

Luis M. Carrascal · Javier Seoane · David Palomino
César L. Alonso · Jorge M. Lobo

Species-specific features affect the ability of census-derived models to map winter avian distribution

Received: 13 June 2005 / Accepted: 3 February 2006 / Published online: 1 April 2006
© The Ecological Society of Japan 2006

Abstract We identify autoecological traits of bird species that influence the accuracy of predictive models of species distribution based on census data obtained from stratified sampling. These models would serve as a complementary approach to the development of regional bird atlases. We model the winter bird abundance of 64 terrestrial bird species in 77 census plots in Central Spain (Madrid province), using regression tree analyses. The predicted distribution of species density derived from statistical models (birds/10 ha) was compared with the published relative abundances depicted by a very accurate regional atlas of wintering birds (birds observed per 10 h). Statistical models explained an average of 41.7% of the original deviance observed in the local bird distribution (range 19.6–79.3%). Significant associations between observed relative abundances (atlas data) and predicted average densities in 1×1 km squares within 10×10 km UTMs were attained for 44 out of 64 species. Interspecific discrepancies between predicted and observed distribution maps decreased with between-year constancy in regional bird distribution and the degree of ecological specialization of species. Therefore, statistical modeling using census localities allowed us to depict geographical variations in bird abundance that were similar to those in the quantitative atlas maps. Nevertheless, bird distributions derived from statistical models are less reproducible in some species than in others, depending on their autoecological traits.

Introduction

Knowledge of environmental resources is the basis for wise ecological management. This knowledge has grown in many “developed” countries since the middle of the twentieth century. As a consequence, several databases with inventories of geographic localities, population levels or distribution maps for some taxa exist (Hagemeijer and Blair 1997; Jalas et al. 1999; Mitchell-Jones et al. 1999; Asher et al. 2001). Among them, some plant groups or vertebrate taxa, such as birds and mammals, are well-known over large geographical areas. Information on distribution ranges, population levels and the way they change over time is the foundation upon which conservation strategies rest. For example, the World Conservation Union coding of endangered species (IUCN 2001) is based on distribution ranges, population sizes and changes in them over a relative short period of time (e.g., 25 years or five generations). Sadly, even in some “developed” countries, many taxa are poorly known, lacking such elemental pieces of information as distribution ranges. For example, complete atlas maps based on a 10×10 km UTM grid have been published only very recently in Spain for popular organisms like birds (Martí and Del Moral 2003), reptiles (Pleguezuelos et al. 2002) or mammals (Palomo and Gisbert 2002). However, it is widely recognized that atlas information is highly dependent on the search effort expended (Donald and Fuller 1998; Dennis and Hardy 1999; Dennis and Thomas 2000; Williams et al. 2002). Thus, the precisions of Spanish atlases are dubious because the absence of a species in many UTM squares does not necessarily mean that the species are not present in them, but that they could not be found with the search effort invested, especially in areas where they are very scarce and localized. This is the usual case for many small, inconspicuous animals in areas where they are very scarce. Obviously, this problem is exaggerated in “underdeveloped” countries.

Distributions and abundances of organisms have been routinely related to environmental attributes in

L. M. Carrascal (✉) · D. Palomino · C. L. Alonso · J. M. Lobo
Departamento de Biodiversidad y Biología Evolutiva,
Museo Nacional de Ciencias Naturales, CSIC, C/José Gutiérrez
Abascal 2, 28006 Madrid, Spain
E-mail: mcnc152@mncn.csic.es
Tel.: +34-91-4111328
Fax: +34-91-5645078

J. Seoane
Departamento Interuniversitario de Ecología,
Facultad de Ciencias, Universidad Autónoma de Madrid,
28049 Madrid, Spain

several research programs like ecological biogeography, wildlife management and analyses of ecological niches (Austin et al. 1990; Iverson and Prasad 1998; Teixeira et al. 2001; Atkinson et al. 2002). A plethora of statistical methods have been used in the last 30 years to carry out this approach (Guisan and Zimmermann 2000; Anderson et al. 2003), although a common pattern has emerged in many of these studies: habitat characteristics and environmental variables are frequently closely related to organism abundance and occurrence if proper models of association are used (these are not always linear relationships: see Osborne et al. 2001; Scott et al. 2002). These statistical models have helped to develop our understanding of the dispersions and abundances of organisms at different spatial scales (e.g., Araújo and Williams 2000; Martínez et al. 2003; Brotons et al. 2004a), and to resolve the question of whether observed ecological patterns arise from causal or casual foundations (Legendre and Legendre 1998). On some occasions, the scope of this approach has been widened and it has been used in attempts to make prospective statements about distribution–abundance in other nonsampled areas (Seoane et al. 2003; Araújo et al. 2005) or to predict the future given several scenarios (climatic change, environmental impacts: Iverson and Prasad 1998; Pearson et al. 2002).

Nevertheless, little work has been done to determine how species-specific traits influence the predictive capabilities of distribution models (but see Boone and Krohn 1999; Stockwell and Peterson 2002; Kadmon et al. 2003). For example, among groups of animal species, the successes of several modeling techniques are inversely related to the spatial variability (mobility and nomadism) and the niche width, but there are some effects that are not consistent across all biological groups (Pearce and Ferrier 2000b; Pearce et al. 2001). Similar effects have been found within particular groups of species, with negative effects of niche width and positive effects of commonness, abundance and detectability (Boone and Krohn 1999; Kadmon et al. 2003). In contrast, Stockwell and Peterson (2002) found that the range size of species had a negative influence on the predictive accuracies of models, and no effects for other species-specific traits such as use of primary habitat, ecological breadth, migratory behavior or body size. Analyzing the association between species biological traits and model accuracy is a useful approach because we could improve the sampling of some species (e.g., by modifying survey protocols) if we knew the effects that specific traits have on modeling results.

Considering this ecological and statistical background, and the growing need for accurate biodiversity inventories in order to foster good environmental decisions or conservation plans, this paper attempts to link distribution mapping to species-specific modeling, taking into account habitat attributes. We obtained census data on avian abundance during the winter in Central Spain (Madrid) in order to model and predict bird distribution. The forecasts were used to evaluate their

congruence with the recently published atlas of wintering birds in the same region (Del Moral et al. 2002), a very geographically diverse area of 8,000 km² with wide climatologic, botanical and land use variations. This is a modern and highly accurate atlas containing distribution maps that provide relative abundances of bird species, expressed in birds observed per 10 h.

Working with very mobile taxa in winter, when it is particularly difficult to produce distribution maps (Bibby et al. 2000), we modeled bird distributions in terms of simple environmental and habitat variables throughout a large region using census inventories, and tried to answer the following questions:

1. Can statistical models obtained with a randomly stratified protocol be used to predict quantitative distribution maps (relative abundances) of species?
2. Are there clear species-specific traits constraining the predictive power of the distribution models and the accuracies of the maps produced?

Interspecific variation in predictive errors

Interspecific discrepancies are expected in the modeling results since species differ in many ecological traits related to their patterns of distribution and abundance (e.g., Kadmon et al. 2003) and to differences in interactions between the species and observers while sampling. Below we summarize the main hypotheses, and their associated predictions, which are used to explain why we expect some of the species to be modeled better than others.

Commonness at the regional scale (the UTM squares occupied), a measure of how widespread a species is, should be linked to the correlation between the atlas maps observed and those predicted by models (Kadmon et al. 2003). This is because the more widespread species often have broader niche widths and are expected to track environmental variability more tightly than the very scarce ones (Gaston and Blackburn 2000). Nevertheless, Stockwell and Peterson (2002) found that the maximum predictive accuracy of a model was not independent of range size, as widespread species were modeled less accurately.

Between-years constancy in regional distribution can be understood as a direct measurement of the spatial stability of the distribution area (Boulinier et al. 1998; Maron et al. 2005). This is an important concern when studying very mobile organisms, like birds in winter, which can perform short migrations tracking benign environmental conditions (e.g., avoiding cold spells in mountainous areas). From this measurement of how stable residency areas are in winter, it is expected that higher between-years constancy in atlas maps is associated with better agreement between the atlas maps and the census-based predicted maps.

Species with specialized *environmental requirements* (stenotopic), and those whose preferred habitats are scarce and clearly identifiable in the study region, should

stand a better chance of being accurately predicted because of their sharply defined distribution patterns (the census samples where the species is absent versus others where the species is present and abundant; Seoane et al. 2005). These remarkable patterns of habitat preference and distribution should be also captured by atlas maps (see also Kadmon et al. 2003; Brotons et al. 2004b).

Maximum ecological densities of species (the highest abundance in the region) are expected to increase the correlation between predicted and observed maps, because the species with the highest abundances in the study region are those with the largest variability in the variables to be modeled (densities in census plots) and checked (relative abundance within atlas squares). This fact should allow us to obtain models that explain a large proportion of the spatial variation in abundance (Seoane et al. 2005). On the other hand, less powerful models and inaccurately predicted maps are expected when the range in the response variable is lower in very rare or relatively scarce species.

Grouping behavior of the species should constrain the correlation between predicted and observed maps. We expect more variation among estimations of relative abundance in quantitative inventory work for species gathering in large flocks than for those living alone or in small flocks. Thus, abundance estimations in census plots and in atlas squares should be more unstable in species living in large groups, precluding the attainment of good predictions.

Inconspicuousness should limit the prediction of accurate distribution maps, because less conspicuous species are more likely to remain undetected in inventory work, especially if they are rare. Therefore, less detectable species should provide unstable estimations of relative abundance in atlas work (Seoane et al. 2005).

Methods

Study area

The study area was located in the center of the Iberian Peninsula, comprising Madrid province and nearby areas (less than 50 km from the province border). This region is very diverse environmentally, including high mountain areas reaching 2,450 m a.s.l. as well as lowland plateaus only 450 m a.s.l. Winter climate is colder in the highest areas of the Guadarrama Range (average winter temperature of 0.5 °C above 2,000 m a.s.l., and regular snowfalls) than in the mild and wet valleys of the Tajo Basin (average winter temperature of 6.9 °C under 600 m a.s.l.). There is a wide variety of habitats in this region: four main autochthonous forests [pine woods (*Pinus sylvestris*), riparian woods, deciduous oak woods (*Quercus pyrenaica*) and evergreen holm-oak woods (*Quercus ilex*)], parklands of ashes (*Fraxinus angustifolia*) or holm-oaks devoted to livestock, several types of shrublands, artificial and natural pasturelands, marshlands, rock outcrops, various agricultural

formations (vineyards, olive plantations, cereal croplands) and urban areas (from small villages to large cities). See Izco (1984) for more details and descriptions of the geographic, climatic and botanic characteristics of the study region.

The data analyses performed in this paper use three independent sources of information: bird census transects, relative abundance in UTM squares derived from the bird atlas, and environmental variables describing geographical location and habitat attributes.

Census data

Bird censuses of terrestrial habitats were performed during winter (December, January and the first fortnight of February) by the authors of this paper and were also obtained from data published in the literature (Santos et al. 1983; Potti 1985a, 1985b; Monreal 1986; Carrascal 1988; Telleria et al. 1988; Carrascal et al. 2000). The locations where censuses were taken by the authors were chosen in order to complement those already available in the literature, covering habitats or geographical areas lacking data. The time period covering the transects spanned from 1981 to 2003. All censuses were carried out following the same methodology: a line transect with survey belts of 25 m at each side of the progression line (Telleria 1986; Bibby et al. 2000). We did not observe any consistent difference between the two datasets (those obtained from the literature and censuses taken by the authors of this study; see below). Although transect locations were not randomly selected, in part because some of them were obtained from the literature, we do not think this sampling design limitation introduces any bias, as they were established throughout the whole study area (see Fig. 1) covering all environments in the region (by habitat structure and altitude). Transects were established in “homogeneous” areas (those that did not have a mixture of different habitats). All censuses were obtained on windless days without precipitation, at 8:00–11:30 and 15:00–17:00 GMT, at a low rate of ~1–3 km/h (slower in wooded habitats and/or places with many species and high densities). Data were obtained for 77 transects (see Fig. 1), covering areas of 8–500 ha (median = 32 ha; a line transect of 1 km samples 5 ha). Ninety-seven species were recorded in 4,173 ha. Due to sample limitations, statistical analyses were only performed with 64 species that appeared in at least five censuses. Bird density was expressed in birds/10 ha.

Each census 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 ranged from 0 to 5: 0—lacking in or containing 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

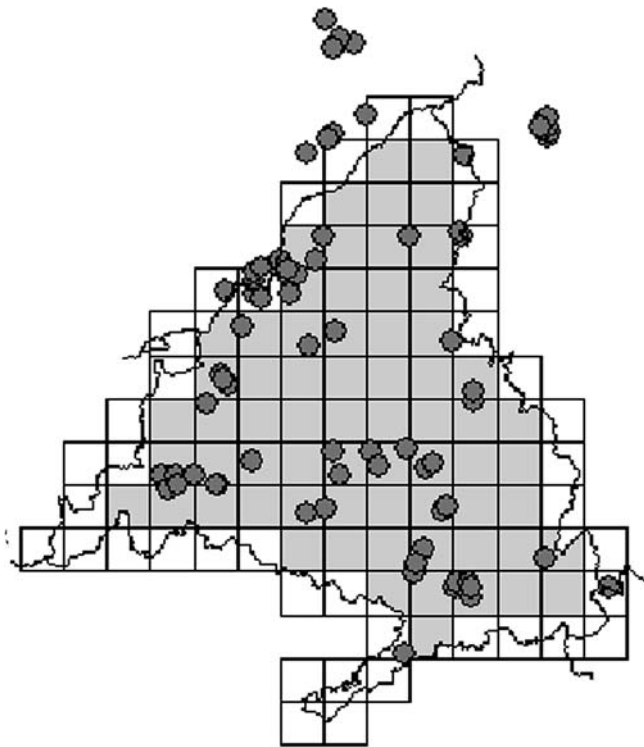


Fig. 1 Location of the study area in Madrid province (Central Spain); 10 km UTM grid squares selected (shown in gray), and location of the 77 bird census transects used (circles)

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.

During the winters of 1999–2003, the authors of this paper also recorded the flock sizes (FS) of bird species in the study region. When a species was sighted, we tried to count all of the individuals observed within a radius of 25 m. Using this procedure we obtained enough data to make a coarse-grained description of the average group sizes of the 64 species included in data analyses (sample sizes for all of the species ranged between 6 and 108 groups; median = 34 groups).

The censuses used in this paper also counted the birds observed outside the census belts. An index of lateral detectability (DET) was constructed (Järvinen and Väisänen 1975) as the ratio between the birds belonging to each species observed inside the transect belt and the total number of birds observed (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 , conspicuousness and mobility. It ranges between high values for inconspicuous species (e.g., >85% of individuals observed less than 25 m away, on both sides of the observer: *Passer montanus* L. and *Regulus regulus* L.) to low figures for more detectable species (e.g., *Buteo*

buteo L. and *Corvus corone* L.; <5% of observations at less than 25 m; see also Järvinen and Väisänen 1976; Järvinen 1978).

Atlas data

Atlas data was obtained from the atlas of wintering birds in Madrid (Del Moral et al. 2002). In this quantitative atlas, the area of study is divided into 115 10×10 km UTM squares that were prospected during the winters of 1999–2001 for a group of 76 experienced volunteers. The observers were asked to cover a minimum of 10 h per square (average time was 0.24 h/km²), sampling along routes in distinct habitats at a low speed. Time was invested in each habitat in proportion to the surface area covered by the habitat in the UTM square (Del Moral et al. 2002). Counts per UTM square were standardized afterwards to birds per 10 h, as an estimate of relative abundance. In the following analysis, we only used the data for 67 UTM squares that had more than 10 h of observation and more than 75% of their surface within the Madrid province (see Fig. 1).

Between-year constancy in regional distribution (BY) was estimated by considering the distribution patterns observed for the studied species over two consecutive winters (1999–2000 and 2000–2001). Spearman rank correlations (r_s) were obtained that related the relative abundance (birds/10 h per UTM) across the 67 squares selected in both winters. The higher the r_s value, the more stable the winter distribution patterns of the species in the study period. Commonness in the study region was estimated using the number of 10×10 km UTM squares (out of a total of 67) occupied by the species.

Geographic information system databases

The statistical models (see below) built with the 77-transect census for the 64 species considered were used to produce predictive distribution maps for the whole study region. IDRISI 32 software (Eastman 1999) was used to obtain several digital raster maps for each of the explanatory variables considered at a spatial resolution of 1,000 m. Latitude and longitude digital maps were also developed using the scores for these spatial variables for each 1 km cell. Altitude scores come from a digital elevational model built using digital cartography (scale 1:200,000) provided by the cartographic division of the Comunidad de Madrid. The presence of water in each 1 km raster cell (pools or streams) was codified to a binomial variable (presence–absence) from a vectorial map of streams, rivers and flooded areas. Cells used for agriculture, cells used as urban land, cells containing rocky outcrops, and cells containing deciduous and evergreen forest or coniferous forest were also mapped using binomial variables from the raster information provided by the European Environment Agency (CORINE programme 1985–1990). The six levels of habitat

complexity used were also derived from the CORINE land use categories.

Analytical procedures

Species density (birds/10 ha) in the census samples was modeled using the geographic positions of the transects and habitat characteristics as explanatory variables, and it was analyzed using regression tree analyses. This is a statistical tool related to general additive models where the response variable undergoes successive splits according to threshold values of the predictor variables that maximize the differences between the two resulting groups of samples. Regression trees can be used to deal with analytical problems linked to variables with non-Gaussian distributions, and they allow the interpretation of datasets with complex nonlinear relationships between response and predictor variables, and/or high-order interactions among predictor variables (Breiman et al. 1984; Venables and Ripley 1999; Boone and Krohn 2000; De'ath and Fabricius 2000). One important mechanism used to prevent trees from overfitting data is tree pruning. Pruning can be employed during tree construction (pre-pruning) based on stopping rules. The stopping rules used in this paper were that (1) nodes should contain at least ten censuses, and (2) the "leaves" (the ends of the tree tips) should include at least five censuses. Tree model significance was tested by means of the residual deviance (a measure of group heterogeneity) attained for the splitting criteria of the whole tree based on a chi-square test. Classification and regression trees were carried out using S-Plus 2000 software (Clark and Pregibon 1993; Crawley 2002).

Dichotomous explanatory predictors were entered as dummy variables. Geographical locations of transects (latitude and longitude) were included in the data analyses for two reasons. First, Carrascal et al. (2002) found clearly defined geographical patterns in avian abundance, diversity and species richness within this region when tracking the combined effects of altitude, temperature and precipitation (mainly from NE to SW). Second, the spatial contiguity of the sample units (i.e. the transects) has to be considered in order to control the spatial autocorrelation and to establish the spatial structure of the data used to analyze the field surveys (Legendre 1993; Legendre et al. 2002).

The models obtained for the 64 species were used to predict their relative abundances in the 1×1 km squares taking into account environmental information obtained from geographical information system (GIS) databases. Final predictions for the 67 10×10 km UTM squares were the average densities for all of the 1×1 km squares in each 10×10 UTM square (100 1×1 km squares). This procedure matches the spatial resolution of both field approaches (survey transects and atlas data). The agreement between the predictions of the statistical models and the quantitative atlas maps was evaluated by correlating both datasets. We used Spearman rank

correlations between the relative abundances obtained from the atlas for the 10×10 UTM squares and the averaged model predictions for the 100 corresponding 1×1 squares. The significance of each correlation is provided in order to help us to identify those species whose predicted maps are in strong agreement with the corresponding atlas maps. We did not transform the data in order to normalize them because it is impossible to do this with the abundance distributions observed for the species in the atlas work. The large number of zeros in the samples of the species studied (where the species was not present or was very unlikely to be detected in 20–30 h of atlas work) ruled out the application of any transformation in order to attain normality. Observed and predicted abundances were in different measurement units, although they refer to the same concept (how many birds per unit of standardized sampling effort; Bibby et al. 2000). Observed bird abundances were expressed in birds/10 h in the atlas data, while model predictions were in units of birds/10 ha for the regression trees.

The atlas work was carried out over two consecutive winters (1999–2001), while census work spanned a substantially longer period (1981–2003). Interannual variations in bird abundance in the study region should constrain the predictive abilities of the models and the correlation between predicted and observed maps. However, this potential bias is only a minor concern because 74% of the 77 censuses were done at the same time as the atlas work ± 1 year. Moreover, a preliminary exploration of the data did not reveal any significant effect of year on the recorded abundance of any species (correlating the residuals of the tree regression models with census year as a lineal predictor— $P > 0.1$ —in the 64 species analyzed).

Differences between the species' environmental preferences and the availability of those environments in the study region (ENVD) were calculated using the Euclidean distance between the means of the explanatory variables (excluding geographical coordinates) using the 77 census plots (the availability sample) and the weighted means of each species in the variables describing these 77 census plots (the preference sample). The weighted means of each species for the variables describing the census plots were obtained by considering densities of species in the censuses as well as the characteristics of the 77 line transects. Before computing the Euclidean distances, each variable was standardized to a mean of 0 and standard deviation of 1 (each variable weighed the same in distance estimations). The maximum abundance in the region (D_{\max}) was estimated as the highest density measured over the whole sample of 77 census plots. Commonness at a regional level (UTMs) was measured as the number of UTM squares occupied in Madrid within the sample of 67 10×10 km squares.

Associations between predictor variables (autoecological species traits in Table 1: BY, FS, DET, ENVD, D_{\max} and UTMs) and interspecific differences in

Table 1 Values of the specific traits considered for the 64 species studied, and the results from models built to predict the densities (birds/10 ha; regression trees) of 64 bird species in 77 census transects made in Central Spain

Species	N	Specific traits						Tree models	
		BY	FS	DET	ENVD	D_{\max}	UTMs	$D^2\%$	r_s
<i>Aegithalos caudatus</i>	28	0.64	5.9	67.1	2.38	11.5	64	75.0***	0.27*
<i>Alauda arvensis</i>	17	0.66	12.5	35.6	4.03	59.5	56	60.5***	0.73***
<i>Alectoris rufa</i>	18	0.70	2.4	50.3	2.31	4.1	63	33.3***	0.43***
<i>Anthus pratensis</i>	23	0.45	4.4	50.8	2.28	67.6	64	21.1***	0.13
<i>Buteo buteo</i>	10	0.57	1.2	3.2	2.55	2.7	60	24.2**	0.22
<i>Carduelis cannabina</i>	25	0.62	31.1	59.7	3.07	64.9	65	19.8***	0.04
<i>Carduelis carduelis</i>	34	0.59	13.4	35.0	1.49	48.3	65	34.8***	0.55***
<i>Carduelis chloris</i>	12	0.58	9.4	43.8	3.24	12.7	65	32.6***	0.61***
<i>Carduelis spinus</i>	11	0.34	11.4	56.9	2.35	32.1	31	32.6***	-0.31**
<i>Certhia brachydactyla</i>	34	0.61	1.2	43.3	2.26	17.4	62	49.5***	0.35**
<i>Cettia cetti</i>	13	0.56	1.1	68.7	3.34	22.1	52	44.5***	0.45***
<i>Cisticola juncidis</i>	11	0.39	1.3	49.4	3.11	4.0	31	38.0***	0.39***
<i>Columba livia</i>	14	0.60	3.0	41.9	4.62	27.8	64	67.0***	0.49***
<i>Columba palumbus</i>	24	0.63	15.5	13.1	3.91	53.8	65	29.1***	0.64***
<i>Corvus corone</i>	17	0.66	2.5	2.1	2.22	5.4	41	30.3***	0.02
<i>Corvus monedula</i>	12	0.69	29.6	35.9	2.95	26.1	54	31.2***	0.12
<i>Cyanopica cyana</i>	9	0.81	10.2	31.9	2.90	12.5	36	27.1***	0.38***
<i>Dendrocopos major</i>	15	0.41	1.1	39.6	2.45	3.5	49	28.9***	0.40***
<i>Emberiza cia</i>	23	0.60	3.5	56.4	1.46	8.1	59	35.6***	0.31**
<i>Emberiza cirius</i>	8	0.42	4.5	51.4	2.16	4.1	36	26.0***	0.07
<i>Emberiza schoeniclus</i>	5	0.57	2.4	53.7	3.29	7.5	25	41.4***	0.40***
<i>Erethacus rubecula</i>	45	0.66	1.0	44.9	1.70	12.1	67	79.3***	0.15
<i>Falco tinnunculus</i>	6	0.60	1.1	20.0	2.60	0.9	59	32.7	-0.26*
<i>Fringilla coelebs</i>	44	0.68	6.1	33.5	1.96	111.6	66	28.2***	0.51***
<i>Galerida cristata</i>	18	0.70	3.9	48.5	2.40	13.8	63	34.5***	0.64***
<i>Galerida theklae</i>	20	0.58	2.8	40.5	1.92	6.1	49	40.3***	-0.17
<i>Garrulus glandarius</i>	7	0.73	1.5	4.8	2.07	0.3	34	11.2	0.34**
<i>Lanius excubitor</i>	23	0.52	1.1	35.4	1.08	1.2	64	32.3	-0.04
<i>Loxia curvirostra</i>	9	0.13	8.8	20.0	4.55	11.0	32	45.4***	0.23
<i>Lullula arborea</i>	13	0.46	3.7	37.7	2.42	6.3	58	27.8***	0.30*
<i>Melanocorypha calandra</i>	6	0.54	26.0	38.6	3.84	28.5	24	29.5***	0.44***
<i>Miliaria calandra</i>	16	0.60	6.2	43.5	2.14	17.8	59	30.3***	0.42***
<i>Motacilla alba</i>	35	0.52	2.2	41.5	2.02	14.5	66	37.2***	0.06
<i>Motacilla cinerea</i>	9	0.23	1.2	75.6	3.64	4.2	48	34.8***	0.02
<i>Parus ater</i>	21	0.74	5.0	54.9	3.56	26.1	48	69.9***	0.64***
<i>Parus caeruleus</i>	42	0.39	1.5	55.8	2.36	26.4	67	65.3***	0.18
<i>Parus cristatus</i>	18	0.84	2.1	48.0	3.60	7.8	38	66.5***	0.66***
<i>Parus major</i>	49	0.65	1.6	45.7	1.73	16.0	66	58.0***	0.34**
<i>Passer domesticus</i>	25	0.65	16.6	63.9	4.44	161.3	67	78.4***	0.46***
<i>Passer montanus</i>	12	0.33	9.7	95.2	2.93	7.5	62	38.6***	0.03
<i>Petronia petronia</i>	7	0.36	9.9	31.1	2.13	4.2	38	24.9***	0.00
<i>Phoenicurus ochruros</i>	21	0.32	1.1	66.8	1.34	5.8	66	37.5***	0.32**
<i>Phylloscopus collybita</i>	37	0.62	1.7	59.8	3.07	55.2	66	39.4***	0.12
<i>Pica pica</i>	40	0.72	2.8	28.1	2.35	23.8	66	44.3***	0.41***
<i>Picus viridis</i>	23	0.49	1.0	11.1	1.74	1.6	66	46.5***	0.32**
<i>Prunella modularis</i>	19	0.56	1.1	40.5	1.13	3.6	44	34.6***	0.30*
<i>Regulus ignicapillus</i>	28	0.41	2.1	64.8	2.75	10.0	59	59.9***	0.29*
<i>Regulus regulus</i>	13	0.31	4.5	85.7	3.96	7.2	31	53.5***	0.25*
<i>Remiz pendulinus</i>	6	0.74	2.6	59.0	3.21	21.3	21	41.1***	0.43***
<i>Saxicola torquata</i>	28	0.38	1.4	48.5	2.42	5.0	62	57.1***	0.34**
<i>Serinus citrinella</i>	12	0.10	3.3	33.9	4.10	13.8	10	20.4***	0.29*
<i>Serinus serinus</i>	35	0.63	5.9	43.4	1.51	20.9	65	57.5***	0.47***
<i>Sitta europaea</i>	15	0.66	1.3	35.2	3.68	3.8	18	37.4***	0.63***
<i>Sturnus unicolor</i>	32	0.44	67.8	14.4	3.32	58.0	67	63.6***	0.18
<i>Sylvia atricapilla</i>	14	0.39	1.1	60.3	3.87	13.0	52	23.2***	0.41***
<i>Sylvia melanocephala</i>	13	0.71	1.0	49.2	4.16	4.8	57	46.3***	0.40***
<i>Sylvia undata</i>	22	0.49	1.2	45.4	2.15	4.7	59	47.8***	0.03
<i>Troglodytes troglodytes</i>	28	0.37	1.1	54.0	2.43	4.8	59	58.5***	0.09
<i>Turdus iliacus</i>	9	0.23	7.6	17.8	3.02	6.3	43	33.5***	0.18
<i>Turdus merula</i>	52	0.69	1.3	44.6	2.04	17.3	67	54.9***	0.29*
<i>Turdus philomelos</i>	34	0.61	2.8	56.8	2.68	40.4	66	25.3***	0.48***
<i>Turdus viscivorus</i>	26	0.82	10.6	17.3	2.04	12.4	51	19.6***	0.34**
<i>Upupa epops</i>	5	0.37	1.1	42.4	3.45	0.4	42	26.4	0.39***

Table 1 Continued

Species	<i>N</i>	Specific traits						Tree models	
		BY	FS	DET	ENVD	D_{\max}	UTMs	$D^2\%$	r_s
<i>Vanellus vanellus</i>	10	0.48	22.2	9.8	3.30	5.7	35	26.1***	0.32**

N number of census transects where the species were present, *BY* between-year constancy in regional distribution (Spearman rank correlation between distribution patterns observed in two consecutive winters), *FS* average flock size, *DET* inconspicuousness measured as the percentage of birds detected in the proximity of the observer (less than 25 m), *ENVD* distance between average environmental preferences and environments available in the study region, D_{\max} maximum abundance measured in the sample of 77 censuses, *UTMs* commonness in the study region, given as the number UTM 10×10 km squares occupied by each species (out of a maximum of 67), $D^2\%$ reduction in original deviance obtained by tree regression models, expressed in percent (the significances of the tree models are expressed with asterisks), r_s Spearman correlations between observed (atlas data) and predicted relative abundances (significance of r_s expressed with asterisks)

*** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$

prediction accuracy (r_s : Spearman rank correlations between observed atlas and predicted relative abundances per UTM 10×10 square) were analyzed by means of multiple regression analysis.

Bird species are related through evolution via a phylogenetic scheme, and so they should not be treated as independent sample units (Felsenstein 1985; Harvey and Purvis 1991). Nevertheless, in some instances ecologists are not interested in patterns of biological diversification across evolutionary time, but only in present-day relationships comprising nonevolutionary associations formed in human-transformed environments (Westoby et al. 1995; Ricklefs and Starck 1996). Since we are studying the relationships between ecological species traits and present-day distributions of species in a transformed landscape, we have simplified the data analysis, avoiding comparative methods.

Results

Model building and prediction of regional distribution

The original deviance explained by the statistical models built using the species densities observed in the sample of 77 census transects was moderate (Table 1; mean = 41.7%, range 19.6–79.3% in models including 2–8 splitting criteria). Nearly all models were

statistically significant ($P < 0.01$ in 60 out of 64 species).

Predictions from regression tree models were significantly associated with observed distributions of relative abundances in the atlas squares in 68.8% of the species (44 out of 64; Table 1). Spearman correlation values between the observed relative atlas abundances and the predictions from the census-based models were highly variable (range –0.31 to 0.73), with an average of 0.29.

Interspecific variation in the predictive accuracies of the statistical models

A significant amount of the interspecific variability in the agreement between predictions from the regression trees and the atlas data was explained by ecological species traits ($R^2 = 0.328$, $F_{(7,56)} = 3.90$, $P = 0.002$; Table 2). Accuracy in predicting the bird atlas distributions was positively and significantly associated with the between-year constancy in regional distribution, which explained the largest proportion of the interspecific variability (Table 2). Differences between environmental preferences and environments available in the study region also (positively) affected the agreement between model predictions and atlas data (Fig. 2). The remaining autoecological traits and the explained deviance of the regression trees did not affect the interspecific differences.

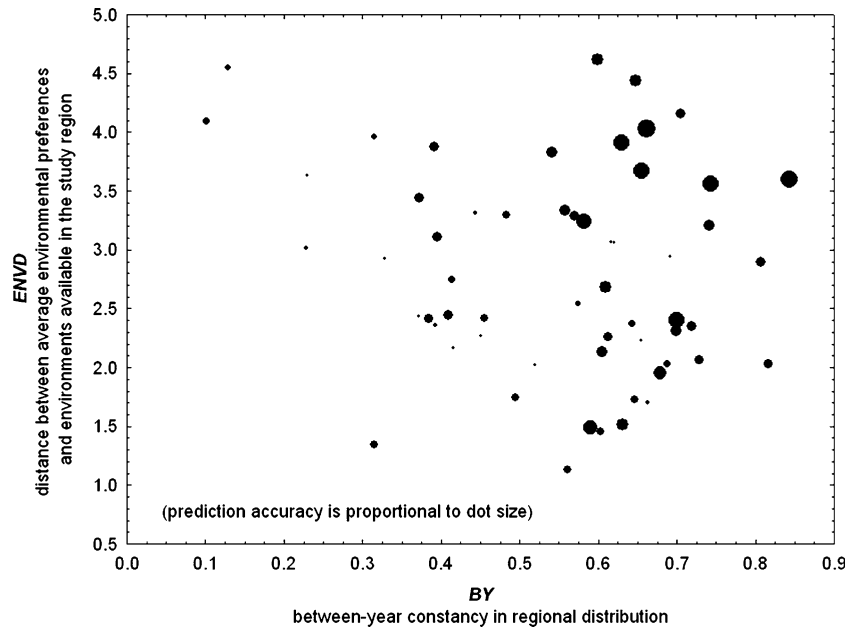
Table 2 Results from multiple regression analysis relating the agreement between atlas data and predictions of regression trees (see r_s in Table 1) and the model characteristics and ecological traits of species in the study region

	β	<i>P</i>	% var
% Deviance explained by the model ($D^2\%$)	–0.013	0.926	0.01
Between-year constancy (BY)	0.418	< 0.001	14.69
Ecological specialization (ENVD)	0.412	0.014	7.73
Maximum abundance (D_{\max} ; in log)	0.154	0.476	0.62
Inconspicuousness (DET)	–0.057	0.671	0.22
Flock size (FS; in log)	–0.156	0.351	1.06
Commonness at the regional scale (UTMS)	0.049	0.816	0.07

See Table 1 for the meanings of the variables. Sample size is 64 species

β standardized partial regression coefficient, % var explained variance (sum of squares of each effect divided by the total sum of squares)

Fig. 2 Relationship between the environmental specialization (*ENVD*) and the between-year constancy in regional distribution (*BY*) along with the prediction accuracy of the regression trees (the agreement between the relative abundances predicted by habitat models and those observed in atlas work; see r_s in Table 1). Prediction accuracy is directly proportional to dot size. Sample size is 64 species



Discussion

Factors affecting accuracy of bird modeling and mapping

We were able to predict the regional winter distributions of most bird species from statistical models of local densities, although there was a considerable interspecific variability in the agreement with quantitative atlas maps. Explaining why the disagreements between atlas distribution and the distribution predicted by the models are higher in some species than in others is more interesting than stating the successes and failures of the models. Thus, disagreements between observed and predicted results were associated with ecological traits related to distribution patterns. This fact is habitually dismissed when modeling species distribution (but see Boone and Krohn 1999; Stockwell and Peterson 2002; Kadmon et al. 2003).

The most important predictor of these disagreements was the between-year constancy in atlas distribution. Nomadic and irruptive species, whose winter quarters lay in northern latitudes, reached our study area in high numbers only after environmental conditions worsened severely in their normal wintering areas [this includes species such as the Redwing (*Turdus iliacus*) and the Siskin (*Carduelis spinus*)]. These species are likely to be recorded randomly in the study region because their wintering in the area is largely mediated by nonlocal factors. Therefore, their occurrence or abundance will be weakly related to the suitability of a particular area, and the same atlas square can attain very different abundance estimates in two consecutive years. In these cases it is unlikely that wintering birds occupy all of the suitable habitat in a region. However, these

disagreements between the relative abundances predicted by statistical models and those observed in quantitative atlas maps are pessimistically counted as a failure of the former.

Environmental specialization, measured as differences between environmental preferences and habitat availability in the study region (*ENVD* in this paper), also explained the disagreements. The distributions of stenotopic species with small available areas in the study region are similarly depicted with both the atlas and the statistical models. On the other hand, ubiquitous species do not show clear relations with the kind of coarse-grained explanatory variables we used, and thus they are difficult to model (Brotons et al. 2004b). Besides, these species appear in most atlas squares because there are probably many suitable patches for a eurytopic species in 100 km². It is worth noting that habitat width seems of little value per se, and may even be misleading when predicting the success of a model (Boone and Krohn 1999; Garrison and Lupo 2002; Hepinstall et al. 2002; but see Stockwell and Peterson 2002), because it is the interaction between habitat preferences and environmental availability that really matters.

Predictions of inconspicuousness, maximum abundance, flock size and commonness were not supported by our data. These variables had no effect on the correlation between model predictions and observed quantitative distributions derived from atlas maps (Table 2). Some of these effects have not been consistently supported in the literature for different taxa. Thus, Stockwell and Peterson (2002) found that widespread species were modeled less accurately than those species with more restricted geographical ranges, while some authors found the opposite (Boone and Krohn 1999; Kadmon et al. 2003). Finally, the species with the more

explanatory habitat models, as judged by the amount of explained deviance, did not show better agreements between distributions derived from statistical models and maps. This fact should prevent us from using the model's percentage of explained deviance as a measure of its predictive power (MacNally 2000).

Statistical models of census inventories as a complement to atlas work

The modeling approach to studying species distributions, based on randomly established census locations or those following a stratified sampling, has several advantages over traditional atlas studies. First, the sampling protocol based on quantitative inventories for census localities yields data on fine-grained patterns of distribution and abundance of species—such as ecological densities (range and maximum), habitat preferences and ecological width—more easily. This information may help us to empirically assess the commonness–rarity of organisms in order to define local conservation priorities at different administrative and political levels (e.g., red lists). Second, statistical models relating occurrence or abundance data to some environmental variables—such as habitat, land use and topographical features—allow us to explain and predict bird distribution. These models can produce maps at higher resolutions than those achieved in atlas work, and can be subjected to more formal criticism when evaluating their degrees of success and their shortcomings (Guisan and Zimmermann 2000; Pearce and Ferrier 2000a).

On the other hand, the effort undertaken for an atlas is typically larger than that required by quantitative inventories on census localities even though they both provide the same kind of information: maps. Considering a sample size of 2 km (10 ha) per locality, the time invested in the Madrid atlas would have allowed for approximately 800 census plots. This huge sample size is one order of magnitude higher than that used in this study to build statistical models. The difference would have been even greater if we follow the recommendations for atlas work stated by other authors (40–45 h for 5×5 km squares in forested habitats, Roberge and Svensson 2003). For these reasons, the statistical models derived from quantitative inventories on census localities could be thought of as an interesting complementary approach to atlas work.

The approach of this paper (random stratified sampling and modeling of species distribution, together with spatially explicit predictions obtained using GIS techniques) can be readily generalized as a complement to bird atlas work. We recommend transects of 2 km censused in 1 h, recording species within 25 m belts and at any distance from the observer (useful for very scarce and highly detectable species). We suggest that this complementary methodology could be used in those environmental heterogeneous territories where sampling the whole ecological variability with a limited number of

qualified observers or economic resources available is a formidable endeavor.

Acknowledgements We would like to thank Juan Carlos del Moral for facilitating access to the database of wintering birds in Madrid, and to all of the volunteers that made it possible to build the atlas. Two anonymous referees considerably improved a first version of the manuscript. Claire Jasinski kindly improved the English of an early draft. This paper is a contribution to the project 07M/0080/2002 funded by the Consejería de Educación de la Comunidad de Madrid, and BOS2000-0993 funded by the Spanish Ministerio de Ciencia y Tecnología. David Palomino was supported by an “El Ventorrillo” Field Station-CSIC grant.

References

- Anderson RP, Lew N, Peterson AT (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol Model* 162:211–232
- Araújo MB, Williams PH (2000) Selecting areas for species persistence using occurrence data. *Biol Conserv* 96:331–345
- Araújo MB, Thuiller W, Williams PH, Reginster I (2005) Down-scaling European species atlas distributions to a finer resolution: implications for conservation planning. *Glob Ecol Biogeogr* 14:17–30
- Asher J, Warren M, Fox R, Harding P, Jeffcoate G, Jeffcoate S (2001) The millennium atlas of butterflies in Britain and Ireland. Oxford University Press, Oxford
- Atkinson PW, Fuller RJ, Vickery JA (2002) Large-scale patterns of summer and winter bird distribution in relation to farmland type in England and Wales. *Ecography* 25:466–480
- Austin MP, Nicholls AO, Margules CR (1990) Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Environ Manage* 60:161–177
- Bibby CJ, Burgess ND, Hill DA, Mustoe SM (2000) Bird census techniques. Academic, New York
- Boone RB, Krohn WB (1999) Modeling the occurrence of bird species: are the errors predictable? *Ecol Appl* 9:835–848
- Boone RB, Krohn WB (2000) Relationship between avian range limits and plant transition zones in Maine. *J Biogeogr* 27:471–482
- Boulinier T, Nichols JD, Hines JE, Sauer JR, Flather CH, Pollock KH (1998) Higher temporal variability of forest breeding bird communities in fragmented landscapes. *Proc Natl Acad Sci USA* 95:7497–7501
- Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Chapman & Hall, New York
- Brotos L, Mañosa S, Estrada J (2004a) Modelling the effects of irrigation schemes on the distribution of steppe birds in Mediterranean farmland. *Biodivers Conserv* 13:1039–1058
- Brotos L, Thuiller W, Araújo MB, Hirzel AH (2004b) Presence–absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography* 27:437–448
- Carrascal LM (1988) Influencia de las condiciones ambientales sobre la organización de la comunidad de aves invernante en un bosque subalpino mediterráneo. *Doñana Acta Verteb* 15:111–131
- Carrascal LM, Alonso CL, 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). Technical Report for the Consejería de Transportes y Obras Públicas de la Comunidad de Madrid-CSIC, Madrid
- Carrascal LM, Palomino D, Lobo JM (2002) Patterns of habitat preference and distribution and abundance of wintering bird fauna in Central Spain. Analysis and prediction of the effect of ecological factors (in Spanish). *Anim Biodivers Conserv* 25:7–40
- Clark LA, Pregibon D (1993) Tree-based models. In: Chambers JM, Hastie TJ (eds) Statistical models in S. Chapman & Hall, London, pp. 377–419

- Crawley MJ (2002) Statistical computing. An introduction to data analysis using S-plus. Wiley, New York
- De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81:3178–3192
- Del Moral JC, Molina B, De la Puente J, Pérez-Tris J (2002) Atlas de las Aves Invernantes de Madrid 1999–2001. SEO-Monticola y Comunidad de Madrid, Madrid
- Dennis RLH, Hardy PB (1999) Targeting squares for survey: predicting species richness and incidence of species for a butterfly atlas. *Glob Ecol Biogeogr* 8:443–454
- Dennis RLH, Thomas CD (2000) Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *J Insect Conserv* 4:73–77
- Donald PF, Fuller RJ (1998) Ornithological atlas data: a review of uses and limitations. *Bird Study* 45:129–145
- Eastman JR (1999) Idrisi32. Reference guide. Clark University, Worcester
- Felsenstein J (1985) Phylogenies and the comparative method. *Am Nat* 125:1–15
- Garrison BA, Lupo T (2002) Accuracy of bird range maps based on habitat maps and habitat relationships models. In: Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (eds) Predicting species occurrences. Issues of scale and accuracy. Island, Washington DC, pp. 367–375
- Gaston KJ, Blackburn TM (2000) Pattern and process in macroecology. Blackwell Science, Oxford
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Model* 135:147–186
- Hagemeyer EJM, Blair MJ (1997) The EBCC atlas of European breeding birds: their distribution and abundance. Poyser, London
- Harvey PH, Purvis A (1991) Comparative methods for explaining adaptations. *Nature* 351:619–624
- Hepinstall JA, Krohn WB, Sader SA (2002) Effects of niche width on the performance and agreement of avian habitat models. In: Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (eds) Predicting species occurrences. Issues of scale and accuracy. Island, Washington DC, pp. 593–606
- IUCN (2001) IUCN Red List Categories and Criteria: Version 3.1. World Conservation Union, Gland, Switzerland
- Iverson LR, Prasad AM (1998) Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecol Monogr* 68:465–468
- Izco J (1984) Madrid Verde. M.A.P.A.-Comunidad Autónoma de Madrid, Madrid
- Jalas J, Suominen J, Lampinen R (1999) Atlas florae Europaeae: distribution of vascular plants in Europe. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki
- Järvinen O (1978) Species-specific census efficiency in line transects. *Ornis Scand* 9:164–167
- Järvinen O, Väisänen RA (1975) Estimating relative densities of breeding birds by line transect method. *Oikos* 26:316–322
- Järvinen O, Väisänen RA (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
- Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* 74:1659–1673
- Legendre P, Legendre L (1998) Numerical ecology, 2nd edn. Elsevier, Oxford
- Legendre P, Dale MRT, Fortin MJ, Gurevitch J, Hohn M, Myers D (2002) The consequences of spatial structure for design and analysis of ecological field surveys. *Ecography* 25:601–615
- MacNally R (2000) Regression and model-building in conservation biology, biogeography and ecology: the distinction between—and reconciliation of—“predictive” and “explanatory” models. *Biodivers Conserv* 9:655–671
- Maron M, Lill A, Watson DM, MacNally R (2005) Temporal variation in bird assemblages: how representative is a one-year snapshot? *Austral Ecol* 30:383–394
- Martí R, Del Moral JC (2003) Atlas de las Aves Reproductoras de España. Ministerio de Medio Ambiente-SEO/BirdLife, Madrid
- Martínez JA, Serrano D, Zuberogoitia I (2003) Predictive models of habitat preferences for the Eurasian eagle owl *Bubo bubo*: a multiscale approach. *Ecography* 26:21–28
- Mitchell-Jones AJ, Amori G, Bogdanowicz W, Krystufek B, Reijnders PJH, Spitzenberger F, Stubbe M, Thissen JMB, Vohralik V, Zima J (1999) The atlas of European mammals. Academic, London
- 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
- Osborne PE, Alonso JC, Bryant RG (2001) Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *J Appl Ecol* 38:458–471
- Palomo LJ, Gisbert J (2002) Análisis de la información obtenida. In: Palomo LJ, Gisbert J (eds) Atlas de los Mamíferos Terrestres de España. Dirección General de la Conservación de la Naturaleza-SECEM-SECEMU, Madrid, pp. 473–492
- Pearce J, Ferrier S (2000a) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol Model* 133:225–245
- Pearce J, Ferrier S (2000b) 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 northeast New South Wales. *J Environ Manage* 62:171–184
- Pearson RG, Dawson TP, Berry PM, Harrison PA (2002) SPECIES: a spatial evaluation of climate impact on the envelope of species. *Ecol Model* 154:289–300
- Pleguezuelos JM, Márquez R, Lizana M (2002) Atlas y Libro Rojo de los Anfibios y Reptiles de España. Dirección General de la Conservación de la Naturaleza-AHE, Madrid
- 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. Tesis de licenciatura. Universidad Complutense de Madrid, Madrid
- Ricklefs RE, Starck JM (1996) Applications of phylogenetically independent contrasts: a mixed progress report. *Oikos* 77:167–172
- Roberge JM, Svensson S (2003) How much time is required to survey land birds in forest-dominated atlas squares? *Ornis Fenn* 80:111–120
- Santos T, Suárez F, Tellería JL (1983) The bird communities of the Spanish juniper woodland (*Juniperus thurifera* L.). In: Purroy FJ (ed) VII Int Conf on Bird Census Work. Universidad de León, León, pp. 79–88
- Scott JM, Heglund PJ, Morrison ML, Haufler JB, Raphael MG, Wall WA, Samson FB (2002) Predicting species occurrences. Issues of scale and accuracy. Island, Washington DC
- Seoane J, Viñuela J, Díaz-Delgado R, Bustamante J (2003) The effects of land use and climate on red kite distribution in the Iberian peninsula. *Biol Conserv* 111:401–414
- Seoane J, Carrascal LM, Alonso CL, Palomino D (2005) Species-specific traits associated to prediction errors in bird habitat suitability modelling. *Ecol Model* 185:299–308
- Stockwell DRB, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. *Ecol Model* 148:1–13
- Teixeira J, Ferrand N, Arntzen JW (2001) Biogeography of the golden striped salamander *Chioglossa lusitanica*: a field survey and spatial modelling approach. *Ecography* 24:618–624
- Tellería JL (1986) Manual para el Censo de los Vertebrados Terrestres. Raíces, Madrid

- Tellería JL, Santos T, Álvarez G, Saiz-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
- Venables WN, Ripley BD (1999) *Modern applied statistics with S-PLUS*, 3rd edn. Springer, Berlin Heidelberg New York
- Westoby M, Leishman MR, Lord MJ (1995) On misinterpreting the “phylogenetic correction”. *J Ecol* 83:531–534
- Williams PH, Margules CR, Hilbert DW (2002) Data requirements and data sources for biodiversity priority area selection. *J Biosci* 27:327–338