



Contents lists available at ScienceDirect

Acta Oecologica

journal homepage: www.elsevier.com/locate/actoec

Original article

Geographical variation in the distributional constraints along a gradient of population aggregation

Sara Vallecillo^{a,*}, Lluís Brotons^a, Patrick E. Osborne^b^a Centre Tecnològic Forestal de Catalunya (CTFC), Àrea de Biodiversita, Ctra. Sant Llorenç de Morunys, km. 2 (Ctra. vella), 25280 Solsona, Catalonia, Spain^b Centre for Environmental Sciences, Faculty of Engineering and the Environment, University of Southampton, Highfield, Southampton SO17 1BJ, UK

ARTICLE INFO

Article history:

Received 3 February 2010

Accepted 13 October 2010

Available online xxx

Keywords:

Species distribution models

Non-stationarity

Occurrence aggregation

Geographical variation

Distributional constraints

ABSTRACT

Factors determining species distributions have frequently been shown to vary geographically, yielding spatially variable species–environment relationships when developing species distribution models. Using occurrence data for four bird species, we analysed to what extent position within the distribution range along a gradient of population aggregation determined geographical variation of distributional constraints. We built generalised linear models for the whole dataset (global models) and separately for each region within the species range with decreasing population aggregation (core, middle and peripheral). For both global and regional models, we compared species responses to habitat conditions showing an overall higher model fit in peripheral regions than in middle and core regions. Better model fit in peripheral regions was also found after including a spatial factor (i.e. an autocovariate estimating the proportion between the presences and total observations within a 10 km radius) into the global and regional models.

The scattered availability of suitable habitat patches in a predominantly hostile landscape matrix seems to be the limiting factor for species distributions in peripheral regions. Conversely, the larger number of spatially dependent occurrences in core and middle region may favour source and sink population dynamics, yielding poorer relationships between habitat conditions and species occurrence. The role of the spatial factor after removing the co-variation with habitat conditions separately for each region suggests that endogenous processes such as dispersal vary throughout the species range but differently for each species. Geographical variation in distributional constraints can be interpreted as a consequence of the inherent spatial character of ecological processes and their interaction with varying environmental conditions throughout the species range. Overlooking the effects of spatial variation in distributional constraints may lead to erroneous management conclusions and inadequate estimation of the species' response to environmental changes.

© 2010 Elsevier Masson SAS. All rights reserved.

1. Introduction

Species' responses to environmental conditions have frequently been shown to vary geographically as a consequence of the different ecological requirements across the range (i.e. niche variation) (Fielding and Howarth, 1995; Murphy and Lovett-Doust, 2007). When niche variation occurs, species distribution models will show differences in the modelled species–environment relationship over space, an example of spatial non-stationarity (Osborne et al., 2007). In spite of its relevance to distribution modelling, the consequences of non-stationarity have rarely been

investigated (Peterson and Holt, 2003). Moreover, the ecological factors causing spatial non-stationarity have scarcely been addressed and no clear conclusions have been drawn (Murphy and Lovett-Doust, 2007; Osborne et al., 2007).

In this context, there is a need to identify factors related to the species' ecology that underpin geographical variation of the species' response to the environment (i.e. non-stationarity). As suggested by Bahn et al. (2006), position within the geographical range plays an important role driving species distributions that has frequently been mistreated when studying species–environment relationships. For example, the species' response to exogenous factors such as habitat availability is likely to be spatially variable throughout the range. Habitat patchiness will increase towards the periphery of the distribution in line with the existence of a less suitable habitat matrix (Guo et al., 2005), yielding a gradient in decreasing overall habitat availability also associated with sharper

* Corresponding author. Tel.: +34 973 48 17 52; fax: +34 973 48 13 92.

E-mail addresses: sara.vallecillo.r@gmail.com (S. Vallecillo), lluis.brotons@ctfc.cat (L. Brotons), p.e.osborne@soton.ac.uk (P.E. Osborne).

contrasts between habitat patches. This variable spatial arrangement of habitat availability may yield different processes of habitat selection (Wiens, 1985). Furthermore, potential sources of colonization decline towards the periphery of the species' range because populations tend to be more sparsely distributed in a patchy environment surrounded by an unsuitable landscape matrix (Holt and Keitt, 2000; Guo et al., 2005). Under these circumstances, endogenous processes such as dispersal become more limiting to species distributions; stochastic, long distance dispersal being the only way of arriving at newly available habitat patches (Brotons et al., 2005). Moreover, intra-specific interactions such as conspecific attraction may also decrease in importance as abundance decreases towards the periphery (Gimona and Brewer, 2006). As a result, position within the geographical range is hypothesized to be associated with important spatial variation in the distributional constraints on species distributions.

In this study, we aimed to assess to what extent position within the distribution range yields geographic variation in distributional constraints for four open habitat bird species showing strong gradients in occurrence aggregation. Specifically, we defined a gradient of population aggregation at a large spatial scale to delineate three functional regions (i.e. core, middle and peripheral) within the distribution range of the species. Then, we analysed how the role of habitat conditions varied throughout the distribution range when building global habitat models. We also assessed whether inclusion of a spatial factor (i.e. an autocovariate estimating the proportion between the number of presences and the total observations within a 10 km radius) improved the explanation of species distribution patterns. Used in this way, an autocovariate attempts to capture neighbourhood occurrence density (Kunin, 1998; Araújo and Williams, 2001), which is a measure of the relative occurrence per unit area in the neighbourhood. Neighbourhood occurrence density often decreases from the core to the periphery of a species' range (Whittaker, 1967; Brown, 1984; Caughley et al., 1988; but also see McGill and Collins, 2003). Finally, if geographical variation in the explanatory power of habitat conditions and the spatial factors along the gradient of population aggregation were to be confirmed, building models separately for each region within the species range would be preferable. Regional models allow better identification of the relative roles of habitat and spatial factors in determining distributions and may therefore provide a clearer assessment of the ecological processes behind species distributions (Osborne and Suarez-Seoane, 2002; Murphy and Lovett-Doust, 2007).

We hypothesized that local habitat conditions are likely to be more important constraints on species distributions as patchiness of habitat availability increase. In peripheral areas, suitable habitat might be considered locally rare and therefore may be an important limiting factor for species distributions at the landscape scale (Brown et al., 1996). On the other hand, in core regions characterized by population aggregation, species may be able to occupy suboptimal habitats at high levels of population density (Sherry and Holmes, 1985; Brown et al., 1995) leading to a laxer response to local habitat conditions. In this case, information on neighbourhood occurrence density (i.e. in the form of an autocovariate) may be essential to explain species distribution patterns.

2. Material and methods

2.1. Bird species data

This work focused on four open habitat bird species with variable preferences for early successional, shrub like habitats and dry, extensive farmlands in Mediterranean landscapes (Estrada et al., 2004): *Galerida theklae* (Thekla Lark), *Lullula arborea* (Wood Lark),

Anthus campestris (Tawny Pipit) and *Emberiza hortulana* (Ortolan bunting). We selected these species because they show strong gradients of population aggregation in Catalonia, NE Spain. Furthermore, these species have been shown in previous studies to be especially abundant on post-fire areas (Herrando et al., 2002), showing a significant expansion in their distributions at the regional level favoured by fire (Vallecillo et al., 2009). However, their response to fire occurrence appears to be spatially variable given the influence of dispersal constraints and the role of the landscape context controlling species-colonization processes (Brotons et al., 2005; Vallecillo et al., 2009). In this context, landscape heterogeneity, as occurs along a gradient of population aggregation, plays a key role in determining the distributions of the studied species (Brotons et al., 2005).

Bird data were obtained from the Catalan Breeding Bird Atlas 1999–2002 (CBBA), (Estrada et al., 2004). The CBBA is a large scale survey covering the whole extent of Catalonia (c. 31,930 km²) using a 10 × 10 km grid. The grid was used as a reference to establish a total of 3077 squares of 1 km² (approximately 9% of the total study area) that were selected to conduct standardized intensive sampling in a stratified fashion to cover the main habitat types present within each of the 10 × 10 km squares (Hirzel and Guisan, 2002). Sampling in the 1 km² squares was based on two 1 h long surveys of species presences and absences during the breeding season (between March and July). From the sampled 1 km² squares, we removed those falling mostly out of the study area or in the Mediterranean Sea, considering finally 3021 squares for analyses.

2.2. Regions within the distribution range of the species

We used a kernel density estimation approach to characterize the relative variation of each species' occurrence over its range in the study area (Fortin et al., 2005). The distance used to calculate the kernel density function was based on the radius at which the species' occurrence showed spatial dependence (i.e. positive autocorrelation). Accounting for the spatially dependent presences, we are considering the species occurrence that are more likely to interact or may show a spatial relationship in the ecological processes behind species distributions. The kernel density function defined a general trend of spatial patterning in the species distributions at a large spatial scale, showing a gradient of spatially dependent occurrence. Assessment of spatial autocorrelation in occurrence data was performed by means of the join count statistic (Fortin and Dale, 2005) with the PASSaGE software (Rosenberg, 2009), considering distance classes of 10 km (Table 1).

The kernel density function of each species was used to define three functional regions (core, middle and peripheral) within the species' range along a decreasing gradient of population aggregation from core to peripheral regions at a large spatial scale (Fig. 1). Limits between regions were defined following the kernel density function (i.e. gradient of population aggregation) by assigning the same number of presences to each region for the purpose of the comparative analyses (Table 1).

2.3. Neighbourhood occurrence density

We estimated the variation in neighbourhood occurrence density by means of an autocovariate (Kunin, 1998; Araújo and Williams, 2001), which is a measure of the relative occurrence within a certain area (i.e. occurrence density in the nearby surroundings). An autocovariate inherently captures spatial autocorrelation originated from spatial processes such as intra-specific attraction and limited dispersal (i.e. endogenous processes), but also from exogenous factors (i.e. environmental factors not included into habitat models) (Lichstein et al., 2002) which may

Table 1

Distances at which species occurrence showed positive autocorrelation according to the join count statistics (standard normal deviate > 1.96). This distance was chosen to estimate the kernel density function and define three functional regions showing a decrease of population aggregation from core to peripheral regions. Within each region, we defined two datasets: the whole dataset and a dataset with the same prevalence between regions (corrected prevalence) to control for likely differences between them.

| Species | Distance (km) | Regions | Whole dataset | | Corrected prevalence | |
|----------------------|---------------|------------|---------------|----------|----------------------|----------|
| | | | Presences | Absences | Presences | Absences |
| <i>G. theklae</i> | 80 | Core | 73 | 265 | 73 | 254 |
| | | Middle | 73 | 254 | 73 | 254 |
| | | Peripheral | 74 | 2282 | 74 | 256 |
| <i>L. arborea</i> | 90 | Core | 303 | 213 | 303 | 213 |
| | | Middle | 303 | 532 | 303 | 213 |
| | | Peripheral | 304 | 1366 | 304 | 214 |
| <i>A. campestris</i> | 60 | Core | 33 | 512 | 33 | 512 |
| | | Middle | 34 | 592 | 34 | 512 |
| | | Peripheral | 34 | 1816 | 34 | 512 |
| <i>E. hortulana</i> | 70 | Core | 39 | 305 | 39 | 305 |
| | | Middle | 39 | 490 | 39 | 305 |
| | | Peripheral | 40 | 2108 | 40 | 313 |

also be variable throughout the species range. The autocovariate was estimated from 1 km² presence/absence bird data by estimating the proportion between the number of neighbouring presences and the total observations within a 10 km radius around a focal location. This distance was chosen to account for the spatial effects of neighbouring locations on the response variable at a focal site (e.g. Augustin et al., 1996; Hubbell et al., 2001). As defined in this work, neighbourhood occurrence density (i.e. the autocovariate) decreased along the gradient of population aggregation from core to peripheral regions (Fig. 2).

By using both the kernel function and the autocovariate, we characterized spatial patterns of species distributions at two spatial scales: population aggregation at a large spatial scale (i.e. >10 km) from the kernel function; and neighbourhood occurrence density at a smaller scale (i.e. <10 km) by means of the autocovariate.

Variation of the spatial patterns in species distributions at both scales showed a common trend; that is, a decrease in relative species occurrence towards the periphery. We chose 10 km as the limit between the two scales of aggregation because this was the grid size used as a reference to locate in a stratified fashion the 1 km² sampled squares and therefore might influence the spatial patterning found in the species distributions.

2.4. Local habitat variables

As explanatory variables, we used habitat conditions recorded at a grid size of 1 km², matching the spatial resolution of the bird data. The variables recorded were climatic and topographic data, the land cover proportions from 2002, available from the web site of *Departament de Medi Ambient i Habitatge*, and other miscellaneous

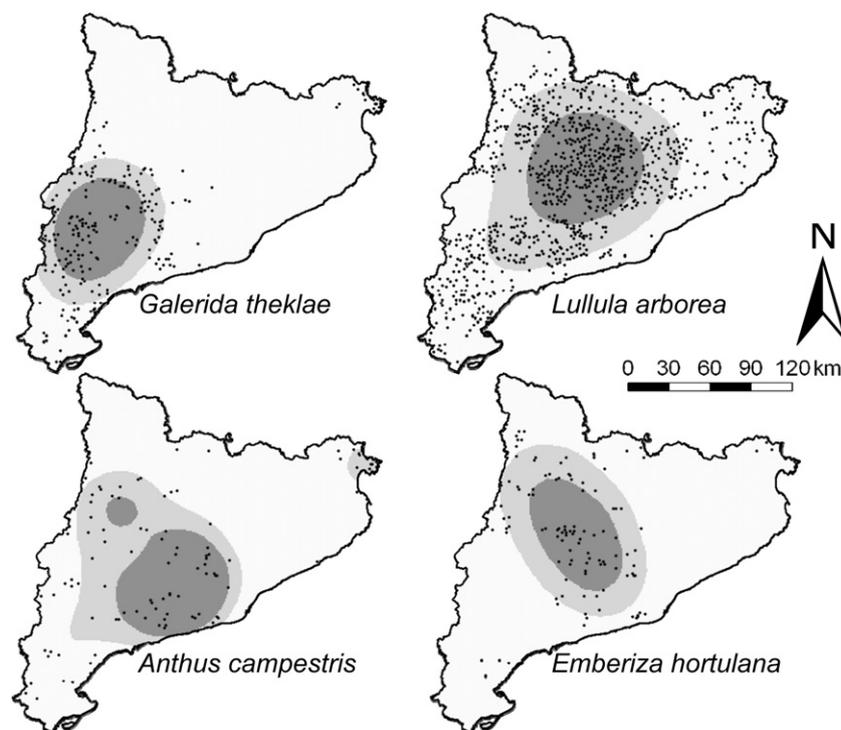


Fig. 1. Map of the functional regions within the distribution range of the species. Core, middle and peripheral regions (from dark grey to white) show a gradient with decreasing population aggregation (black points) estimated by means of the kernel density function. The bandwidth selected to define this function was the radius of species occurrence showing positive autocorrelation.

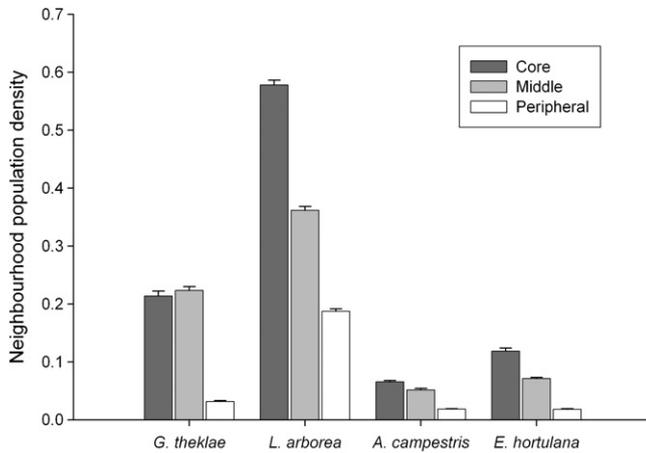


Fig. 2. Neighbourhood occurrence density estimated for the different regions of the distribution range (core, middle and peripheral) for the studied bird species. Neighbourhood occurrence density was estimated as the proportion between the number of neighbouring presences and the total observations within a 10 km radius (i.e. autocovariate). Differences between regions were significant ($p < 0.001$) except for *G. theklae* between core and middle regions.

variables (Table 2). The land cover classification was somewhat coarse for some land cover classes of interest such as shrubland, which included different open habitats (i.e. maquis brushwood, mountain grassland, pastures, abandoned crops, burnt areas and areas of transition between forest and secondary vegetation) (Viñas and Baulies, 1995). Since this land cover class is, in fact, the most important habitat for the species studied, we sub-classified the shrubland class by reference to a temporal series of Normalized Difference Vegetation Index (NDVI), (Mather, 1999) data derived from Landsat imagery from 2002 for February, April, May and

Table 2

Variables used in the global habitat models at 1 km², classified in four main groups: climatic, topographic, land cover (log transformed proportion of each land cover class) and miscellaneous variables.

| | |
|---------------|---|
| Climatic | Minimum temperature from January (modelled values according to the <i>Atlas Climàtic de Catalunya</i> (ACC), (Ninyerola et al., 2000)) Total precipitation (mm year ⁻¹) (modelled values according to the ACC (Ninyerola et al., 2000)) |
| Topographic | Slope (in %) Proportion of sunny surface (based on the slope aspect) Mean solar radiation (kJm ² day ⁻¹) (estimated by means of solar position and digital elevation model in ACC (Ninyerola et al., 2000)) |
| Land covers | Non vegetated bare: bare land, rocky outcrops, sparse vegetation, badlands, mineral extraction and stream beds Non irrigated herbaceous crops Non irrigated fruit trees Irrigated crops and fruit trees Non irrigated vineyards Shrubland: sub-classified in four categories based on NDVI values for different seasons as a surrogate of vegetation density (Chuvieco, 1999) and its phenological variation Coniferous forest: <i>Pinus halepensis</i> , <i>P. pinea</i> , <i>P. pinaster</i> , <i>P. nigra</i> , <i>P. sylvestris</i> , <i>P. uncinata</i> and <i>Abies alba</i> Deciduous forest: <i>Quercus faginea</i> , <i>Q. pubescens</i> , <i>Q. petraea</i> , <i>Q. robur</i> , <i>Fagus sylvatica</i> , <i>Populus</i> sp., <i>Salix</i> sp., <i>Castanea sativa</i> Sclerophyllous forest: <i>Quercus ilex</i> , <i>Q. suber</i> Built up areas: urban areas, industrial areas and infrastructures |
| Miscellaneous | Distance to rivers (log transformed) Distance to forest (log transformed) Distance to roads (log transformed) Mean and standard deviation of NDVI from the most representative month of vegetation during the breeding season (May). |

August. Accounting for different temporal NDVI images improves the shrubland classification by considering likely phenological changes in vegetation from the end of winter (February) until almost the end of summer (August). Since ground truth data were lacking, we used the iterative self organizing algorithm for clustering based on computing the minimum Euclidean distance when assigning each cell to a cluster (ISO cluster function from ArcGis; ESRI, 2005). In this way, we defined four shrubland sub-classes based on NDVI values that may be considered as a surrogate of vegetation density (Chuvieco, 1999).

2.5. Hypotheses evaluation

We hierarchically developed three different types of models (i.e. global habitat models, global spatial models and regional spatial models) to evaluate different hypothesis for the assessment of the geographical variation in distributional constraints along the distribution range of a species (Fig. 3).

2.5.1. Global habitat models

We first developed global habitat models using generalized linear models (GLM), with a logistic link function for binomial bird data. To detect non-linear relationships between the species and environmental variables, we examined the response curves at two degrees of freedom obtained by means of Generalized Additive Models (GAM). When the GAM response curves were significant, the quadratic term of the explanatory variable was included into the GLM. Model selection was performed using backward stepwise regression, including in the global models of each species all the variables that significantly ($p < 0.05$) contributed to minimizing the residual deviance of species distributions (Zar, 1998).

Global habitat models were built to analyse to what extent position throughout a gradient of population aggregation at a large spatial scale yielded spatially variable goodness of model fit (step 1, Fig. 3). We assessed the estimation–observation agreement of the global models separately for each region of the distribution range (core, middle and peripheral regions) by means of Receiver Operator Characteristic (ROC) curves. By running the model globally for the whole dataset, the number of variables included (i.e. model complexity) was the same for all regions making evaluation of model fit directly comparable between them. A ROC curve and its associated Area Under the Curve (AUC) is a graphic representation of the trade off between the estimated false negative and false positive rates for every possible cut off point (Fielding and Bell, 1997).

Since AUC is likely to be influenced by data prevalence (Jiménez-Valverde and Lobo, 2006) which varies from the core to the peripheral regions, we contrasted the results from the whole dataset for each region with those derived from a subset with differences of prevalence corrected by randomly removing the excess of absences to attain equal prevalence in all regions (Table 1). Global models were built for this new dataset using the same variables as in the previous models. Hence, we were able to obtain comparable AUC values without potentially interfering statistical artefacts regarding data prevalence. For the models built with equal prevalence between regions, we performed a second evaluation of the model fit to avoid likely inflated AUC values arising from the inclusion of a large geographical extent outside the species environmental domain (Lobo et al., 2008). Large AUC values could be obtained in peripheral regions simply by a large proportion of absences being well explained in areas outside the species niche. We partially controlled for this drawback by evaluating model fit from the corrected prevalence dataset only for the potentially suitable area for each species, i.e. within the geographic area above the 5% habitat suitability threshold for the species. Using this

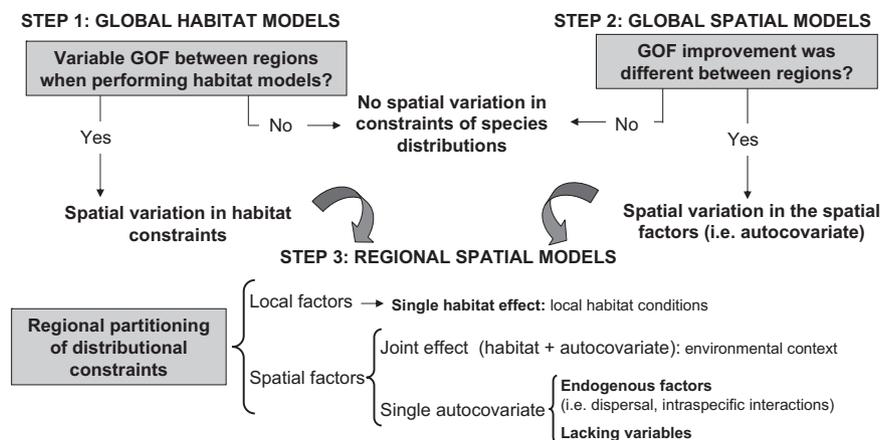


Fig. 3. Flow chart of the steps followed for the assessment of geographical variation in distributional constraints on species distributions. Step 1: Global habitat models were performed to assess whether goodness of fit was variable between regions of the species range by mean of Generalized Linear Models (GLM). Step 2: Global spatial models included an autocovariate to the previous models (i.e. autologistic models) allowing to test whether the influence of the spatial factor was variable between regions. Step 3: Regional spatial models were built to assess whether differences in the goodness of model fit between regions were confirmed. We partitioned the explained variation of species distributions into three different factors related to the local and spatial factors underpinning species distributions.

suitability threshold, we mostly removed observations where the species is absent in low suitable habitat conditions.

The application of these three different approaches for the evaluation of model fit (i.e. for the whole dataset, for the corrected prevalence dataset and the restricted AUC), allowed us to assess with more certainty the variation of model fit between regions.

2.5.2. Global spatial models

Afterwards, we evaluated whether including a spatial factor (i.e. using an autocovariate) improved the explanation of habitat models and whether this improvement was variable throughout the range (step 2, Fig. 3). Specifically, we performed autologistic models, including the same variables as in global habitat models and, subsequently, an autocovariate (see above). The quadratic term of the autocovariate was also included in models when significant. Separately for each region, we evaluated the significance of improvement in model fit using the S-PLUS roc library (Atkinson and Mahoney, 2004). Spatially variable improvement in goodness of model fit between regions after including the autocovariate indicates a geographic variation in the influence of neighbourhood occurrence density (step 2, Fig. 3). This suggests that the spatial factors driving species distributions may vary spatially.

2.5.3. Regional spatial models

If spatial variation in the constraints on species distributions were confirmed for both habitat conditions and the spatial factor, then regional GLMs were built by splitting the data into the different regions along the distribution range (step 3, Fig. 3). In this case, model selection was performed separately for each region, including the variables that significantly contributed to minimizing residual deviance of species distributions within each region according to backward stepwise regression. Regional spatial models allow a better assessment of regional variation in the dataset and provide a means to investigate ecological processes within each region (Osborne and Suarez-Seoane, 2002; Murphy and Lovett-Doust, 2007).

We assessed the role of local and spatial factors as drivers of species distributions by partitioning out the variation in species occurrence. Variation partitioning is a quantitative method in which the variation in species occurrence can be separated into components reflecting the relative role of different groups of explanatory variables and their shared effect (Borcard et al., 1992; Bahn et al., 2006; Carrete et al., 2007). We partitioned the

explained variation into three different components: a) the single effect of habitat conditions not shared by the autocovariate. This component represents the influence of local habitat conditions at a specific location after removing the likely influence of the spatial factor. b) The single effect of the autocovariate not shared by the habitat conditions, representing the relevance of neighbourhood occurrence density after removing the co-variation with habitat conditions. This may be indicative of interactions between neighbourhood occurrences related to the species' behaviour as a surrogate of endogenous factors affecting the spatial pattern of species aggregation (i.e. intra-specific interactions, dispersal). However, the single effect of the autocovariate might be also indicative of some other variables not included into the models. c) The joint effect of the habitat conditions and the autocovariate, indicative of the relevance of spatial factors matching the spatial pattern of habitat conditions. Thus, the shared effect of both habitat conditions and the autocovariate may be considered as a surrogate of environmental context in a broad sense, where conditions of occurrence density and habitat conditions in the surrounding environment are integrated. In this way, we decomposed the explanatory power of the spatial factor into two different components to account for the spatial effect of neighbouring locations at a focal site (step 3, Fig. 3) (e.g. Augustin et al., 1996; Hubbell et al., 2001).

3. Results

3.1. Habitat constraints throughout the species range

When comparing goodness of fit from global habitat models along the gradient of population aggregation (step 1, Fig. 3), we observed better fitting of the species–habitat relationships in peripheral regions, even after accounting for potential interfering differences in prevalence and restricting model evaluation to suitable areas for the species (restricted AUC). The comparison of AUC values by means of three different approaches evaluating the goodness of fit of global models (with the whole dataset, the dataset with corrected prevalence and the restricted AUC) showed consistent results for all four studied species (Table 3). We found a marked tendency for spatial variation in habitat constraints to become more important drivers of species distributions in areas with lower population aggregation (i.e. peripheral regions). However, while the fit of species–habitat relationships generally decreased gradually towards the core regions, this was not the case

Table 3
Goodness of fit (AUC value and standard error between brackets) of global habitat models within the different regions along the gradient of population aggregation. Models were performed using the whole dataset and the same dataset after correcting for differences in prevalence between regions. For these last models, a second correction was applied (restricted AUC), redrawing the ROC curve for each region using only potentially suitable areas for the studied species (area with habitat suitability above the 5th percentile for species presences). The largest AUC values for each species and for each approach of model evaluation appear in bold.

| Species | Regions | AUC whole dataset | AUC corrected prevalence | Restricted AUC |
|----------------------|------------|--------------------|--------------------------|--------------------|
| <i>G. theklae</i> | Core | 0.81 (0.03) | 0.80 (0.03) | 0.79 (0.03) |
| | Middle | 0.73 (0.03) | 0.72 (0.03) | 0.70 (0.03) |
| | Periphery | 0.93 (0.01) | 0.91 (0.02) | 0.91 (0.02) |
| <i>L. arborea</i> | Core | 0.65 (0.03) | 0.65 (0.03) | 0.63 (0.03) |
| | Middle | 0.74 (0.02) | 0.75 (0.02) | 0.72 (0.02) |
| | Peripheral | 0.82 (0.01) | 0.81 (0.02) | 0.78 (0.02) |
| <i>A. campestris</i> | Core | 0.68 (0.05) | 0.68 (0.05) | 0.67 (0.05) |
| | Middle | 0.80 (0.04) | 0.80 (0.04) | 0.76 (0.04) |
| | Peripheral | 0.83 (0.03) | 0.82 (0.03) | 0.75 (0.05) |
| <i>E. hortulana</i> | Core | 0.80 (0.04) | 0.80 (0.04) | 0.79 (0.04) |
| | Middle | 0.77 (0.04) | 0.80 (0.04) | 0.79 (0.04) |
| | Peripheral | 0.84 (0.03) | 0.82 (0.03) | 0.81 (0.04) |

for all species. *E. hortulana* showed similar goodness of fit between core and middle regions and *G. theklae* showed a better fit in the core than in the middle region (Table 3).

3.2. Influence of neighbourhood occurrence density throughout the species range

Fitting global models including the autocovariate (step 2, Fig. 3) allowed us to account for the likely spatial influence of neighbourhood occurrence density, which was largely variable throughout the distribution range of the studied species (Fig. 2). However, it was not enough to explain the differences in goodness of model fit between regions, which still showed weaker relationship between species and distributional constraints (habitat and spatial factors) in the middle and core regions compared to the peripheral regions (Table 4).

Neighbourhood occurrence density contributed to a better explanation of the species distributions, but not for all regions of the distribution range (Table 4). As predicted, the role of neighbourhood occurrence density was important in regions with large population aggregation (i.e. core regions), but only for two species (the *G. theklae* and the *L. arborea*). These species showed significant response to the negative quadratic term of the autocovariate (Table 4). This indicates that neighbourhood occurrence density increased the probability of presence at a focal site (i.e. aggregation) until reaching a certain occurrence threshold when the probability

of presence levelled off or even started to decrease. Conversely, *A. campestris* and *E. hortulana* only showed significant improvement in goodness of model fit after including the autocovariate in peripheral regions. For these last species, the probabilities of presence increased as the neighbourhood occurrence density rose, because only a positive linear term was significant (Table 4). Overall, the spatial influence of neighbourhood occurrence density on species distributions was the lowest in the middle regions for all four species except *G. theklae*, which showed non-significant improvement of goodness of model fit after including the autocovariate (p -value = 0.955, Table 4).

3.3. Regional spatial models

The spatial variation in the goodness of fit of global models throughout the species range confirmed the need to perform models separately for each region along the distribution range for a better assessment of the regional variation in the dataset (step 3, Fig. 3). Goodness of fit of the regional spatial models showed the same pattern of variation as the global spatial models, i.e. a better fit of the species–habitat relationship in peripheral regions. Importantly, even after building the models regionally, we did not find major improvements in goodness of model fit. Only for *A. campestris* was there an overall increase in AUC for all regions (Tables 4 and 5).

Partialling-out the ecological variation into three different components showed a common trend regarding local factors for all species (Table 5). The single effect of local habitat conditions was more important in the middle regions, showing for the studied species that habitat conditions (after removing likely spatial factors) were more important at intermediate levels of population aggregation. Decomposition of the spatial factor into two different components (the single effect of the autocovariate and the joint effect of habitat conditions and the autocovariate) showed that the role of environmental context was much more important at the periphery of the distribution for all studied species (Table 5). Conversely, the single effect of the autocovariate was much more variable, showing no common pattern among species (Table 5). However, in a similar way to the results shown for global spatial models, the single effect of the autocovariate had only a minor role in explaining species distributions in the middle regions, except for *L. arborea*.

4. Discussion

With this study we have shown how distributional constraints on species distributions are geographically variable throughout

Table 4
AUC from global spatial models and AUC increase (with standard error between brackets) when comparing with global habitat models for the whole dataset. When significant, the quadratic term of the autocovariate was included. Significant differences of AUC increase (p -value) were evaluated using S-PLUS roc library (Atkinson and Mahoney, 2004). Numbers in bold indicate the largest increase in goodness of model fit.

| Species | Autocovariate | Regions | Global spatial model | | |
|----------------------|----------------|------------|----------------------|--------------|------------|
| | | | AUC | % increase | p -value |
| <i>G. theklae</i> | quadratic term | Core | 0.84 (0.02) | 3.22 | 0.037 |
| | | Middle | 0.73 (0.03) | 0.11 | 0.955 |
| | | Peripheral | 0.94 (0.01) | 1.19 | 0.016 |
| <i>L. arborea</i> | quadratic term | Core | 0.74 (0.02) | 14.17 | <0.001 |
| | | Middle | 0.79 (0.02) | 6.02 | <0.001 |
| | | Peripheral | 0.84 (0.01) | 2.65 | 0.002 |
| <i>A. campestris</i> | linear term | Core | 0.69 (0.05) | 1.60 | 0.370 |
| | | Middle | 0.78 (0.04) | −2.69 | 0.059 |
| | | Peripheral | 0.86 (0.02) | 4.05 | 0.002 |
| <i>E. hortulana</i> | linear term | Core | 0.81 (0.04) | 2.01 | 0.454 |
| | | Middle | 0.76 (0.04) | −0.92 | 0.543 |
| | | Periphery | 0.89 (0.02) | 5.76 | <0.001 |

Table 5

Results from the variation partitioning procedure decomposing the proportion of explained variation (in %) in three different components: single effect of habitat conditions not shared by the autocovariate (single habitat); the shared effect between the habitat and the autocovariate (joint effect); and the single effect of the autocovariate not shared by the habitat conditions (single autocovariate, for the linear or quadratic term). Analyses were performed separately for each region of the distribution range: core, middle and peripheral. We also show the goodness of fit (AUC and standard error between brackets) from regional spatial models, where habitat and spatial factors (i.e. autocovariate) were included. In bold are the largest values for each species and components from the variation partitioning.

| Species | Regions | Single habitat | Joint effect | Single autocovariate | AUC |
|----------------------|------------|----------------|--------------|--------------------------|--------------------|
| <i>G. theklae</i> | Core | 57.79 | 37.04 | 5.17 (linear) | 0.87 (0.02) |
| | Middle | 80.57 | 19.40 | 0.02 n.s. | 0.74 (0.03) |
| | Peripheral | 37.09 | 54.08 | 8.83 (linear) | 0.94 (0.01) |
| <i>L. arborea</i> | Core | 24.20 | 33.23 | 42.57 (linear) | 0.75 (0.02) |
| | Middle | 42.64 | 39.74 | 17.61 (quadratic) | 0.80 (0.02) |
| | Peripheral | 30.33 | 57.53 | 12.14 (quadratic) | 0.84 (0.01) |
| <i>A. campestris</i> | Core | 49.93 | 16.18 | 33.89 (quadratic) | 0.78 (0.04) |
| | Middle | 82.64 | 16.90 | 0.46 n.s. | 0.84 (0.03) |
| | Peripheral | 64.01 | 28.05 | 7.95 (linear) | 0.91 (0.02) |
| <i>E. hortulana</i> | Core | 62.20 | 28.90 | 8.89 (linear) | 0.86 (0.03) |
| | Middle | 94.57 | 4.95 | 0.48 n.s. | 0.79 (0.04) |
| | Peripheral | 30.45 | 35.23 | 34.32 (quadratic) | 0.91 (0.02) |
| Average | Core | 48.53 | 28.84 | 22.63 | 0.81 |
| | Middle | 75.11 | 20.25 | 4.64 | 0.79 |
| | Peripheral | 40.47 | 43.72 | 15.81 | 0.90 |

a gradient of population aggregation. This appears to be a consequence of the inherent spatial character of ecological processes and its interaction with the varying environmental conditions throughout the species range (Wagner and Fortin, 2005).

Changes in overall habitat availability throughout the species range appeared to mediate changes in the goodness of fit of global habitat models, resulting in a stronger fit of the species–habitat relationship towards the periphery of species distributions (i.e. at low levels of population aggregation). Although some critical issues of using AUC could still arise, the different approaches used for the evaluation of model fit (i.e. for the whole dataset, for the corrected prevalence dataset and the restricted AUC) showed consistent results and are ecologically sound. As suggested by Luoto et al. (2005), a sparse fragmented distribution pattern may reflect tight habitat specificity of a species, increasing goodness of model fit at the periphery of the distribution range. Similar patterns of geographical variation in goodness of fit when performing global models were also found by Osborne and Suarez-Seoane (2002), showing larger explanatory power of models in peripheral than in core regions. The increasing environmental heterogeneity and decrease in overall habitat availability towards the periphery appear to be a limiting factor for species distributions at the landscape scale (Brown et al., 1996), where habitat conditions become the main constraint confining species to suitable patches (Wagner and Fortin, 2005).

Nonetheless, lower AUC values might also have been expected when considering dispersal limitations in areas with scattered populations (i.e. peripheral regions). If dispersal limits the species' movement into newly available habitats, unoccupied suitable habitats might be frequent, yielding a mismatch between species and habitat conditions. However, this effect is unlikely to influence the general pattern of goodness of model fit throughout the species range. More accurate determination of habitat suitability under field conditions would be needed for the proper assessment of unoccupied suitable habitats (Pulliam, 2000; Titeux et al., 2007), and this was not within the scope of our study.

Throughout the species range, we found a poorer fit of the species–environment relationship in middle and core regions. However, contrary to our expectations, goodness of model fit did not always decrease as population aggregation increased (towards the core) (Table 3). Explanations for this might lie beyond the factors considered in this study. For instance, the core region of *E. hortulana*, as defined in this work, has been largely displaced during the last 20 years through its association with fire

disturbance (Brotons et al., 2008). Thus for species showing distribution changes due to external influences, alternative definitions of core areas (Osborne and Suarez-Seoane, 2007) would be required to assess the spatial variability in distributional constraints across the species range. Although we were not able to derive generalizations about the differences between core and middle regions for all species, we found an overall laxer response of species distributions to habitat factors in these regions when compared to peripheral regions.

While the use of an autocovariate in global models allowed us to account for likely differences in contextual conditions throughout the species range (Araújo and Williams, 2001), it was not enough to explain differences in model fit between regions. The role of neighbourhood occurrence density in explaining species distributions was spatially variable along the gradient of population aggregation but the pattern of variation was different among species (Table 4). Therefore, the role of neighbourhood occurrence density within each region when performing global spatial model appears to be species specific. Contrary to our expectation, information on neighbourhood occurrence density did not always become more important at high levels of population aggregation (i.e. in core regions). The significance of the quadratic term of the autocovariate for *L. arborea* and *G. theklae* (which are more abundant in the study area than *E. hortulana* and *A. campestris*; see Table 1) shows that the probability of presence at a focal site does not always increase linearly with occurrence density. This suggests that more complex responses than linear ones might arise at certain levels of occurrence density as a likely result of the intra-specific interactions.

4.1. Ecological processes behind species distributions

Since the goodness of fit of global models was variable between regions of the distribution range, the need to optimise the explanatory power of models separately for each region was confirmed (Osborne and Suarez-Seoane, 2002; Murphy and Lovett-Doust, 2007). However, regional spatial models showed the same pattern of variation in model fit as the global models, with an overall poorer model fit in core and middle regions, suggesting the omission of some factors contributing to current species distributions for these areas. The larger number of spatially dependent occurrences in core and middle region may favour source and sink population dynamics (Pulliam, 2000), yielding a poor relationship between habitat conditions and species distributions (Gimona and Brewer, 2006). Furthermore, the high occurrence density in core

regions as estimated in this work by means of the autocovariate (Fig. 2) may lead to density-dependent species distribution patterns (Diez and Pulliam, 2007). This suggests that more detailed information on species density may be needed to model regions with large population aggregations (i.e. core regions). In addition, in core areas demographic parameters (birth, immigration, death and emigration rates) would also be needed to understand the role of source and sink population dynamics.

The superior explanatory power of regional models in peripheral regions may also be explained by the important role played by environmental context, as shown by the joint effect between habitat and the autocovariate (Table 5). Thus, where populations are sparsely distributed, the species presence at a focal site is not only a response to local habitat conditions, but also to the joint effect of habitat conditions and neighbourhood occurrence density at larger spatial scales (i.e. up to 10 km as estimated in this work) (Table 5). In other words, occurrence densities in peripheral areas largely co-vary with the spatial pattern of habitat conditions, suggesting that such conditions in the surroundings may also determine species distributions. In areas characterized by a sharp contrast between habitat patches, there is likely to be a large influence of surrounding habitat conditions on species distributions (Brotons et al., 2005). This may be the case in peripheral areas, where a hostile landscape matrix separating populations becomes an important constraint on neighbourhood occurrence density and, therefore, of species occurrence at a focal site.

Although we did not find any clear pattern in the role of endogenous processes in determining spatial pattern, the results from our global and regional models suggest that they are likely to be very important at both extremes of distribution ranges (at low and high levels of population aggregation) but less so in middle regions. This was corroborated by the large explanatory power of the pure effect of habitat conditions in the middle regions, showing no influence of spatial factors at intermediate levels of population aggregations. Although further research would be needed to confirm this, we envisage that different endogenous processes may be acting at each extreme of the gradient of population aggregation, which cancel out in middle regions where spatial processes appear to be negligible. In core regions, intra-specific interactions might be the dominant processes, whereas in peripheral regions the effective spatial isolation of populations (Holt and Keitt, 2000) might mean that dispersal limitations are the dominant behavioural process underpinning patterns of species distributions. Although dispersal limitation as a driver of local aggregation has been described for plants (Leps, 1990) such a process might also be expected for birds (Brotons et al., 2005).

Although in this work we discuss the common pattern found in the geographical variation in distributional constraints along a gradient of population aggregation for the four studied species, we stress that results might be different for other species, especially if we consider the likely role of numerous factors affecting geographical variation of species niches such as evolutionary processes and interspecific interactions (Holt, 2003; Mönkkönen et al., 2004).

4.2. Conclusion

In this study, we have stressed the importance of considering geographical variation in distributional constraints and its interaction with spatial environmental heterogeneity throughout the species range in explaining species distributions. The geographical variation in the explanatory power of habitat conditions and the spatial factor along a gradient of population aggregation described here help us to understand the determinants of species distributions throughout the species range. The results shown here, based

on explanatory models, may be considered as a starting point to evaluate the ecological causes of non-stationarity, which should be considered when making predictions of species distributions based on global models. Although assessment of predictive performance of global and regional models was not within the scope of this work, results from Osborne and Suarez-Seoane (2002) assessing the predictive power of models showed the same pattern of variation as model fit. Therefore our results based on explanatory models jointly with those from Osborne and Suarez-Seoane (2002) suggest that confidence in species distribution models, especially at large spatial extents, should be lower at higher levels of population aggregation (normally away from the periphery of the range). A better understanding of the determinants of species distributions provides valuable insights into the likely species' responses to environmental changes and habitat perturbations. Based on our results, predictions of species distributions into the future might be improved by accounting for geographical variation in distributional constraints. In addition, further knowledge in the processes driving changes in species distributions in core areas will be also required, for instance, density-dependent processes and dispersal.

Therefore, we need to consider explicitly the geographical variation in distributional constraints when making management decisions. As shown in this work, areas with low population aggregation (the peripheries of ranges) appear to be highly constrained by both habitat conditions and the environmental context. This, coupled with the higher probability of local extinction in peripheral regions (Brown et al., 1996), may make these populations more vulnerable to environmental change (Guo et al., 2005). Conversely, in order to assess better the influence of distributional constraints on species distributions for areas with high population aggregation (i.e. core regions), information about abundance processes and demographic parameters should be considered. Finally, spatial processes related to the species' behaviour may be very important at both extremes of the distribution range. Since the processes favouring aggregated distributions are difficult to distinguish from only distributional data (van Teeffelen and Ovaskainen, 2007), further studies are needed to assess how processes of dispersal and intra-specific interaction co-vary with spatial environmental heterogeneity throughout the distribution range. Overlooking the effects of spatial variation in distributional constraints can lead to erroneous management conclusions and inadequate estimation of the species' response to environmental changes.

Acknowledgements

We thank all volunteers that participated in the CBBA. Thanks to A. Gil-Tena for her useful comments and help on spatial autocorrelation analyses. This work has received financial support from projects CGL2005-2000031/BOS, CGL2008-05506/BOS and CONSOLIDER-MONTES CSD2008-00040 granted by the Spanish Ministry of Education and Science. L.B. benefited from a Ramon y Cajal contract from the Spanish government and S.V. (FI fellowship) received financial support from the CUR of the DIUE from the Generalitat de Catalunya and the European Social Fund.

References

- Araújo, M.B., Williams, P.H., 2001. The bias of complementarity hotspots toward marginal populations. *Conserv. Biol.* 15, 1710–1720.
- Atkinson, B., Mahoney, D., 2004. Roc: S-PLUS ROC functions. Mayo Clinic. <http://mayoresearch.mayo.edu/mayo/research/biostat/splusfunctions.cfm>.
- Augustin, N.H., et al., 1996. An autologistic model for the spatial distribution of wildlife. *J. Appl. Ecol.* 33, 339–347.
- Bahn, V., et al., 2006. Importance of spatial autocorrelation in modelling bird distributions as a continental scale. *Ecography* 29, 835–844.

- Borcard, D., et al., 1992. Partialling out the spatial component of ecological variation. *Ecology* 73, 1045–1055.
- Brotons, L., et al., 2008. Wild fires and large scale changes in distribution of open habitat species: the ortolan bunting (*Emberiza hortulana*) in Mediterranean landscapes. *J. Appl. Ecol.* 45, 1059–1066.
- Brotons, L., et al., 2005. Colonization of dynamic Mediterranean landscapes: where do birds come from after fire? *J. Biogeogr.* 32, 789–798.
- Brown, J.H., 1984. On the relationship between abundance and distribution of species. *Am. Nat.* 124, 255–279.
- Brown, J.H., et al., 1995. Spatial variation in abundance. *Ecology* 76, 2028–2043.
- Brown, J.H., et al., 1996. The geographic range: size, shape, boundaries and internal structure. *Annu. Rev. Ecol. Syst.* 27, 597–623.
- Carrete, M., et al., 2007. Habitat, human pressure, and social behavior: partialling out factors affecting large-scale territory extinction in an endangered vulture. *Biol. Conserv.* 136, 143–154.
- Caughley, G., et al., 1988. The edge of the range. *J. Anim. Ecol.* 57, 771–785.
- Chuvieco, E., 1999. Measuring changes in landscape patterns from satellite images: effects of large fire disturbances. *Int. J. Remote Sens.* 20, 2331–2346.
- Departament de Medi Ambient i Habitatge. Avd/Diagonal, 523-525. Barcelona <http://mediambient.gencat.net/cat/inici.jsp> (accessed January 2006).
- Diez, J.M., Pulliam, H.R., 2007. Hierarchical analysis of species distributions and abundance across environmental gradients. *Ecology* 88, 3144–3152.
- ESRI, 2005. ArcGIS 9.0. Environmental Systems Research Institute, Redlands, CA.
- Estrada, J., et al., 2004. Catalan Breeding Bird Atlas (1999–2002). Lynx Editor.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38–49.
- Fielding, A.H., Howarth, P.F., 1995. Testing the generality of bird-habitat models. *Conserv. Biol.* 9, 1466–1481.
- Fortin, M.J., Dale, M., 2005. *Spatial Analysis: A Guide for Ecologists*. Cambridge University Press.
- Fortin, M.J., et al., 2005. Species' geographic ranges and distributional limits: pattern analysis and statistical issues. *Oikos* 108, 7–17.
- Gimona, A., Brewer, M.J., 2006. Local environmental effects and spatial effects in macroecological studies using mapped abundance classes: the case of rook *Corvus frugilegus* in Scotland. *J. Anim. Ecol.* 75, 1140–1146.
- Guo, Q., et al., 2005. Spatial-temporal population dynamics across species range: from centre to margin. *Oikos* 108, 47–57.
- Herrando, S., et al., 2002. Bird community succession after fire in a dry Mediterranean shrubland. *Ardea* 90, 303–310.
- Hirzel, A., Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modeling. *Ecol. Modell.* 157, 331–341.
- Holt, R.D., 2003. On the evolutionary ecology of species' range. *Evol. Ecol. Res.* 5, 159–178.
- Holt, R.D., Keitt, T.H., 2000. Alternative causes for range limits: a metapopulation perspective. *Ecol. Lett.* 3, 41–47.
- Hubbell, S.P., et al., 2001. Local neighborhood effects on long-term survival of individual trees in a neotropical forest. *Ecol. Res.* 16, 859–875.
- Jiménez-Valverde, A., Lobo, J., 2006. The ghost of unbalanced species distribution data in geographical model predictions. *Divers. Distrib.* 12, 521–524.
- Kunin, W.L., 1998. Extrapolating species abundance across spatial scales. *Science* 281, 1513–1515.
- Leps, J., 1990. Can underlying mechanisms be deduced from observed patterns? In: Krahulec, F., Agnew, A.D.Q., Willems, J.H. (Eds.), *Spatial Processes in Plant Communities*. SPB Academic Publishing, The Hague, pp. 1–11.
- Lichstein, J.W., et al., 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecol. Monogr.* 72, 445–463.
- Lobo, J.M., et al., 2008. AUC: misleading measure of the performance of predictive distribution models. *Glob. Ecol. Biogeogr.* 17, 145–151.
- Luoto, M., et al., 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Glob. Ecol. Biogeogr.* 14, 575–584.
- Mather, P.M., 1999. *Computer Processing of Remotely Sensed Images*. John Wiley, New York.
- McGill, B., Collins, C., 2003. A unified theory for macroecology based on spatial patterns of abundance. *Evol. Ecol. Res.* 5, 469–492.
- Mönkkönen, M., et al., 2004. Qualitative geographical variation in interspecific interactions. *Ecography* 27, 112–118.
- Murphy, H.T., Lovett-Doust, J., 2007. Accounting for regional niche variation in habitat suitability models. *Oikos* 116, 99–110.
- Ninyerola, M., et al., 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *Int. J. Climatol.* 20, 1823–1841.
- Osborne, P.E., et al., 2007. Non-stationarity and local approaches to modelling the distributions of wildlife. *Divers. Distrib.* 13, 313–323.
- Osborne, P.E., Suarez-Seoane, S., 2002. Should data be partitioned spatially before building large-scale distribution models? *Ecol. Modell.* 157, 249–259.
- Osborne, P.E., Suarez-Seoane, S., 2007. Identifying core areas in a species' range using temporal suitability analysis: an example using little bustards *Tetrax tetrax* L. in Spain. *Biodivers. Conserv.* 16, 3505–3518.
- Peterson, A.T., Holt, R.D., 2003. Niche differentiation in Mexican birds: using point occurrence to detect ecological innovation. *Ecol. Lett.* 6, 774–782.
- Pulliam, H.R., 2000. On the relationship between niche and distribution. *Ecol. Lett.* 3, 349–361.
- Rosenberg, M.S., 2009. PASSAGE2. Pattern Analysis, Spatial Statistics, and Geographic Exegesis. School of life science, Arizona State University.
- Sherry, T.W., Holmes, R.T., 1985. Dispersion patterns and habitat responses of birds in northern hardwoods forests. In: Cody, M.L. (Ed.), *Habitat Selection in Birds*. Academic Press, pp. 283–309.
- Titeux, N., et al., 2007. Fitness-related parameters improve presence-only distribution modelling for conservation practice: the case of the red-backed shrike. *Biol. Conserv.* 138, 207–223.
- Vallecillo, S., et al., 2009. Dangers of predicting bird species distributions in response to land-cover changes. *Ecol. Appl.* 19, 538–549.
- van Teeffelen, A.J., Ovaskainen, O., 2007. Can the cause of aggregation be inferred from species distributions? *Oikos* 116, 4–16.
- Viñas, O., Baulies, X., 1995. 1:2 50 000 Land-use map of Catalonia (32 000 km²) using multitemporal landsat-TM data. *Int. J. Remote Sens.* 16, 129–146.
- Wagner, H.H., Fortin, M.J., 2005. Spatial analysis of landscape: concepts and statistics. *Ecology* 86, 1975–1987.
- Whittaker, J., 1967. Gradient analysis of vegetation. *Biol. Rev. Camb. Philos. Soc.* 42, 207–264.
- Wiens, J.A., 1985. Habitat selection in variable environments: shrub-steppe birds. In: Cody, M.L. (Ed.), *Habitat Selection in Birds*. Academic Press, pp. 227–251.
- Zar, J.H., 1998. *Biostatistical Analysis*. Prentice-Hall.