ORIGINAL ARTICLE

Predicting the distribution of four species of raptors (Aves: Accipitridae) in southern Spain: statistical models work better than existing maps

Javier Bustamante* and Javier Seoane

Department of Applied Biology, Estación Biológica de Doñana, CSIC, Sevilla, Spain

ABSTRACT

Aim To test the effectiveness of statistical models based on explanatory environmental variables vs. existing distribution information (maps and breeding atlas), for predicting the distribution of four species of raptors (family Accipitridae): common buzzard *Buteo buteo* (Linnaeus, 1758), short-toed eagle *Circaetus gallicus* (Gmelin, 1788), booted eagle *Hieraaetus pennatus* (Gmelin, 1788) and black kite *Milvus migrans* (Boddaert, 1783).

Location Andalusia, southern Spain.

Methods Generalized linear models of 10×10 km squares surveyed for the presence/absence of the species by road census. Statistical models use as predictors variables derived from topography, vegetation and land-use, and the geographical coordinates (to take account of possible spatial trends). Predictions from the models are compared with current distribution maps from the national breeding atlas and leading reference works.

Results The maps derived from statistical models for all four species were more predictive than the previously published range maps and the recent national breeding atlas. The best models incorporated both topographic and vegetation and land-use variables. Further, in three of the four species the inclusion of spatial coordinates to account for neighbourhood effects improved these models. Models for the common buzzard and black kite were highly predictive and easy to interpret from an ecological point of view, while models for short-toed eagle and, particularly, booted eagle were not so easy to interpret, but still predicted better than previous distribution information.

Main conclusions It is possible to build accurate predictive models for raptor distribution with a limited field survey using as predictors environmental variables derived from digital maps. These models integrated in a geographical information system produced distribution maps that were more accurate than previously published ones for the study species in the study area. Our study is an example of a methodology that could be used for many taxa and areas to improve unreliable distribution information.

*Correspondence: Javier Bustamante, Department of Applied Biology, Estación Biológica de Doñana, CSIC, Avda María Luisa s/n, 41013 Sevilla, Spain. E-mail: busta@ebd.csic.es

INTRODUCTION

Resource managers need to know how species are distributed, how abundant they are in the landscape and the relative suitability of different habitats for a given species. Distribution maps in reference books and field guides have been compiled traditionally from the records of localities where a species is known to be present (or was present in the past) plus a certain

degree of interpolation, expert knowledge, and guess, usually in unknown proportions (see e.g. Harrison, 1982). Distribution atlases provide maps that are built in a more systematic way, the study region is divided into regular areas (usually squares) of a certain size that are all visited and the species present (or breeding) are recorded. Atlases are costly to produce, because they require lots of fieldwork, and are not necessarily detailed enough for all applications (see e.g. Hagemaijer & Blair, 1997). One problem of atlases is that it is not always possible to distinguish unequivocally between real absences and areas that have not been well covered with fieldwork. In addition, in most atlases, areas where a species is abundant are not necessarily distinguished from those where the species is rare or accidental (Purroy, 1997).

Predictive models provide an alternative way to build distribution, abundance and/or habitat suitability maps for a species (see a review in Guisan & Zimmermann, 2000). They are based on the knowledge that species are habitat selective (Cody, 1985), and they assume it is possible to find environmental correlates of their distribution or abundance (Nicholls, 1989; Buckland & Elston, 1993). Considering that abundance is a rough indicator of habitat suitability (although this is not always true, see Van Horne, 1983; Vickery et al., 1992) it is possible to build habitat suitability maps with predictive models.

Geographical information systems (GIS) provide tools that allow one to measure easily environmental variables that are available in a digital format for any point where the distribution of the species has been surveyed. These variables can be tested statistically as potential predictors of the distribution of a given species. The resulting statistical models can generate predictive maps of the distribution of the species, again with the help of a GIS, provided that we have digital maps of the predictors in the study area (Pereira & Itami, 1991; Guisan et al., 1998; He et al., 1998; Rico Alcázar et al., 2001). A detailed cartography for the predictors, that can be updated easily and at low cost, must be available. Models based on predictors that are as difficult or costly to update as to survey the distribution of the species itself are of limited use for a resource manager.

It is expected that vegetation will be a better predictor of bird distribution than topography, because bird species tend to be associated with certain vegetation types more than with certain topographic features (Cody, 1985). However, it is easier and cheaper to generate a digital elevation model (DEM) for an area than an up-to-date vegetation map. In addition, land-use/land-cover maps are very dependent on the criteria used to generate them, so that two land-use/land-cover maps of the same area can be very different if they have been generated for different purposes or by different agencies (Cherrill & McClean, 1999). As vegetation and land-use over a territory is not independent of topography, we expect that topographic variables measured on a DEM would have a certain predictive ability of bird distribution. Finally, it is being increasingly recognized that the spatial autocorrelation should be taken into account to model species distribution (Keitt et al., 2002). This is because the occupancy pattern for a species may be partially driven by historical reason, intraspecific and inter-specific interactions (e.g. conspecific attraction, natal dispersal), and because adjacent, potentially occupied, areas may share a similar environment (Legendre et al., 2002).

In this study we tested the possibility of building predictive models for four species of raptors (Accipitridae): common buzzard Buteo buteo (Linnaeus, 1758), booted eagle Hieraaetus pennatus (Gmelin, 1788), short-toed eagle Circaetus gallicus (Gmelin, 1788) and black kite Milvus migrans (Boddaert, 1783) in southern Spain. We used field data from a sample covering 23% of the study area that was the by-product of a road census conducted to survey the breeding distribution of another raptor that is currently rare in the area (the red kite Milvus milvus Linnaeus, 1758). We used as predictors environmental variables easily derived from available digital maps (topographic variables derived from a DEM, broad vegetation variables derived from a land-use/land-cover digital map, and spatial coordinates to correct for neighbourhood effects). The four species considered are abundant, well known, and easy to watch and identify. And the study area is a well-known region. We think that published information available on the distribution of the four species in southern Spain, could be considered relatively reliable (certainly above the average), and much better than that of other less well-known taxa or that from remote regions. Models can be used either to provide an index of habitat suitability or to predict the distributions in areas not covered by a survey. We have no independent data that allow us to validate our models as an index of habitat suitability, but we can compare their ability in predicting the result of and independent survey vs. that of existing maps. The main point is neither to test if existing maps or atlases are wrong, nor if traditional ways of generating range maps or producing distribution atlas have flaws, but if statistical models using environmental predictors can be a productive way of deriving distribution information more reliable than the one currently available. We address the following five specific issues: (1) whether the information contained in a DEM is sufficient to predict the distribution of these four species, (2) whether the models derived from a DEM are better or worse than those derived only from a vegetation map, (3) whether models derived from one group of variables (topography or vegetation) can be improved with variables from the other group, (4) whether, by considering spatial effects, we can improve the predictive ability of habitat models, and (5) whether maps generated by the statistical models are better predicting distribution than the most up-to-date distribution maps or breeding atlas.

METHODS

Study area

Our study area was all the area covered by natural and semi-natural Mediterranean forest and scrubland in the Autonomous Community of Andalusia (southern Spain). We divided the area in 10×10 km squares using the UTM grid (the same grid used in the breeding atlas, see below), and selected all

continuous squares where dominant vegetation was Mediterranean forest or Mediterranean scrubland. The study area was divided in nine zones that covered a surface of 37,700 km². We selected a sample of 10 × 10 km squares in each zone to be censused for raptors (Fig. 1). Census work was initially designed to estimate the red kite breeding population – a species not considered in this paper, so the sample was stratified between zones, and the number of squares sampled within each zone was proportional to expected breeding density of red kites (Seoane *et al.*, 2003). Within each zone the squares to be censused were selected at random. After the censuses were carried we selected the four species of forest raptors that resulted more abundant: common buzzard, booted eagle, short-toed eagle and black kite, and tried to build predictive models for them. As we had excluded *a priori* those zones where dominant land-use is agriculture we did not make predictions from our models in those areas.

Raptor census

A total of 88 squares of 10 × 10 km were censused (23% of the study area, Fig. 1). In each square we covered *c.* 40 km of road census with a vehicle, driving along dirt roads or roads with little traffic allowing us to census at a speed of 20 km h^{-1} (a method shown to be adequate for censusing raptors such as the study species, Bibby *et al.*, 1992; Viñuela, 1997). Two persons carried out each census, one driving and the other recording all raptors. All squares were censused in spring 1996, between May and July. Although observers recorded number of individuals of each species and the coordinates of each contact with a raptor, for our models we only used the presence/absence of each species in each square. Prevalence (the ratio of positive squares to total sample) was similar for the four species: common buzzard (0.45), short-toed eagle (0.43), booted eagle (0.42) and black kite (0.39).

Current distribution maps

Current distribution data for the four species of raptors were taken from the published maps in the works of Perrins (1998), Clark (1999) and the New Spanish Atlas of Breeding Birds (named breeding atlas hereafter, Martí & Del Moral, 2003). The Handbook of the Birds of Europe, the Middle East and North Africa (Cramp & Simmons, 1980) can be considered the reference work more widely used to obtain information about European birds of prey. Its maps have been updated in the electronic abridged version (Perrins, 1998). Maps in Clark's (1999) raptor field guide are based on Perrins' synthesis, plus a certain degree of expert knowledge from the author. The New Spanish Atlas is supposed to be the most current and detailed source of information for breeding bird ranges in Spain. In this collective work data has been gathered by volunteers with field work during spring in the period 1998–2001 and can be considered a source of distribution information completely independent from published maps. Recordings are classified as probable, possible and confirmed breeding, and referenced to the 10 × 10 km UTM grid.

Predictive variables

Estimates of all environmental variables within each square were obtained in a GIS, using IDRISI for Windows v.2.0 (Eastman, 1997) and IDRISI32 v.1.01 (Eastman, 1999). Topographic variables (T) were estimated from a DEM of the study area (50 m horizontal resolution, 20 m vertical resolution) that had been derived from interpolation of 1 : 50,000 topographic maps. Overall DEM accuracy was checked by comparing a random sample of point coordinates with the altitude measured from 1 : 50,000 topographic maps. This gave an error of < 20 m (or one contour line).
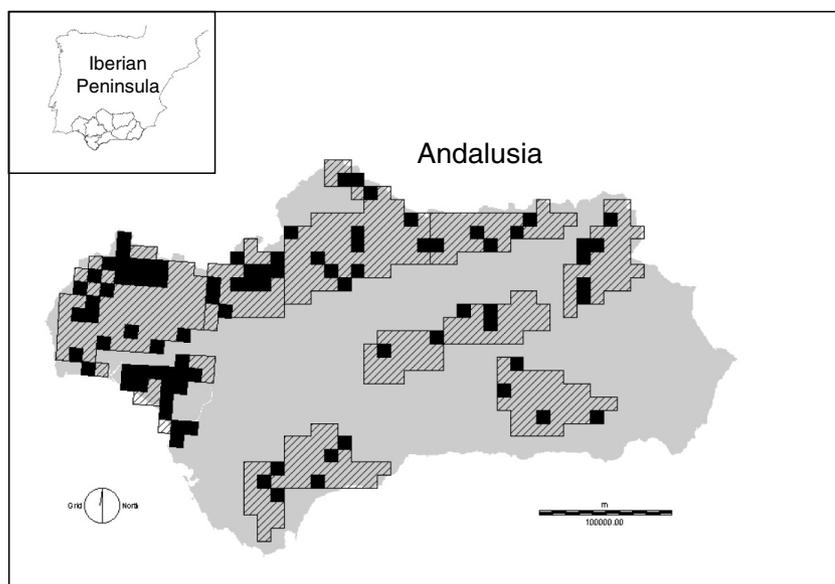
Figure 1 Location of study area (black striped polygons) and censused squares (solid squares) in the Autonomous Community of Andalusia (Spain).

Land-use/land-cover variables (U) were estimated from the SinambA (1995) digital land-use/land-cover map for Andalusia (Consejería de Medio Ambiente, Junta de Andalucía. Information on the 1991 map has been published in: Moreira & Fernández-Palacios, 1995). The map has 112 land-use/land-cover classes that have been updated with satellite images and aerial photographs and records all land-use/land-cover polygons that have more than 25 ha. The original coverage was rasterized at 50 m resolution, and all our variables were estimated on this raster image. The estimated vegetation index for each square was derived from the Experimental Calibrated Global Vegetation Index from NOAA-AVHRR (NOAA, 1992). Topographic and land-use/land-cover variables tested as predictors in the models are given in Table 1.

Statistical models

We used generalized linear models (GLM) (McCullagh & Nelder, 1989) with a binomial error and a logistic link to model the presence/absence of each raptor species in each 10×10 km square. We built different statistical models using as predictors a set of topographic variables (T models), a set of vegetation and land-use/land-cover variables (U models), both sets of variables (TU models), and both sets of variables plus spatial coordinates to correct for possible neighbourhood effects (TUC models). These final TUC models were also simplified with more stringent statistical criterion looking for possible causal relationships between environmental variables and species distribution (TUCS models).

Table 1 Description of explanatory variables tested in predictive models

Acronym	Description and source
Topographic variables	
Altitude	Mean altitude a.s.l. estimated from a DEM*
Slope	Mean slope (%) in the 10×10 km square for a 50-m pixel as estimated from the DEM using the SURFACE module of IDRISI32 that uses a rook's case procedure (Monmonier, 1982; Eastman, 1999)
Southern Orientation	Fraction of 50-m pixels showing a south-east to south-west orientation in the 10×10 km square. Orientation calculated from the DEM with SURFACE module of IDRISI32. Flat pixels are considered to have a southern orientation
Rivers	Fraction of 50-m pixels that are crossed by a river or a stream obtained converting from vector to raster the 1 : 50,000 hydrology coverage of Andalusia†
Land-use/land-cover variables	
Urban	Fraction of 50-m pixels classified in the land-use/land-cover raster map‡ in the urbanized and infrastructure classes
Agricultural	Fraction of 50-m pixels classified in the land-use/land-cover raster map in the agricultural classes
Natural	Fraction of 50-m pixels classified in the land-use/land-cover raster map as natural vegetation
Dense Forest	Fraction of 50-m pixels included in classes with tree coverage >50% in the land-use/land-cover raster map
Dispersed Forest	Fraction of 50-m pixels included in classes with tree coverage from 5% to 50% in the land-use/land-cover raster map
Dense Scrubland	Fraction of 50-m pixels included in classes with a scrub coverage >50% in the land-use/land-cover raster map
Disperse Scrubland	Fraction of 50-m pixels included in classes with a scrub coverage from 20% to 50% in the land-use/land-cover raster map
Pine Forest	Fraction of 50-m pixels included in classes of natural or planted coniferous forest (tree coverage >5%) in the land-use/land-cover raster map
Eucalyptus Forest	Fraction of 50-m pixels included in classes of planted *Eucalyptus* spp. forest (tree coverage >5%) in the land-use/land-cover raster map
Broad-leaved Forest	Fraction of 50-m pixels included in classes of natural broad-leaved forest (mostly *Quercus* spp. forest, tree coverage >5%) in the land-use/land-cover raster map
Olive/fruit Groves	Fraction of 50-m pixels included in classes of cultivated trees (mainly olive, fruit and almond groves) in the land-use/land-cover raster map
Forest Perimeter	Border in meters between forested and non-forested classes divided by number of 50-m pixels with land-use information
MSSVI	Mean spring–summer vegetation index (March to August). First a mean image for each month for the period April 1985–August 1991 was obtained from monthly values of the Experimental Calibrated Vegetation Index (version 2, monthly values from the US Geological Survey) (NOAA, 1992). Then the mean index was computed from mean monthly values. The NDVICOMP module of IDRISI32 was used to compute mean values as quadratic means of original values according to the formula $x' = \sqrt{\Sigma x_j^2 / n}$. The mean index (original data at *c.* 15 km resolution) was reprojected to UTM coordinates at 1 km resolution, filtered three times with a 3×3 mean filter, and mean values were extracted for 10×10 km squares

Source: SinambA (1999), Consejería de Medio Ambiente, Junta de Andalucía (unpubl. data).

*Digital elevation model of Andalusia obtained by interpolation of 20-m contours from the 1 : 50,000 topographic maps. Horizontal resolution: 50 m.

†Hydrology cover of Andalusia digitized from 1 : 50,000 topographic maps. All permanent rivers and streams are considered irrespective of their size.

‡Land-use/land-cover 1995 digital map of Andalusia [see Moreira & Fernández-Palacios (1995) for a description of land-use/land-cover 1991 digital map].

Model fitting was performed using S-Plus 2000 (MathSoft Inc., 1999). To find the best T, U, TU model for each species, we chose an automatic forward–backward stepwise variable selection procedure (procedure step.gam) testing in turns polynomial fits of all the explanatory variables in a set up to third degree. The step.gam procedure uses a stepwise search to select the best model in terms of Akaike's information criterion (AIC) that takes into account both the information explained by the model and its complexity (the lesser the AIC, the better the model Sakamoto *et al.*, 1986).

We obtained for each species (1) the best predictive model derived from the topographic set (T model), (2) the best predictive model derived from the land-use/land-cover set (U model), and (3) the best predictive model derived from all topographic or land-use/land-cover variables (TU model).

The best TU model was corrected for spatial effects fitting a nonparametric surface of latitude and longitude (a bivariate local regression surface, LOESS, with span equal to 0.5), and then we tested by removal (procedure step.gam) if T variables, U variables and the spatial surface remained significant. The resulting model was the TUC model.

In order to avoid overparametrization and to obtain a more parsimonius model, we further modified the TUC model by a backward stepwise procedure using the chi-square statistic (more conservative than AIC, Ludden *et al.*, 1994; Burnham & Anderson, 1998) and produced the TUCS models. We only kept variables (or terms of variables) when removing them yielded a significant increment of residual deviance. A variable was removed from the model if $P > 0.01$, but the order of a polynomial was reduced if $P > 0.05$. We believe that the TUCS models although usually having lower predictive power than the TUC models provide us with more appropriate cues of actual habitat selection by the species.

Success of predictions was measured by the area under the curve (AUC) of a receiver operating characteristic (ROC) plot (Hanley & McNeil, 1982; Swets, 1988), that measures the probability of correctly ranking any pair of squares, one with presence and the other with absence ($AUC = 0.5$ for chance success). Model validation was carried out using a leave-one-out (LOO) resampling technique. Each square was left out in turns and the model was refitted with the remaining 87 squares, the probability of the square left out was estimated from the model not containing it. We compared the result of the road census (presence/absence) for each square with the prediction of the model generated by resampling and calculated the AUC. The stability of each model was analysed by comparing the percentage of agreement between presence/absence predictions of the original model with those of the corresponding LOO model for each square. This measures the average effect of a single observation in the model predictions.

Finally, we compared how effective were statistical models vs. existing distribution data in predicting the results of the road census. The presence/absence recorded for each species in each 10×10 km square was predicted from the statistical model fitted using the other 87 squares. This cross-validation procedure allows us to compare the predictive ability of

models vs. that of maps, because the model used to predict the probability of occurrence in each square is not influenced by the recorded presence/absence in that square. The atlas records for a square were transformed into presence/absence by using three increasingly stringent criteria. The species was considered present: (1) if at least a possible breeding was recorded, (2) if at least a probable breeding was recorded, and (3) only if confirmed breeding was recorded. Breeding range maps for Andalusia from Perrins (1998) and Clark (1999) were projected to the UTM grid. We extracted presence/absence information for each 10×10 UTM square (here the criterion to consider a square holding the species was that more than 50% of its land surface was coloured as breeding range in the maps).

RESULTS

Common buzzard

The models that contained both topographic and land-use variables (TU model) predicted better that those based on one type of variables (T or U models) (Table 2). The spatial coordinates did not improve the TU model. All models predicted better than chance, but those with a single set of variables had a poor predictive ability ($AUC < 0.7$, Fig. 2). Models were quite stable, as measured by the percentage of coincidences with LOO models that were above 90% in every case (Fig. 3). The spatial predictions of the best model (TUC) are represented in Fig. 4. According to the simplified model (TUCS) the probability of presence of the common buzzard increases with the variables Dense Forest and Forest Perimeter while decreases with Slope and Eucalyptus Forest. These four variables were present in all the mixed models (TU and TUC model), but Altitude entered instead of Slope in the T model, and Pine Forest entered instead of Eucalyptus Forest in the U model (Table 2). However, within these pairs of variables the relationship with the response was very similar (a strong linear decrease in both cases). Altitude is positively correlated with Slope and Eucalyptus Forests are to a great extent correlated with Pine Forests so these pairs of explanatory variables probably indicate the same habitat and land-use features.

Neither the atlas nor the distribution maps from Perrins (1998) or Clark (1999) predicted well the results of the road census (mean $AUC \pm 2$ SE $= 0.56 \pm 0.11$, Fig. 2a). The value is not significantly different from 0.5 that is the result of a random classification.

Short-toed eagle

The models that contained both topographic and land-use variables (TU and TUC models) predicted better that those based on one type of variables (T or U models) (Table 2). All models predicted better than chance, having AUCs varying from a poor (T and U model, 0.70 and 0.65, respectively) to a fair discrimination ability (TU, TUC and TUCS models, 0.76, 0.77 and 0.75, respectively, Fig. 2). The model with highest corrected classification rate was the TUC model (Fig. 5).

Table 2 Best models for each species, based on topography (T), land-use/land-cover (U), topography and land-use/land-cover (TU), topography, land-use/land-cover and coordinates (TUC), and simplified TUC model (TUCS)

Species	Model	Variables	Residual deviance	d.f.
Common buzzard	T	Altitude + Rivers	112.1	85
	U	$(Dense\ Forest)^2 + (Disperse\ Scrubland)^2 + Pine\ Forest$	101.6	82
	TU	$Slope + (Dense\ Forest)^2 + Dense\ Scrubland + Eucalyptus\ Forest + Forest\ Perimeter + MSSVI$	80.6	80
	TUC	$Slope + (Dense\ Forest)^2 + Dense\ Scrubland + Eucalyptus\ Forest + Forest\ Perimeter + MSSVI$	80.6	80
	TUCS	$Slope + Dense\ Forest + Eucalyptus\ Forest + Forest\ Perimeter$	90.6	83
Short-toed eagle	T	$(Altitude)^2 + (Rivers)^2 + Southern\ Orientation$	97.8	82
	U	$(Dense\ Forest)^2 + (Pine\ Forest)^2 + (MSSVI)^2$	97.3	81
	TU	$(Rivers)^2 + Southern\ Orientation + (Dense\ Forest)^3 + (Pine\ Forest)^2 + (Eucalyptus\ Forest)^2 + (MSSVI)^2$	73.7	75
	TUC	$(Rivers)^2 + Southern\ Orientation + (Dense\ Forest)^3 + (Pine\ Forest)^2 + (Eucalyptus\ Forest)^2 + (MSSVI)^2 + LOESS(Latitude, Longitude)$	56.3	67
	TUCS	$(Rivers)^2 + MSSVI + LOESS(Latitude, Longitude)$	81.0	76
Booted eagle	T	$Slope + Southern\ Orientation$	112.8	85
	U	$(Dense\ Forest)^2$	109.3	85
	TU	$Southern\ Orientation + (Dense\ Forest)^2 + (Disperse\ Scrubland)^2 + (Broad\text{-}leaved\ Forest)^2$	89.7	80
	TUC	$Southern\ Orientation + (Dense\ Forest)^2 + (Disperse\ Scrubland)^2 + (Broad\text{-}leaved\ Forest)^2 + LOESS(Latitude, Longitude)$	86.5	78
	TUCS	$(Dense\ Forest)^2 + Disperse\ Scrubland + Southern\ Orientation$	96.6	83
Black kite	T	$(Slope)^3$	90.4	84
	U	$(Agricultural)^2 + (Natural)^2 + Disperse\ Scrubland + Olive/fruit\ Groves + (Pine\ Forest)^3 + (Eucalyptus\ Forest)^3$	68.7	75
	TU	$Slope + (Natural)^2 + (Disperse\ Scrubland)^2 + (Pine\ Forest)^3$	69.8	79
	TUC	$(Natural)^2 + (Pine\ Forest)^3 + LOESS(Latitude, Longitude)$	87.1	74
	TUCS	$Natural + (Pine\ Forest)^2 + LOESS(Latitude, Longitude)$	94.8	76
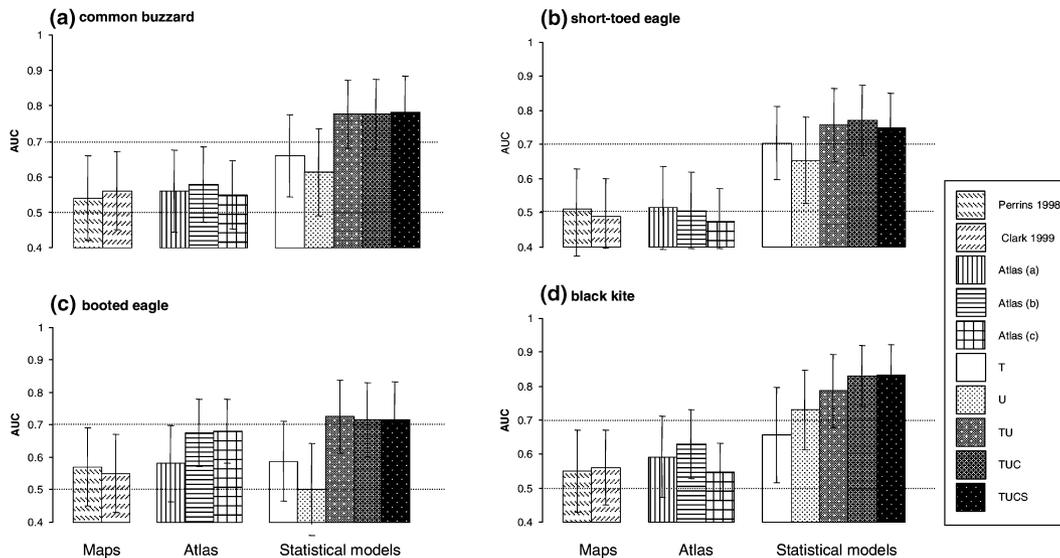
Figure 2 Predictive ability of recent distribution maps (Perrins, 1998; Clark, 1999), breeding atlas (Martí & Del Moral, 2003) and statistical models (resampled by a leave-one-out procedure) assessed with road census data. In the breeding atlas, we consider as presence three possible conditions: (a) at least a possible breeding was recorded, (b) at least a probable breeding was recorded, and (c) only if confirmed breeding was recorded. Variable sets used in statistical models: T, topography; U, land-use/land-cover; TU, topography and land-use/land-cover; TUC, topography, land-use/land-cover and coordinates; TUCS, simplified TUC model. Error bars are ±2 SE (estimated by bootstrapping).
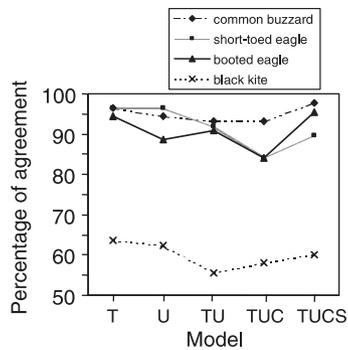
Figure 3 Percentage of agreement between predictions of the original models and those generated by the leave-one-out procedure.

Coincidence of prediction between each model and those generated by the LOO procedure were above 95% for the T and U models, and between 84% and 92% for the mixed TU, TUC and TUCS models (Fig. 3). The spatial predictions derived from all models showed a good spatial agreement.

The simplified (TUCS) model retained only two of the six topographic and habitat variables of the TU and TUC model (Table 2): Rivers and MSSVI. Both entered with a similar form in the single T and U models. Coordinates entered the TUC model without removing variables or modifying noticeably their form in the TU model. The shape of the spatial surface in the model indicates a contagious distribution and that the probability of presence of short-toed eagle increases in the north and central parts of the study area.

Neither the breeding atlas nor the distribution maps from Perrins (1998) or Clark (1999) predicted well the results of the road census for the short-toed eagle. They were not significantly better that a null model (mean $AUC \pm 2$ $SE = 0.50 \pm 0.11$, Fig. 2b).

Booted eagle

The models that contained both topographic and land-use variables (TU and TUC models) predicted better that those

based on one type of variables (T or U models) (Table 2). However, the latter did not predict clearly better than chance ($AUC = 0.59$ and 0.50, respectively) and the validation AUCs were relatively low for the TU,TUC and TUCS models (0.73, 0.72 and 0.72, respectively) (Fig. 2). Coincidences with predictions of LOO models varied between 84% (TUC) and 95% (TUCS, Fig. 3). The spatial predictions derived from the different models agree in some areas but disagree in others. Spatial prediction from the TUC model are given in Fig. 6.

The explanatory variables Southern Orientation and Dense Forest that were included in the simplified TUCS model, were the more stable of all the variables entering models for this species (single and mixed, Table 2). The probability of presence of the eagle increased with Southern Orientation following a positive linear function and also increased as a quadratic function of Dense Forest. The TUCS model also indicated an increase in the probability of presence with Disperse Scrubland that had not entered in U model. The spatial coordinates surface of the TUC model indicated a higher probability of presence in the north of the study area but it was no longer significant in the TUCS model. The coordinates entered the TUC model without removing or changing the form of the variables that were previously in the TU model.

The breeding atlas and the distribution maps for the booted eagle predicted the results of the road census slightly better than for other species, but only the atlas resulted significantly better than a null model (Fig. 2c). AUC values of maps and atlas were on average (mean $AUC \pm 2$ $SE = 0.61 \pm 0.11$) similar to the validation AUC from simple models (T and U) but were worse that the best statistical model (TU, TUC, Fig. 2c).

Black kite

Models for the black kite that contained both topographic and land-use variables (TU and TUC models) performed better
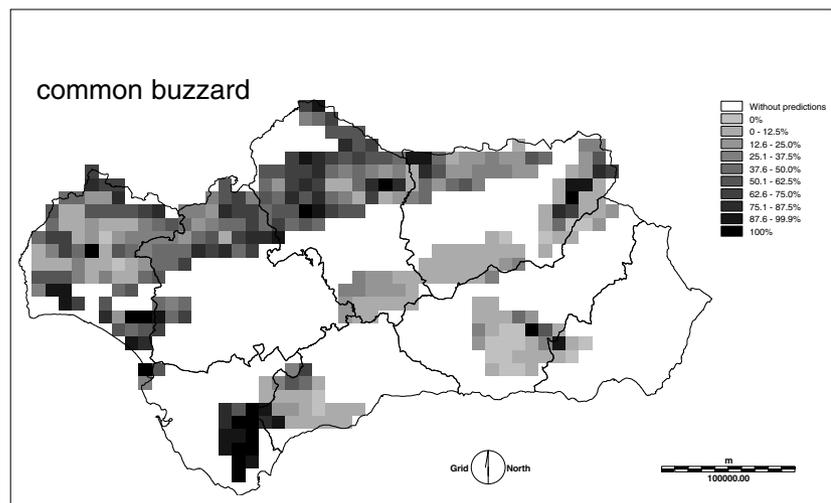
Figure 4 Spatial predictions of the TUC model (including topography, vegetation and coordinates as explanatory variables) for the common buzzard. White areas indicate where the model is not applicable.
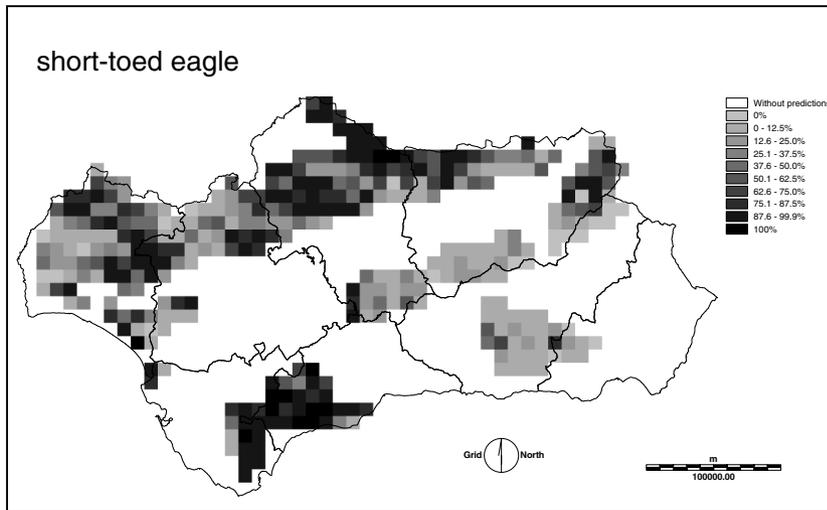
Figure 5 Spatial predictions of the TUC model (including topography, vegetation and coordinates as explanatory variables) for the short-toed eagle. White areas indicate where the model is not applicable.
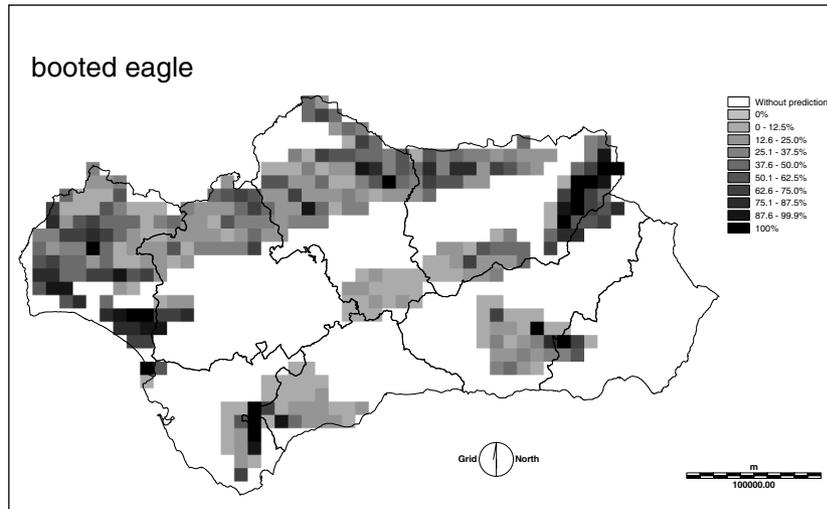
Figure 6 Spatial predictions of the TUC model (including topography, vegetation and coordinates as explanatory variables) for the booted eagle. White areas indicate where the model is not applicable.

than those based on one type of variables (T and U models), although the difference between the U and the TU model was only slight (6%, Table 2). Mixed models (TU, TUC and TUCS) reached the highest AUC in this study (0.79, 0.83 and 0.83, respectively). However, the stability of the models was low, as coincidences with predictions of LOO models were only between 55% and 65% (Fig. 3). The spatial predictions from the TUC model are given in Fig. 7.

The explanatory variables Natural and Pine Forest were present in both the single and the mixed models. In the TUCS model the probability of presence of the black kite decreased linearly with Natural and increased with Pine Forest. The spatial coordinates surface improved the TU model and indicated an increase in the probability of presence to the north and west of the study area.

The recordings from the breeding atlas and the published distribution maps of Perrins (1998) and Clark (1999) performed poorly in predicting the road census observations (mean $AUC \pm 2$ SE = 0.58 $\pm$ 0.11, Fig. 2d).

DISCUSSION

Common buzzard

Only the statistical mixed models (those including both topography and land-use/land-cover: TU, TUC and TUCS) for the common buzzard are satisfactory in terms of predictive ability (the threshold of practical utility is $AUC =$ 0.75 or $AUC = 0.70$ as given, respectively, by Elith, 2000; Harrell, 2001). Although, all of the statistical models generated better predictive maps than both the recordings of the breeding atlas and the distribution maps of Perrins (1998) and Clark (1999).

The variables that entered in the simplified TUCS model are easily interpreted and are in agreement with what we expected before building the models: the common buzzard seems to favour forested areas interdigitated with other open habitats, and avoids poor homogeneously reforested land (areas dominated by eucalyptus and pine trees). This kind of
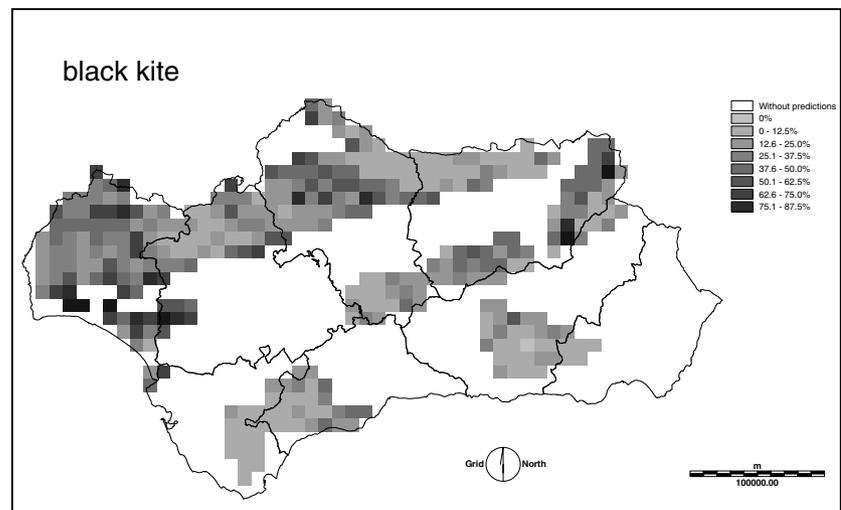
Figure 7 Spatial predictions of the TUC model (including topography, vegetation and coordinates as explanatory variables) for the black kite. White areas indicate where models are not applicable.

habitat is most common at low and medium altitude (where Slope tends to be moderate). This preference with forested habitat has also been seen in other Mediterranean areas in Spain (Sánchez-Zapata & Calvo, 1999), and a strong relationship between breeding distribution and border between forest and open habitats was also apparent in the previous study and in another study in Scotland (Austin *et al.*, 1996).

Short-toed eagle

The statistical models for the short-toed eagle were fairly accurate and predicted better than both the recordings of the atlas and the distribution maps (that resulted particularly poor). However, the best models (the TU and TUC models) have numerous variables and are difficult to interpret. Some of these may have entered by chance, as our more strict procedure of variable removal led to a very simplified TUCS model. This simple model is still difficult to interpret from the point of view of the ecology of the species. It suggests a selection for the more productive areas regarding vegetation and a relevant contagious distribution identified by the significance of the spatial coordinates.

Booted eagle

The statistical models for the booted eagle were relatively unsatisfactory. The predictive ability of the best models (TU, TUC and TUCS) was only slightly above the threshold of practical utility. *AUC* values were above those of the distribution maps and the atlas, but confirmed breeding in the atlas (Fig 2c) was only 5% worse that the TU model. The models were also difficult to interpret. They suggest a slight spatial effect, as coordinates entered the TUC model (but were rejected in the stricter TUCS model). TUCS model suggest also an expected preference for more forested squares and a positive relation (although slight) with Disperse Scrubland and

Southern Orientation. A similar preference for forested areas and areas of scrubland was found by Sánchez-Zapata & Calvo (1999) in south-eastern Spain.

Black kite

Statistical mixed models for this species have the greatest predictive ability among those generated in this work (and single models for this species were the best among single models). This was to be expected, as the black kite is the species with a more localized distribution in the west of the study area (it is abundant mainly in the Doñana National Park and surrounding areas), and thus it is *a priori* easy to model: any variable that identifies the clumped zone of distribution will have a high predictive ability. This seems to be the case for the positive relation between the presence of the species and Pine Forest (TUCS model), as the habitat the black kite occupies in Doñana is mainly pine-tree forest. However, TUCS model shows a negative relation with Natural that could be closer to the actual habitat selection of this species, known to breed and feed in humanized areas more frequently than the rest of the raptors analysed in this work. Finally, a clumped distribution could also facilitate making accurate maps in both atlas work (because main breeding places are easily located) and reference guides (because breeding areas are supposed to be known to experts), but it does not seem to be the case for the black kite.

GENERAL CONCLUSIONS

The models for the common buzzard and the black kite were good, and were based on environmental variables with a possible causal relationship with each species distribution, although the ones for the black kite were not robust. The models for the short-toed eagle and booted eagle, although good in their predictive ability, were difficult to interpret ecologically and are probably based on variables with no causal

relationship with the distribution of the species. These models provide only a limited insight on the factors that affect the distribution of three of the four species (black kite, short-toed eagle and booted eagle), so one should not expect that they can make accurate predictions outside the study area.

Our results indicate that topography and vegetation (as derived from a land-use/land-cover map) have a certain predictive ability on the distribution of forest raptors, but neither of them alone seems to be able to provide accurate predictions of the distribution of the species we studied. Mixed models were needed to obtain a fair (40–70%) improvement over chance in predictions, indicating that both set of variables provide a different information content. Probably topographic variables complement the information on changes in habitat/vegetation that are not adequately covered in the land-use/land-cover map. Our results are in agreement with those of Beard *et al.* (1999) who found that models based on vegetation, climate or spatial autocorrelation were better than null models to predict the distribution of birds in Idaho, but that the best predictive models were those combining variables from two sets. In addition, as it has been shown in other studies (Smith, 1994; Augustin *et al.*, 1996; Chou & Soret, 1996; Merrill *et al.*, 1999; Osborne *et al.*, 2001), there is much to gain in incorporating terms in the model building that take account of spatial effects. These can be caused by historical reasons or geographical barriers limiting the distribution a species, by the habitat being more similar between neighbouring areas in an environmental factor not considered within the predictors, or because the probability of finding a individual in a place may not be independent of the probability of finding individuals in neighbouring places (Augustin *et al.*, 1996). Raptors are very mobile and their distribution may be less influenced by local habitat features than in other animal groups (Chou & Soret, 1996).

Our results clearly show that it is possible to obtain statistical models for all four species (TUC maps, Figs 4–7) that predict their distribution in the study area better than chance and, perhaps with the exception of those for the booted eagle, were fairly accurate and significantly better than published distribution maps (Perrins, 1998; Clark, 1999) and recent breeding atlas (Martí & Del Moral, 2003). It could be argued that published distribution maps are not intended to predict accurately the probability of finding the species in the conditions of extent and resolution of our study. But we are afraid that those distribution maps are currently used as the best information available on bird distribution in environmental impact assessment studies (Díaz *et al.*, 2001), commonly with smaller extents and greater spatial resolutions. Even the breeding atlas that is based on recent fieldwork at the same resolution of our study (10 × 10 km), was not very accurate in predicting the results of the road census. It could be argued the atlas will record the areas in which a species breeds and our models predict the areas in which a species can be observed in spring (there may be a different selection for breeding and activity areas, or non-breeders could have a different spatial distribution or habitat selection than

breeders, see e.g. Bustamante *et al.*, 1997). We think, however, that our comparison is reasonable because the atlas considers also observations of individuals as signals for breeding in an area (e.g. an individual seen during the breeding period in an apparently suitable habitat is taken as a possible breeding record, Martí & Del Moral, 2003). Moreover, at the broad resolution of our study (100 km^2) we would not expect such a decoupling between breeding and activity areas, or between breeders and non-breeders. In our opinion, the low predictive power of the atlas is probably due to spatial heterogeneity in field effort and that differences in abundance are not properly reflected in the results.

Finally, our study shows that it is possible to produce accurate predictive maps for the distribution of raptors using variables derived from available digital environmental maps elaborated for other purposes. This is a pilot study with a limited survey on a small study area, but for all four species it was possible to obtain models that predicted consistently better the distribution than the best information published. Caution should be taken when using these models as their good predictive accuracy does not necessarily imply that they are good explanatory models of the habitat selection of each species (MacNally, 2000), and the reported predictive accuracy only applies to the study area.

ACKNOWLEDGMENTS

REFERENCES

Augustin, N.H., Mugglestone, M.A. & Buckland, S.T. (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, **33**, 339–347.

Austin, G.E., Thomas, C.J., Houston, D.C. & Thompson, D.B.A. (1996) Predicting the spatial distribution of buzzard *Buteo buteo* nesting areas using a geographical information system and remote sensing. *Journal of Applied Ecology*, **33**, 1541–1550.

Beard, K.H., Hengartner, N. & Skelly, D.K. (1999) Effectiveness of predicting breeding bird distributions using probabilistic models. *Conservation Biology*, **13**, 1108–1116.

Bibby, C.J., Burgess, N.D. & Hill, D.A. (1992) *Bird census techniques*. Academic Press, London.

Buckland, S.T. & Elston, D.A. (1993) Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology*, **30**, 478–495.

Burnham, K.P. & Anderson, D.R. (1998) *Model selection and inference: a practical information-theoretic approach*. Springer-Verlag, New-York.

Bustamante, J., Donázar, J.A., Hiraldo, F., Ceballos, O. & Travaini, A. (1997) Differential habitat selection by immature and adult Grey Eagle-buzzards *Geranoaetus melanoleucus*. *Ibis*, **139**, 322–330.

Cherrill, A. & McClean, C. (1999) Between-observer variation in the application of a standard method of habitat mapping by environmental consultants in the UK. *Journal of Applied Ecology*, **36**, 989–1008.

Chou, Y.-H. & Soret, S. (1996) Neighborhood effects in bird distributions, Navarre, Spain. *Environmental Management*, **20**, 675–687.

Clark, W.S. (1999) *A field guide to the raptors of Europe, the Middle East, and North Africa*. Oxford University Press, Oxford.

Cody, M.L. (1985) *Habitat selection in birds*. Academic Press, Inc., San Diego, CA.

Cramp, S. & Simmons, K.E.L. (1980) *Handbook of the birds of Europe, the Middle East, and North Africa*. Oxford University Press, Oxford.

Díaz, M., Illera, J.C. & Hedo, D. (2001) Strategic environmental assessment of plans and programs: a methodology for estimating effects on biodiversity. *Environmental Management*, **28**, 267–279.

Eastman, J.R. (1997) *Idrisi for Windows. User's guide*. Clark Labs, Clark University, Worcester.

Eastman, J.R. (1999) *Idrisi32. Reference guide*. Clark Labs, Clark University, Worcester.

Elith, J. (2000) Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. *Quantitative methods for conservation biology* (ed. by M. Burgman), pp. 39–58. Springer, New York.

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

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

Hagemaijer, E.J.M. & Blair, M.J. (1997) *The EBCC atlas of European breeding birds: their distribution and abundance*. T & A D Poyser, London.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

Harrell, F.E. (2001) *Regression modeling strategies*. Springer, New York.

Harrison, C. (1982) *An atlas of the birds of the Western Palearctic*. Collins, London.

He, H.S., Mladenoff, D.J., Radeloff, V.C. & Crow, T.R. (1998) Integration of GIS data and classified satellite imagery for regional forest assessment. *Ecological Applications*, **8**, 1072–1083.

Keitt, T.H., Bjornstad, O.N., Dixon, P.M. & Citron-Pousty, S. (2002) Accounting for spatial pattern when modeling organism–environment interactions. *Ecography*, **25**, 616–625.

Legendre, P., Dale, M.R.T., Fortin, M.-J., Gurevitch, J., Hohn, M. & Myers, D. (2002) The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography*, **25**, 601–615.

Ludden, T.M., Beal, S.L. & Sheiner, L. (1994) Comparison of the Akaike information criterion, the Schwarz criterion and the F test as guides to model selection. *Journal of Pharmacokinetics and Biopharmaceutics*, **2**, 431–445.

MacNally, R. (2000) Regression and model-building in conservation biology, biogeography and ecology: the distinction between – and reconciliation of – 'predictive' and 'explanatory' models. *Biodiversity and Conservation*, **9**, 655–671.

Martí, R. & Del Moral, J.C. (2003) *Atlas de las aves reproductoras de España*. Dirección General de la Conservación de la Naturaleza – Sociedad Española de Ornitología, Madrid.

MathSoft Inc. (1999) *S-Plus 2000 user's guide*. Data Analysis Products Division, Seattle, WA.

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*. Chapman & Hall/CRC, London.

Merrill, T., Mattson, D.J., Wright, R.G. & Quigley, H.B. (1999) Defining landscape suitable for restoration of grizzly bears *Ursus arctos* in Idaho. *Biological Conservation*, **87**, 231–248.

Monmonier, M. (1982) *Computer-assisted cartography: principles and prospects*, pp. 76–80. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Moreira, J.M. & Fernández-Palacios, A. (1995) *Usos y coberturas vegetales del suelo en Andalucía. Seguimiento a través de imágenes de satélite*. Consejería de Medio Ambiente, Junta de Andalucía, Sevilla.

Nicholls, A.O. (1989) How to make biological surveys go further with generalised linear models. *Biological Conservation*, **50**, 51–75.

NOAA (1992) *Experimental calibrated global vegetation index from NOAA's advanced very high resolution radiometer, 1985–1991 (KGRD: 28)*. NOAA, Boulder, CO.

Osborne, P.E., Alonso, J.C. & Bryant, R.G. (2001) Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology*, **38**, 458–471.

Pereira, J.M.C. & Itami, R.C. (1991) GIS-based habitat modelling using logistic multiple regression: a study of the Mt. Graham red squirrel. *Photogrammetric Engineering and Remote Sensing*, **57**, 1475–1486.

Perrins, C. (1998) *The complete birds of the Western Palearctic on CD-ROM.* Oxford University Press, Oxford.

Purroy, F.J. (1997) *Atlas de las aves de España (1975–1995).* Lynx editions, Barcelona.

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.

Sakamoto, Y., Ishiguro, M. & Kitagawa, G. (1986) *Akaike information criterion statistics.* KTK Scientific Publishers, Tokyo.

Sánchez-Zapata, J.A. & Calvo, J.F. (1999) Raptor distribution in relation to landscape composition in semi-arid Mediterranean habitats. *Journal of Applied Ecology*, **36,** 254–262.

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. *Biological Conservation*, **111,** 401–414.

Smith, P.A. (1994) Autocorrelation in logistic regression modelling of species' distributions. *Global Ecology and Biogeography Letters*, **4,** 47–61.

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

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

Vickery, P.D., Hunter, J.M.L. & Wells, J.V. (1992) Is density an indicator of breeding success? *The Auk*, **109,** 706–710.

Viñuela, J. (1997) Road transects as a large-scale census method for raptors: the case of the red kite *Milvus mivus* in Spain. *Bird Study*, **44,** 155–165.

BIOSKETCHES

Javier Bustamante is a Tenured Scientist at the Conservation Biology Group of the Department of Applied Biology at Doñana Biological Station, and currently member of the Board of Governors of the Society for Conservation Biology – European Section. His current research interests include modelling species distribution and using remote sensing to derive environmental predictors for these models. He is expert in the ecology of birds of prey.

Javier Seoane has recently finished a Doctoral thesis on 'Predictive cartography of land-bird distribution: a pilot study in Western Andalusia'.