

Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards

P.E. OSBORNE†, J.C. ALONSO‡ and R.G. BRYANT†*

†Department of Environmental Science, University of Stirling, Stirling FK9 4LA, UK; and ‡Museo Nacional de Ciencias Naturales, CSIC, José Gutiérrez Abascal 2, 28006, Madrid, Spain

Summary

1. Many species are adversely affected by human activities at large spatial scales and their conservation requires detailed information on distributions. Intensive ground surveys cannot keep pace with the rate of land-use change over large areas and new methods are needed for regional-scale mapping.

2. We present predictive models for great bustards in central Spain based on readily available advanced very high resolution radiometer (AVHRR) satellite imagery combined with mapped features in the form of geographic information system (GIS) data layers. As AVHRR imagery is coarse-grained, we used a 12-month time series to improve the definition of habitat types. The GIS data comprised measures of proximity to features likely to cause disturbance and a digital terrain model to allow for preference for certain topographies.

3. We used logistic regression to model the above data, including an autologistic term to account for spatial autocorrelation. The results from models were combined using Bayesian integration, and model performance was assessed using receiver operating characteristics plots.

4. Sites occupied by bustards had significantly lower densities of roads, buildings, railways and rivers than randomly selected survey points. Bustards also occurred within a narrower range of elevations and at locations with significantly less variable terrain.

5. Logistic regression analysis showed that roads, buildings, rivers and terrain all contributed significantly to the difference between occupied and random sites. The Bayesian integrated probability model showed an excellent agreement with the original census data and predicted suitable areas not presently occupied.

6. The great bustard's distribution is highly fragmented and vacant habitat patches may occur for a variety of reasons, including the species' very strong fidelity to traditional sites through conspecific attraction. This may limit recolonization of previously occupied sites.

7. We conclude that AVHRR satellite imagery and GIS data sets have potential to map distributions at large spatial scales and could be applied to other species. While models based on imagery alone can provide accurate predictions of bustard habitats at some spatial scales, terrain and human influence are also significant predictors and are needed for finer scale modelling.

Key-words: autologistic models, distributions, fragmentation, logistic regression, satellite imagery.

Journal of Applied Ecology (2001) **38**, 458–471

Introduction

Many wild species are adversely affected by human-induced changes in land use that operate over very large spatial scales. For example, in Europe agricultural

policy change and its consequent effects on farming practice have profoundly influenced many bird species (O'Connor & Shrubbs 1986; Pain & Pienkowski 1997). Among these is the great bustard *Otis tarda* L., a globally threatened species that has suffered dramatic declines (Collar & Andrew 1988; Heredia, Rose & Painter 1996). Although the reasons for such declines are not completely understood, it seems that agriculture intensification and habitat fragmentation due to human activities have played a decisive role. The species' stronghold is

Correspondence: Dr P.E. Osborne (fax 01786 467843; e-mail peo1@stir.ac.uk).

*Present address: Department of Geography, University of Sheffield, Winter Street, Sheffield S10 2TN, UK.

now the agricultural landscape of Spain, with an estimated population of 20 000 birds (Alonso & Alonso 1996), more than half the world total (Hidalgo de Trucios 1990; Del Hoyo, Elliot & Sargatal 1996). The population was probably declining until hunting was outlawed in Spain in 1980 and is now thought to be stable at best. Its conservation is still threatened by habitat fragmentation over most of the Iberian peninsula. Recent dispersal studies using individual marking and radiotracking techniques have shown that although the species is capable of performing considerable seasonal migration, individuals display a marked site fidelity to their breeding areas (Alonso, Morales & Alonso 2000; Morales *et al.* 2000). Once traditional sites are lost, this behaviour may restrict the potential to establish new populations elsewhere.

Monitoring national or regional changes in great bustard distributions and numbers through field surveys cannot realistically keep pace with the rate of agricultural and infrastructure development. This is equally true for large numbers of other species that require assessments of the reasons for population decline. Such knowledge is essential for compiling conservation management action plans under international conventions and legislation such as the Biodiversity Convention and the Birds Directive. There is thus an urgent need to develop ways for mapping threatened species at large spatial scales with reduced field effort (Gaston & Blackburn 1995; Williams *et al.* 1997). In this paper we present the results from a pilot study that attempted to model the breeding distribution of great bustards in central Spain from remotely sensed data and digitally mapped data layers. Large-scale studies continue to pose major challenges in applied ecology and model development may provide ecological insight at scales where manipulation is not possible (Ormerod, Pienkowski & Watkinson 1999; Caldow & Racey 2000).

Both bustard sexes are highly aggregated in early spring, when the surveys were conducted (males usually in a single flock and females in just a few flocks). Until recently great bustards in Iberia were considered sedentary in the vicinity of breeding leks (areas for male sexual exhibition and copulation). However, work on radio-marked birds has demonstrated that both sexes behave as partial migrants between the lek site and post-breeding or wintering areas (Alonso, Morales & Alonso 2000; Morales *et al.* 2000) and generally show strong interannual fidelity to lek sites in spring. Females nest close to the lek where they copulate, and take over all brood caring duties. Thus surveys conducted in spring are likely to reveal consistent breeding distributions, but these may differ from wintering sites which are not addressed here.

Data availability is a constraint in building large-scale models of species' distributions, and two basic approaches seem to be emerging to make best use of available resources. Interpolation methods, ranging from simple linear interpolation (Farina 1997) to kriging (Palma, Beja & Rodrigues 1999), estimate species' occurrences

between sample points based on their spatial arrangement. This is likely to be most successful where habitat discontinuities are few, but we know of no published studies that assess model performance. The alternative approach, which may generally be called correlative mapping, relates species' occurrences at points to a suite of predictor variables that are available across the whole study area (Osborne & Tigar 1992; Buckland & Elston 1993; Augustin, Muggleston & Buckland 1996). Derived equations are then used to predict occurrences across the species' range. This is a data-hungry approach because environmental features are needed for every grid square or pixel covering the species' distribution. However, it is likely to detect more subtle changes in distributions than interpolation methods, providing the predictor variables are reasonably correlated with the habitat features chosen by the species being mapped. Fortunately, the digital data sets now available reasonably approximate some ecological requirements of the great bustard and probably other species too. Our emphasis throughout was on the use of readily available data sets that later would permit scaling-up to the national or regional scale.

Our starting premise was that vegetation type, terrain characteristics and human disturbance determine bustard distributions in Spain, factors that may apply equally to other species. Great bustards favour open, steppe-like, landscapes comprising cereal-fallow rotations, a habitat that is particularly under threat of intensification through irrigation under European Union agricultural policy. Numerous studies (Lyon 1983; Avery & Haines-Young 1990; Austin *et al.* 1996; Lavers, Haines-Young & Avery 1996) have related bird distributions to habitats using remotely sensed imagery such as LANDSAT thematic mapper (TM) and multispectral scanner (MSS), but the high cost precludes their use over extensive areas. Meteorological satellite data from the advanced very high resolution radiometer (AVHRR) operated by the National Oceanic and Atmospheric Administration are much cheaper and more readily available, but suffer from coarse spatial resolution (*c.* 1 km) which may limit discrimination of land-use types. However, the high temporal resolution of these data can offer an alternative route. Several workers (Kremer & Running 1993; Reed *et al.* 1994; Paruelo & Lauenroth 1995) have utilized time series of AVHRR normalized difference vegetation indices (NDVI) to map and monitor a range of habitats. NDVI is calculated from the near infra-red (NIR) and red (R) spectral bands as $NDVI = (NIR - R) / (NIR + R)$, exploiting the fact that vigorous vegetation reflects strongly in the NIR and absorbs radiation in the red band (Mather 1999). Rogers & Williams (1994) and Rogers *et al.* (1997) have taken this approach one step further by using NDVI to discriminate wildlife habitats, while others (Walker *et al.* 1992; Fjeldsa *et al.* 1997) have utilized the same data for identifying and understanding biogeographic patterns. We used a short time series of AVHRR data to predict the vegetation component of great bustard distribution in central Spain.

Few studies have actually demonstrated a preference for flat to slightly undulating terrain (although see Alonso & Alonso 1990; Onrubia *et al.* 1998), but it is a commonly stated habitat requirement of great bustards (Johnsgard 1991). The birds probably prefer sites with good horizontal visibility both to watch for predators and because the breeding system of dispersed leks involves strong visual cues over long distances. In our study region the species is distributed around 11 leks. Between late winter and early spring, males concentrate at these traditional arenas where they display and fight to establish a hierarchical rank. Females also gather at these arenas to mate between late March and early April. If lekking bustards select certain terrain characteristics, these should be discernible in currently available digital terrain models (DTM) with resolutions under 100 m.

Evidence for an effect of human disturbance on bustards is similarly anecdotal, although it is generally assumed that the presence of human infrastructures affects the distribution negatively. However, Lane, Alonso & Martin (2001) reported the absence of flocks within a band of about 1 km around villages and busy roads. For other birds, strong effects of roads on breeding density and performance have been noted (Reijnen & Foppen 1994; Reijnen *et al.* 1995). Analysis of the effects of roads and buildings on distributions should be straightforward with any accurate vector data source.

Methods

STUDY SITE

The pilot study area measured *c.* 126 × 132 km, centred on Madrid province, Spain, with the lower right coordinate at 2°44' W, 39°49' N (Fig. 1). It was chosen because accurate great bustard census data, geographic information system (GIS) data coverages and satellite imagery were readily available. The area comprised 55.9% agricultural land, 29.2% natural scrub or wooded cover, 8.2% forestry, 5.6% built environment and the remaining 1.1% bare ground or open water (European Union Corine Land Cover Project). The majority of analyses



Fig. 1. Approximate location of the study site (boxed area) measuring 126 × 132 km in central Spain, centred on Madrid province.

were confined to Madrid province itself because this was where bustard censuses were conducted but, where possible, models were extrapolated to the full study area (see later).

GIS COVERAGES

Digitized infrastructure maps were available from Autonomous Community of Madrid Cartographic Service at 1 : 100 000 scale. The data were separated into four layers (roads, buildings, railways and river systems) using ArcView software (ESRI 1996) and then rasterized to 80-m pixels in Idrisi (Eastman 1995). This resolution was chosen because a digital terrain model (DTM) was also available at this scale for the province. We created new variables from each of the infrastructure layers by replacing the central pixel of a 13 × 13-cell moving window with the proportion of pixels recording the feature of interest (Table 1). This is equivalent to calculating the percentage land cover of the feature at 80-m resolution within a *c.* 1-km² quadrat.

The DTM was used to derive the altitude of each pixel and to calculate terrain variability as bustards appear to choose open, gently undulating, landscapes. Using moving windows of 5 × 5, 9 × 9 and 13 × 13 pixels we calculated the coefficient of variation (CV) in altitude and placed this value in the central pixel. The three window sizes were selected to examine the effect of scale on terrain variability.

SATELLITE IMAGERY

A range of cloud-free AVHRR images at 1.1-km resolution was obtained for each month during 1996 to calculate the temporal NDVI signature for each pixel. Several factors can affect the reliability of NDVI values extracted from these data: changing illumination and viewing conditions within a single image and between images on different days; the presence of cloud cover; variations in atmospheric constituents such as water vapour and aerosols (Marçal & Wright 1997). The use of maximum value composites (MVC) is a means of partial correction of AVHRR data for the effects of these different factors (Holben 1986). Given a large number of images throughout a year, the assumption behind this approach is that the maximum NDVI values of the image set will correspond to ideal conditions, i.e. low solar zenith and viewing angle, low water vapour and aerosol concentrations and cloud-free conditions (Marçal & Wright 1997). However, for time series generated over large spatial areas, the MVC approach can have some limitations, as variations resulting from off-nadir viewing geometry cannot always be accommodated or corrected (Stoms, Bueno & Davis 1997). For this study, which concentrated on a relatively small study area, viewing geometries for MVC data were assumed to be near-nadir. NDVI values were therefore calculated for each image and used to form a MVC time series from the best cloud-free images from each month.

Table 1. Predictor variables used to compare pixels occupied by bustards with random locations

Variable	Definition
GIS layers	
Roads	Proportion of 80-m pixels in a 13 × 13-array containing roads. Equivalent to the density of roads at 1.1-km resolution
Buildings	Proportion of 80-m pixels in a 13 × 13-array containing buildings or large built structures such as airfields
Railways	Proportion of 80-m pixels in a 13 × 13-array containing railway tracks
Rivers	Proportion of 80-m pixels in a 13 × 13-array containing rivers
Altitude	The altitude in m recorded on the 80-m digital terrain model
Terrain variability 25	Coefficient of variation in altitude in a 5 × 5-pixel array of 80-m pixels. Measures variation in altitude at 0.16-km ² resolution
Terrain variability 81	Coefficient of variation in altitude in a 9 × 9-pixel array of 80-m pixels. Measures variation in altitude at 0.52-km ² resolution
Terrain variability 169	Coefficient of variation in altitude in a 13 × 13-pixel array of 80-m pixels. Measures variation in altitude at 1.1-km ² resolution
Slope	% slope. The maximum of either the north–south or east–west slopes across a 3 × 3-pixel array
Satellite imagery	
NDVI (month)	The value of the normalized difference vegetation index for each month based on a maximum value composite of AVHRR imagery at 1.1-km ² resolution. Scaled 0–255

BIRD CENSUS DATA

A complete great bustard census was conducted in March 1997 for the whole of Madrid province. The distribution of the species in the study area was known from previous censuses (Alonso & Alonso 1990, 1996). During 1 week three teams each of two experienced observers counted bustard flocks at the known breeding areas and also searched all other potential sites. In practice, birds sighted just beyond the province boundary were also recorded. One-hundred and three flocks (960 birds) were first marked on field maps, then digitized and the point coverage rasterized to 80-m and 1.1-km resolutions, recording the presence of bustards in the pixel. This resulted in 71 pixels with one or more flocks at 1.1-km resolution, and 92 pixels at 80-m resolution. For comparison, we generated equivalent random point coverages, stratifying them geographically both to sample the whole province and to reduce spatial autocorrelation (i.e. to reduce the probability of using adjacent pixels). The number of random points selected is important because prevalence (i.e. the ratio of positive to negative pixels) affects the outcome of model performance testing in logistic regression as used here (Fielding & Bell 1997; Manel *et al.* 1999).

ANALYSES

Analyses were based on the comparison of landscape features at the random points and at the locations used by bustards. In grid-cell mapped data the appropriate test to use depends on the spatial autocorrelation in the sampling points because this in effect reduces the degrees of freedom and thus increases the chance of type I errors (Cliff & Ord 1981; Legendre & Legendre 1998). We assessed spatial autocorrelation using Moran's I and then for univariate analyses followed the advice

of Cliff & Ord (1981) for modifying *t*-tests based on the results.

Multivariate analyses were carried out using forward stepwise logistic regression (SPSS 1997) to contrast pixels used by bustards with the random set. Analyses were carried out separately on the GIS data at 80-m resolution and the AVHRR imagery at 1.1-km resolution (Table 1). For the latter, models included the 12-monthly NDVI values and two NDVI contrasts (April minus July, and April minus January) derived empirically by inspection of mean NDVI temporal signatures (Fig. 2). We also included quadratic terms for the predictor variables to allow for the possibility of optimum vegetation conditions being selected by the birds (quadratic terms within logistic regression model Gaussian responses).

As in univariate analysis, spatial autocorrelation affects significance tests on logistic regression coefficients and, as no satisfactory method exists at present to correct for this, caution is needed in their interpretation. Conventional statistical modelling on spatial data ignores spatial autocorrelation in the residuals due to the ecological likelihood that neighbouring pixels will have dependent probabilities of use. To overcome this, we adopted the approach of Augustin, Muggleston & Buckland (1996) by incorporating an autologistic term in the models based on the modified Gibbs sampler. A probability surface was first generated by logistic regression in the usual way. Then a moving window of 9 × 9 pixels was used to calculate the mean of the probabilities assigned to the 80 neighbouring cells, weighted by Euclidean distance. This autologistic term was entered into the regression and the model rerun. The procedure is iterated to stability to produce the final probability surface (for further explanation see Augustin, Muggleston & Buckland 1996). In image-processing terms, the autologistic term acts as a smoothing filter, removing isolated pixels and consolidating habitat patches defined as suitable.

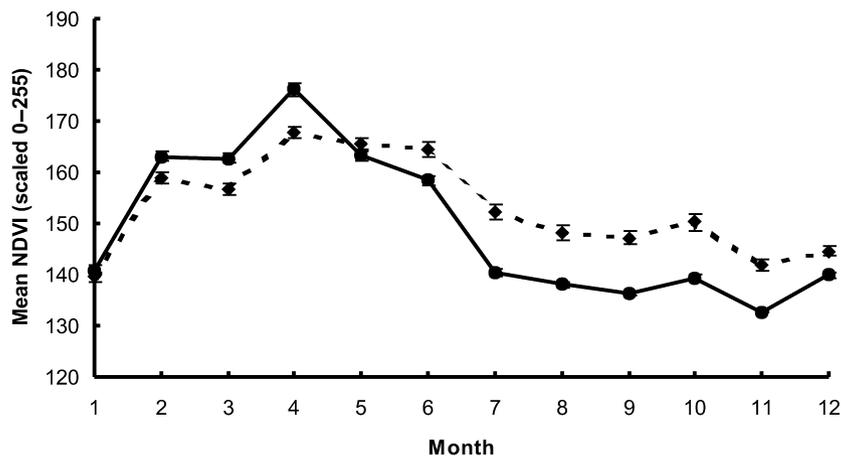


Fig. 2. Monthly means and standard errors along the temporal NDVI signatures for sites with great bustards (solid line) and those without (dashed line). Months are January = 1 to December = 12. $n = 71$ for both curves.

Results of logistic regression models are often judged as successful if predicted probabilities > 0.5 correspond with observed occurrences and values < 0.5 with absences. However, this dichotomy is arbitrary and lacks any ecological basis; patches rated with a 0.6 probability of occurrence may in fact be unsuitable. The more powerful approach used here is to assess model success across the full range of dichotomies using receiver operating characteristics (ROC) plots. ROC plots are widely used in clinical chemistry (Beck & Shultz 1986; Zweig & Campbell 1993) but rarely by ecologists (Fielding & Bell 1997). A ROC plot depicts on the y -axis sensitivity, i.e. $a/(a + c)$ in a 2×2 confusion matrix of the model prediction against the observations. This is plotted against $1 - \text{specificity}$, i.e. $1 - (d/(b + d))$ from the same confusion matrix. The chance performance of a model lies on the positive diagonal of a ROC plot, whereas models that out-perform chance follow a curve lying in the upper left half. The area under the ROC curve (AUC) is a convenient measure of overall fit and varies from 0.5 (for a chance performance) to 1.0 for a perfect fit. We generated ROC plots using SPSS software and calculated the AUC and its standard error using a non-parametric approach. The results are reported here as the $\text{AUC} \pm$ its standard error together with the significance of a test that the area = 0.5, i.e. that the model results do not differ from chance.

The probability surfaces derived from the two independent logistic regression models were combined using Bayesian inference after first resampling the 1.1-km surface to 80-m resolution. The technique permits prior probabilities (for example derived from one model) to be revised on the basis of new probabilities calculated from a second model. The appropriate formula (from Pereira & Itami 1991) is:

$$\text{combined probability} = \frac{1}{1 + \exp[\log(1 - P_{\text{NDVI}}/P_{\text{NDVI}}) - \log(P_{\text{GIS}}/1 - P_{\text{GIS}})]}$$

where P_{NDVI} is the probability derived from the NDVI model, and P_{GIS} is the probability derived from the

model based on GIS data layers. Bayesian approaches to decision-making have previously been used as here by Pereira & Itami (1991), and in other ways in wildlife distribution modelling by Aspinall & Veitch (1993) and Tucker *et al.* (1997).

Results

UNIVARIATE ANALYSES

At 80-m resolution, neither the bustard locations nor the random points exhibited significant spatial autocorrelation (Moran's $I < -0.0001$ in both cases). Following Cliff & Ord (1981) we therefore compared site characteristics using standard t -tests adjusted for unequal variance.

Sites occupied by bustards had significantly lower densities of roads, buildings, railways and rivers than random points (Table 2). Of these, the effect of buildings was particularly strong; bustards occurred at sites with a mean of only 0.8% land cover by buildings, whereas random sites averaged 10.5% ($P < 0.001$). Table 2 gives the ranges for each variable within which bustards occurred and these are combined as a threshold mask in Fig. 3a. This clearly shows the elimination of the radiating network of roads and buildings from Madrid City.

Bustards occurred within a narrow range of elevations from 566 to 780 m a.s.l., whereas random points covered the range 520–2194 m a.s.l. The terrain surrounding bustard sites was also significantly less variable than that around random sites at all three scales examined (Table 2). There was no obvious trend for a difference between the scales, although the significance of the difference between bustard and random sites increased within window size. Using 13×13 cells (about 1 km²), bustards occurred at sites with up to 2.8% coefficient of variation in altitude (mean 1.2%), whereas random sites had up to 5.9% variability (mean 2.0%). Combining the results for altitude and terrain variability with the mask in Fig. 3a yields Fig. 3b. This indicates that only 2133 km² of the 6540 km² (32.6%) studied meets the characteristics of infrastructure, elevation and river

Table 2. Comparison of features around 92 sites occupied by great bustards and 93 random points (except for terrain variability and slope based on 83 and 87 sites, respectively, to eliminate edge effects). Values are means \pm standard deviations and all *t*-tests are adjusted for significant unequal variance

Variable	Bustard sites	Random points	Adjusted <i>t</i> -test	Range used by bustards
Roads	0.022 \pm 0.045	0.043 \pm 0.080	2.28*	0–16.0%
Buildings	0.008 \pm 0.026	0.105 \pm 0.216	4.32***	0–18.3%
Railways	0.006 \pm 0.022	0.018 \pm 0.052	2.02*	0–11.2%
Rivers	0.028 \pm 0.044	0.053 \pm 0.066	3.11**	0–16.6%
Altitude	673.9 \pm 56.9	799.7 \pm 285.4	4.17***	566–780 m
Terrain variability 25	0.540 \pm 0.413	0.949 \pm 0.814	4.16***	0–1.50%
Terrain variability 81	0.871 \pm 0.555	1.518 \pm 1.027	5.15***	0–2.37%
Terrain variability 169	1.153 \pm 0.649	1.996 \pm 1.270	5.48***	0.03–2.81%
Slope	3.08 \pm 3.05	6.60 \pm 7.92	3.85***	0–12.9%

networks at sites used by bustards. Furthermore, this area is fragmented into blocks by the radiating networks from Madrid City.

PREDICTIVE MODELLING USING LOGISTIC REGRESSION

Building threshold masks (e.g. Fig. 3) is a convenient way to define areas meeting criteria but suffers from failing to take into account interactions between variables. Using the GIS variables in Table 2 (selecting terrain variability 169 for variation in elevation), we built predictive models for bustard presence using forward

stepwise logistic regression. Only railways and slope were not included in the model (at $P < 0.05$) and all other variables were significant at $P < 0.01$ (Table 3). The greatest contributions came from terrain variability ($P < 0.0001$) and housing density ($P < 0.0002$). Overall the ROC plot for the model (Fig. 4a) had an AUC of 0.898 ± 0.023 and was highly significant ($P < 0.001$). A simplified probability surface for bustard occurrence based on the significant GIS variables is shown in Fig. 5. There are obvious similarities with Fig. 3b but the probability plot shows far more texture and greater weighting in the east-central and extreme south-west parts of the province.

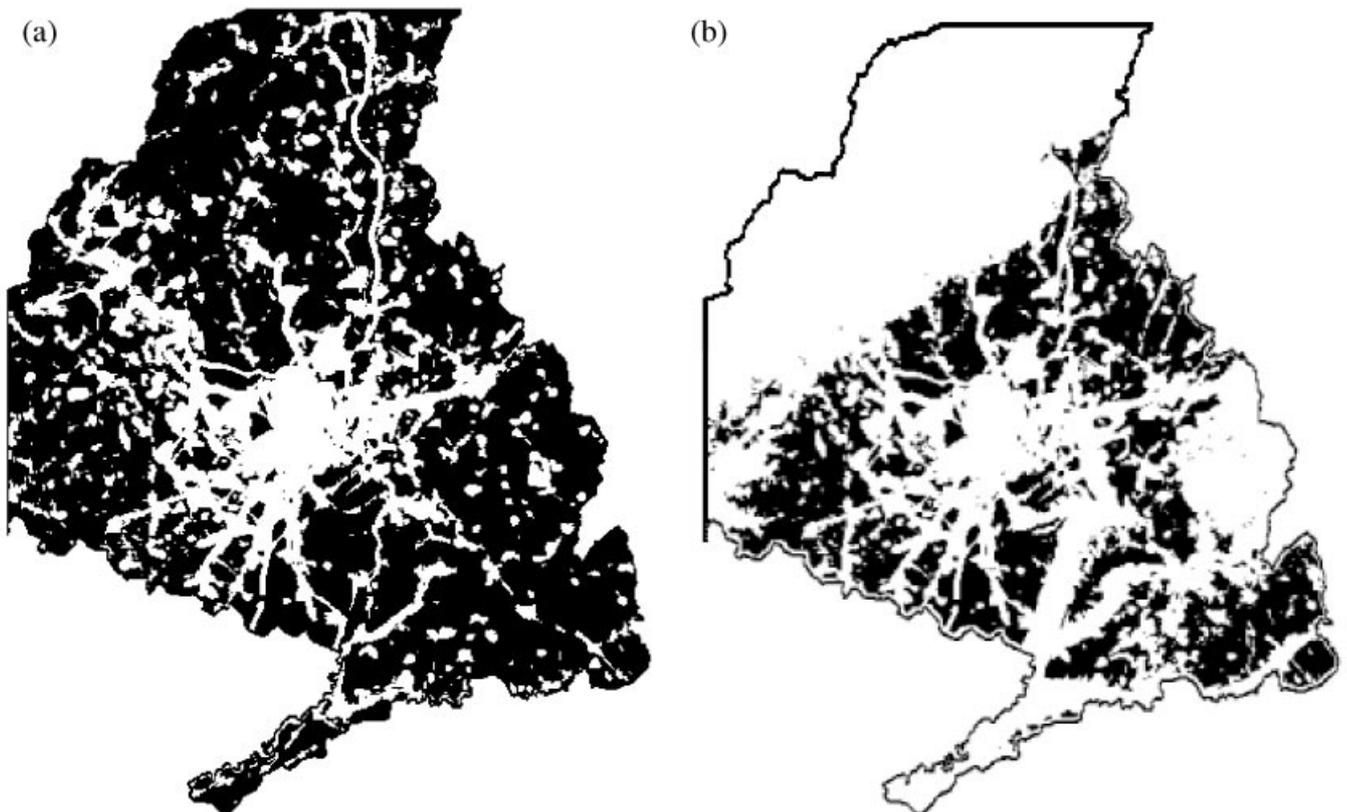
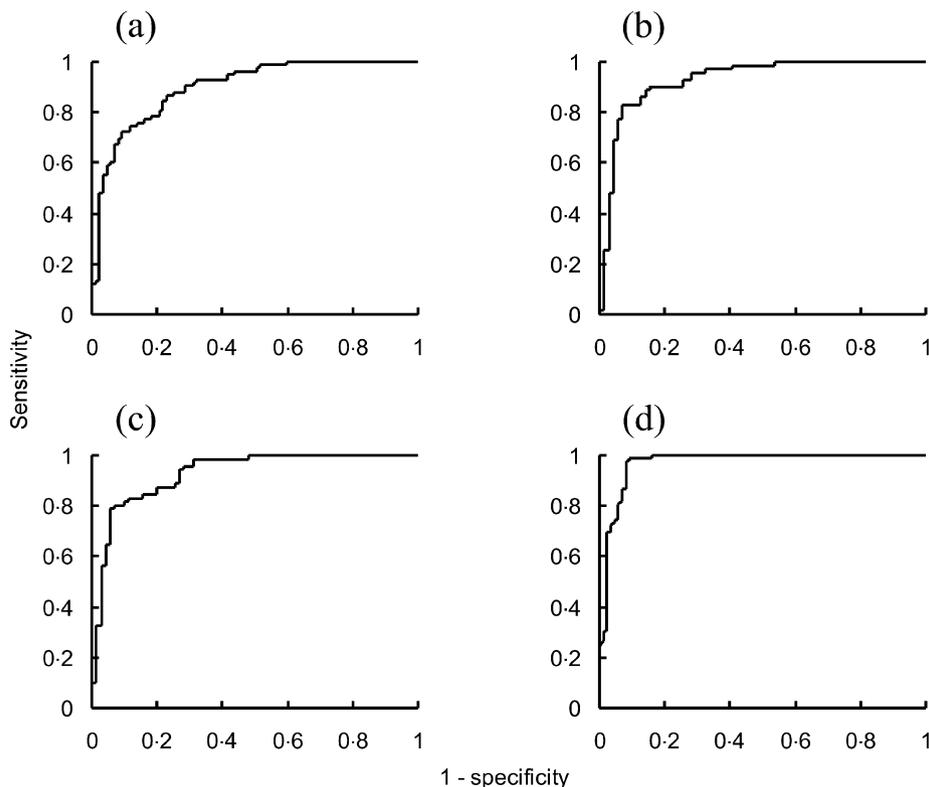


Fig. 3. Threshold masks for areas suitable for bustards (in black) based on (a) constraints of roads, buildings, railways and rivers, and (b) with the addition of altitude and terrain variability.

Table 3. Summary results of the logistic regression analyses. The significance of the coefficients was assessed using the Wald statistic

Model	Predictor variable	Coefficient	Standard error
GIS data layers (80-m resolution)	Roads	-10.97**	3.78
	Buildings	-26.90***	7.13
	Rivers	-9.86**	3.74
	Altitude	-0.009***	0.003
	Terrain variability 169	-1.14***	0.29
NDVI data (1.1-km resolution)	(Constant)	9.25***	2.08
	NDVI (January)	0.12*	0.05
	NDVI (November)	-0.31***	0.08
	NDVI (April) – NDVI (July)	0.08**	0.03
	(Constant)	23.26**	8.92

**Fig. 4.** ROC plots for (a) logistic regression model with five GIS variables, (b) logistic regression model based on NDVI, (c) autologistic regression model based on NDVI, and (d) Bayesian integrated model incorporating both the GIS variables and the autologistic NDVI model.

We built a separate model based on the satellite data at 1.1-km resolution. The stepwise inclusion of the explanatory variables resulted in the selection of three NDVI variables for January, November and the contrast April minus July (Table 3). None of the quadratic terms was significant. The model ROC plot (Fig. 4b) had an AUC of 0.93 ± 0.022 and was highly significant ($P < 0.001$). The resultant probability surface is shown in Fig. 6. Although this model of habitat suitability was derived by analysing only the presence or absence of bustards, the probabilities also relate to the numbers of birds present (Fig. 7). Most points lie beneath the diagonal from bottom left to top right, indicating that large numbers of birds only occur in pixels with high predicted probabilities.

We took account of autocorrelation in the model based on the satellite data by including an autologistic term; the regression coefficients stabilized within five iterations to produce Fig. 8. Comparing this with Fig. 7 reveals the down weighting of isolated pixels previously defined as suitable, and the consolidation of the larger suitable habitat blocks. The ROC plot for the autologistic model (Fig. 4c) differed very little from the standard NDVI model (AUC = 0.93 ± 0.022 , $P < 0.001$). The 1.1-km² resolution probability surface in Fig. 8 was resampled to 80-m resolution in order to combine it with the thresholds imposed by infrastructure and natural features (i.e. Fig. 3). Taking an arbitrary lower threshold of 50% probability of occurrence from the habitat suitability map (Fig. 8) alone yielded an area of 972 km² or

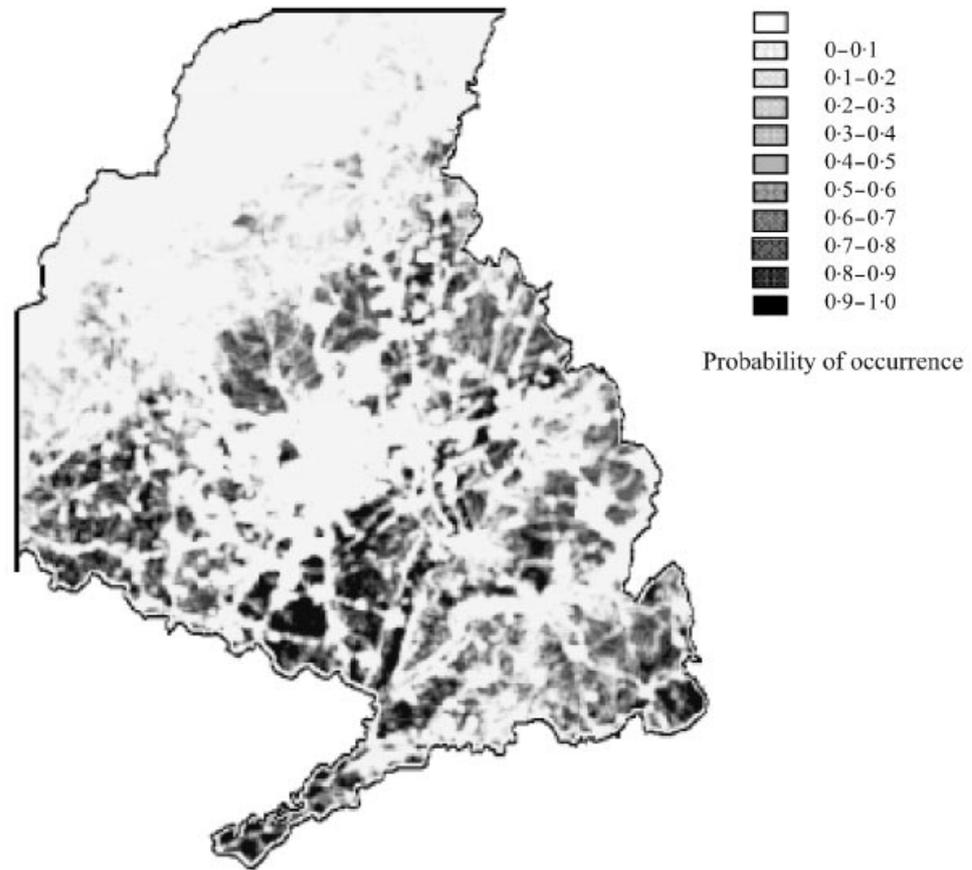


Fig. 5. Simplified probability surface for the occurrence of great bustards based on logistic regression analysis of five GIS variables.

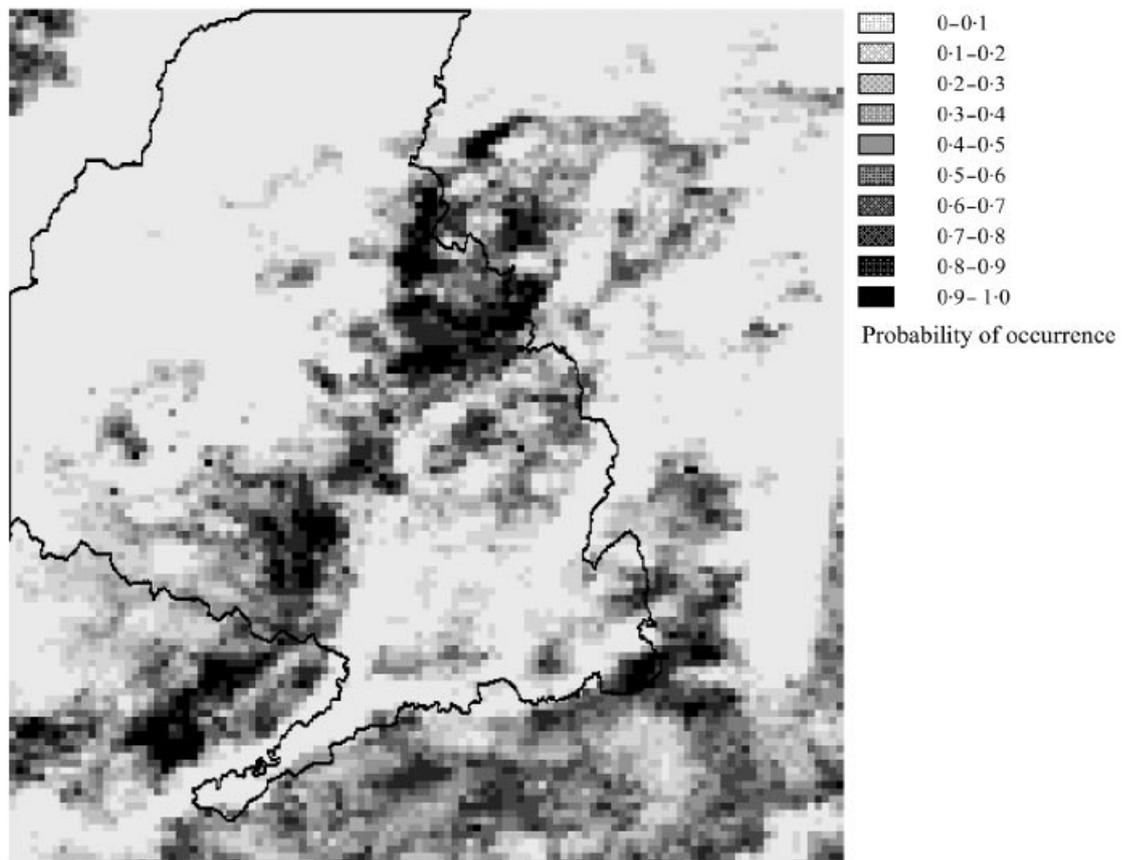


Fig. 6. Simplified probability surface for the occurrence of great bustards based on logistic regression analysis of temporal NDVI signatures.

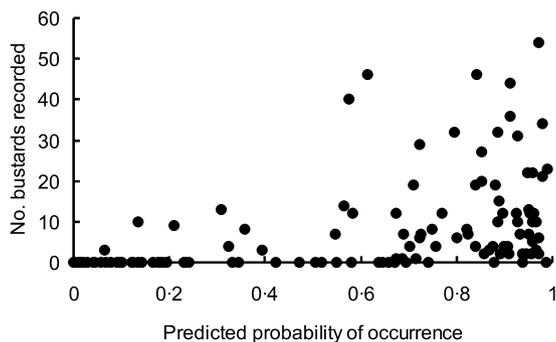


Fig. 7. Relationship between the predicted probability of occurrence generated by logistic regression analysis of presence-absence data and the number of bustards recorded.

14.9% of the province. This provides an estimate of the area of the province with vegetation index signatures comparable to those used by bustards. When combined with Fig. 3b, however, this falls to 570 km² or 8.7% of the province (Fig. 9). The difference between these two values (400 km²) is the amount of apparently suitable habitat that bustards are unlikely to use due to proximity to unfavourable landscape elements. Figure 9 (area 570 km²) is the area of Madrid province that cannot be distinguished from sites occupied by bustards using GIS and remote sensing, based on simple threshold mapping.

An alternative approach is to combine the two probability surfaces based on feature data (Fig. 5) and NDVI (Fig. 8) using Bayesian integration (see the Methods). As Fig. 5 represents largely fixed features of the landscape today, we regarded these as prior probabilities that may be refined by consideration of Fig. 8 which represents land cover based on green biomass. The Bayesian integrated probability model (Fig. 10) showed an excellent agreement with the original census data (simplified in Fig. 11 for ease of display) and had an ROC AUC value of 0.969 ± 0.013, *P* < 0.001 (Fig. 4d).

Discussion

We started with a premise gleaned from the literature that great bustard distributions may be related to vegetation, topography and human influence, and sought digital data sets that could characterize these features. At the landscape scale, our analyses successfully predicted the occurrence of bustards around Madrid province and all these features were significant predictors.

The negative effect of human disturbance on bustard occurrence is not surprising but is interesting because at first sight it appears that the birds are well integrated into the urban setting in Madrid province. There is growing evidence that roads impact on large mammals

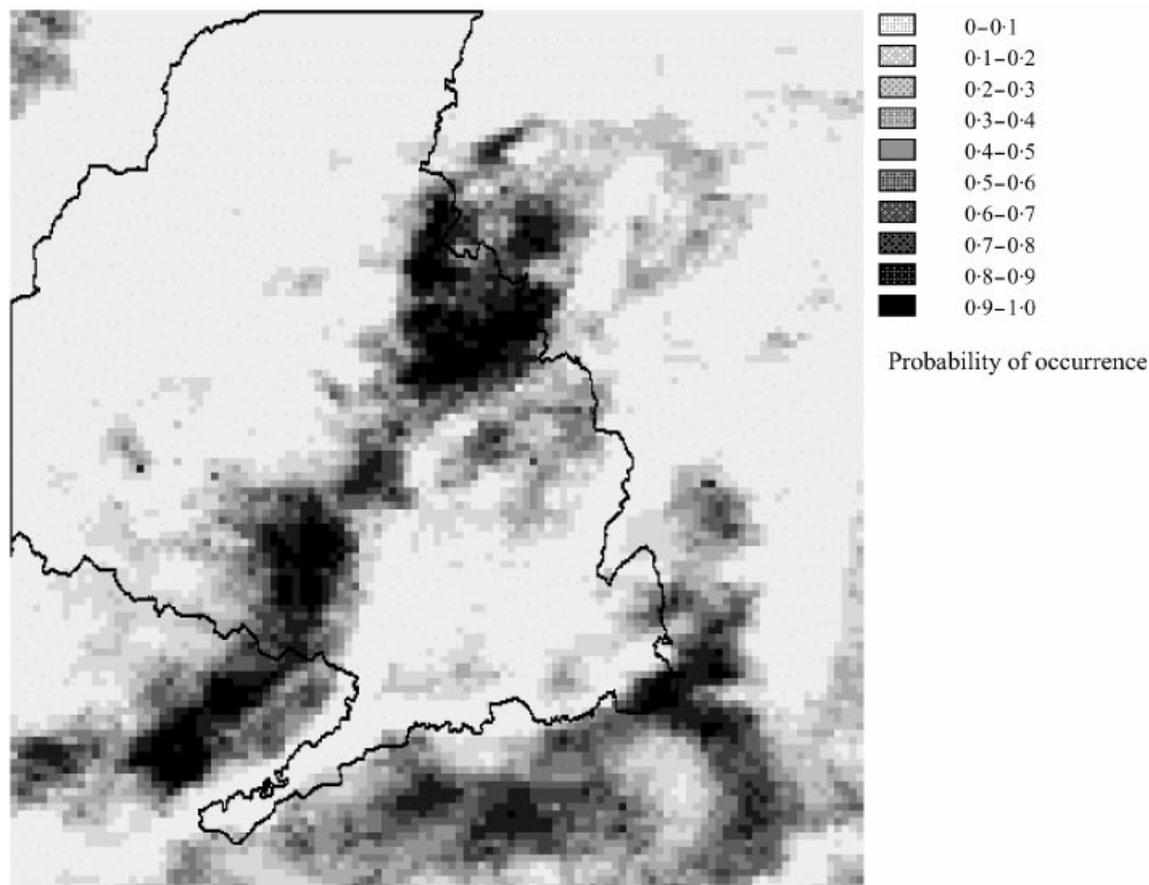


Fig. 8. Simplified probability surface for the occurrence of great bustards based on autologistic regression analysis of temporal NDVI signatures. Note how in comparison with Fig. 6 the image shows better definition of core areas.



Fig. 9. Threshold map for the occurrence of great bustards based on the limits from Fig. 3b and habitat suitability > 0.5 from Fig. 7. Defined in this way, 8.7% of the province could be used by great bustards.

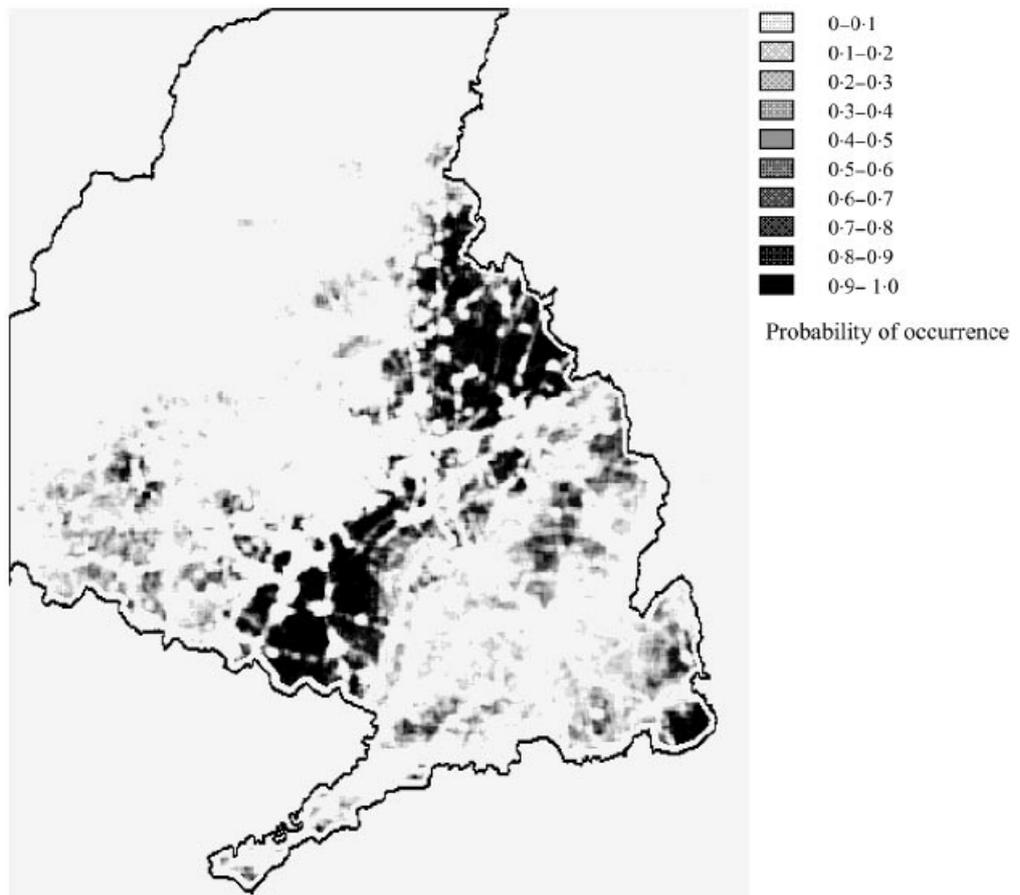


Fig. 10. Bayesian integrated probability map for the occurrence of great bustards in central Spain.

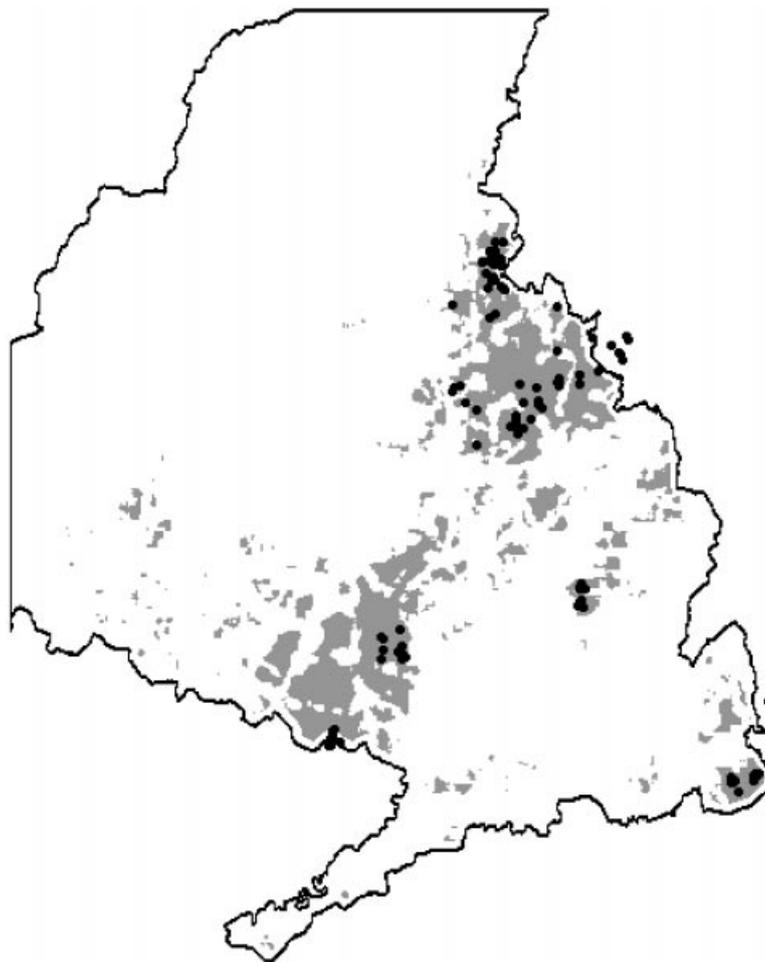


Fig. 11. Predicted probabilities of occurrence > 0.7 from the Bayesian integrated model overlaid with the recorded locations of great bustard flocks. Note that flocks observed beyond the province boundary are adjacent to areas predicted to be highly suitable for bustards.

(Mech 1989; Mladenoff *et al.* 1995; Mace *et al.* 1996; Palma, Beja & Rodrigues 1999) and birds (Reijnen & Foppen 1994; Reijnen *et al.* 1995) and this is important in the context of environmental impact analysis (Hill *et al.* 1997). The negative effect of buildings was even stronger and independent of the impact of roads in our multiple logistic regression analyses. A similar analysis on grey geese (*Anser anser* and *A. brachyrhynchus*) in Scotland found the same result (C. Urquhart & P.E. Osborne, unpublished data) and it may be important that future studies consider the possible effects of roads and housing independently.

We do not believe that the negative impact of rivers and their tributaries demonstrates an actual avoidance of these features but rather that irrigated crops are frequent along watercourses and these are avoided. Bustards occupied only a part of the altitude range available and, more significantly, occurred at sites where the surrounding terrain was less variable than at random sites. This accords with their assumed preference for relatively open terrain. Our analysis with three window sizes (400 m to 1 km) did not detect the point where terrain variability at bustard and random sites

became the same, although logically with a large enough window this would be true. This suggests an avenue for further research; indeed, finding the scales at which impacts of impaired visibility and disturbance are greatest would provide valuable data for conservation.

As Fig. 2 illustrates, there was a marked difference between the mean NDVI signatures for sites used by bustards and those that were not. Although AVHRR data are relatively coarse-grained, the use of temporal signatures provided enough resolution to discriminate occupied and unoccupied sites and indicates the potential of using NDVI time series to predict avian habitats. We believe that our model largely detected the sharp change in NDVI brought about by rapid crop growth in cereal fields in spring followed by biomass loss through harvesting in July. However, based on local knowledge, it appears that the model is not simply predicting all cereal areas, because some large areas were excluded. How the subset of more suitable fields is recognized is less clear. Maurer (1994) has previously shown a relationship between avian abundance and NDVI, but the correlation of abundance with probabilities of occurrence generated by temporal NDVI curves is new. The relationship here

indicates that sites defined as suitable have the potential to hold most birds, although they do not necessarily do so. Conversely, sites defined as unsuitable appear capable of holding only few birds.

Despite the coarse-grained nature of AVHRR data, models based on NDVI alone do provide significant predictions of bustard occurrence. In some situations, remotely sensed data may be all that are available and our results indicate that they are a useful first step in model building. However, we conclude that while models based on vegetation alone can provide accurate predictions of bustard habitats at some spatial scales, terrain and human influence are also significant predictors and are needed for finer scale modelling. For example, the breeding site in the centre of the south-east part of the province (Fig. 11) was not predicted by the NDVI probability surface (Fig. 8) but appeared after integration with the landscape GIS variables. We thus concur with Manel *et al.* (1999) that prediction success in distribution models may be enhanced when local data are available, despite the apparent success of coarse-grained models.

The use of an autologistic term based on a modified Gibbs sampler (Augustin, Muggleston & Buckland 1996) in the logistic regression models proved very effective in sharpening the definition between occupied and unoccupied patches. Although similar visual effects can be achieved more simply in image processing through Gaussian or median filters, they lack an ecological basis, whereas the Gibbs sampler uses information on the proximity of occupied pixels. We chose to consider neighbours within 4.4 km of the target pixel (i.e. a 9×9 window at 1.1-km resolution) because this seemed an appropriate scale for breeding great bustards, comparable with the average area of influence of a lek. Augustin, Muggleston & Buckland (1996) tried a range of window sizes to 4 km distance for red deer *Cervus elaphus* but found that 3 km gave optimum results in terms of the balance between model accuracy and computation time. Further work is needed on the selection of optimum neighbourhood distances for a range of species when trying to account for spatial autocorrelation in distribution models.

Generally, the prediction of occupied sites was more successful than the prediction of absences. For example, the Bayesian integrated model (Fig. 10) correctly predicted 93.0% of occupied sites and 78.9% of unoccupied sites, based on a 50% probability dichotomy. This is the opposite result to that reported by Manel *et al.* (1999) for six bird species and reinforces their view that more work is needed on model performance indicators and their determinants. Differences in performance at predicting presence and absence may be due to a number of factors (Fielding & Bell 1997) but we offer the following ecological explanations. First, the bustard census data were a single-day snapshot of occurrences, taken when flocking was at its maximum at lek sites. Under these conditions, adjacent areas could be mistakenly regarded as unsuitable whereas with greater flock dispersion they would be occupied. Secondly, the

Spanish great bustard population is fragmented into numerous habitat patches and, due to high site fidelity (Alonso *et al.* 1995), movement between patches for breeding may be scarce. Tilman, Lehman & Kareiva (1997) predict that these conditions will lead to the non-occupancy of some suitable sites even at population equilibrium. Thus we might expect some unoccupied sites to be ecologically inseparable from occupied sites. Thirdly, recent field studies by Lane, Alonso & Martin (2001) could not discriminate occupied from unoccupied but apparently suitable areas for bustards in central Spain, and it is unreasonable to expect large-scale models to outperform intensive field investigations. Their study compared 13 areas occupied by great bustards with a matched set of 12 unoccupied areas nearby. These sites could not be distinguished using discriminant analysis on a set of variables defining relevant habitat characteristics such as crop type, substrate heterogeneity, field size, presence of roads, villages and powerlines. Both their study and ours indicate that not all potentially suitable areas are occupied and that great bustards show fidelity to sites regardless of the availability of suitable habitat elsewhere. We believe that this may be due to a combination of a series of local extinction processes in recent decades due to human-induced habitat deterioration and hunting, and the very low re-colonization capability of the species which arises from its complicated lek breeding system. Settlement patterns are probably determined by the presence of conspecifics rather than habitat cues (J.C. Alonso, unpublished data). This means that conservation efforts must be directed towards protecting traditional lek sites, and that once a lek is extinct the site will probably remain empty in the future.

The limited number of historically documented extinctions (e.g. the small patch predicted as suitable west of Madrid city was occupied 20 years ago; Figs 10 and 11) and circumstantial evidence (e.g. place names in the larger area with high probability of occurrence south of the city) suggest that Fig. 11 shows the likely distribution of great bustards some 20–40 years ago. Since then many of the predicted areas have become vacant due to local extinction processes. It is also important to remember that the locations in Fig. 11 were based on breeding birds and that these perform seasonal movements and expand the occupied area throughout the year. Conservation based on breeding sites alone would almost certainly be ineffective, and a landscape-scale approach incorporating seasonal changes in locations is required.

Our study was preliminary in developing methods for landscape-scale models but the intention is to examine distributions at the national scale. We envisage two major challenges in relation to scaling-up to larger study areas. First, as mentioned earlier, the selection of maximum value NDVI data may not lead to adequate standardization over large areas due to a level-of-view and solar-angle bias. In practice, this means that a given surface may generate a different NDVI–MVC score at different locations on the image. The main

factor causing this is the inherent variability of surfaces at different look angles. A traditional solution to this problem is to first standardize images against a common target such as a large body of water or bare ground (Sannier *et al.* 1998). A better approach, however, may be to take into account the bidirectional reflectance distribution function (BRDF) of the surfaces being sensed (Cihlar, Manak & Voisin 1994). Correction for these effects can be undertaken using linear semi-empirical kernel-driven models (Roujean, Leroy & Deschamps 1992) that adjust image data for BRDF variability and extract surface information. Initial results indicate that the new models provide enormous scope for improving data quality and also provide useful additional information (Chopping 1998).

The second challenge to scaling-up is that animals may not choose habitats according to absolute needs but may adopt a comparative approach. Particularly on the edge of a species' range, occupied sites are likely to be far from the ideal habitat with poorer breeding performance to match (Lawton *et al.* 1994). Also, several habitat types with different spectral signatures may be equally suitable at large geographical scales. Large-scale models must therefore include samples from across the species' range and analysis may be better partitioned spatially prior to modelling.

Acknowledgements

We are grateful to Dr Antonio Yagüe of Infocarto, S.A., for supplying the NOAA data, and the Autonomous Community of Madrid Cartographic Service for the GIS coverages. John McArthur and Alison Martin helped prepare the data for analysis. Dr Enrique Martín, Dr Manuel Morales, Carlos Martín, Dr Simon Lane and Dr Javier A. Alonso assisted with the collection of bird data. This research was jointly funded by British Council/Ministry of Education and Culture Accion Integrada HB96-82 and HB97-69 awards, DGICYT project PB94-0068, and by the Dirección General del Medio Natural of Madrid Community.

References

Alonso, J.C. & Alonso, J.A. (1990) *Parámetros demográficos, selección de hábitat y distribución de la avutarda en tres regiones españolas*. ICONA, Madrid, Spain.

Alonso, J.C. & Alonso, J.A. (1996) The great bustard *Otis tarda* in Spain: present status, recent trends and an evaluation of earlier censuses. *Biological Conservation*, **77**, 79–86.

Alonso, J.C., Alonso, J.A., Martín, E. & Morales, M. (1995) Range and patterns of great bustard movements at Villafáfila, NW Spain. *Ardeola*, **42**, 69–76.

Alonso, J.C., Morales, M.B. & Alonso, J.A. (2000) Partial migration, lek fidelity and nesting area fidelity in female great bustards *Otis tarda*. *Condor*, **102**, 127–136.

Aspinall, R. & Veitch, N. (1993) Habitat mapping from satellite imagery and wildlife survey data using a Bayesian modeling procedure in a GIS. *Photogrammetric Engineering and Remote Sensing*, **59**, 537–543.

Augustin, N.H., Muggleston, 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 geographic information system and remote sensing. *Journal of Applied Ecology*, **33**, 1541–1550.

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

Beck, J.B. & Shultz, E.K. (1986) The use of relative operating characteristic (ROC) curves in test performance evaluation. *Archives of Pathology and Laboratory Medicine*, **110**, 13–20.

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

Caldow, R.W.G. & Racey, P.A. (2000) Introduction. Large-scale processes in ecology and hydrology. *Journal of Applied Ecology*, **37** (Supplement 1), 6–12.

Chopping, M.J. (1998) Evaluation of the performance of linear semi-empirical BRDF models in a semi-arid grassland biome. *Proceedings of the 24th Annual Conference of the Remote Sensing Society* (eds P.J.A. Burt, C.H. Power & P.M. Zukowskyj), pp. 568–574. Remote Sensing Society, Nottingham, UK.

Cihlar, J., Manak, D. & Voisin, N. (1994) AVHRR bidirectional reflectance effects and compositing. *Remote Sensing of the Environment*, **48**, 77–88.

Cliff, A.D. & Ord, J.K. (1981) *Spatial Processes: Models and Applications*. Pion, London, UK.

Collar, N.J. & Andrew, P. (1988) *Birds to Watch: the ICBP World Check-list of Threatened Birds*. ICBP Technical Publication 8. ICBP, Cambridge, UK.

Del Hoyo, J., Elliot, A. & Sargatal, J. (1996) *Handbook of the Birds of the World*, Vol. 3. Lynx, Barcelona, Spain.

Eastman, J.R. (1995) *Idrisi for Windows. User's Guide*, Version 1.0. Clark Labs, Clark University, Worcester, USA.

ESRI (1996) *ArcView GIS*. Environmental Systems Research Institute, Redlands, CA.

Farina, A. (1997) Landscape structure and breeding bird distribution in a sub-Mediterranean agro-ecosystem. *Landscape Ecology*, **12**, 365–378.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Fjeldsa, J., Ehrlich, D., Lambin, E. & Prins, E. (1997) Are biodiversity hotspots correlated with ecoclimatic stability? A pilot study using the NOAA-AVHRR remote sensing data. *Biodiversity and Conservation*, **6**, 401–422.

Gaston, K.J. & Blackburn, T.M. (1995) Mapping biodiversity using surrogates for species richness – macro-scales and New World birds. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, **262**, 335–341.

Heredia, B., Rose, L. & Painter, M. (1996) *Globally Threatened Birds in Europe*. Council of Europe, Strasbourg, France.

Hidalgo de Trucios, S.J. (1990) World status of the great bustard (*Otis tarda*) with special attention to the Iberian peninsula populations. *Miscelanea Zoologica*, **14**, 167–180.

Hill, D., Hockin, D., Price, G., Tucker, R., Morris, R. & Treweek, J. (1997) Bird disturbance: improving the quality and utility of disturbance research. *Journal of Applied Ecology*, **34**, 275–289.

Holben, B. (1986) Characteristics of maximum value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, **7**, 1471–1434.

Johnsgard, P.A. (1991) *Bustards, Hemipodes and Sandgrouse. Birds of Dry Places*. Oxford University Press, Oxford, UK.

Kremer, R.G. & Running, S.W. (1993) Community type differentiation using NOAA/AVHRR data within a sagebrush-steppe ecosystem. *Remote Sensing of the Environment*, **46**, 311–318.

Lane, S.J., Alonso, J.C. & Martin, C.A. (2001) Habitat preferences of great bustard *Otis tarda* flocks in the arable steppes

- of central Spain: are potentially suitable areas unoccupied? *Journal of Applied Ecology*, **38**, 193–203.
- Lavers, C.P., Haines-Young, R.H. & Avery, M.I. (1996) The habitat associations of dunlin *Calidris alpina* in the Flow Country of northern Scotland and an improved model for predicting habitat quality. *Journal of Applied Ecology*, **33**, 279–290.
- Lawton, J.H., Nee, S., Letcher, A.J. & Harvey, P.H. (1994) Animal distributions: patterns and processes. *Large-Scale Ecology and Conservation Biology* (eds P.J. Edwards, R.M. May & N.R. Webb), pp. 41–58. Blackwell Scientific Publications, Oxford, UK.
- Legendre, P. & Legendre, L. (1998) *Numerical Ecology. Developments in Environmental Modelling 20*. Elsevier, Amsterdam, the Netherlands.
- Lyon, J.L. (1983) LANDSAT-derived land-cover classifications for locating potential kestrel nesting habitat. *Photogrammetric Engineering and Remote Sensing*, **49**, 245–250.
- Mace, R.D., Walker, J.S., Manley, T.L., Lyon, L.J. & Zuuring, H. (1996) Relationships among grizzly bears, roads and habitat in the Swan Mountains, Montana. *Journal of Applied Ecology*, **33**, 1395–1404.
- Manel, S., Dias, J.M., Buckton, S.T. & Ormerod, S.J. (1999) Alternative methods for predicting species distribution: an illustration with Himalayan river birds. *Journal of Applied Ecology*, **36**, 734–747.
- Marçal, A.R.S. & Wright, G.G. (1997) The use of overlapping NOAA-AVHRR NDVI maximum value composites for Scotland and initial comparisons with the land cover census on a Scottish regional and district basis. *International Journal of Remote Sensing*, **18**, 491–503.
- Mather, P.M. (1999) *Computer Processing of Remotely Sensed Images. An Introduction*. John Wiley and Sons, Chichester, UK.
- Maurer, B.A. (1994) *Geographical Population Analysis Tools for the Analysis of Biodiversity*. Blackwell Scientific Publications, Oxford, UK.
- Mech, L.D. (1989) Wolf population survival in an area of high road density. *American Midland Naturalist*, **121**, 387–389.
- Mladenoff, D.J., Sickley, T.A., Haight, R.G. & Wydeven, A.P. (1995) A regional landscape analysis and prediction of favorable gray wolf habitat in the Northern Great Lakes region. *Conservation Biology*, **9**, 279–294.
- Morales, M.B., Alonso, J.C., Alonso, J.A. & Martin, E. (2000) Migration patterns in male great bustards. *Auk*, **117**, 493–498.
- O'Connor, R.J. & Shrubbs, M. (1986) *Farming and Birds*. Cambridge University Press, Cambridge, UK.
- Onrubia, A., Saenz de Buruaga, M., Osborne, P., Baglione, V., Purroy, F.J., Lucio, A.J. & Campos, M.A. (1998) *Viabilidad de la población Navarra de avutardas*. Consultora de Recursos Naturales, Vitoria, Spain.
- Ormerod, S.J., Pienkowski, M.W. & Watkinson, A.R. (1999) Communicating the value of ecology. *Journal of Applied Ecology*, **36**, 847–855.
- Osborne, P.E. & Tigar, B.J. (1992) Interpreting bird atlas data using logistic models: an example from Lesotho, Southern Africa. *Journal of Applied Ecology*, **29**, 55–62.
- Pain, D.J. & Pienkowski, M.W. (1997) *Farming and Birds in Europe*. Academic Press, London, UK.
- Palma, L., Beja, P. & Rodrigues, M. (1999) The use of sighting data to analyse Iberian lynx habitat and distribution. *Journal of Applied Ecology*, **36**, 812–824.
- Paruelo, J.M. & Lauenroth, W.K. (1995) Regional patterns of normalized difference vegetation index in North American shrublands and grasslands. *Ecology*, **76**, 1888–1898.
- Pereira, J.M.C. & Itami, R.M. (1991) GIS-based habitat modeling using logistic multiple regression: a study of the Mt. Graham red squirrel. *Photogrammetric Engineering and Remote Sensing*, **57**, 1475–1486.
- Reed, B.C., Brown, J.F., Vanderzee, D., Loveland, T.R., Merchant, J.W. & Ohlen, D.O. (1994) Measuring phenological variability from satellite imagery. *Journal of Vegetation Science*, **5**, 703–714.
- Reijnen, R. & Foppen, R. (1994) The effects of car traffic on breeding bird populations in woodland. I. Evidence of reduced habitat quality for willow warblers *Phylloscopus trochilus* breeding close to a highway. *Journal of Applied Ecology*, **31**, 85–94.
- Reijnen, R., Foppen, R., ter Braak, C. & Thissen, J. (1995) The effects of car traffic on breeding bird populations in woodland. III. Reduction in density in relation to the proximity of main roads. *Journal of Applied Ecology*, **32**, 187–202.
- Rogers, D.J. & Williams, B.G. (1994) Tsetse distribution in Africa: seeing the wood and the trees. *Large-Scale Ecology and Conservation Biology* (eds P.J. Edwards, R.M. May & N.R. Webb), pp. 247–272. Blackwell Scientific Publications, Oxford, UK.
- Rogers, D.J., Hay, S.I., Packer, M.J. & Wint, G.R.W. (1997) Mapping land-cover over large areas using multispectral data derived from the NOAA-AVHRR: a case study of Nigeria. *International Journal of Remote Sensing*, **18**, 3297–3303.
- Roujean, J.L., Leroy, M. & Deschamps, P.Y. (1992) A bidirectional reflectance model of the earth's surface for the correction of remote sensing data. *Journal of Geophysical Research*, **97**, 20 455–20 420 468.
- Sannier, C.A.D., Taylor, J.C., Du Plessis, W. & Campbell, K. (1998) Real-time vegetation monitoring with NOAA-AVHRR in Southern Africa for wildlife management and food security assessment. *International Journal of Remote Sensing*, **19**, 621–639.
- SPSS (1997) *SPSS for Windows, 8.0.0*. SPSS Inc., Michigan, USA.
- Stoms, D.M., Bueno, M.J. & Davis, F.W. (1997) Viewing geometry of AVHRR image composites derived using multiple criteria. *Photogrammetric Engineering and Remote Sensing*, **63**, 681–689.
- Tilman, D., Lehman, C.L. & Kareiva, P. (1997) Population dynamics in spatial habitats. *Spatial Ecology* (eds D. Tilman & P. Kareiva), pp. 3–20. Princeton University Press, Princeton, NJ.
- Tucker, K., Rushton, S.P., Sanderson, R.A., Martin, E.B. & Blaiklock, J. (1997) Modelling bird distributions – a combined GIS and Bayesian rule-based approach. *Landscape Ecology*, **12**, 77–93.
- Walker, R.E., Stoms, D.M., Estes, J.E. & Cayocca, K.D. (1992) Improved modelling of biological diversity with multitemporal vegetation index data. *Technical Papers of the 1992 Annual Meeting of the ASPRS/ACSM*, pp. 562–571. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland, USA.
- Williams, P., Humphries, C., Vanewright, D. & Gaston, K. (1997) Descriptive and predictive approaches to biodiversity measurement. *Trends in Ecology and Evolution*, **12**, 444–445.
- Zweig, M.H. & Campbell, G. (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**, 561–577.

Received 3 February 2000; revision received 29 November 2000