Available online at www.sciencedirect.com

SCIENCE DIRECT®

Ecological Modelling 175 (2004) 137–149

ECOLOGICAL MODELLING

www.elsevier.com/locate/ecolmodel

Are existing vegetation maps adequate to predict bird distributions?

Javier Seoane, Javier Bustamante*, Ricardo Díaz-Delgado

Department of Applied Biology, Estación Biológica de Doñana, CSIC, Avda. María Luisa s/n, 41013 Sevilla, Spain

Received 29 May 2002; received in revised form 21 October 2003; accepted 31 October 2003

Abstract

Bird species are selective on the vegetation types in which they are found but predictive models of bird distribution based on variables derived from land-use/land-cover maps tend to have limited success. It has been suggested that accuracy of existing maps used to derive predictors is in part responsible for the limited success of bird distribution models. In two areas of 4900 km^2 of Western Andalusia, Spain, we compared the predictive ability of bird distribution models derived from two existing general-purpose land-use/land-cover maps, which differ in their resolution and accuracy: a coarse scale vegetation map of Europe, the CORINE land-cover map, and a detailed regional map, the 1995 land-use/land-cover map of Andalusia from the SINAMBA (Consejería de Medio Ambiente, Junta de Andalucía). We compared the bird distribution models derived from these general-purpose vegetation maps with models derived from two more accurate structural vegetation maps built considering directly variables that influence bird habitat selection, one built from satellite images for this study and another obtained by improving the resolution and accuracy of the SINAMBA map with satellite data. We sampled the presence/absence of bird species at 857 points using 15-min point surveys. Predictive models for 54 bird species were built with generalised additive models (GAMs), using as potential predictors the same set of landscape and vegetation structure variables measured on each map. We compared for each bird species the predictive accuracy of the best model derived from each map. Vegetation structure measured at bird sample points was used as ground-truth for comparing the accuracy of vegetation maps. Although maps differed in their resolution and accuracy, the results show that all of them produced similarly accurate bird distribution models, with a mixed map produced with both thematic and satellite information being the best. The models derived from the more accurate vegetation structure maps obtained from satellite data were not more accurate than those derived directly from the SINAMBA or CORINE maps. Our results suggest that some general-purpose land-use/land-cover maps are accurate enough to derive bird distribution models. There is a certain limit to improve vegetation maps above which there is no effect in their power to predict bird distribution.

Keywords: Map quality; Wildlife-habitat models; Bird distribution

1. Introduction

Bird species are selective on the vegetation types in which they inhabit (Cody, 1985). It is considered that the vegetation holds a great predictive potential for the distribution of birds, and several ongoing projects are using vegetation types to map potential distribution of bird species, for example the GAP project in USA

* Corresponding author. Tel.: +34-95-4232340; fax: +34-95-4621125.
E-mail addresses: seoane@ebd.csic.es (J. Seoane), busta@ebd.csic.es (J. Bustamante), rdiaz@ebd.csic.es (R. Díaz-Delgado).

0304-3800/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2003.10.011

(Scott et al., 1993), or are using vegetation variables to build predictive models of bird distribution (Pearce and Ferrier, 2000b). Statistical models of bird distribution using as predictors variables derived from vegetation are rarely able to explain perfectly observed distributions. There are several possible reasons for this fact (Fielding and Bell, 1997; Beutel et al., 1999):

(a) Statistical reasons (e.g. when using logistic regression to model presence/absence data the predicted values are in a continuous scale from 0 to 1 while observed data are discrete presences and absences).
(b) Historical reasons (e.g. an species has not occupied all potential adequate habitat because of geographical barriers, or because it has been extirpated by man from otherwise suitable habitat).
(c) Unsaturated habitats (small populations are not able to occupy all suitable habitats, but also demographic stochasticity and localised dispersal generate an imperfect correlation between habitat suitability and species distribution; Tyre et al., 2001).
(d) Poor quality of the response variable (e.g. an inadequate census method for a species difficult to detect may render a distribution pattern of observed presences that does not reflect the real pattern of distribution or abundance).
(e) Poor quality of the predictive variables (e.g. when the predictors we are measuring are not adequate to explain the distribution of the species or they are measured with too much error; Guisan and Zimmermann, 2000).

Land-use/land-cover or vegetation maps can be used as the source of predictive variables in statistical models of the distribution of bird species (Tobalske and Tobalske, 1999; Guisan and Zimmermann, 2000; Pearce and Ferrier, 2000b). Maps are themselves models of reality and as such they are always a simplification. Available vegetation maps may not represent adequately the vegetation variables relevant for the species of bird whose distribution we want to predict, or may have not the adequate spatial resolution (Pearce et al., 2001). We may be able to measure directly at bird sampling points those vegetation variables that an expert on the species would consider more relevant, but the final model obtained will not be useful to map the potential habitat for the species if the maps of these vegetation variables are not available. Vegetation, land-use and land-cover maps are currently built by governmental agencies at different resolutions and for different purposes. Existing environmental maps are cheap predictors for mapping potential habitat for birds while the best potential predictors we might think about may never be mapped. On the other hand, remote sensing is a potential tool for mapping the vegetation variables that we might consider more relevant for the distribution of a species of bird (Palmeirin, 1988; Avery and Haines-Young, 1990; Franklin and Steadman, 1991; Andries et al., 1994; Paruelo and Golluscio, 1994; Wu and Strahler, 1994; Roy et al., 1996; Trodd, 1996; Ormerod and Watkinson, 2000). Remote sensing imagery sensu lato (airborne sensors, aerial photography and satellite images) is the tool most widely used nowadays to create new land-cover maps (CORINE project in Europe, MIOMBO project in Southern Africa, etc.), to improve thematic maps accuracy (Stehman, 1996), or final map spatial resolution (Defries and Belward, 2000).

In this paper we compare the capacity to predict the distribution of 54 species of birds of variables derived from two existing vegetation maps (a coarse scale vegetation map of Europe and a more detailed regional land-use/land-cover map), a vegetation map derived from satellite data, and a vegetation map obtained by improving the accuracy of the existing regional map with satellite data. It has been suggested that higher accuracy and resolution of input maps is necessary to improve predictions of plants and animal distributions models (Guisan and Zimmermann, 2000). As we were more interested in the effect of map accuracy in predictive capacity than in the potential of different maps to measure different predictors, we measured the same set of predictors in all vegetation maps.

2. Study area and methods

We performed 1144 unlimited-distance bird point surveys in two 70 km × 70 km squares in Western Andalusia, Southern Spain (centres: 6°21′W and 37°39′N; 5°28′W and 36°44′N), during the springs of 1999 and 2000 (Fig. 1). Both areas have a similar proportion of different land-cover types and have approximately 20% of cropland (mainly wheat, sunflower and olive groves), 70% of scrubland and forests (mainly Mediterranean scrubland, evergreen

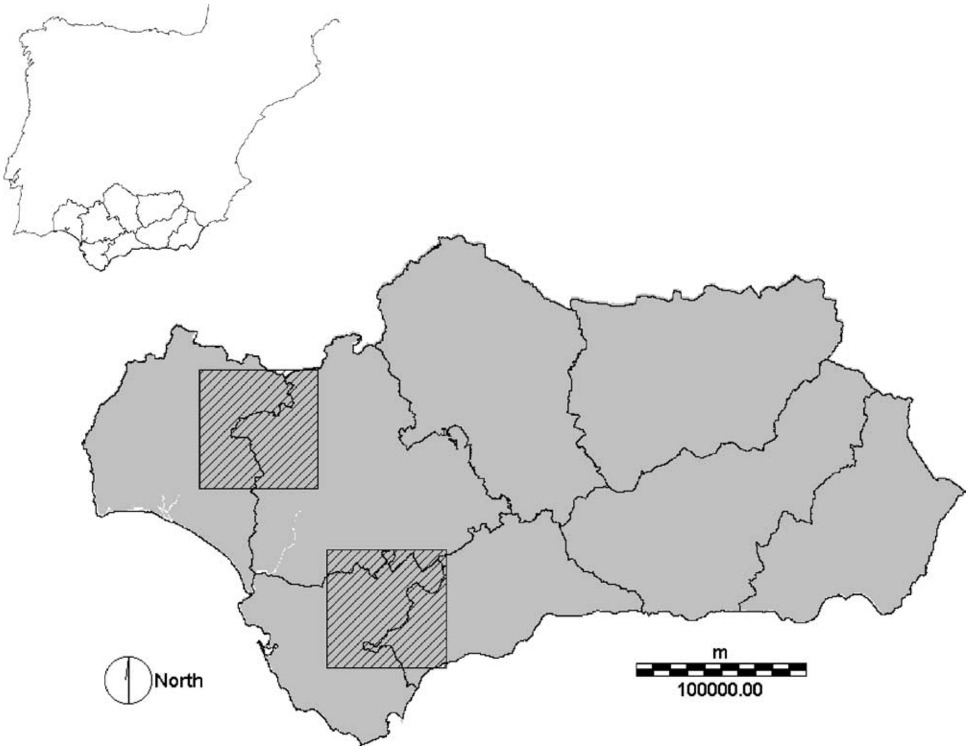

Fig. 1. Location of study areas.

oak *Quercus rotundifolia* and cork oak *Quercus suber* forests and open oak forest with pastures called "dehesas"). At each survey point the presence/absence of breeding bird species was recorded during 15 min. For subsequent modelling, we selected 857 points in natural and seminatural areas (that means, excluding those in agricultural and urban areas), and we selected 54 bird species that appeared in more than 5% of these sampling points (Appendix A).

2.1. Vegetation maps

We used seven different vegetation maps of each study area to derive the predictive variables to model bird distribution: Three of them were independent vegetation maps: (1) CORINE land-cover map of Europe from the European Environmental Agency (CORINE_250). Original data in raster format at 250 m resolution (v.12/1989) were obtained from the European Topic Center on Land Cover, Kiruna, Sweden. The CORINE map legend has 44 land-cover classes for the whole Europe. Source data correspond nominally to the period 1989–1991. (2) The SINAMBA land-use/land-cover digital map of Andalusia (SINAMBA_50) from the Environmental Department of the Junta de Andalucía. Original data in vector format were rasterised to 50 m resolution. The map legend has 112 classes. Source data correspond nominally to 1995. (3) A vegetation structure map derived from satellite images from 1999 to 2000 (SATELLITE_30, see below for details). Original data were in raster format at 30 m resolution and consisted of two maps in a continuous scale: degree of tree cover and degree of shrub cover. We generated another vegetation map (4) by combining information of SINAMBA map and satellite images (MIXED_30). This map at 30 m resolution consisted also in a tree cover and a shrub cover map and had a significantly greater accuracy than SATELLITE_30.

To test whether differences in predictive accuracy were due to differences in data quality, predictors used, or map resolution we generated another three maps: (5) SINAMBA_250 was obtained resampling the SINAMBA_50 map at 250 m resolution. (6) SINAMBA_250R was also obtained by resampling the SINAMBA_50 map at 250 m resolution but then using only the reduced set of predictors that could be measured in CORINE_250. (7) SATELLITE_50 was generated by resampling the SATELLITE_30 map at 50 m, the same resolution as SINAMBA_50.

In each vegetation map the original legend categories (or the values in a continuous scale) were reclassified into three categories of shrub cover and three categories of tree cover: (1) no cover, (2) disperse cover, and (3) dense cover (Fig. 2). With the help of a GIS we derived from each vegetation map at each bird survey point the set of vegetation structure and landscape predictors indicated in Table 1. These predictors included variables descriptive of vegetation structure in circles of 350 m in diameter centred in bird survey points and variables indicating distances to landscape features of different sizes. The same predictors were measured in each map with a few exceptions. The CORINE_250 map did not distinguish clearly between disperse and dense tree cover and between disperse and dense shrub cover in its legend, and some predictors involving these variables could not be measured (see Table 1). For this reason we generated the SINAMBA_250R map that had the same predictors as CORINE_250. Also, a few predictors changed their range of possible values when measured at 250 m resolution.

Extraction of variables from the GIS was done using IDRISI 32 (Eastman, 1999), IDRISI for Windows (Eastman, 1997) and MIRAMON (Pons, 2000).

2.2. Vegetation maps derived from satellite data

At each bird survey point the observer recorded structural attributes of vegetation that we had

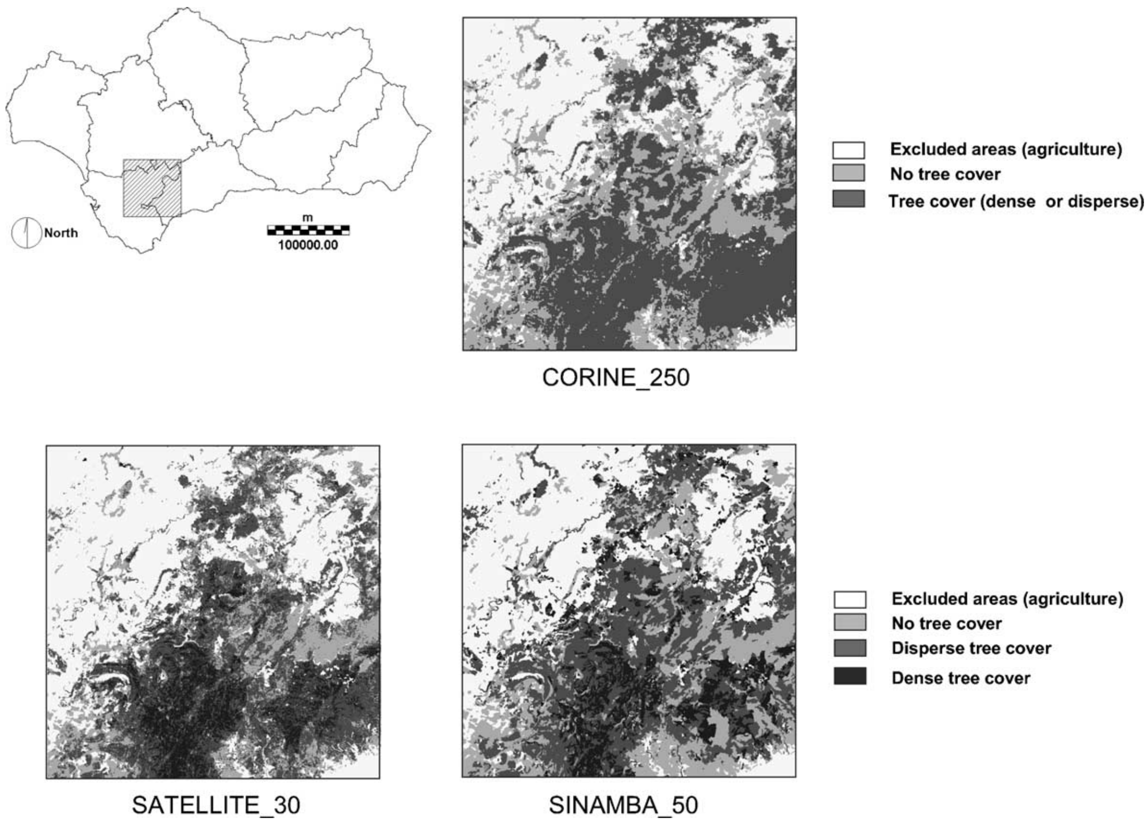

Fig. 2. Tree cover coverages from the CORINE_250, SINAMBA_50 and SATELLITE_30 maps for one of the study areas.

Table 1
List of variables measured in each map and used in the initial sets of potential predictors of bird distribution models

Variable description
Proportion of area covered by shrubland[a]
Proportion of area covered by forest (including both dense and disperse forest)
Length of boundaries between forested landcover categories and the rest of vegetation categories[b]
Length of boundaries between forest and shrubland
Compactness ratio of dense forest areas (an indirect estimate of surface-perimeter ratio)[b]
Proportion of area covered by disperse forest[b]
Proportion of area covered by dense forest[b]
Proportion of area covered by disperse shrubland[b]
Proportion of area covered by dense shrubland[b]
Distance to the nearest patch of shrubland (either dense or disperse)[c]
Distance to the nearest patch of forest (either dense or disperse)[c]
Distance to the nearest patch of dense forest[b,c]

[a] Variable values are estimated in circles of 350 m radius centred in bird point surveys. For example, "proportion of area covered by shrubland" is the fraction of 30 or 50 m pixels in the circle that have dense or disperse shrub cover.

[b] Variables that could not be measured in the CORINE_250 map and consequently were also excluded as potential predictors in the SINAMBA_250R models.

[c] Each distance was four variables: (1) distance to the nearest patch of any size, (2) distance to the nearest patch <2 ha, (3) distance to the nearest patch 2–10 ha in size, (4) distance to the nearest patch 10–100 ha in size.

considered a priori important for bird distribution (Table 2). We performed a principal components analysis (PCA) on these structural variables (excluding the survey points in agricultural or urban areas). The two first components explained 32.9 and 23.2%, respectively of the original variance (a total of 56.2%). The first component defined a gradient of tree cover (high loads for variables: mean tree DBH, and cover of trees >6 m high). The second component defined a gradient of shrub cover (high positive loads for variables like cover of shrubs <0.50 m tall, and cover of shrubs 0.5–2 m tall, and high negative load for cover of herbaceous vegetation). Then we used the first two components of the PCA as the response variable in a generalised additive model (GAM; Hastie and Tibshirani, 1990) with normal errors and identity link. We used as predictors reflectance values of bands 1–7, and NDVI of three Landsat scenes (TM and ETM+) for each study area corresponding to early spring, mid spring and summer of the years 1999 and 2000. Images were geometrically corrected with the aid of a Digital Elevation Model (Palà and Pons, 1995) and radiometrically calibrated according to Pons and Solé-Sugrañes model (1994). GAM models (J. Bustamante and R. Díaz-Delgado, unpublished data) explained 37–40% variance of the tree cover gradient (each study area respectively) and 21–30% variance of the shrub cover gradient. GAM models predicted tree cover and shrub cover values in a continuous scale for each 30 m pixel in the study area. We selected cut-points in this gradient to recode tree and shrub cover in three classes (no cover, disperse cover, and dense cover), so that surface covered by each tree cover and shrub cover class was as close as possible to that of the SINAMBA_50 map. The resulting coverages defined the SATELLITE_30 vegetation map. The tree cover models and shrub cover models improved significantly if the land-use/land-cover class of the SINAMBA map at the location of each sampling point was included as a factor. We refitted the GAM models for the tree cover and shrub cover gradients of each study area using the SINAMBA class as a factor, satellite reflectance values and NDVI values. These new GAM models explained 55–56% of the variance in tree cover and 26–49% of the variance in shrub cover. The gradients were reclassified to three discrete classes and generated the MIXED_30 map.

Table 2
Vegetation variables measured in circles 50 m in radius centred in bird survey points

Variables	Possible values
Cover of herbaceous vegetation (%)	<10, 10–50, >50
Cover of shrubs <0.5 m tall (%)	Absence, <25, >25
Cover of shrubs 0.5–2 m tall (%)	Absence, <25, >25
Cover of trees 2–6 m tall (%)	Absence, <25, >25
Cover of trees >6 m tall (%)	Absence, <25, >25
Mean diameter at breast height (DBH) of the five biggest trees (m)	Continuous
Number of trees with DBH >0.2 m in a circle 25 m radius	Integer

Each sampling point was classified into one of nine exclusive categories (Table 3) using the coordinates in the tree cover and shrub cover gradients of the PCA and the cut-points selected for the satellite vegetation maps. These points were used as ground-truth for all

Table 3
Categories used in vegetation maps (structural categories) to compare map quality (accuracy)

Categories
No tree cover, no shrub cover
No tree cover, disperse shrub cover
No tree cover, dense shrub cover
Disperse tree cover, no shrub cover
Disperse tree cover, disperse shrub cover
Disperse tree cover, dense shrub cover
Dense tree cover, no shrub cover
Dense tree cover, disperse shrub cover
Dense tree cover, dense shrub cover

vegetation maps. A confusion matrix was generated comparing ground-truth classification with classification from each map.

2.3. Predictive models for birds

We built a GAM (Hastie and Tibshirani, 1990) for the presence/absence of each bird species in each study area with binomial errors and logit link using as predictors the variables in Table 1. Seven models were generated for each bird species with the predictors derived from each one of the seven vegetation maps. We selected the variables to include in the models with a forward–backward stepwise selection from the complete set of predictive variables measured from each map (with the step.gam procedure of S-PLUS 2000; MathSoft, 1999). We started from a null model and tested each predictor sequentially as a smoothing spline with 3 degrees of freedom. The predictor that reduced the most the residual deviance was included in the model and the procedure was repeated until no more predictors improved the model. Then, we tried to simplify the resulting model by decreasing the complexity of each of the predictors included (by means of a smoothing spline with 2 degrees of freedom and a linear term). The criterion to enter, remove or simplify a term was the Akaike's Information Criterion (AIC; Sakamoto et al., 1986), that takes into account the reduction both in residual deviance and in residual degrees of freedom due to a certain predictor. Automatic procedures for selection of predictors have been criticised because they can yield ecologically implausible models (Greenland, 1989; James and McCulloch, 1990); but it is a method that allows for rapid development of models (Pearce and Ferrier, 2000b), and it has been shown empirically that frequently performs better than tedious manual selection techniques incorporating opinion of experts (Pearce et al., 2001). In our study, the random inclusion of spurious correlations in the predictive models could affect equally the models derived from each map and would not bias the comparison between models.

2.4. Comparison of predictive accuracy of maps

The predictive ability of each model was assessed by the area under the curve (AUC) of receiver operating characteristics (ROC) plots (Murtaugh, 1996; Pearce and Ferrier, 2000a). We evaluated the models with a data-splitting strategy by which models were built with a random selection of 75% of data and their predictive accuracy (AUC) was estimated on the other 25%. Data were reshuffled and the procedure was repeated 10 times to give a mean estimate of AUC (Appendix A). Differences among model types were tested with a repeated measures factorial ANOVA (with an error term due to species to control for the between-species variation, of no interest in this study). Preplanned comparisons (Montgomery, 2001) were carried out to test differences between particular models.

First we compared the different vegetation maps to see if they differed in their accuracy regarding the structural vegetation classes defined, using as ground-truth the vegetation data measured at the 857 bird survey points. To this aim, we followed a data-splitting strategy (building set with 75% of survey points, validation set with 25%) repeated 100 times. Second, we tested if there were differences in predictive accuracy of bird distribution related to the original map source of predictors: CORINE_250, SINAMBA_50 or SATELLITE_30. We explored this further testing if any differences in predictive ability could be related specifically to the map source (comparing CORINE_250 versus SINAMBA_250R), to the predictors used (SINAMBA_250R versus SINAMBA_250), or to the map spatial resolution (SINAMBA_250 versus SINAMBA_50, and SATELLITE_30 versus SATELLITE_50). Finally, we tested if a more accurate vegetation map derived from two sources (MIXED_30) had a greater predictive accuracy than the original maps (SINAMBA_50 and SATELLITE_30).

Table 4
Percentage of agreement between map categories and ground-truth data ($n = 857$) for each map

Map	Percentage of agreement	κ	S.E. of κ	Z value (P)
CORINE_250	30.0[a]	7.0	3.5	2.02 (0.02)
SINAMBA_250R	38.6	12.7	3.8	3.33 (0.0004)
SINAMBA_50	22.1	13.1	2.5	5.27 (<0.0001)
SATELLITE_50	29.1	19.2	2.9	6.54 (<0.0001)
SATELLITE_30	29.9	20.0	3.2	6.17 (<0.0001)
MIXED_30	31.6	21.6	3.5	6.17 (<0.0001)

The κ values and their significance that indicate the percentage improvement over a random classification.
[a] Percentage agreement value of CORINE_250 can only be compared with SINAMBA_250R that has a reduced set of four classes also: (1) no tree cover no shrub cover, (2) no tree cover disperse or dense shrub cover, (3) disperse or dense tree cover no shrub cover, and (4) disperse or dense tree cover disperse or dense shrub cover. (Values are means for 100 repetitions of a data-splitting evaluation, with a random sample of 75% of survey points to build the models in SATELLITE_50, SATELLITE_30 and MIXED_30 and 25% to evaluate them. For comparison, the estimates for the remaining maps have been calculated with 100 repetitions of a random sample of 25% of the points.)

3. Results

3.1. Accuracy of vegetation maps

Percentage of agreement and κ values (classification rate corrected for chance; Titus et al., 1984) indicated that greatest map quality (or accuracy) corresponded to the MIXED_30 map. Map quality declined in this order MIXED_30 > SATELLITE_30 > SATELLITE_50 > SINAMBA_50 > SINAMBA_250R > CORINE_250 (Table 4).

3.2. Bird distribution models

It was possible to build predictive models significantly better than a null model for 83% of the bird species ($n = 54$) with maps with 250 m resolution, and this figure increased to 91% of the species when predictors were derived from maps at spatial resolution of 50 or 30 m (Appendix A). Mean AUC for each map ranged from 0.65 (S.E. = 0.043) for SINAMBA_250R to 0.70 (S.E. = 0.044) for MIXED_30 (Table 5).

It is considered that predictive models should have an AUC > 0.7 to be of practical utility (Harrell, 2001). We tabulated the number of bird species with models with AUC > 0.7 for each map (Table 6), and tested if increasing map accuracy rendered more species with good models. There is a significant trend for a greater proportion of species with good models with increasing map accuracy, when maps are ordered according to their accuracy—κ value in Table 4 ($\chi^2 = 7.525$, d.f. $= 1$, $P = 0.0061$).

Contrary to what we expected, there were not significant differences in bird predictive ability (AUC values) when comparing the models derived from different data sources (CORINE_250 versus SINAMBA_50 versus SATELLITE_30) (Table 7). CORINE_250 differed from SINAMBA_50 in map accuracy, the number of predictors derived, and the spatial resolution of the source map.

Table 5
Bird prediction accuracy, mean AUC (and S.E.) values, for the models generated with the different maps

Model (map source)	AUC (S.E.)
CORINE_250	0.69 (0.043)
SINAMBA_250R	0.65 (0.043)
SINAMBA_250	0.69 (0.045)
SINAMBA_50	0.69 (0.043)
SATELLITE_50	0.69 (0.039)
SATELLITE_30	0.69 (0.042)
MIXED_30	0.70 (0.044)

Table 6
Number of bird species ($n = 54$) that had models with AUC > 0.7

Map source	AUC > 0.7
CORINE_250	19
SINAMBA_250R	11
SINAMBA_250	22
SINAMBA_50	21
SATELLITE_50	23
SATELLITE_30	23
MIXED_30	28

Maps are in order of increasing accuracy.

Table 7
Results of repeated measures ANOVA testing the effect in bird prediction accuracy (model AUC) of original independent map sources (CORINE_250, SINAMBA_50 and SATELLITE_30), and planned comparisons between them

Variable	d.f.	SS	MS	*F*	*P*
Error: species					
Residuals	53	0.901	0.017	–	–
Error: within					
Map source	2	<0.001	<0.001	0.069	0.933
CORINE_250 vs. SINAMBA_50	1	<0.001	<0.001	0.120	0.730
SATELLITE_30 vs. SINAMBA_50	1	<0.001	<0.001	0.019	0.891
Residuals	106	0.130	0.001		

Table 8
Results of repeated measures ANOVA testing the effect in bird prediction accuracy (model AUC) of map source type, and planned comparisons between them to test the effect of map source, spatial resolution and set of predictors derived

Variable	d.f.	SS	MS	*F*	*P*
Error: species					
Residuals	53	0.964	0.018	–	–
Error: within					
Map source type	3	0.050	0.167	14.17	<0.0001
Source (CORINE_250 vs. SINAMBA_250R)	1	0.030	0.030	25.78	<0.0001
Resolution (SINAMBA_50 vs. SINAMBA_250)	1	<0.001	<0.001	0.29	0.588
Predictors (SINAMBA_250 vs. SINAMBA_250R)	1	0.019	0.019	16.43	<0.0001
Residuals	159	0.188	0.0012		

To study the relative effect of each of these factors independently we compared the models derived from CORINE_250, SINAMBA_250R, SINAMBA_250 and SINAMBA_50 (Table 8). Reducing the number of predictors diminished significantly the predictive ability of models (SINAMBA_250 versus SINAMBA_250R: $F_{1,159} = 16.43$, $P < 0.0001$). There were not significant differences in predictive ability due to map resolution (SINAMBA_50 versus SINAMBA_250: $F_{1,159} = 0.29$, $P = 0.6$). There was also a very significant difference attributable to map source (CORINE_250 versus SINAMBA_250R: $F_{1,159} = 25.78$, $P < 0.0001$) but in a sense contrary to what we expected. The supposedly less accurate map source (CORINE) gave significantly more accurate bird maps.

Table 9
Results of repeated measures ANOVA testing the effect in bird prediction accuracy (model AUC) of map source type, and planned comparisons between them to test the effect of map source, resolution and map improvement

Variable	d.f.	SS	MS	*F*	*P*
Error: species					
Residuals	53	1.25	0.23	–	–
Error: within					
Map source type	3	0.004	0.001	1.30	0.275
Source (SINAMBA_50 vs. SATELLITE_50)	1	<0.001	<0.001	<0.01	0.957
Resolution (SATELLITE_30 vs. SATELLITE_50)	1	<0.001	<0.001	0.11	0.742
Improvement (MIXED_30 vs. SINAMBA_ 50)	1	0.004	0.004	3.80	0.053
Residuals	159	0.171	0.001		

Improving the quality of the maps by generating a MIXED map had a positive significant effect on the predictive ability of bird models derived from them, but degrading the spatial resolution did not have any significant effect (Table 9). Planned comparisons indicated that there was a statistically significant (but very slight) difference attributed to map quality improvement (SINAMBA_50 versus MIXED_30: $F_{1,159} = 3.80$, $P = 0.053$; SATELLITE_30 versus MIXED_30: $F_{1,159} = 3.94$, $P = 0.048$), while there were not statistically significant differences due to either the source (SINAMBA_50 versus SATELLITE_50: $F_{1,159} = 0.004$, $P = 0.96$) or the resolution (SATELLITE_30 versus SATELLITE_50: $F_{1,159} = 0.11$, $P = 0.74$).

4. Discussion

Our original vegetation maps show a gradient in quality (accuracy) for several reasons. The CORINE map has a coarser resolution (250 m), has a reduced set of land-cover classes (44 for the whole Europe), it does not reflect well differences in vegetation structure (for example, most classes do not distinguish between dense and disperse tree and shrub cover) and is 10 years older than our ground-truth data. The SINAMBA map has a finer spatial resolution (50 m), has more land-cover classes (that are easier to reclassify as disperse or dense tree and shrub cover), but is 5 years older than ground-truth data. The SATELLITE map has the finer spatial resolution (30 m), is contemporaneous with ground truth data and with bird surveys (1999–2000) and uses models to discriminate directly the structural variables we were interested in (degree of tree cover and degree of shrub cover). The confusion matrix of map classification and ground-truth data for sampling points indicates that map accuracy increases in a gradient: CORINE $<$ SINAMBA $<$ SATELLITE.

When the complete set of maps is considered there is a certain trend of increasing bird predictive ability with increasing map accuracy (Table 6), but this trend is not associated to the three different map sources (although they differ in accuracy). Measuring the same predictors on each map—that reflect characteristics of vegetation structure in a 350 m radius around the bird survey points, and distances to landscape features—we found that the three map sources (CORINE, SINAMBA, SATELITE) can predicted bird distribution equally well (mean AUC for the 54 species around 0.7 for each source map). There is only a slight effect in bird predictive ability if the map is improved combining two sources (MIXED map) or map quality is artificially degraded by reducing the number of classes (SINAMBA_250R).

This apparently surprising result may be due to two facts. First, we could have reached the maximum accuracy achievable with only broad-grained vegetation type data (Seoane et al., in press), as indeed there may be a limit in the accuracy of empirical models based on indirect variables to predict a response at a high spatial resolution (Fielding and Haworth, 1995; Manel et al., 1999; Rico Alcázar et al., 2001). Second, it is common that bird species select their habitats according to landscape features (rather than patch measures; Saab, 1999; Bailey et al., 2002), which may blur the distinction between maps made at a small range of spatial resolutions (30 m versus 250 m in our study) or even thematic resolutions (that is, vegetation maps with a very detailed legend may not render models more accurate than vegetation maps with a few—but relevant—general categories).

Thus, even the coarse grained CORINE map seems good enough to build broad habitat models. Before the analysis, we thought that CORINE map would be unsuccessful and, in an effort to study the effect of the quality of maps we made a new map by transforming the SINAMBA map to the resolution of CORINE map, and reducing the set of predictive variables extracted from it to match those available from CORINE. To our surprise resolution had no effect on models predictive ability and the resulting map equivalent to CORINE (SINAMBA_250R) was significantly worse. A factor that may be responsible for this is the selection of limits for land-cover classes (e.g. no shrub cover, disperse shrub cover, dense shrub cover) as many distance variables, that have a high predictive ability (Seoane et al., in press), are very sensitive on where these limits are set. Apparently our grouping of disperse and dense vegetation categories in the SINAMBA map was not equivalent to the land-cover classes of forest and shrubland in the CORINE map, and this had a stronger effect on the result than map accuracy.

Improving the SINAMBA map with satellite data (MIXED_30) improved slightly the predictive ability of models. The MIXED_30 map derives from models that are significantly better than those from the SATELLITE map, and also shows a better agreement with ground-truth data (Table 4). However, the small gain in predictive ability of the SATELLITE map (average of 1%) is certainly surpassed by the much greater effort needed to develop it (satellite imagery must be georeferenced and radiometrically corrected before the extraction of predictive variables). The improvement of mixed maps (combining satellite data and thematic information) over thematic maps could be greater for species whose habitats are poorly mapped. For example, in our study the riparian habitats are not very satisfactorily included in the thematic cartography (SINAMBA), and, accordingly, the models built with the mixed map tend to be better than those built with the former for riparian species (SINAMBA versus MIXED_30; *Cettia cetti*: 0.65 versus 0.76, *Hippolais polyglotta*: 0.66 versus 0.71, *Oriolus oriolus*: 0.53 versus 0.60).

Indeed, our results are encouraging for the development of habitat models because the available vegetation maps produced with a general purpose can show a fair predictive ability of bird distribution. It has to be considered that demographic stochasticity and dispersal can prevent a perfect adjustment between predictive models and wildlife distribution as Tyre et al. (2001) have shown with computer simulation models. Also, it is interesting to note that when general-purpose land-use/land-cover maps of enough accuracy and resolution are not available there is the alternative of deriving vegetation structural characteristics from satellite images resulting in maps with similar predictive ability of bird distribution.

Acknowledgements

This work is a contribution to the project "Predictive cartography of land birds: A pilot study in Western Andalusia", funded by the Dirección General de Enseñanza Superior e Investigación Científica (Ministry of Science and Technology) and FEDER funds from the European Union (EU), project # 1DF-97-0648. During the work, J.S. had a predoctoral fellowship from the Ministry of Education and Culture. The extensive field work presented in this paper could not have been done without the help and sense of humour of Daniel López Huertas, Luis M. Carrascal, and Mario Díaz who also brought intellectual insight to the project. We are grateful to the Environmental Department of the Junta de Andalucía for providing access to the 1995 SINAMBA land-use/land-cover map of Andalusia and to the European Environmental Agency (EEA) and the European Topic Center on Land Cover (ETC/LC) for providing access to the CORINE Land Cover databases.

Appendix A

Predictive accuracy (AUC) for each bird species and map source tested. Mean AUC values (and standard deviation between parentheses) where calculated with a data-splitting procedure (models were built with 75% of the observations and predictions made on the other 25%) the process was repeated 10 times reshuffling the data. Also shown is the number of presences of each species in the sample ($n = 857$).

	CORINE_250	SINAMBA_250R	SINAMBA_250	SINAMBA_50	SATELLITE_50	SATELLITE_30	MIXED_30	Presences
Aegithalos caudatus	0.61 (0.07)	0.60 (0.07)	0.57 (0.07)	0.57 (0.06)	0.61 (0.04)	0.62 (0.06)	0.62 (0.08)	51
Carduelis cannabina	0.77 (0.03)	0.79 (0.03)	0.79 (0.03)	0.73 (0.03)	0.73 (0.03)	0.74 (0.03)	0.78 (0.04)	225
Carduelis carduelis	0.70 (0.02)	0.66 (0.04)	0.66 (0.03)	0.66 (0.04)	0.66 (0.02)	0.65 (0.03)	0.68 (0.03)	365
Carduelis chloris	0.67 (0.04)	0.65 (0.03)	0.66 (0.04)	0.68 (0.03)	0.70 (0.03)	0.67 (0.03)	0.69 (0.02)	456
Certhia brachydactyla	0.69 (0.03)	0.72 (0.02)	0.72 (0.03)	0.70 (0.03)	0.70 (0.03)	0.72 (0.02)	0.72 (0.03)	297
Cettia cetti	0.60 (0.04)	0.62 (0.09)	0.64 (0.11)	0.65 (0.11)	0.70 (0.05)	0.65 (0.07)	0.76 (0.08)	47
Columba palumbus	0.59 (0.06)	0.59 (0.04)	0.59 (0.04)	0.68 (0.04)	0.62 (0.05)	0.59 (0.03)	0.61 (0.04)	141
Cuculus canorus	0.57 (0.04)	0.64 (0.05)	0.64 (0.05)	0.59 (0.03)	0.60 (0.04)	0.54 (0.03)	0.56 (0.02)	131
Cyanopica cyana	0.79 (0.04)	0.62 (0.04)	0.76 (0.04)	0.72 (0.05)	0.77 (0.02)	0.78 (0.03)	0.77 (0.03)	96
Delichon urbica	0.64 (0.08)	0.65 (0.06)	0.62 (0.07)	0.61 (0.05)	0.55 (0.03)	0.61 (0.05)	0.57 (0.04)	67
Dendrocopos major	0.81 (0.06)	0.76 (0.06)	0.83 (0.05)	0.81 (0.05)	0.80 (0.07)	0.75 (0.07)	0.76 (0.05)	52
Emberiza cia	0.62 (0.05)	0.60 (0.03)	0.64 (0.04)	0.61 (0.06)	0.59 (0.04)	0.59 (0.05)	0.59 (0.05)	88
Emberiza cirlus	0.62 (0.06)	0.54 (0.05)	0.62 (0.06)	0.62 (0.06)	0.58 (0.06)	0.62 (0.05)	0.67 (0.07)	47
Erithacus rubecula	0.84 (0.01)	0.74 (0.03)	0.83 (0.01)	0.86 (0.03)	0.89 (0.02)	0.91 (0.02)	0.85 (0.04)	153
Fringilla coelebs	0.79 (0.02)	0.79 (0.02)	0.81 (0.03)	0.79 (0.02)	0.79 (0.02)	0.77 (0.02)	0.81 (0.02)	485
Galerida theklae	0.63 (0.05)	0.63 (0.04)	0.70 (0.03)	0.68 (0.03)	0.74 (0.03)	0.71 (0.02)	0.65 (0.04)	128
Garrulus glandarius	0.62 (0.05)	0.65 (0.04)	0.62 (0.05)	0.60 (0.04)	0.64 (0.02)	0.65 (0.05)	0.66 (0.05)	80
Hippolais polyglotta	0.63 (0.02)	0.61 (0.06)	0.60 (0.06)	0.66 (0.04)	0.65 (0.02)	0.69 (0.06)	0.71 (0.04)	82
Hirundo daurica	0.65 (0.05)	0.54 (0.05)	0.57 (0.04)	0.56 (0.03)	0.63 (0.04)	0.59 (0.05)	0.57 (0.06)	78
Hirundo rustica	0.68 (0.04)	0.63 (0.03)	0.67 (0.03)	0.71 (0.03)	0.64 (0.03)	0.65 (0.04)	0.66 (0.02)	158
Jynx torquilla	0.77 (0.03)	0.61 (0.04)	0.62 (0.05)	0.73 (0.07)	0.72 (0.06)	0.72 (0.06)	0.73 (0.04)	68
Lanius senator	0.68 (0.03)	0.59 (0.03)	0.70 (0.03)	0.70 (0.06)	0.74 (0.03)	0.75 (0.04)	0.71 (0.04)	163
Lullula arborea	0.74 (0.04)	0.64 (0.03)	0.67 (0.05)	0.69 (0.03)	0.72 (0.03)	0.71 (0.04)	0.71 (0.03)	185
Luscinia megarhynchos	0.66 (0.02)	0.62 (0.02)	0.63 (0.03)	0.66 (0.03)	0.68 (0.04)	0.65 (0.04)	0.63 (0.03)	325
Merops apiaster	0.68 (0.04)	0.58 (0.02)	0.63 (0.03)	0.69 (0.03)	0.61 (0.03)	0.59 (0.02)	0.60 (0.04)	258
Miliaria calandra	0.76 (0.02)	0.68 (0.03)	0.77 (0.02)	0.78 (0.03)	0.80 (0.02)	0.81 (0.02)	0.78 (0.02)	368
Monticola solitarius	0.79 (0.07)	0.80 (0.05)	0.80 (0.04)	0.77 (0.06)	0.76 (0.05)	0.74 (0.06)	0.81 (0.05)	52
Oenanthe hispanica	0.63 (0.08)	0.69 (0.06)	0.71 (0.09)	0.76 (0.07)	0.76 (0.05)	0.72 (0.08)	0.72 (0.08)	66
Oriolus oriolus	0.58 (0.05)	0.55 (0.02)	0.55 (0.03)	0.53 (0.02)	0.55 (0.03)	0.56 (0.04)	0.60 (0.05)	73
Parus caeruleus	0.69 (0.04)	0.70 (0.03)	0.69 (0.03)	0.67 (0.03)	0.70 (0.03)	0.66 (0.04)	0.67 (0.03)	310
Parus cristatus	0.80 (0.04)	0.72 (0.04)	0.79 (0.04)	0.75 (0.03)	0.78 (0.03)	0.82 (0.03)	0.83 (0.03)	80
Parus major	0.59 (0.03)	0.59 (0.04)	0.59 (0.04)	0.59 (0.03)	0.60 (0.03)	0.59 (0.03)	0.60 (0.03)	353
Passer domesticus	0.68 (0.03)	0.67 (0.06)	0.71 (0.03)	0.65 (0.02)	0.65 (0.02)	0.67 (0.04)	0.66 (0.03)	142
Petronia petronia	0.73 (0.05)	0.61 (0.05)	0.72 (0.08)	0.76 (0.06)	0.72 (0.06)	0.70 (0.05)	0.74 (0.07)	85
Phoenicurus phoenicurus	0.78 (0.05)	0.69 (0.09)	0.75 (0.06)	0.76 (0.09)	0.80 (0.04)	0.87 (0.05)	0.91 (0.02)	41
Phylloscopus bonelli	0.72 (0.04)	0.65 (0.03)	0.73 (0.04)	0.77 (0.04)	0.72 (0.07)	0.73 (0.07)	0.77 (0.05)	90
Phylloscopus collybita	0.74 (0.03)	0.73 (0.03)	0.76 (0.02)	0.75 (0.04)	0.82 (0.04)	0.84 (0.04)	0.71 (0.04)	55

Appendix A (*Continued*)

	CORINE_250	SINAMBA_250R	SINAMBA_250	SINAMBA_50	SATELLITE_50	SATELLITE_30	MIXED_30	Presences
Picus viridis	0.65 (0.05)	0.64 (0.05)	0.70 (0.04)	0.65 (0.06)	0.59 (0.05)	0.60 (0.05)	0.61 (0.06)	74
Pyrrhocorax pyrrhocorax	0.89 (0.06)	0.78 (0.05)	0.85 (0.05)	0.90 (0.04)	0.84 (0.04)	0.83 (0.06)	0.81 (0.06)	56
Regulus ignicapillus	0.72 (0.04)	0.68 (0.04)	0.77 (0.04)	0.76 (0.05)	0.75 (0.06)	0.83 (0.04)	0.78 (0.07)	80
Saxicola torquata	0.74 (0.05)	0.75 (0.05)	0.74 (0.06)	0.73 (0.03)	0.72 (0.04)	0.74 (0.05)	0.78 (0.05)	105
Serinus serinus	0.60 (0.02)	0.58 (0.03)	0.60 (0.03)	0.60 (0.03)	0.59 (0.03)	0.62 (0.03)	0.61 (0.04)	408
Sitta europaea	0.73 (0.02)	0.75 (0.05)	0.75 (0.03)	0.75 (0.03)	0.71 (0.04)	0.73 (0.03)	0.74 (0.03)	170
Streptopelia turtur	0.58 (0.04)	0.54 (0.05)	0.62 (0.05)	0.64 (0.07)	0.58 (0.06)	0.56 (0.04)	0.59 (0.04)	112
Sturnus unicolor	0.69 (0.03)	0.57 (0.04)	0.64 (0.04)	0.64 (0.04)	0.65 (0.05)	0.59 (0.05)	0.71 (0.05)	119
Sylvia atricapilla	0.69 (0.02)	0.62 (0.03)	0.75 (0.03)	0.74 (0.04)	0.81 (0.03)	0.83 (0.03)	0.73 (0.03)	161
Sylvia cantillans	0.69 (0.08)	0.61 (0.08)	0.62 (0.07)	0.61 (0.07)	0.61 (0.04)	0.62 (0.05)	0.65 (0.07)	39
Sylvia hortensis	0.65 (0.08)	0.67 (0.04)	0.65 (0.08)	0.56 (0.04)	0.66 (0.06)	0.70 (0.05)	0.75 (0.06)	61
Sylvia melanocephala	0.62 (0.04)	0.66 (0.04)	0.73 (0.03)	0.68 (0.02)	0.64 (0.02)	0.60 (0.03)	0.68 (0.02)	441
Sylvia undata	0.63 (0.06)	0.70 (0.05)	0.69 (0.05)	0.67 (0.06)	0.61 (0.05)	0.68 (0.04)	0.73 (0.03)	90
Troglodytes troglodytes	0.64 (0.04)	0.55 (0.03)	0.68 (0.05)	0.69 (0.04)	0.66 (0.05)	0.65 (0.05)	0.60 (0.05)	169
Turdus merula	0.69 (0.03)	0.69 (0.03)	0.69 (0.03)	0.66 (0.03)	0.71 (0.03)	0.71 (0.03)	0.70 (0.03)	570
Turdus viscivorus	0.57 (0.05)	0.62 (0.07)	0.68 (0.08)	0.57 (0.05)	0.57 (0.05)	0.58 (0.05)	0.62 (0.06)	44
Upupa epops	0.76 (0.05)	0.68 (0.06)	0.74 (0.06)	0.74 (0.05)	0.69 (0.05)	0.69 (0.04)	0.71 (0.05)	162

References

Andries, A.M., Gulinck, H., Herremans, M., 1994. Spatial modelling of the barn owl *Tyto alba* habitat using landscape characteristics derived from SPOT data. Ecography 17, 278–287.

Avery, M.I., Haines-Young, R.H., 1990. Population estimates for the dunlin *Calidris alpina* derived from remotely sensed satellite imagery of the Flow Country of northern Scotland. Nature 344, 860–862.

Bailey, S.A., Haines-Young, R.H., Watkins, C., 2002. Species presence in fragmented landscapes: modelling of species requirements at the national level. Biol. Conserv. 108, 307–316.

Beutel, T.-S., Beeton, R.J.S., Baxter, G.S., 1999. Building better wildlife-habitat models. Ecography 22, 219–223.

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

Defries, R.S., Belward, A.S., 2000. Global and regional land cover characterization from satellite data: an introduction to the Special Issue. Int. J. Remote Sens. 21, 1083–1092.

Eastman, J.R., 1997. Idrisi for Windows: User's Guide. Clark Labs, Worcester.

Eastman, J.R., 1999. Idrisi32. Reference Guide. Clark Labs, Worcester.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.

Fielding, A.H., Haworth, P.F., 1995. Testing the generality of bird-habitat models. Conserv. Biol. 9, 1466–1481.

Franklin, J., Steadman, D.W., 1991. The potential for conservation of Polynesian birds through habitat mapping and species translocation. Conserv. Biol. 5, 506–519.

Greenland, S., 1989. Modelling variable selection in epidemiologic analysis. Am. J. Public Health 79, 719–732.

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

Harrell, F.E., 2001. Regression Modeling Strategies. Springer, New York.

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. Chapman & Hall, London.

James, F.C., McCulloch, C.E., 1990. Multivariate analysis in ecology and sistematics: panacea or Pandora's box? Ann. Rev. Ecol. Systemat. 21, 129–166.

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.

MathSoft, I., 1999. S-Plus 2000 User's Guide. Data Analysis Products Division, Seattle.

Montgomery, D.C., 2001. Design and Analysis of Experiments, fifth ed. John Wiley & Sons, New York.

Murtaugh, P.A., 1996. The statistical evaluation of ecological indicators. Ecol. Appl. 6, 132–139.

Ormerod, S.J., Watkinson, A.R., 2000. Editors' introduction: birds and agriculture. J. Appl. Ecol. 37, 699–705.

Palà, V., Pons, X., 1995. Incorporation of relief into a geometric correction based on polynomials. Photogrammet. Eng. Remote Sens. 61, 935–944.

Palmeirin, J.M., 1988. Automatic mapping of avian species habitat using satellite imagery. Oikos 52, 59–68.

Paruelo, J.M., Golluscio, R.A., 1994. Range assessment using remote sensing in Northwest Patagonia (Argentina). J. Range Manage. 47, 498–502.

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.L., Cherry, K., Drielsma, M., Ferrier, S., Whish, G., 2001. Incorporating expert opinion and fine-scale vegetation mapping into statistical models of faunal distribution. J. Appl. Ecol. 38, 412–424.

Pons, X., 2000. MiraMon: Geographic Information System and Remote Sensing Software. Universidad Autónoma de Barcelona, Barcelona.

Pons, X., Solé-Sugrañes, L., 1994. A simple radiometric correction model to improve automatic mapping of vegetation from multispectral satellite data. Remote Sens. Environ. 48, 191–204.

Rico Alcázar, L., Martínez, J.A., Morán, S., Navarro, J.R., Rico, D., 2001. Preferencias de hábitat del Águila-azor Perdicera (*Hieraaetus fasciatus*) en Alicante (E de España) a dos escalas espaciales. Ardeola 48, 55–62.

Roy, P.S., Sharma, K.P., Jain, A., 1996. Stratification of density in dry deciduous forest using satellite remote sensing digital data—an approach based on spectral indices. J. Biosci. 21, 723–734.

Saab, V., 1999. Importance of spatial scale to habitat use by breeding birds in riparian forests: a hierarchical analysis. Ecol. Appl. 9, 135–151.

Sakamoto, Y., Ishiguro, M., Kitagawa, G., 1986. Akaike Information Criterion Statistics. KTK Scientific Publishers, Tokyo.

Scott, J.M., Davies, F., Csuti, B., Noss, R., Butterfield, B., Groves, C., Anderson, H., Caicco, S., D'Erchia, F., Edwards Jr., T.C., Ulliman, J., Wright, R.G., 1993. GAP Analysis: a geographic approach to protection of biological diversity. Wildlife Monogr. 123, 1–41.

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

Stehman, S.V., 1996. Use of auxiliary data to improve the precision of estimators of thematic map accuracy. Remote Sens. Environ. 58, 169–176.

Titus, K., Mosher, J.A., Williams, B.K., 1984. Chance-corrected classification for use in discriminant analysis: ecological applications. Am. Midland Natur. 111, 1–7.

Tobalske, C., Tobalske, B.W., 1999. Using atlas data to model the distribution of woodpecker species in the Jura, France. The Condor 101, 472–483.

Trodd, N.M., 1996. Analysis and representation of heathland vegetation from near-ground level remotely-sensed data. Global Ecol. Biogeogr. Lett. 5, 206–216.

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

Wu, Y.C., Strahler, A.H., 1994. Remote estimation of crown size stand density and biomass on the Oregon transect. Ecol. Appl. 4, 299–312.