Heglund, P.J., Morrison, M.L., Haufler, J.B., Raphael, M.G., Wall, W.A., Samson, F.B. (Eds.), 2002. Predicting species occurrences. Issues of scale and accuracy. Island Press, Washington.

Segurado, P., Araújo, M.B., 2004. An evaluation of methods for modelling species distributions. J. Biogeogr. 31, 1555–1568.

Seoane, J., Bustamante, J., Díaz-Delgado, R., 2004. Competing roles for landscape, vegetation, topography and climate in predictive models of bird distribution. Ecol. Model. 171, 209–222.

StatSoft, I. 2001. STATISTICA (data analysis software system), 6.0 ed. StatSoft Inc., Tulsa, Oklahoma.

Tellería, J.L., 1986. Manual para el censo de los vertebrados terrestres. Raíces, Madrid.

Tellería, J.L., Santos, T., Álvarez, G., Sáez-Royuela, C., 1988. Avifauna de los campos de cereales del interior de España. In: Bernis, F. (Ed.), Aves de los medios urbano y agrícola en las mesetas españolas. SEO, Madrid, pp. 173–319.

Thuiller, W., Aráujo, M.B., Lavorel, S., 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. J. Veg. Sci. 14, 669–680.

Tobias, R.D., 2003. An Introduction to Partial Least Squares Regression. URL: http://www.ats.ucla.edu/stat/sas/library/pls.pdf. Last accessed: 16 December 2003.

Walter, H., 1973. Vegetation of the Earth in Relation to Climate and the Eco-Physiological Conditions. Springer-Verlag, New York.

Westoby, M., Leishman, M.R., Lord, M.J., 1995. On misinterpreting the 'phylogenetic correction'. J. Anim. Ecol. 83, 531–534.

Yen, P.P.W., Huettmann, F., Cooke, F., 2004. A large-scale model for the at-sea distribution and abundance of Marbled Murrelets (*Brachyramphus marmoratus*) during the breeding season in coastal British Columbia, Canada. Ecol. Model. 171, 395–413.