

ORIGINAL
ARTICLE

Do we need land-cover data to model species distributions in Europe?

Wilfried Thuiller^{1,2*}, Miguel B. Araújo^{1,2} and Sandra Lavorel^{1,3}

¹Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de Recherche Scientifique, Montpellier cedex 5, France, ²Macroecology and Conservation Unit, University of Évora, Estrada dos Leões, Évora 7000-730, Portugal and ³Laboratoire d'Ecologie Alpine, CNRS, Université J. Fournier, Grenoble Cedex 9, France

ABSTRACT

Aim To assess the influence of land cover and climate on species distributions across Europe. To quantify the importance of land cover to describe and predict species distributions after using climate as the main driver.

Location The study area is Europe.

Methods (1) A multivariate analysis was applied to describe land-cover distribution across Europe and assess if the land cover is determined by climate at large spatial scales. (2) To evaluate the importance of land cover to predict species distributions, we implemented a spatially explicit iterative procedure to predict species distributions of plants (2603 species), mammals (186 species), breeding birds (440 species), amphibian and reptiles (143 species). First, we ran bioclimatic models using stepwise generalized additive models using bioclimatic variables. Secondly, we carried out a regression of land cover (LC) variables against residuals from the bioclimatic models to select the most relevant LC variables. Finally, we produced mixed models including climatic variables and those LC variables selected as decreasing the residual of bioclimatic models. Then we compared the explanatory and predictive power of the pure bioclimatic against the mixed model.

Results (1) At the European coarse resolution, land cover is mainly driven by climate. Two bioclimatic axes representing a gradient of temperature and a gradient of precipitation explained most variation of land-cover distribution. (2) The inclusion of land cover improved significantly the explanatory power of bioclimatic models and the most relevant variables across groups were those not explained or poorly explained by climate. However, the predictive power of bioclimatic model was not improved by the inclusion of LC variables in the iterative model selection process.

Main conclusion Climate is the major driver of both species and land-cover distributions over Europe. Yet, LC variables that are not explained or weakly associated with climate (inland water, sea or arable land) are interesting to describe particular mammal, bird and tree distributions. However, the addition of LC variables to pure bioclimatic models does not improve their predictive accuracy.

Keywords

Bioclimatic models, climate gradients, land-use effects, large-scale patterns, species distributions.

*Correspondence: Wilfried Thuiller, Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de Recherche Scientifique, 1919 route de Mende, 34293 Montpellier cedex 5, France. E-mail: wilfried.thuiller@cefe.cnrs-mop.fr

INTRODUCTION

It is widely accepted that distributions of plants and animals are broadly constrained by their physiological tolerances to climatic factors (Woodward, 1987, 1990). This generalization

is held to be true for species at a variety of spatial scales (Whittaker *et al.*, 2001), although there is a wide recognition that the importance of climate is best expressed at large spatial scales (Rahbek & Graves, 2001; Willis & Whittaker, 2002). Following this recognition, bioclimatic models are being used

W. Thuiller *et al.*

to predict the distribution of plant and animal species at large spatial scales (for a review see Guisan & Zimmermann, 2000). However, bioclimatic models may produce inaccurate predictions when important local or regional factors are missing from input data in the models (Iverson *et al.*, 1999; Thuiller *et al.*, 2003b). As bioclimatic models assume species distributions to be at equilibrium with current environmental conditions, they are bound to over-predict distributions of species that are highly affected by human activities and that, consequently, may be missing from potentially suitable areas (Loehle & LeBlanc, 1996; Araújo, 2003). This problem should be particularly severe in regions where a positive relationship between species occurrence and human population density has been reported (e.g. Araújo, 2003).

For example, in Europe, land use changes, in particular agricultural intensification, have led to widespread changes in the distribution and abundance of many different taxa (Benton *et al.*, 2003). More specifically, numerous studies have revealed how important land use and other human activities affect mammal (Mladenoff *et al.*, 1999), bird (Chamberlain *et al.*, 1999), herptiles (Shenbrot *et al.*, 1991) and plant (Duckworth *et al.*, 2000) species distributions at regional scales.

As human management activities alter natural dynamics of species within ecosystems, bioclimatic models are expected to provide increasingly inefficient predictions of species distributions (e.g. Araújo *et al.*, 2001). A possibility to compensate for such human-induced factors is to include within-species distribution models additional variables expected to summarize important factors affecting local distributions of species. The inclusion of these variables should be expected to improve the accuracy of bioclimatic models (Loehle & LeBlanc, 1996; Midgley *et al.*, 2003). In this paper, we analyse the usefulness of land cover (LC) variables to predict plant, bird, mammal and herptile species distributions over Europe. We also distinguish between the explanatory and predictive power of land cover. Variables with high explanatory value help to understand the determinants of species distribution, whereas variables with predictive value play an important role for model development without necessary being ecologically relevant. Some variables can play both roles, whereas others are more orientated toward one role. The distinction between both components has thus serious implication in practice.

We address the following questions:

1. How much of land-cover distribution is explained by climate?
2. How does land cover increase the explanatory power of bioclimatic models at large spatial scales?
3. How does land cover increase the predictive power of bioclimatic models at large spatial scales?

METHODS

Species data

Original species data included 2294 plant (Jalas & Suominen, 1972–1996), 186 mammal (Mitchell-Jones *et al.*, 1999), 143

amphibian and reptile (Gasc *et al.*, 1997) and 440 breeding bird species (Hagemeyer & Blair, 1997). Terrestrial vertebrates include all known species (Araújo *et al.*, 2001) whereas plants comprise only c. 20% of the European flora (Humphries *et al.*, 1999). The grid used is based on the Atlas Florae Europaeae (AFE; Lahti & Lampinen, 1999), with cells boundaries typically following the 50 km lines of the Universal Transvers Mercator (UTM) grid. The remaining atlases use slightly different grid-systems, including different rules to represent data on islands and coasts. Hence, vertebrate data were converted to the AFE grid system by identifying unique correspondence between cells in these grids (Williams *et al.*, 2000). The mapped area (2434 grid cells) includes western, northern and southern Europe, but exclude most of the eastern European countries where recording effort was both less uniform and less intensive (Williams *et al.*, 2000).

As plants are by far the largest including a great variety of responses, we divided them into three life-form-based groups according to Tutin *et al.* (1964–1993): herbs, shrubs and trees.

Bioclimatic data

We used a comprehensive set of bioclimatic variables for analyses (T.D. Mitchell, 2002, unpublished data): mean annual, winter and summer precipitation, mean annual temperature and minimum temperature of the coldest month, growing degree days (> 5°) and an index of humidity (mean ratio of annual actual evapotranspiration over annual potential evapotranspiration). Mean values are averaged from 1960 to 1990. These data were supplied on a 10° grid, covering Europe and then aggregated by averaging to 50 × 50 km UTM in order to match with the resolution of species data.

Land-cover data

The land-cover data were developed at a spatial resolution of 10° for grid cells based on the ATEAM project geographical window (<http://www.pik-potsdam.de/ateam/>). The 10° grid cells were derived by aggregation from the PELCOM land-cover data base. PELCOM is a 1-km pan European land cover data base developed mainly from remotely sensed data. The classification methodology is based on a regional and integrated approach of the NOAA-AVHRR satellite data and ancillary information such as topographic features (Mücher, 2000). Although finer spatial resolution data bases exist such as CORINE (CEC, 1993), PELCOM was selected due to its complete spatial coverage of the European window, and because of the homogeneity of the methodology used for the land-cover classification. PELCOM is also the most up-to-date of pan-European land-cover data bases.

The percentages of each land-cover class were calculated for the individual 10° grid cells from the 1 km PELCOM cells. The data comprised four classes of land cover: forest, agriculture, urban (perurban) and others. Forests were further subdivided into three classes: percentage of deciduous, coniferous and mixed forest (respectively named perdec,

percon and permtf). Agriculture was also divided into three classes: the percentage of arable lands, grasslands and permanent crops (respectively named perara, pergrass and preperme). The 'other' land-cover class comprised: percentage of semi-natural areas (divided into areas of shrubland, and barrenland, respectively named pershrub and perbarren), inland waters (perinwater), wetlands (perwetland), permanent ice and snow (perpice) and sea (persea). Similarly to climate, we aggregated 10° grid data in 50 × 50-km grid to match with species data by calculating the percentage of each class in the new grid system.

Model

Bioclimatic distribution of land cover in Europe

Given that climate governs global patterns of land cover (Dale, 1997), land cover and climate are not fully independent. Hence, before implementing predictive models of species distributions we investigated patterns of co-variation between land-cover and climate variables. This was performed with the Outlying Mean Index (OMI), a co-inertia analysis recently developed by Doledec *et al.* (2000). This analysis is similar to the well-known canonical correspondence analysis (CCA: ter Braak, 1986), but allows to separate land-cover classes and to measure the distance between the mean bioclimatic conditions used by each land-cover class and the mean bioclimatic conditions across Europe. Moreover, unlike CCA, this method

has the advantage of making no assumptions about the shape of land-cover response curves to the bioclimate. We applied OMI to separate land-cover classes in Europe by performing a normalized PCA on the bioclimatic table and linking the land-cover table to the PCA output table using the adequate diagonal and metric matrices (Thuiller *et al.*, 2003c).

The first two selected axes explained 96% of the total inertia, or variation in the data. The first axis was mainly related to temperature and the second to precipitation (Fig. 1). Projecting land cover (LC) variables on the two-dimensional bioclimatic space allows describing their distributions (Fig. 1). For instance, permanent crops occurred mainly at cooler temperatures, whereas grasslands were mainly located in mid-altitude mountains less intensively disturbed by human management. However, some LC variables were found to be widespread within climatic space. This was the case, for example, of percentage of sea, inland water, arable lands, urban area or coniferous forests (Fig. 1).

Given these patterns of co-variation between land cover and bioclimatic variables, if we modelled species distribution with both kinds of variables and with a stepwise variable-selection procedure (Chambers & Hastie, 1997) we would rarely select LC variables as bioclimatic variables explains the most important components of their distributions. However, LC variables may still contribute significantly to model species distribution in Europe when more regional patterns of LC, corresponding to the residual part of the OMI, may influence distributions of particular species.

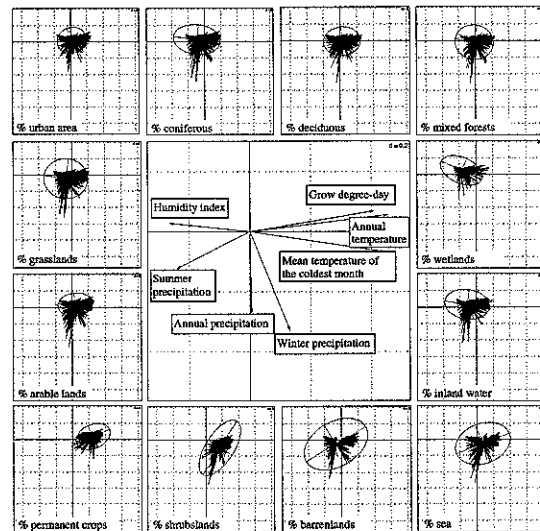


Figure 1 Centre: Canonical weight of bioclimatic variables and relationships with the first two axes of the OMI. Sixteen plots show the distribution of land-cover variables on the first two axes of the OMI analysis. Crosses identify the position of the land-cover variables in the bioclimatic space using the canonical weights of bioclimatic variables. Ellipses represent the land cover inertia representing at least 90% of points for each variable. The dashed lines emanating from the centroids of each ellipse represent the major and minor axes of the ellipse.

Rule of procedure

Step	Rule of procedure
Step 1	Stepwise GAM using only bioclimatic variables (GAM-clim)
Step 2	Stepwise GAM on residuals from step 1 using only land-cover variables
Step 3	GAM using bioclimatic variables retained in step 1 and land-cover variables retained in step 2 (GAM-clim-LC)
Step 4	Evaluation of GAM-clim accuracy on evaluation data using ROC curve
Step 5	Comparative ANOVA between GAM-clim and GAM-clim-LC (test)
Step 6	If test is significant: evaluation of GAM-clim-LC on evaluation data using ROC curve. Else, evaluation of GAM-clim on evaluation data (same as step 3)
Step 7	Averaging of ROC curve from GAM-clim on all species
Step 8	Averaging of ROC curve from GAM-clim-LC (if test is significant) and GAM-clim (if test is not significant)

Table 1 Descriptive steps of the iterative procedure

Including land-cover variables into bioclimatic models

In order to explore for the potential residual contribution of land cover explaining current regional distributions of species, we implemented an iterative framework (Table 1) based on a particular implementation of Generalized Additive Models (GAM). Here we first run stepwise GAM with bioclimatic variables and then assessed if the inclusion of LC variables improved the explanatory and predictive performance of bioclimatic models.

Original data were randomly divided into a calibration (70%) and an evaluation (30%) data set. Species with more than 10 occurrences in the calibration data were selected for model calculations (165 mammals, 117 herptiles, 387 breeding birds and 1527 plants remaining). Stepwise GAMs were then run using selected bioclimatic variables. Then we ran stepwise GAM of residuals against LC variables. Selected LC variables for each species were then added to the set of bioclimatic variables to produce mixed bioclimatic models including land cover. To assess the explanatory power of land cover, comparative ANOVAs were performed between the pure bioclimatic and the mixed models. To assess the importance of land cover as predictive variables a test of accuracy was performed on the evaluation data using the relative operating characteristic (ROC) curve (Pearce & Ferrier, 2000; Thuiller *et al.*, 2003a). The ROC curve was applied to pure bioclimatic models and to the mixed models. Subjective guidelines (Swets, 1988) suggest very good agreement for AUC above 0.9 (Table 1). All analyses were performed with Splus (Anon., 1999).

RESULTS

How does land cover improve the explanatory power of bioclimatic models at a large spatial scale?

The inclusion of land cover increased the explanatory power of bioclimatic models for species of all groups (Table 2). This was particularly true for mammals, birds and trees, where the inclusion of land cover increased the explained deviance of bioclimatic models by more than 60%.

Table 2 Explanatory power of land-cover variables. Number of species corresponds to the number of species studied in this analysis. Step 2 corresponds to the number of species where land-cover variables were retained as decreasing residuals of bioclimatic models (step 2, Table 1). Step 5 corresponds to the number of species for which mixed models had a higher explained deviance than pure bioclimatic models (step 5, Table 1). Percentage corresponds to the percentage of cases where the inclusion of land cover increased the explanatory power of bioclimatic models

	No. of species	Step 2	Step 5	Percentage
Mammals	165	162	107	65
Birds	387	377	241	62
Trees	71	71	43	61
Shrubs	78	78	43	55
Herptiles	117	110	62	53
Herbs	1378	1265	624	45

In order to explore the additional effects of land cover after bioclimatic models have been adjusted, we selected a few species with established relationships with land cover at lower spatial scales. This was the case for wolf *Canis lupus*, whose populations are known to be inversely correlated with human population and activities (Breitenmoser, 1998; Massolo & Meriggi, 1998). In our analysis, the LC variables retained for this species were percentage of urban areas and percentage of grasslands. The response curves to these two variables were exponential negative and linearly negative, respectively, showing that even at large spatial scales important LC variables can be detected by models. Overall, the three LC variables most often retained by models for mammals were the percentages of arable land, permanent crops and mixed forests (Fig. 2, mammals). For birds, we analysed models for the grey partridge *Perdix perdix*. Robinson *et al.* (2001) showed that the grey partridge has declined in Britain as a consequence of the intensification of agricultural practices and the decrease of available arable land. In support of this study, our analysis selected arable land as the most important LC variable for the species (Table 1). More generally, the inclusion of the percentage of arable land into models of distributions for

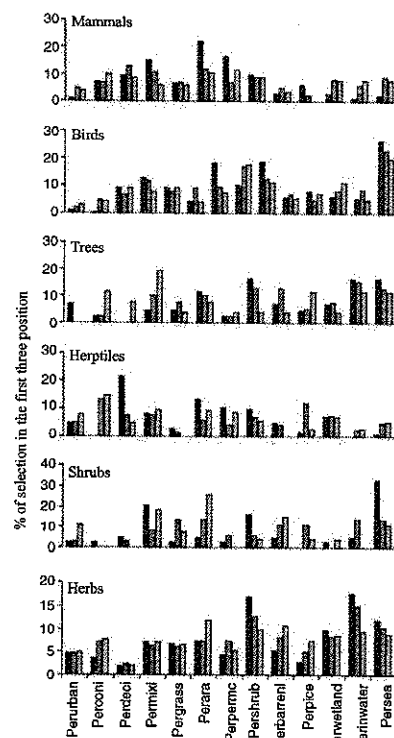


Figure 2 Percentage of models using each land-cover variable as one of the first three land-cover variables selected in step 2 for decreasing the residuals of pure bioclimatic models. Dark grey, grey and clear grey bars represent, respectively, the most, the second most and the third most explanatory selected variables during the process for each group of taxa.

farmland birds increased significantly the explanatory power of models. As a whole, agriculture (percentage of arable land and permanent crops) and presence of sea were the most important LC variables accounting for residuals in bird distributions (Fig. 2, birds). For trees, we focused on *Quercus petraea*, a Euro-Siberian species with wide distribution (Tutin *et al.*, 1964–1993). The only selected LC variable was the percentage of deciduous forest, with a strong positive relationship. This is, indeed, a deciduous tree dominant in central European deciduous forests. Another example is *Q. pedunculiflora*, which was negatively correlated to permanent crops and positively related to the percentage of shrublands. For trees, in general,

there were no unequivocally dominant LC variables explaining residuals, although presence of sea, inland water and shrubs appeared often as useful to explain distributions (Fig. 2, plants-trees).

For herptiles and shrubs, the inclusion of land cover increased the explanatory power of bioclimatic models for more than half of the species (Table 2). Herptiles responded mainly to percentage of deciduous, coniferous forests and arable land (Fig. 2, herptiles), while shrubs were more related to percentage of sea, arable land and mixed forests (Fig. 2, plants-shrubs). These observations are consistent with other published records. For instance, Martin & Lopez (2002) showed that the lizard *Lacerta lepida* occurred mainly under *Quercus* forest and understory low bushes and avoided croplands. Here, the two retained variables were percentage of mixed forest (positive relation) and permanent crops (negative relation). For shrubs, we chose to investigate the response of the pioneer white willow *Salix alba* and found that the percentage of arable land (positive relationship) was the only additional variable selected for the species.

Finally, herbs were the group for which the inclusion of land cover produced the least improvement over pure bioclimatic models (Table 2). The most relevant variables selected were the percentage of inland water, sea and shrubs (Fig. 2, plants-herbs). We focused on the woodland herbs group, which are mainly correlated to percentage of deciduous and mixed forest. For instance, the yellow anemone *Anemone ranunculoides* that typically occurs under beech forests (Tutin *et al.*, 1964–1993) has a positive unimodal relationship with percentage of deciduous and a negative relationship with mixed forests and permanent crops. The superb pink *Dianthus superbus*, which occurs principally under open forests (Tutin *et al.*, 1964–1993), was as expected positively related to mixed forests and negatively to arable land.

How does land cover improve the predictive power of bioclimatic models at a large spatial scale?

The predictive power of models was assessed using AUC on evaluation data. We concluded that models, on average, provided very good results for all groups (Table 3).

Shrubs and herptiles were the species best predicted by models with an average AUC of 0.95 for evaluation. Conversely, mammals had the lowest (although still rather high) mean AUC values with 0.914. The lowest values of AUC were generally comprised 0.66 and 0.80 while maximum values of AUC were equal or very close to 1 (Table 3).

Predictions using bioclimatic and LC data showed very similar results to predictions using bioclimatic data alone. Although there were no significant differences between AUC for both analyses, some general trends can be outlined. First, the overall minimum AUC was generally higher (but not for trees) for models using LC data than for models that did not use it (Table 3). Secondly, standard deviations of AUC for models using bioclimatic and LC data were lower than for models using only bioclimatic variables (Table 3). The last two

Table 3 Statistical summary of AUC from the ROC curve procedure displaying the standard deviation (SD), the minimum (min), the average (mean) and maximum (max) for each group according to the models (bioclimatic model against mixed including land-cover variables). Difference correspond to the net difference between mean AUC from bioclimatic and mixed models

	Bioclimatic model				Mixed model				Difference
	SD	Min	Mean	Max	SD	Min	Mean	Max	
Herptiles									
Cali	0.0467	0.8017	0.9573	0.9999	0.0413	0.8099	0.9627	0.9999	-0.0054
Eval	0.0505	0.7791	0.9497	0.9988	0.0487	0.7933	0.9504	0.9988	-0.0007
Mammals									
Cali	0.0582	0.7457	0.9277	0.9992	0.0522	0.7812	0.9358	0.9992	-0.0081
Eval	0.0632	0.7267	0.9140	0.9987	0.0603	0.7552	0.9133	0.9975	0.0007
Birds									
Cali	0.0459	0.7853	0.9388	0.9990	0.0421	0.8227	0.9447	0.9990	-0.0059
Eval	0.0525	0.7592	0.9254	0.9980	0.0512	0.7592	0.9260	0.9980	-0.0006
Herbs									
Cali	0.0419	0.7341	0.9568	1.0000	0.0419	0.7341	0.9567	1.0000	0.0001
Eval	0.0503	0.6628	0.9313	1.0000	0.0503	0.6628	0.9313	1.0000	0.0000
Shrubs									
Cali	0.0314	0.8692	0.9708	1.0000	0.0256	0.8950	0.9743	1.0000	-0.0035
Eval	0.0421	0.8041	0.9530	1.0000	0.0406	0.8046	0.9558	1.0000	-0.0028
Trees									
Cali	0.0277	0.8950	0.9642	0.9993	0.0327	0.8482	0.9615	0.9993	0.0027
Eval	0.0347	0.8188	0.9436	1.0000	0.0384	0.8136	0.9414	1.0000	0.0032

points support the idea that the inclusion of LC variables may help improve predictions for species where bioclimatic models yielded poorer predictions.

Amongst the examples provided, there is no species for which the inclusion of land cover increased significantly the predictive accuracy of bioclimatic models. However, land cover improved bioclimatic models (on evaluation data) only for a limited number of species: 34% of mammals, 40% of herptiles, 34% of birds and 21% of plants. Animals appeared to be more related to land cover, but differences between AUC on evaluation data are too narrow to draw clear conclusions on the benefit from including land cover to predict animal species distributions at large scales.

DISCUSSION

Distribution of land cover in Europe

The OMI multivariate analysis of environmental data revealed that, over Europe, the spatial variation of land cover is highly correlated to spatial variation among bioclimatic variables. This is an important result given that it explains why including both types of information into species distribution models may bring redundant information. The degree to which this is the consequence of the large spatial scale and coarse resolution of data remains unknown, but it is possible that models of species distributions using a smaller spatial extent and using finer resolution data might produce a different pattern (Franklin,

1995; Collingham *et al.*, 2000). Indeed, a hierarchical scheme of environmental controls on species distributions has been suggested (but not demonstrated), in which climatic variables are large-scale determinants, followed by geology, land cover and topography, which moderate many of the effects of macroclimatic variables (Franklin, 1995). In our analysis, although a large proportion of the variability of land-cover distribution was described by climate, there were residuals related to regional land cover patterns that could not be explained by climate. There were indeed some LC variables either weakly or not related to climatic factors, such as inland water, sea, arable land and coniferous forests. Sea is present around a large proportion of Europe and is obviously not dependent on climate. Similarly to sea, in Europe, inland water is not restricted by climate and is present across all regions. More surprisingly the distribution of coniferous forests did not appear to be influenced by climate. The reason for this is that this class brings together a variety of species with a large range of climatic and ecological requirements, which makes it difficult to find a general pattern (Richardson, 1998). For instance, the *Pinus* genus is present in the Siberian plains (*Pinus sibirica*), in the Alps mountains (*P. cembra*), in restricted areas of the Mediterranean region (*P. halepensis*) and in the Euro-Siberian region (*P. sylvestris*). Such a variety of chorotypes explains why the conifer distributions do not appear to be controlled by climate in our analysis, although the species making-up this class are indeed related to climate at European scale (Richardson, 1998). Arable lands occur across

the whole of Europe mainly because this class embeds adequate lands suitable for agriculture independently of the type of cultures.

These widespread LC variables that are little related, or even unrelated, to climate at the European scale are, however, influenced by multiple regional or local factors such as soil type, slope, aspect or groundwater distribution and availability, that were not taken into account in our analysis.

Explanation of current species distributions

As widespread LC variables are not restricted by climate, it was not surprising to find that they were often retained by GAM models (Table 1, step 2), and decreased significantly the residual from pure bioclimatic models for many species. Arable lands and permanent crops improved significantly the explanatory power of bioclimatic models for several breeding birds and mammals. Our results corroborate previous studies showing the importance, positive or negative, of arable land for birds, mammals and plants from regional to local scales (e.g. Robinson *et al.*, 2001; Robinson & Sutherland, 2002; Benton *et al.*, 2003). These authors showed that the intensification of agriculture and the contraction of arable cultivations have produced important local extinctions of birds and mammals (Robinson *et al.*, 2001). Our study emphasized that even at large spatial scales and coarse resolutions, human-related variables could be useful to describe and explain species distributions (e.g. *Perdix perdix*).

However, our approach is only correlational and the ecological meaning of the observed relationships between land cover and species distributions is not always obvious. Although bioclimatic models identify correlational relationships, there have been a number of studies investigating causal relationships between species and bioclimatic variables (at least for plants) (Woodward, 1987, 1992). This kind of analysis has not been developed, as far as we are aware, for species and land cover at large spatial scales. The causal relationships between variables such as land cover and species distributions are not always easy to uncover since they are often indirect. This is the case for plants as land cover can be seen more as a limiting factor (at least for human-oriented variables) than a factor having direct physiological impact. However, for vertebrates the relationship can be more direct. For instance, high quality habitats for winter farmland birds (stubble, game cover and game feeders) occur principally in arable lands (Benton *et al.*, 2003). Granivorous passerines and small rodents are obviously strongly related to permanent crops and arable land where they find refuges and food.

Prediction of species distribution ranges

From a predictive standpoint, LC variables did not improve significantly the predictive performance of our models built on evaluation data. Several non-exclusive hypotheses can be proposed to explain this pattern:

(1) Goodness-of-fit does not make an assessment of predictive performance (although their failure underscores aspects of the relationship between the modelled species distributions and the explanatory variables used in the model) (Hosmer & Lemeshow, 1989). Even if some LC variables improve the explanatory power of bioclimatic models, this does not imply that they would also improve the predictive performance.

(2) Climate explains almost all LC variation and the residuals are not large enough to improve the predictive performance of bioclimatic models. The residuals of the climate-LC relationships may have a strongly regional distribution and therefore do not provide a strong underlying gradient affecting species distributions. Hence they may not be relevant for predictive purposes.

(3) In modelling species against bioclimatic variables it is often assumed that variables can be grouped into three important groups: direct, resources and indirect variables (Austin & Smith, 1989). Direct or resource gradients provide mechanistic and physiological explanations for distributions of species. Models that use such variables are supposed to have greater predictive performance than models using indirect variables, hence they should be more accurate when predicting distributions over large areas or at other times in the future (Guisan & Zimmermann, 2000). Indirect variables, such as land cover, are supposed to have little direct physiological relevance for species. Hence they should only be applied within a limited geographical extent without significant errors, because in a different region, or time, the same LC variables can correspond to a different combination of direct and resource gradients (Guisan & Zimmermann, 2000).

(4) Using the 50 × 50-km grid cell resolution across Europe, we demonstrated that LC distribution was driven mainly by climatic gradients. However, inside each grid cell there is an underlying high heterogeneity of LC that is lost (see the PELCOM data base), while this heterogeneity is a crucial factor for predicting current species occurrences (Cowling & Lombard, 2002; Benton *et al.*, 2003). This constrain by the data combines two potential problems. First, the problem of using data with insufficient resolution to explain observed patterns (Robinson *et al.*, 1992; Pearson, 1993). Secondly, the possibility that LC variables used may not include all relevant factors. For example, variables representing the indices of agricultural intensification (Mader, 1984), habitat heterogeneity (Bascompte & Solé, 1996; Allen & O'Connor, 2000; Gaston *et al.*, 2002), frequency and intensity of perturbation (Croonquist & Brooks, 1993), density of roads (Mader, 1984) could be more relevant to improve the predictive ability of bioclimatic models for some species, even at a 50 × 50-km resolution.

ACKNOWLEDGMENTS

MBA thanks the national atlas representatives and the many volunteer fieldworkers who contributed to the atlas records, species distribution data was kindly supplied by J.P. Gasc (Societas Europaea Herpetologica), W.J.M. Hagemeijer

(European Bird Census Council), Raino Lampinen (Atlas Flora Europaea), and A.J. Mitchell-Jones (Societas Europaea Mammalogica). Tim Mitchell (Tyndall Centre for Climate Research) generated climate data, Markus Erhard and Sönke Zaehle ('Potsdam Institut für Klimafolgenforschung') aggregated climate surfaces to the AFE grid. Marc Rounsevell and Isabelle Reginster (University of Louvain-La-Neuve) for aggregated PELCOM land-cover data to the AFE grid. Current research is funded by the European Commission ATEAM (EVK2-CT-2000-00075) project, MBA also thanks FCT (SPRH/BPD/5547/2001) for funding.

REFERENCES

- Allen, A.P. & O'Connor, R.J. (2000) Interactive effects of land cover and other factors on regional bird distributions. *Journal of Biogeography*, **27**, 889–900.
- Anon. (1999) *S-PLUS 2000 user's guide*. MathSoft, Seattle, WA.
- Araújo, M.B. (2003) The coincidence of people and biodiversity in Europe. *Global Ecology and Biogeography*, **12**, 5–12.
- Araújo, M.B., Humphries, C.J., Densham, P.J., Lampinen, R., Hagemeijer, W.J.M., Mitchell-Jones, A.J. & Gasc, J.P. (2001) Would environmental diversity be a good surrogate for species diversity? *Ecography*, **24**, 103–110.
- Austin, M.P. & Smith, T.M. (1989) A new model for the continuum concept. *Vegetatio*, **83**, 35–47.
- Bascompte, J. & Solé, R.V. (1996) Habitat fragmentation and extinction thresholds in spatially explicit models. *Journal of Animal Ecology*, **65**, 465–473.
- Benton, T.G., Vickery, J.A. & Wilson, J.D. (2003) Farmland biodiversity: is habitat heterogeneity the key? *TREE*, **18**, 182–188.
- ter Braak, C.J.F. (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167–1179.
- Breitenmoser, U. (1998) Large predators in the Alps: the fall and rise of man's competitors. *Biological Conservation*, **83**, 279–289.
- CEC (1993) *CORINE land cover technical guide*, European Union. Directorate-General Environment, Nuclear Safety and Civil Protection. Office for Official Publications of the European Communities, Luxembourg.
- Chamberlain, D.E., Wilson, A.M., Browne, S.J. & Vickery, J.A. (1999) Effects of habitat type and management on the abundance of skylarks in the breeding season. *Journal of Applied Ecology*, **36**, 856–870.
- Chambers, J.M. & Hastie, T.J. (1997) *Statistical models in S*. Chapman & Hall, London.
- Collingham, Y.C., Wadsworth, R.A., Huntley, B. & Hulme, P.E. (2000) Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology*, **37**, 13–27.
- Cowling, R.M. & Lombard, A.T. (2002) Heterogeneity, speciation/extinction history and climate: explaining regional plant diversity patterns in the Cape Floristic Region. *Diversity and Distributions*, **8**, 163–179.
- Croonquist, M.J. & Brooks, R.P. (1993) Effects of habitat disturbance on bird communities in riparian corridors. *Journal of Soil and Water Conservation*, **48**, 65–70.
- Dale, V.H. (1997) The relationship between land-use change and climate change. *Ecological applications*, **7**, 753–769.
- Doledec, S., Chessel, D. & Gimaret-Carpentier, C. (2000) Niche separation in community analysis: a new method. *Ecology*, **81**, 2914–2927.
- Duckworth, J.C., Bunce, R.G.H. & Malloch, A.J.C. (2000) Vegetation-environment relationships in Atlantic European calcareous grasslands. *Journal of Vegetation Science*, **11**, 15–22.
- Franklin, J. (1995) Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474–499.
- Gasc, J.-P., Cabela, A., Crnobrnja-Isailovic, J., Dolmen, D., Grossenbacher, K., Häfner, P., Lescure, J., Martens, H., Martínez Rica, J.P., Maurin, H., Oliveira, M.E., Sofianidou, T.S., Veith, M. & Zuidervijk, A. (1997) *Atlas of amphibians and reptiles in Europe*. Societas Europaea Herpetologica & Museum National d'Histoire Naturelle, Paris.
- Gaston, K.J., Pressey, R.L. & Margules, C.R. (2002) Persistence and vulnerability: retaining biodiversity in the landscape and in protected areas. *Journal of Biosciences*, **27**, 361–384.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Hagemeijer, W.J.M. & Blair, M.J. (1997) *The EBCC atlas of European breeding birds, their distribution and abundance*. Poyser, London.
- Hosmer, D.W. & Lemeshow, S. (1989) *Applied logistic regression*. Wiley, New York.
- Humphries, C., Araújo, M.B., Williams, P., Lampinen, R., Lahti, T. & Uotila, P. (1999) Plant diversity in Europe: Atlas Florae Europaeae and WORLDMAP. *Acta Botanica Fennica*, **162**, 11–21.
- Iverson, L.R., Prasad, A. & Schwartz, M.W. (1999) Modelling potential future individual tree-species distributions in the Eastern United States under climate change scenario: a case study with *Pinus virginiana*. *Ecological Modelling*, **115**, 77–93.
- Jalas, J. & Suominen, J. (1972–1996) *Atlas Florae Europaeae*. The Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, Helsinki.
- Lahti, T. & Lampinen, R. (1999) From dot maps to bitmaps – Atlas Florae Europaeae goes digital. *Acta Botanica Fennica*, **162**, 5–9.
- Loehle, C. & LeBlanc, D. (1996) Model-based assessments of climate change effects on forests: a critical review. *Ecological Modelling*, **90**, 1–31.
- Mader, H.J. (1984) Animal habitat isolation by roads and agricultural fields. *Biological Conservation*, **29**, 81–96.
- Martin, J. & Lopez, P. (2002) The effect of Mediterranean dehesa management on lizard distribution and conservation. *Biological Conservation*, **108**, 213–219.
- Massolo, A. & Meriggi, A. (1998) Factors affecting habitat occupancy by wolves in northern Apennines (northern Italy): a model of habitat suitability. *Ecography*, **21**, 97–107.

- Midgley, G.F., Hannah, L., Millar, D., Thuiller, W. & Booth, A. (2003) Developing regional and species-level assessments of climate change impacts on biodiversity: a preliminary study in the Cape Floristic Region. *Biological Conservation*, **112**, 87–97.
- Mitchell-Jones, A.J., Amori, G., Bogdanowicz, W., Krystufek, B., Reijnders, P.J.H., Spitzenberger, F., Stubbe, M., Thissen, J.M.B., Vohralik, V. & Zima, J. (1999) *Atlas of European mammals*. Academic Press, London.
- Mladenoff, D.J., Sickley, T.A. & Wydeven, A. (1999) Predicting gray wolf landscape recolonization: logistic regression models vs. new field data. *Ecological Applications*, **9**, 37–44.
- Mücher, C.A. (ed.) (2000) *PELCOM project 'development of a consistent methodology to derive land cover information on a European scale from remote sensing for environmental modelling'*. Final report, February 2000, Part 2, EU DG XII-D, European Commission.
- Pearce, J. & Ferrier, S. (2000) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127–147.
- Pearson, S.M. (1993) The spatial extent and relative influence of landscape-level factors on wintering bird populations. *Landscape Ecology*, **8**, 3–18.
- Rahbek, C. & Graves, G.R. (2001) Multiscale assessment of patterns of avian species richness. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4534–4539.
- Richardson, D.M. (1998) *Ecology and biogeography of Pinus*. Cambridge University Press, Cambridge.
- Robinson, R.A. & Sutherland, W.J. (2002) Post-war changes in arable farming and biodiversity in Great Britain. *Journal of Applied Ecology*, **39**, 157–176.
- Robinson, G.R., Holt, R.D., Gaines, M.S., Hamburg, S.P., Johnson, M.L., Fitch, H.S. & Martinko, E.A. (1992) Diverse and contrasting effects of habitat fragmentation. *Science*, **257**, 524–526.
- Robinson, R.A., Wilson, J.D. & Crick, H.Q.P. (2001) The importance of arable habitat for farmland birds in grassland landscapes. *Journal of Applied Ecology*, **38**, 1059–1069.
- Shenbrot, I.G., Rogovin, K.A. & Surov, A.V. (1991) Comparative analysis of spatial organization of desert lizard communities in Middle Asia and Mexico. *Oikos*, **61**, 157–168.
- Swets, K.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- Thuiller, W., Araújo, M.B. & Lavorel, S. (2003a) Generalized models versus classification tree analysis: a comparative study for predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, **14**, 669–680.
- Thuiller, W., Vaydiera, J., Pino, J., Sabaté, S., Lavorel, S. & Gracia, C. (2003b) Large-scale environmental correlates of forest tree distributions in Catalonia (NE Spain). *Global Ecology and Biogeography*, **12**, 313–325.
- Thuiller, W., Lavorel, S., Midgley, G.F., Laverne, S. & Rebelo, A.G. (2003c) Relating plant traits and species distributions along bioclimatic gradients for 88 *Leucadendron* species in the Cape Floristic Region. *Ecology* (in press).
- Tutin, T.G., Burges, N.A., Chater, A.O., Edmondson, J.R., Heywood, V.H., Moore, D.M., Valentine, D.H., Walters, S.M. & Webb, D.A. (1964–1993) *Flora Europaea*. Cambridge University Press, Cambridge.
- Whittaker, R.J., Willis, K.J. & Field, R. (2001) Scale and species richness: towards a general, hierarchical theory of species diversity. *Journal of Biogeography*, **28**, 453–470.
- Williams, P., Humphries, C., Araújo, M.B., Lampinen, R., Hagemeyer, W., Gasc, J.-P. & Mitchell-Jones, T. (2000) Endemism and important areas for representing European biodiversity: a preliminary exploration of atlas data for plants and terrestrial vertebrates. *Belgian Journal of Entomology*, **2**, 21–46.
- Willis, K.J. & Whittaker, R.J. (2002) Species diversity-scale matters. *Science*, **295**, 1245–1248.
- Woodward, F.I. (1987) *Climate and plant distribution*. Cambridge University Press, Cambridge.
- Woodward, F.I. (1990) The impact of low temperatures in controlling the geographical distribution of plants. *Philosophical Transactions of the Royal Society of London B*, **326**, 585–593.
- Woodward, F.I. (1992) Predicting plant responses to global environmental change. *New Phytologist*, **122**, 239–251.

BIOSKETCHES

The authors of this paper participated in the European Union's ATEAM project. **Wilfried Thuiller**, **Miguel B. Araújo** and **Sandra Lavorel** are involved with ATEAM in the 'Biodiversity and nature conservation' sector. The main interest of the lead author is to study the determinants of species diversity at the biogeographical level in order to forecast the impacts of future global change on distributions of diversity. He is currently working on a PhD examining patterns and causes of diversity distribution in Europe and South Africa. The main interests of Miguel B. Araújo's concern the development of reserve selection techniques that are more robust to local extinctions of species. Sandra Lavorel's research interests focus on the dynamics of plant diversity in landscapes, especially in relation to natural and land-use disturbances.