Journal of Biogeography, **30**, 1907–1920

Prediction of butterfly diversity hotspots in Belgium: a comparison of statistically focused and land use-focused models

Dirk Maes[1]*, Marius Gilbert[2], Nicolas Titeux[3], Philippe Goffart[4] and Roger L. H. Dennis[5]
[1]*Institute of Nature Conservation, Brussels,* [2]*Laboratoire de Biologie Animale et Cellulaire, Université Libre de Bruxelles, Brussels,* [3]*Unité d'Environnemétrie et de Géomatique, Faculté d'Ingénierie Biologique, Agronomique et Environnementale, Université catholique de Louvain, Louvain-la-Neuve,* [4]*Observatoire de la Faune, de la Flore et des Habitats (OFFH), Centre de Recherche de la Nature des Forêts et du Bois, Gembloux, Belgium and* [5]*School of Biological and Molecular Sciences, Oxford Brookes University, Headington, Oxford, UK*

Abstract

Aim We evaluate differences between and the applicability of three linear predictive models to determine butterfly hotspots in Belgium for nature conservation purposes.

Location The study is carried out in Belgium for records located to Universal Transverse Mercator (UTM) grid cells of 5×5 km.

Methods We first determine the relationship between factors correlated to butterfly diversity by means of modified *t*-tests and principal components analysis; subsequently, we predict hotspots using linear models based on land use, climate and topographical variables of well-surveyed UTM grid cells ($n = 197$). The well-surveyed squares are divided into a *training set* and an *evaluation set* to test the model predictions. We apply three different models: (1) a 'statistically focused' model where variables are entered in descending order of statistical significance, (2) a 'land use-focused' model where land use variables known to be related to butterfly diversity are forced into the model and (3) a 'hybrid' model where the variables of the 'land use-focused model' are entered first and subsequently complemented by the remaining variables entered in descending order of statistical significance.

Results A principal components analyses reveals that climate, and to a large extent, land use are locked into topography, and that topography and climate are the variables most strongly correlated with butterfly diversity in Belgium. In the statistically focused model, biogeographical region alone explains 65% of the variability; other variables entering the statistically focused model are the area of coniferous and deciduous woodland, elevation and the number of frost days; the statistically focused model explains 77% of the variability in the training set and 66% in the evaluation set. In the land use-focused model, biogeographical region, deciduous and mixed woodland, natural grassland, heathland and bog, woodland edge, urban and agricultural area and biotope diversity are forced into the model; the land use-focused model explains 68% of the variability in the training set and 57% in the evaluation set. In the hybrid model, all variables from the land use-focused model are entered first and the covariates elevation, number of frost days and natural grassland area are added on statistical grounds; the hybrid model explains 78% of the variability in the training set and 67% in the evaluation set. Applying the different models to determine butterfly diversity hotspots resulted in the delimitation of spatially different areas.

Main conclusions The best predictions of butterfly diversity in Belgium are obtained by the hybrid model in which land use variables relevant to butterfly richness are entered first after which climatic and topographic variables were added on strictly statistical

*Correspondence: Dirk Maes, Institute of Nature Conservation, Kliniekstraat 25, B-1070 Brussels, Belgium. E-mail: dirk.maes@instnat.be

© 2003 Blackwell Publishing Ltd

grounds. The land use-focused model does not predict butterfly diversity in a satisfactory manner. When using predictive models to determine butterfly diversity, conservation biologists need to be aware of the consequences of applying such models. Although, in conservation biology, land use-focused models are preferable to statistically focused models, one should always check whether the applied model makes sense on the ground. Predictive models can target mapping efforts towards potentially species-rich sites and permits the incorporation of un-surveyed sites into nature conservancy policies. Species richness distribution maps produced by predictive modelling should therefore be used as pro-active conservation tools.

Keywords
Modelling, species richness, biodiversity hotspots, conservation, butterflies, Belgium.

INTRODUCTION

Species distribution databases are the primary source material used in nature conservation (Lobo *et al.*, 1997; Dennis *et al.*, 1999). Ideally, such databases consist of an equal number of visits applied, in a standardized manner, to all the mapping units within the geographical frame where species are recorded. However, most databases are adversely affected by unequal sampling effort in both time and space (Dennis & Thomas, 2000) and by differences in the ability of recorders to detect and identify species accurately (Dennis & Hardy, 1999), even those of taxonomic groups generally considered as well studied (e.g. birds, mammals and plants; Williams & Gaston, 1994). Nature conservancy policies are mostly based on these incomplete and biased distribution databases (Lobo *et al.*, 1997). This may lead to non-optimal use of limited resources in nature conservation by wrongly prioritizing the designation or acquisition of areas for conservation (Myers *et al.*, 2000; Pullin & Knight, 2001).

The recently accelerated decline in butterfly diversity in north-west Europe (Maes & Van Dyck, 2001; Warren *et al.*, 2001; Dennis & Shreeve, 2003) calls for a rapid, accurate and cost-effective assessment of species richness over large regions. Recently, several authors have used predictive modelling as a conservation tool, both in poorly investigated taxonomic groups (e.g. dung beetles; Lobo & Martín-Piera, 2002) as in more 'popular' groups in countries or regions where large areas have been under-surveyed (Sparks *et al.*, 1995; Dennis *et al.*, 2000, 2002; Fleishman *et al.*, 2001b). Predictive modelling permits targeting of recorders towards potentially or predicted species-rich areas (Dennis & Hardy, 1999), can delimit priority sites for conservation (so called 'hotspots', i.e. sites with a large number of species; Myers *et al.*, 2000) and facilitates decision making on the impact of land-use changes in un-recorded sites (Fleishman *et al.*, 2001b). This greatly extends the value of collected records in distribution databases and increases the efficiency of mapping schemes that usually have limited logistical and financial resources.

Within the large group of invertebrates, butterflies are certainly the most intensively recorded organisms worldwide and butterfly mapping schemes exist for most of the north-west European countries (Tax, 1989; Stoltze, 1996; Settele *et al.*, 1999; Lafranchis, 2000; Asher *et al.*, 2001). But, even with a relatively large number of volunteers, butterfly distribution databases do not overcome the problem of biases in mapping (Dennis & Thomas, 2000). Furthermore, most distribution atlases do not indicate mapping intensity on species distribution maps that enable the reader to interpret distributions (Dennis & Hardy, 1999). High quality data on butterfly distribution and on biotopes, topography and climate are readily available in Belgium (Goffart & De Bast, 2000; Maes & Van Dyck, 2001). Some Belgian regions, however, remain poorly surveyed and others are almost certainly under-recorded. Furthermore, recorders generally prefer to visit sites that are known to have a high species richness than to survey new sites where species richness is unknown (Dennis & Thomas, 2000).

In this paper (1) we determine land use, topographic and climate factors that correlate with butterfly diversity in Belgium, (2) we develop three predictive models: a statistically focused, a land use-focused and a hybrid model to predict butterfly diversity using linear modelling and (3) we predict butterfly diversity hotspots using these three models. We compare the results of the three models and discuss their applicability for nature conservancy policy. Although the models are geographically limited to Belgium, we believe that this – from a biodiversity point of view – strongly impoverished north-west European country can be taken as a model area that has some representative character for many other industrialized regions elsewhere in Europe and in the rest of the world.

MATERIALS AND METHODS

Study area

Belgium is a strongly industrialized north-west European country with a high human population density (335 inhabitants per km^2; Van Goethem, 2001) and, consequently, intense pressure on nature (OECD, 1998). The general landscape and topography differ considerably between the two administrative regions of Belgium: Flanders and Wallonia. Flanders, the northern part, is a lowland zone

(average elevation = 38m) and only has a limited total area of nature reserves (1.6% of the territory; Van Goethem, 2001); the most butterfly-rich Flemish biotopes are heathlands and woodlands in the Campine region. Wallonia, the southern part and comparatively an upland region (average elevation = 310m) has a similar total area of nature reserves (*c.* 1% of the territory; Van Goethem, 2001); here, the most species-rich butterfly biotopes are found on nutrient-poor grasslands and in large woodlands in the Fagne-Famenne-Calestienne and in the Lorraine regions.

The study extends to the whole territory of Belgium (Fig. 1). We used the Universal Transverse Mercator (UTM) projection as this mapping grid is used in all invertebrate recording schemes in Belgium. Units of distribution for the present analyses have a grid size of 25 km^2 (5×5 km), hereafter called 'squares' ($n = 1374$). For the rest of the analyses, we only consider those squares that have an area of at least 24 km^2 and have >90% of their area within Belgium ($n = 1108$). The squares of the correction zone of the UTM projection are therefore excluded from the present analyses (see Fig. 1).

Butterfly data

Butterfly distribution data were obtained from two separate databases managed by the two regional butterfly working groups: (1) the Flemish database consists of about 210 000 records covering 95% of all Flemish squares ($n = 644$; Maes & Van Dyck, 2001) and (2) the Walloon database contains about 50 000 records covering 63% of all Walloon squares ($n = 802$; Goffart & De Bast, 2000). The organization of Flemish and Walloon mapping schemes is separated for several reasons: nature conservancy policy was regionalized in 1980 which means that since then both the Flemish and Walloon government can decide autonomously on nature conservation matters; additionally, species composition differs considerably between the two regions: Flanders has 64 indigenous species of which 46 are still present (Maes & Van Dyck, 2001), Wallonia has 103 native species of which 87 have been recorded since 1990 (Goffart & De Bast, 2000). We excluded all migrant species (*Vanessa atalanta*, *V. cardui*, *Colias croceus*, *C. hyale*) and, where possible, also observations of vagrants (observations not indicative of a breeding population) for the present analyses.

Environmental variables

Three types of 'environmental' data were collected: (1) land use data were derived from the Belgian Corine Land Cover vector map (CEC, 1994), (2) topographic variables were derived from a digital elevation model for Belgium (resolution 20 m, National Geographical Institute) and (3) climate point data were made available by the Royal Meteorological Institute of Belgium for the period 1996–2001 (Table 1). The areas of land use and topographic variables were estimated per square using ArcView3.2 (ESRI, Redlands, CA, USA). In addition, a biotope diversity index (BDI) was estimated per square using only terrestrial biotopes (Shannon diversity index; Magurran, 1988). The length of the edges between grasslands and heathlands on the one hand and deciduous and mixed woodlands on the other was also estimated per square using ArcView GIS 3.2 and Corine Land Cover maps. As Belgium is located in two European biogeographical regions (EEA, 2002), a binary variable 'region' was incorporated into the analyses: the Atlantic biogeographical region, north of the rivers Meuse and Sambre, and the Continental biogeographical region, south of these rivers (Dufrêne & Legendre, 1991; Fig. 1). Climate data were interpolated to the squares by universal kriging (Isaaks & Srivastava, 1989) when a clear spatial structure could be modelled or, alternatively, using the 'inverse weighted distance' interpolation method. Universal kriging

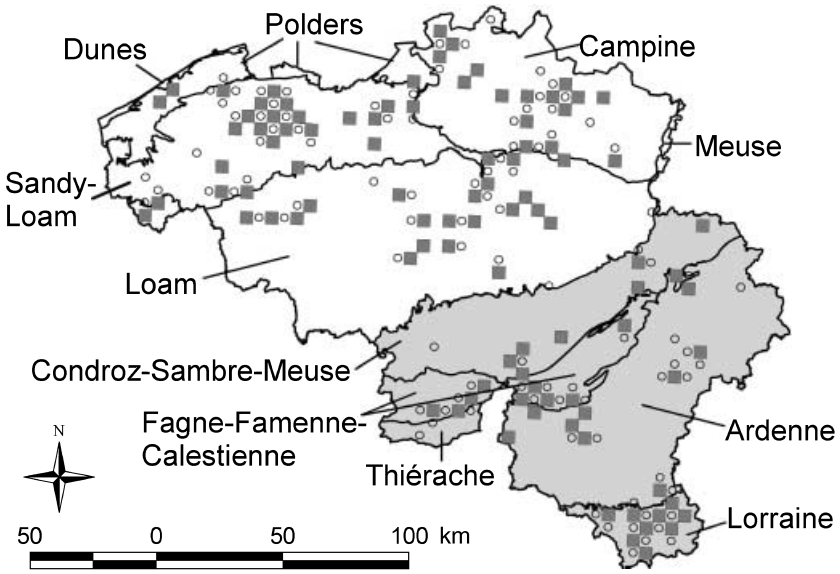

Figure 1 Squares used to build three predictive models for butterfly diversity in Belgium: grey squares are used as training set, empty circles as an evaluation set; the eleven ecological regions in Belgium are illustrated. The Atlantic and the Continental biogeographical regions are shown in white and grey, respectively.

Table 1 Environmental variables, their abbreviation (Abbr.), applied transformation prior to analyses (Transf.) and the source of the data

Name	Abbr.	Transf.	Source
Land use data			
Urban area	Ur	Sqrt	CORINE codes beginning with 1
Arable land	Ar	no	CORINE codes beginning with 2
Deciduous woodland	Dw	Log	CORINE code 311
Mixed woodland	Mw	Log	CORINE code 313
Coniferous woodland	Cw	Log	CORINE code 312
Natural grassland	Gr	Log	CORINE code 321
Heathland and bog	Hb	Log	CORINE code 322 + 412
Shrub	Sh	Log	CORINE code 324
Salt marshes	Sm	Log	CORINE code 421 + 423
Dunes	Du	Log	CORINE code 331
Water	Wa	Log	CORINE code 511 + 512 + 522
Marsh	Ma	Log	CORINE code 411
Edge	Ed	Sqrt	length of the transition zone between Gr and Hb on the one hand and Dw and Mw on the other
Biotope Diversity Index	BDI	no	Shannon index of terrestrial biotopes
Climate data			
Frost	F	Log	Interpolated point data of the RMIB
Rain	R	Log	Interpolated point data of the RMIB
Sun	S	Log	Interpolated point data of the RMIB
Temperature	T	Sqrt	Interpolated point data of the RMIB
Topographic data			
Maximum elevation	El	Sqrt	Digital elevation model Belgium
X coordinate	X	no	Lambert Belgium 1972 projection (increases from east to west)
Y coordinate	Y	no	Lambert Belgium 1972 projection (decreases from north to south)

RMIB, Royal Meteorological Institute of Belgium; no, no transformation; Sqrt, square root transformation; Log, Log_{10}-transformation.

with a linear drift was used to interpolate the point data of yearly cumulated rainfall (mm; $n = 186$; mean minimum distance between data points 8 km), yearly average maximum temperature (°C; $n = 114$) and yearly cumulated number of frost days ($n = 114$; mean minimum distance between data points 11 km); 'inverse weighted distance' was used to interpolate the point data of yearly cumulated sunshine exposure (hours, $n = 22$, mean minimum distance between data points 23.5 km).

Analyses

The different predictive models are based on the best-surveyed squares in Belgium during the period 1991–2002. As mapping intensity differed considerably between Wallonia and Flanders (mean number of visits per square are 10.6 and 66.7, respectively), we applied different selection criteria to determine sufficient recording effort in both regions. To build the predictive models, we selected the best-surveyed squares (75 squares in Wallonia and 122 squares in Flanders) based on the number of visits. As most of the predictor variables are not normally distributed, we transformed the variables using either log_{10} or square root functions (resulting in the lowest skewness value) prior to analyses (Table 1). The analysis involved four steps: (1) exploratory analysis of the spatial structure of butterfly diversity, (2) exploratory analysis of the relationships between the predictors in relation to butterfly diversity, (3) the design of three linear models relating butterfly diversity to environmental covariates and (4) application of the linear models to predict butterfly diversity patterns and hotspots in Belgium.

Correlations and linear models statistics are affected by spatial autocorrelation in the response (butterfly diversity) and environmental variables, i.e. the tendency for the value of neighbouring points to be more similar than distant points. In general terms, spatial autocorrelation is important in spatial data analysis for the insight it provides in the data under study (Rossi *et al.*, 1992). It contradicts the assumption of independence among samples replicated through space (Clifford *et al.*, 1989; Lennon, 2000). Because of this, modelling the structure of the spatial autocorrelation allows spatial interpolation by the method known as *kriging* (Isaaks & Srivastava, 1989).

We have used the experimental co-variogram to quantify spatial autocorrelation in butterfly diversity, which is a function that estimates the level of covariance for points separated by increasingly greater intervals of distance (Rossi *et al.*, 1992). Typically, it is a rising curve (points close by have fairly similar values and a low covariance estimate) that levels off at a given distance known as the 'range' (distance over which sample points are independent), while the height is known as the 'sill'. Points separated by a null distance have

a covariance equal to zero, so the curve should start at the origin of the two axes. This is rarely the case with ecological data, and the value at which the experimental semi-variogram intercepts the *Y*-axis is termed the 'nugget' and represents experimental error or variability at a smaller scale than the smallest distance interval. In the presence of a large-scale trend, the co-variogram is biased and tends to increase above the limit value of 1. In such a case, the large-scale trend is modelled by a linear or quadratic function of spatial coordinates, and the co-variogram is estimated on the basis of the large-scale trend model residuals. This was the case with butterfly diversity and the co-variogram was estimated using the residuals of a large-scale linear trend model. The co-variogram was modelled by a spherical model using a combination of 'fit-by-eye' and least squares approach, and values for the scale, nugget, range and R^2 of the fit were obtained. Spatial statistics were carried out using the software Surfer 8.0 (Golden software, Inc., Golden, Colorado, USA).

The second step was to explore the relationship between environmental variables and butterfly diversity using two approaches. First, the correlation between butterfly diversity and the land use, topographic and climatic factors was estimated. Unbiased correlation levels of significance were obtained using the method proposed by Clifford *et al.* (1989) modified by Dutilleul (1993) that quantify the reduction in degrees of freedom according to spatial autocorrelation observed in the two variables. Secondly, a principal components analysis was carried out using the set of environmental variables (Table 1); four variables were entered as supplementary to the analysis [butterfly diversity and the number of visits (both $\log_{10}$ values), *X* and *Y* coordinates distinguishing eastings and northings of grid squares].

The third step was to build linear models relating butterfly diversity to environmental predictors such as frequently applied in similar research (McCullagh & Nelder, 1989; Nicholls, 1989; Lobo & Martín-Piera, 2002; Luoto *et al.*, 2002). Linear models generally assume a constant variance among observations [$\mathrm{Var}(e_i) = \sigma^2$], and a covariance among observations equal to zero [$\mathrm{Cov}(e_i, e_j) = 0$], which is clearly violated in the presence of spatial autocorrelation. For each linear model, the co-variogram of the residuals was estimated to check if it exhibited evidence of spatial autocorrelation. In such a case the covariance among residuals due to spatial autocorrelation was modelled using the SAS MIXED procedure by the function (Littell *et al.*, 1996)

$$\mathrm{Cov}(e_i, e_j) = \sigma^2 f(d_{ij})$$

where e_i is the error corresponding to the *i*th observation, d_{ij} is the distance between the spatial location of the *i*th and *j*th residual and *f* is the spatial covariance function. The spatial covariance function was estimated by modelling the experimental co-variogram of the multiple regression residuals using the spherical model (Isaaks & Srivastava, 1989). The spatial covariance model parameters were identified using a combination of 'fit-by-eye' and least squares approaches, selecting the model providing the best fit. The presence of curvilinear relationships between each environmental variable and butterfly diversity was assessed by incorporating the quadratic terms of the environmental variables (Nicholls, 1989) and the best function (linear or quadratic) was retained using 'Akaike's Information Criterion'; this criterion compromises between model fit (the ability to explain the observed variation on the dependent variable) and model complexity (the number of parameters to estimate; Akaike, 1978). The (linear or quadratic) function of the environmental variable that accounted for the largest reduction in deviance (*F*-ratio test with *P* level of 0.05) was first incorporated into the model (Nicholls, 1989; Crawley, 1993). Next, all the remaining environmental variables were tested in the same way until inclusion was no longer significant. At each step, all previously entered variables were tested for their significance and removed from the model if they were no longer significant. Three multiple regression models were built by splitting the set of 197 squares into two, randomly selected, subsets (cf. Pearson & Carroll, 1999; Luoto *et al.*, 2002): a 'training set' of 98 squares and an 'evaluation set' of 99 squares (Fig. 1) that was used to test the models performances. In the first model, termed 'statistically focused model', environmental variables were entered in descending order of statistical significance. The second model, termed 'land use-focused model', was based on *a priori* knowledge of the relationship between butterfly species richness and the following land use variables (Mac Nally, 2000, 2002):

- biogeographical region: butterfly diversity is markedly higher in the continental region than in the Atlantic region (Goffart & De Bast, 2000);
- BDI: butterfly diversity increases with increasing biotope diversity (e.g. Sparks *et al.*, 1995; Weibull *et al.*, 2000; Kerr *et al.*, 2001; Kerr, 2001; Hawkins & Porter, 2003) and the length of the edges between mixed and deciduous woodlands on the one hand and grasslands, heathlands and bogs on the other;
- deciduous and mixed woodland, natural grasslands, heathlands and bogs: all these biotope types are inhabited by typical species giving rise to a higher butterfly diversity (van Swaay & Warren, 1999);
- urban and agricultural area: both types of land use have a negative impact on butterfly diversity (Blair & Launer, 1997; Hardy & Dennis, 1999; Dennis & Hardy, 2001).

The choice of this modelling approach stemmed from the assumption that factors chosen according to established relationships that include causal pathways produce more robust models. In the third model, termed 'hybrid model', the same variables as in the land use-focused model were forced in the model in a first step, and additional environmental variables were subsequently added in a stepwise selection procedure such as described in the statistically focused model (cf. Luoto *et al.*, 2002).

The last step was to apply these three models to the whole Belgian territory in order to predict the spatial distribution of butterfly diversity and to determine butterfly diversity hotspots in Belgium (i.e. the 5% most diverse squares;

Prendergast *et al.*, 1993a). The spatial distribution of the predicted hotspots was compared with the observed hotspots and field knowledge to determine the most adequate best model for nature conservancy policies in Belgium.

RESULTS

Factors determining butterfly diversity

Relationships among variables determining butterfly diversity in Belgium were analysed using PCA (Fig. 2). Five components have eigenvalues greater than 1, each accounting for more than 5% of the variance. Cumulatively these five axes account for more than 72% of the variance in the predictor variable set. The first two axes account for 39% and 15% of the variance, respectively, and are used to illustrate relationships between the variables.

Altogether twelve and four variables respectively load modestly ($\alpha > 0.50$) on axes 1 and 2 (Fig. 2). Variables are polarized on both axes. Axis 1 distinguishes eight variables with positive sign ($\alpha > 0.50$, deciduous, mixed and coniferous woodland, shrub, BDI, number of frost days, rainfall and elevation) from four with negative signs ($\alpha > 0.50$, biogeographical region, urban area, sunshine and temperature). Axis 2 distinguishes the area of arable land (positive) from most of the other land use variables (heathlands and bogs, shrubs, BDI, edges, water, marshes, all negative). Only four variables have their highest and meaningful loadings on additional axes (natural grasslands, salt marshes, dunes and edges).

Communalities on the first five axes indicate the existence of a substantial amount of unique variance in variables; only one variable (elevation) has communalities higher than 90% on the first five axes and only five variables have substantial variance accounted for (>70%) on the first two axes (biogeographical region, number of frost days, rainfall, temperature and elevation). Butterfly diversity and the number of visits have 53% and 38% of their variances accounted for on the first five axes. Thus, on this variable set, component scores do not provide an adequate substitute for the original variables in determining variance in butterfly diversity. The plot of variables in the first two axes (Fig. 2) indicates a strong geographical patterning to the environmental variables, with *X* and *Y* coordinates polarized on axis 1. Because of increasing altitude in southern and eastern Belgium the usual environmental trends in northern latitudes is reversed (Dennis & Williams, 1986; Dennis, 1993; Kerr *et al.*, 1998; Kerr, 2001). Conditions become warmer and sunnier to the north and colder and wetter to the south-east. Many natural biotopes also increase to the south-east, especially woodland biotopes. Butterfly diversity and the number of visits are also polarized. Butterfly diversity increases to the south-east on higher ground and decreases to the north, whereas visits decrease eastwards and increase northwards.

A strong spatial autocorrelation in butterfly diversity was identified at the scale of this study (Fig. 3a). A range distance of 96 km was observed which means that only observations separated by distances >96 km can be considered statistically independent. This high level of autocorrelation had a

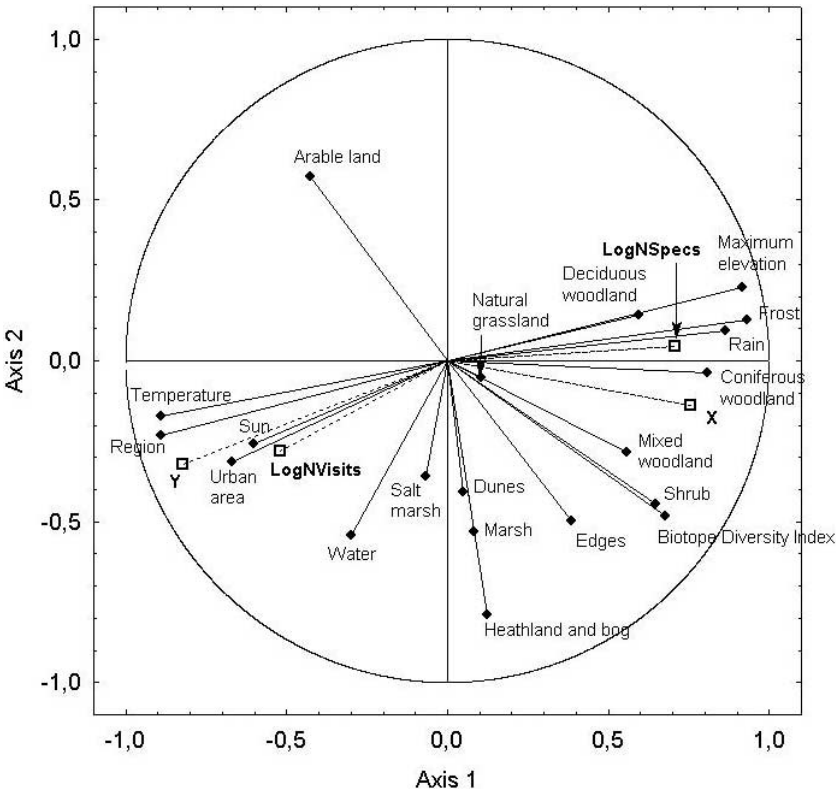

Figure 2 Principal components analysis of the environmental variables in the well-surveyed squares ($n = 197$) in Belgium. Axes 1 and 2 explain 39% and 15% of the variation in the data, respectively. The number of visits (LogNVisits), the number of species (LogNSpecs) (both $\log_{10}$-transformed) and the *X* and *Y* coordinates are entered as supplementary variables.

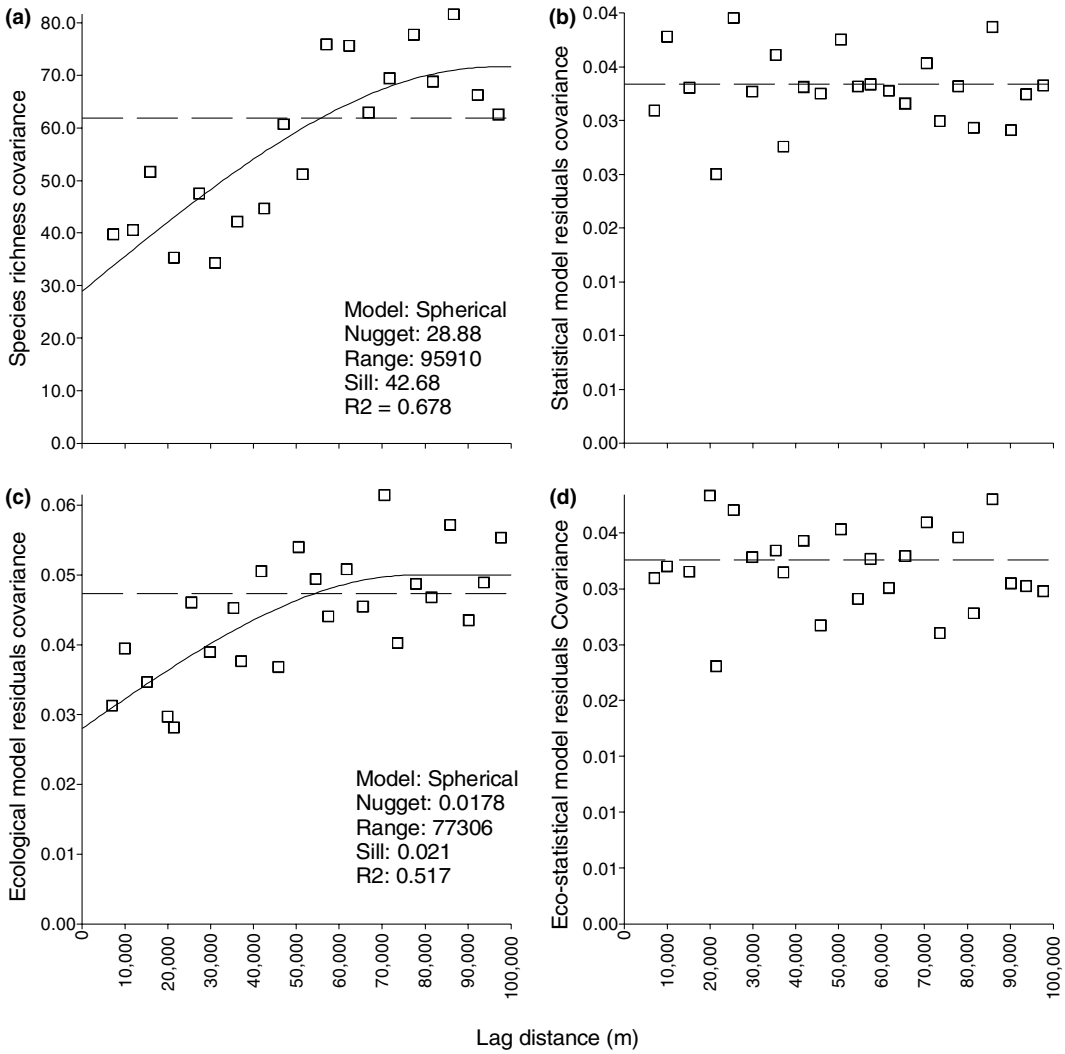

Figure 3 Co-variograms of (a) butterfly diversity showing a high degree of spatial autocorrelation up to 95.9 km; (b) the statistically focused model residuals showing no spatial autocorrelation; (c) the land use-focused model residuals showing spatial autocorrelation up to 77.3 km; and (d) the hybrid model residuals showing no spatial autocorrelation.

strong impact on the level of significance of the correlations between butterfly diversity and environmental factors (Table 2). This test should be interpreted cautiously, as it does not necessarily imply that these correlations are spurious, but only that they could not be statistically proven within the scale of Belgium. Similarly, butterfly diversity in the well-surveyed squares was not significantly correlated with environmental variables when corrected for spatial autocorrelation; only deciduous and coniferous woodland showed a positive trend with butterfly diversity. Mutually significant correlations are mainly found between land use variables, but not between climatic and topography variables. Arable land is negatively correlated with most other biotope types, while the BDI is positively correlated with most land use types, except for arable land (Table 2).

Modelling butterfly species richness

When entered separately, 'biogeographical region' accounted for the greatest change in partial R^2 and was entered first into the statistically focused model. Variable selection ended after the inclusion of deciduous woodland in step 6. Estimates and standard errors of the parameters for the statistically focused model are given in Table 3a. The statistically focused model explained 77.2% of the variability in the training set (Table 3a) and 66.3% of the variability in the evaluation set (Spearman r correlation between observed and expected butterfly diversity in evaluation set = 0.80; $P < 0.001$). The residuals of the statistically focused model showed no evidence of spatial autocorrelation (Fig. 3b) indicating that most of the spatial structure in butterfly diversity (Fig. 3a) was

Table 2 Pearson *r* correlation (lower left part of the table) between the 21 variables and *P*-value (upper right part of the table) for the well-surveyed grid cells in Belgium ($n = 197$). *P*-values are based on the modified *t*-test of Dutilleul (1993). Significant correlations ($P < 0.05$) are in bold. Freq. = number of squares in which the biotope type occurs. Abbreviations of the environmental variables are given in Table 1

	Freq.	Ur	Ar	Dw	Mw	Cw	Gr	Hb	Sh	Sm	Du	Wa	Ma	Ed	BDI	F	R	S	T	El	X	Y	SR
Ur	197	–	0.258	0.109	0.107	0.189	0.288	0.409	0.143	0.304	0.672	0.052	0.720	0.247	0.286	0.136	0.124	0.274	0.129	0.172	0.449	0.146	0.129
Ar	197	−0.175	–	0.057	**0.005**	**0.008**	0.088	**<0.001**	**0.001**	0.116	**0.013**	0.083	0.051	**<0.001**	**0.002**	0.222	0.157	0.359	0.186	0.267	0.101	0.454	0.137
Dw	141	−0.470	−0.312	–	0.570	0.104	0.279	0.160	0.309	0.210	0.904	0.449	0.765	0.779	0.064	0.207	0.207	0.445	0.298	0.213	0.364	0.132	0.067
Mw	112	−0.345	**−0.377**	0.116	–	**0.029**	0.934	0.158	**0.005**	0.351	**0.010**	0.278	0.154	0.061	**<0.001**	0.111	0.080	0.464	0.102	0.283	0.114	0.588	0.361
Cw	142	−0.453	**−0.464**	0.507	**0.481**	–	0.299	0.818	0.080	0.174	0.752	0.102	0.487	0.057	**0.010**	0.126	0.181	0.168	0.158	0.172	0.204	0.257	0.088
Gr	4	0.098	−0.134	0.103	−0.007	0.096	–	0.684	0.300	0.888	0.829	0.837	0.507	0.189	0.110	0.922	0.874	0.509	0.574	0.245	0.402	0.444	0.212
Hb	44	0.100	**−0.400**	−0.163	0.149	0.028	−0.030	–	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.003**	**<0.001**	**0.003**	0.798	0.664	0.923	0.769	0.923	0.196	0.457	0.986
Sh	65	−0.347	**−0.480**	0.243	**0.496**	0.461	0.087	**0.350**	–	0.087	**0.005**	0.910	**0.004**	**0.005**	**0.001**	0.169	0.096	0.269	0.085	0.180	0.123	0.350	0.272
Sm	1	0.072	−0.112	−0.089	−0.066	−0.093	−0.010	**0.269**	0.120	–	0.912	**<0.001**	**0.003**	0.770	0.621	0.358	0.580	0.156	0.377	0.160	0.693	0.222	0.600
Du	3	−0.038	**−0.178**	0.010	**0.198**	0.028	−0.016	**0.286**	**0.215**	−0.008	–	0.466	0.605	**0.011**	**0.008**	0.987	0.716	0.204	0.324	0.509	0.489	0.241	0.921
Wa	57	0.344	−0.170	−0.122	−0.132	−0.285	0.017	**0.383**	0.015	**0.282**	0.056	–	**0.007**	0.932	0.925	0.223	0.162	0.136	0.089	0.145	0.821	0.143	0.369
Ma	25	0.038	−0.166	0.030	0.146	0.073	−0.049	**0.252**	**0.268**	**0.208**	0.038	**0.227**	–	**0.037**	**<0.001**	0.773	0.713	0.097	0.703	0.770	0.284	0.778	0.560
Ed	21	−0.016	**−0.357**	0.039	0.211	0.287	0.099	**0.677**	**0.346**	−0.021	**0.183**	−0.008	**0.163**	–	**0.009**	0.199	0.109	0.238	0.067	0.159	0.106	0.385	0.202
BDI	197	−0.280	**−0.479**	0.457	**0.617**	**0.673**	0.136	**0.328**	**0.643**	0.035	**0.214**	0.013	0.418	**0.340**	–	0.147	0.163	0.439	0.204	0.259	0.123	0.385	0.166
F	197	−0.664	−0.289	0.545	0.475	0.729	0.010	0.038	0.493	−0.060	0.002	−0.267	0.036	0.255	0.551	–	0.124	0.184	0.136	0.183	0.269	0.224	0.149
R	197	−0.598	−0.296	0.477	0.465	0.572	0.016	0.060	0.508	−0.038	−0.034	−0.277	0.045	0.277	0.468	0.842	–	0.258	0.057	0.182	0.337	0.165	0.225
S	197	0.331	0.155	−0.220	−0.164	−0.435	−0.057	0.012	−0.269	0.098	0.105	0.222	0.178	−0.168	−0.205	−0.575	−0.435	–	0.090	0.092	0.235	0.215	0.338
T	197	0.614	0.269	−0.398	−0.430	−0.612	−0.056	−0.039	−0.520	0.060	0.093	0.345	0.042	−0.312	−0.426	−0.834	−0.845	0.608	–	0.100	0.323	0.182	0.274
El	197	−0.638	−0.259	0.548	0.328	0.685	0.120	−0.014	0.481	−0.093	−0.066	−0.330	−0.034	0.276	0.444	0.901	0.783	−0.685	−0.885	–	0.337	0.170	0.231
X	197	−0.344	−0.395	0.398	0.476	0.614	0.078	0.194	0.543	−0.026	0.063	−0.046	0.137	0.319	0.580	0.795	0.598	−0.520	−0.604	0.722	–	0.473	0.359
Y	197	0.652	0.176	−0.636	−0.167	−0.580	−0.082	0.109	−0.342	0.082	0.113	0.326	0.034	−0.171	−0.343	−0.852	−0.792	0.533	0.780	−0.908	−0.568	–	0.167
SR	197	−0.540	−0.264	0.587	0.220	0.616	0.125	0.002	0.299	−0.037	0.009	−0.171	0.069	0.196	0.404	0.716	0.550	−0.323	−0.506	0.636	0.477	−0.696	–

Table 3 Estimates obtained by the three multiple regression models for butterfly diversity in Belgium: (a) statistically focused model (R^2 for the training set = 0.772; n = 98; P < 0.001), (b) land use-focused model (R^2 for the training set = 0.676; n = 98; P < 0.001) and (c) hybrid model (R^2 for the training set = 0.777; n = 98; P < 0.001)

	Estimate	SE	d.f.	F value	P
(a) Statistically focused model					
Region	−0.7044	0.1168	91	36.40	<0.001
Cw	0.1337	0.0422	91	10.05	0.002
El^2	−0.0023	0.0004	91	36.03	<0.001
F^2	6.8930	2.5816	91	7.13	0.009
F	−22.4377	9.1307	91	6.04	0.016
Dw	0.0765	0.0383	91	4.00	0.049
Intercept	21.9321				
(b) Land use-focused model					
Region	−0.4987	0.1430	88	12.16	<0.001
Dw	−0.0062	0.0619	88	0.01	0.920
Mw	−0.0548	0.0583	88	0.88	0.350
Gr	0.0273	0.3975	88	<0.01	0.945
Hb	−0.0526	0.0952	88	0.31	0.582
Ed	0.0002	0.0020	88	0.01	0.916
Ur	−0.0269	0.0147	88	3.34	0.071
Ar	−0.0011	0.0007	88	2.47	0.119
BDI	0.0576	0.0857	88	0.45	0.503
Intercept	3.8925				
(c) Hybrid model					
Region	−0.7794	0.1344	85	33.64	<0.001
Dw	0.0265	0.0540	85	0.24	0.625
Mw	−0.0091	0.0515	85	0.03	0.860
Gr	4.9373	2.2077	85	5.00	0.028
Hb	−0.0891	0.0921	85	0.94	0.336
Ed	0.0024	0.0020	85	1.50	0.224
Ur	−0.0278	0.0132	85	4.42	0.039
Ar	−0.0012	0.0007	85	3.50	0.065
BDI	0.0512	0.0803	85	0.41	0.526
El^2	−0.0021	0.0004	85	407.06	<0.001
F^2	0.4999	0.1625	85	162.23	0.003
Gr^2	−11.0284	5.0436	85	144.77	0.032
Intercept	2.8360				

accounted for by the environmental factors. It also means that the statistically focused model did not need to be modified to account for spatial autocorrelation in the residuals.

The land use-focused model explained 67.6% of the variability in the training set (Table 3b) and 56.7% of the variability in the evaluation set (Spearman r correlation between observed and expected butterfly diversity in evaluation set = 0.66; P < 0.001). Residuals of this model showed a high degree of spatial autocorrelation (Fig. 3c), and the linear model therefore included a model of covariance among residuals to obtain unbiased factor estimates and levels of significance.

Finally, the hybrid model explained 77.7% of the variability in the training set (Table 3c) and 67.4% of the variability in the evaluation set (Spearman r correlation between observed and expected butterfly diversity in evaluation set = 0.80; P < 0.001). Here, residuals showed no evidence of spatial autocorrelation (Fig. 3d) indicating that most of the spatial structure in butterfly diversity was accounted for by the complementary environmental factors (the squared terms of elevation, number of frost days and natural grassland area), and that no adjustment was required to account for spatial autocorrelation.

Plotting residuals against observed butterfly diversity revealed that the three models all overestimated the number of species for squares with a low observed butterfly diversity and underestimated diversity in species-rich squares (Pearson r between observed species number and the residuals in the training set for the statistically focused model = 0.47, P < 0.001; for the land use-focused model = 0.56, P < 0.001; for the hybrid model = 0.45, P < 0.001).

Butterfly diversity and diversity hotspots

Observed butterfly diversity and hotspots

In South Belgium, the Fagne-Famenne-Calestienne and the Lorraine region have the highest butterfly diversity, while the Campine region is the most species-rich region in North Belgium (although absolute numbers are lower in the north; Fig. 4a). Observed butterfly diversity hotspots (i.e. the top 5% of the 1108 analysed squares, n = 57 with 35 species) were all situated in the continental region of Belgium (Fig. 5), particularly in the regions Fagne-Famenne-Calestienne (n = 23), Lorraine (n = 18), Ardennes (n = 10), Condroz (n = 5) and Thiérache (n = 1).

Statistically focused predicted butterfly diversity and hotspots

The predicted butterfly diversity ranged from 12 to 60 species in the statistically focused model. Extrapolating the model to the whole of Belgium predicted high butterfly diversity in the Fagne-Famenne-Calestienne, the region around the river Meuse in the Condroz region, in the Lorraine regions and, to a lesser degree, in the Campine region in the north (Fig. 4b). Statistically focused predicted butterfly diversity hotspots (squares were the model predicted at least 42.4 species) were all situated in the continental region of Belgium (Fig. 5), particularly in the regions Fagne-Famenne-Calestienne (n = 25), Ardennes (n = 13), Lorraine (n = 10) and Condroz (n = 9).

Land use-focused predicted butterfly diversity and hotspots

The predicted butterfly diversity ranged from 18 to 48 species in the land use-focused model. Extrapolating the model to the whole of Belgium predicted high butterfly diversity in the Ardennes, the Lorraine, the Thiérache region and, to a lesser degree, in the Campine region in the north (Fig. 4c). Land use-focused predicted butterfly diversity hotspots (i.e. squares where the model predicted at least 41.6 species) were all situated in the continental region of Belgium (Fig. 5), particularly in the Ardennes (n = 39), Fagne-Famenne-Calestienne (n = 7), Lorraine (n = 7) and Thiérache (n = 4).

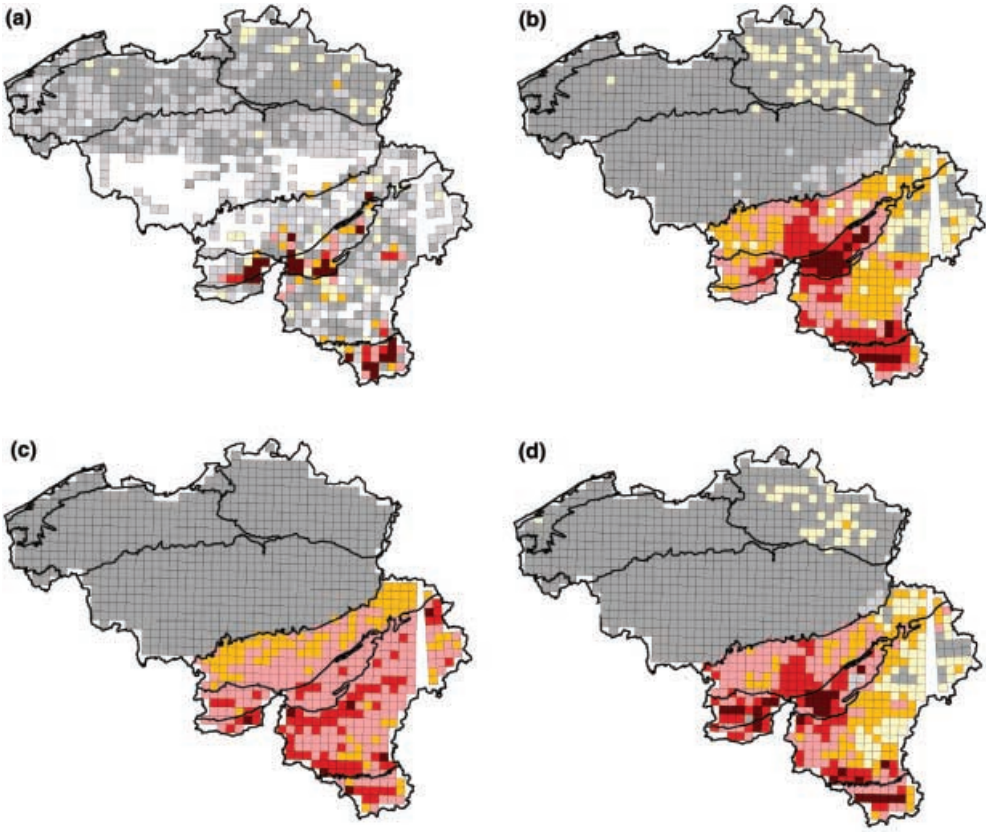

Figure 4 Observed (a, unvisited squares are not shown) and predicted butterfly diversity in Belgium using the statistically focused model (b), the land use-focused model (c) and the hybrid model (d). Light grey = 1–15 species; dark grey = 15–25 species; yellow = 25–30 species; orange = 30–35 species; rose = 35–40 species; red = 40–45 species and brown = 45 or more species. The blank wedge in the east of Belgium is the correction zone of the UTM projection; the squares in this correction zone are not included in the analyses because their area is much smaller than that of the other squares.

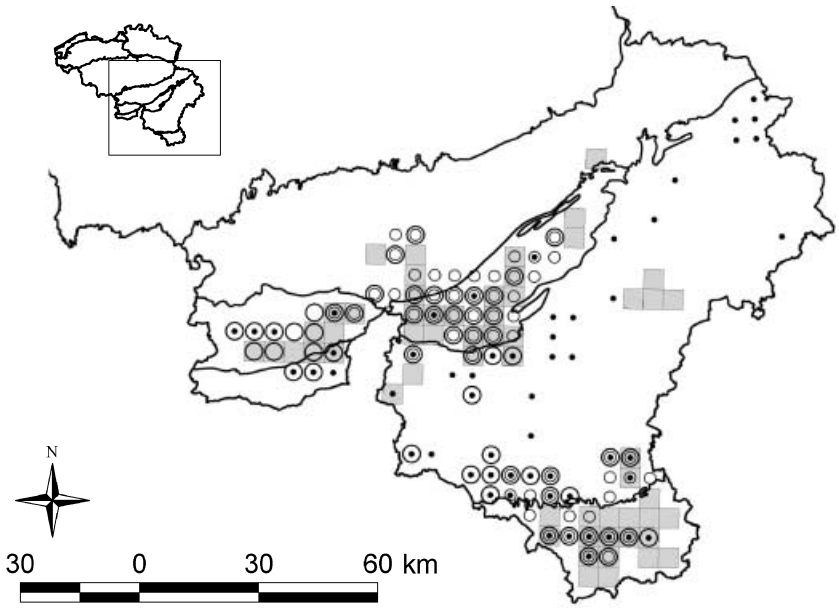

Figure 5 Observed (grey squares) and predicted hot spots in the Continental region of Belgium using the hybrid model (large circles), the statistically focused (intermediate-sized circles) and the land use-focused model (small black dots).

Hybrid model predicted butterfly diversity and hotspots
The predicted butterfly diversity ranged from 0 to 63 species in the hybrid model. Extrapolating the hybrid model to the whole of Belgium predicted a high butterfly diversity in the Fagne-Famenne-Calestienne, Lorraine, the central part of the Condroz region, in the south of the Ardenne region and, to a lesser degree, in the Campine region in the north (Fig. 4d). Predicted butterfly diversity hotspots using the hybrid model (i.e. squares where the model predicted at least 42.6 species) were all situated in the continental region of Belgium (Fig. 5), particularly in Fagne-Famenne-Calestienne ($n = 25$), the Ardennes ($n = 17$), Lorraine ($n = 8$), Condroz ($n = 4$) and Thiérache ($n = 3$).

DISCUSSION

Factors explaining butterfly diversity

The dominant environmental factors explaining butterfly diversity in Belgium are topography and climate as is often the case in predictive models for butterfly diversity at a large spatial scale (e.g. Sparks *et al.*, 1995; Kerr *et al.*, 1998; Fleishman *et al.*, 2001a). Although some of the land use variables (different woodland types, shrub, biotope diversity and urban area) have modest to high loadings on the first PCA-axis, they contribute less to the explanation of butterfly diversity in Belgium.

The strong correlation of butterfly diversity with coniferous woodland is unexpected, because this biotope type is not known for its butterfly diversity (van Swaay & Warren, 1999) and has an entirely artificial origin in Belgium; an explanation for this correlation can be that conifers were planted in formerly butterfly-rich sites (e.g. heathlands, dry calcareous and wet grasslands, moors; Goffart *et al.*, 2000) that have now become too small to be distinguished by CORINE land cover maps.

Modelling butterfly species richness

Mapping the observed butterfly diversity in Belgium (Fig. 4a) indicates that not all regions have been surveyed. Figure 4a also shows that North Belgium has been surveyed more completely than South Belgium where a large part of the Loam region and, to a lesser degree, the western part of the Condroz region, are completely un-surveyed. For modelling butterfly species richness in Belgium, we selected the best-surveyed squares in the Flemish and in the Walloon region separately. This implied different thresholds for the number of visits in both regions. The application of a single, common threshold for the whole Belgian territory would have strongly over-represented the Flemish region, which was not appropriate when modelling species diversity for the whole Belgian territory. Furthermore, the predicted number of species per square in both approaches (different thresholds vs. one single threshold) was very strongly correlated reassuring us that the difference in the selection criteria between Flanders and Wallonia did not affect the outcome of the spatial patterns.

The squares to develop and evaluate the model are well spread over the different ecological regions in Belgium (Fig. 1); only three ecoregions are not represented in the training set (Dunes and Meuse are too narrow to have squares that fall completely within Belgium and Thiérache only covers eleven complete squares). The training set of 98 well-surveyed squares differs significantly from the other squares in a number of variables and is not a random sample of squares in Belgium ($n = 1108$; MANOVA, $F = 3.01$; $P < 0.001$). The training set has a smaller area of arable land ($F = 5.01$; $P = 0.025$), but has higher values for salt marshes ($F = 4.86$; $P = 0.028$), marshes ($F = 13.22$; $P < 0.001$) and BDI ($F = 7.01$; $P = 0.008$). The training set was selected on the basis of the number of visits and recorders visited these squares more often because they are known to have a high butterfly diversity. This is related to limited areas of arable land and a high biotope diversity (Dennis & Thomas, 2000). Splitting the data set into a training set and an evaluation set is often applied in modelling research, because it provides a more robust estimate of the model appropriateness (e.g. Pearson & Carroll, 1999; Luoto *et al.*, 2002). However, when both sets come from the same larger data set, the evaluation set cannot be considered truly independent of the training set (Guisan & Zimmermann, 2000), mainly because of spatial autocorrelation (points of the evaluation set are very close to those from the training set). This was confirmed by the fact that the training and evaluation set did not differ significantly in the environmental variables used or in the number of species or visits (MANOVA: $F = 0.35$; $P = 0.99$). In a small country like Belgium other types of data set subdivisions would be difficult to achieve. The only way to have independent training and evaluation sets would be to find sampling locations separated by >96 km, which is only possible along a NW–SE axis (choosing the southern sample points from Wallonia on the one hand and the northern sample points from Flanders on the other). Such a subdivision is not appropriate because it coincides with two different biogeographical regions with different ecological relationships.

All three models explain high percentages of the variability in the evaluation set compared with similar studies (e.g. Lobo & Martín-Piera, 2002; Luoto *et al.*, 2002). The hybrid model explains variability in the evaluation set only slightly better than the statistically focused model but uses more variables resulting in a higher Akaike's Information Criterion value (565.06 vs. 557.84). In particular, 'biogeographical region' is an important factor explaining butterfly diversity in Belgium (cf. Bio *et al.*, 2002). All three models showed a significant trend when observed diversity was plotted against model residuals. This can be caused by the fact that the model was unable to fully fit the complex interaction between butterfly diversity and environmental variables or by the absence of other predictor variables. Candidates for such missing variables are: (1) interaction terms between variables, (2) biotope quality (in the present analyses only biotope quantity is entered) or (3) higher order terms of the environmental variables (Legendre & Legendre, 1998). Inclusion of higher order terms or interaction terms increases the models complexity

and makes model interpretations difficult and/or spurious (Bio *et al.*, 2002). Evapotranspiration, one of the frequently used variables in species richness analyses (e.g. Hawkins & Porter, 2003), was not incorporated in our analyses, because variation in evapotranspiration is very limited within the extent of Belgium (220 × 270 km; Bultot & Dupriez, 1974).

Application of predictive modelling for butterfly conservation in Belgium

Although butterfly distribution is relatively well known in Belgium (Goffart & De Bast, 2000; Maes & Van Dyck, 2001), predictive modelling can considerably increase the efficiency of butterfly mapping schemes and incorporate un-surveyed regions into nature conservancy policy making. But, the choice of the most accurate predictive model is of major importance when financial and personal resources are limited. Although land use-focused models are usually preferred for their interpretability and logical link with the studied organisms (Lennon, 2000; Mac Nally, 2000, 2002; Luoto *et al.*, 2002), the hybrid and the statistically focused models predicted butterfly diversity better than the land use-focused one in Belgium. As regression models aim to minimize residual variance and many of the land use variables covary simultaneously with climatic or topographic variables (Table 2), the latter are often better predictors for butterfly diversity than the different land use variables separately (Dennis & Williams, 1986). Both the statistically focused and the hybrid model explained about 10% more of the variability in the evaluation set than the land use-focused model and are therefore preferred to the latter when used for nature conservation purposes in Belgium.

Hotspots predicted by the statistically focused, the land use-focused and the hybrid model represent 78, 75 and 77 of 88 present-day indigenous butterfly species, respectively, while observed hotspots cover 82 of the present-day indigenous species. Six of the present-day indigenous species are not present in the predicted or observed hotspots: *Heteropterus morpheus*, *Maculinea alcon*, *M. arion*, *Coenonympha glycerion*, *C. hero* and *Erebia ligea*. Five of these species are extremely rare and occupy only one (*H. morpheus*, *M. arion* and *C. hero*) or three squares (*C. glycerion* and *E. ligea*) in Belgium, mostly situated along the borders of the country; *M. alcon* is limited to North Belgium (Maes & Van Dyck, 1999).

Some authors have predicted species diversity on very large scales using large grid cells (e.g. Kerr *et al.*, 1998; Kerr, 2001; Lobo & Martín-Piera, 2002). Such analyses can provide insight to large-scale differences in species diversity but have practical limitations for nature conservation. Predicting species richness on scales at which species interact with their environment and that are closer to biotope size is therefore more useful in species conservation (Prendergast *et al.*, 1993b; Dennis & Hardy, 1999; Pearson & Carroll, 1999). This is more likely to be the case using relatively small grid cells (25 km^2 in our case); further analyses are to be undertaken to determine whether it is possible to extend the present analyses to even smaller grid cells (e.g. 1 km^2).

Predictive modelling is a very useful tool in mapping individual species or species diversity distributions over large and unequally surveyed areas. Land use data, derived from CORINE land cover maps (restricted to Europe) or from satellite images, and other environmental data (e.g. climate, topography) are nowadays readily available for many countries and on different scales. A relatively small set of well-surveyed squares could suffice to apply predictive modelling in under-surveyed regions or countries. For example, applying predictive modelling to the recently published distribution atlas of European butterflies (Kudrna, 2002) to indicate potential species distributions, could greatly extend its applicability for European wide nature conservation purposes. We believe that future atlases of butterflies and other organisms, should make more use of predictive modelling to produce predicted distribution maps as a more pro-active conservation tool; predictive models should, off course, always be validated and based on similar well-surveyed regions to produce valuable models that meet minimum standards.

ACKNOWLEDGMENTS

Marius Gilbert is a post-doctoral fellow with the National Fund of Scientific Research – Belgium (FNRS). We kindly thank Violaine Fichefet and Yvan Barbier for compiling the Walloon butterfly distribution data. We also thank all the recorders that have participated in the mapping schemes. We are grateful to Piet De Becker, Dirk Bauwens, Luc De Bruyn and Ralf Gyselings for useful discussions on statistical analyses. We thank the Royal Meteorological Institute of Belgium for making the climate data available. Comments of two anonymous referees improved the manuscript considerably.

REFERENCES

Akaike, H. (1978) A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics*, **30**, 9–14.

Asher, J., Warren, M., Fox, R., Harding, P., Jeffcoate, G. & Jeffcoate, S. (2001) *The millenium atlas of butterflies in Britain and Ireland*. Oxford University Press, Oxford.

Bio, A.M.F., De Becker, P., De Bie, E., Huybrechts, W. & Wassen, M. (2002) Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. *Biodiversity and Conservation*, **11**, 2189–2216.

Blair, R.B. & Launer, A.E. (1997) Butterfly diversity and human land use: species assemblages along an urban gradient. *Biological Conservation*, **80**, 113–125.

Bultot, F. & Dupriez, G.L. (1974) L'évapotranspiration potentielle des bassins hydrographiques en Belgique. *Royal Meteorological Institute of Belgium Publications, Serie A*, **85**, 1–61.

CEC (1994) *CORINE land cover technical guide*. European Commission, Luxemburg.

Clifford, P., Richardson, S. & Hémon, D. (1989) Assessing the significance of the correlation between two spatial processes. *Biometrics*, **45**, 123–134.

Crawley, M.J. (1993) *GLIM for ecologists*. Blackwell Scientific Publications, Oxford.

Dennis, R.L.H. (1993) *Butterflies and climate change*. Manchester University Press, Manchester.

Dennis, R.L.H. & Hardy, P.B. (1999) Targeting squares for survey: predicting species richness and incidence of species for a butterfly atlas. *Global Ecology and Biogeography Letters*, **8**, 443–454.

Dennis, R.L.H. & Hardy, P.B. (2001) Loss rates of butterfly species with urban development. A test of atlas data and sampling artefacts at a fine scale. *Biodiversity and Conservation*, **10**, 1831–1837.

Dennis, R.L.H. & Shreeve, T.G. (2003) Gains and losses of French butterflies: tests of predictions, under-recording and regional extinction from data in a new atlas. *Biological Conservation*, **110**, 131–139.

Dennis, R.L.H. & Thomas, C.D. (2000) Bias in butterfly distributions maps: the influence of hot spots and recorder's home range. *Journal of Insect Conservation*, **4**, 73–77.

Dennis, R.L.H. & Williams, W.R. (1986) Butterfly 'diversity': regressing and a little latitude. *Antenna*, **10**, 108–112.

Dennis, R.L.H., Sparks, T.H. & Hardy, P.B. (1999) Bias in butterfly distribution maps: the effects of sampling effort. *Journal of Insect Conservation*, **3**, 33–42.

Dennis, R.L.H., Shreeve, T.G., Olivier, A. & Coutsis, J.G. (2000) Contemporary geography dominates butterfly diversity gradients within the *Aegean archipelago* (Lepidoptera: Papilionoidea, Hesperioidea). *Journal of Biogeography*, **27**, 1365–1384.

Dennis, R.L.H., Shreeve, T.G., Sparks, T.H. & Lhonore, J.E. (2002) A comparison of geographical and neighbourhood models for improving atlas databases. *Biological Conservation*, **108**, 143–159.

Dufrêne, M. & Legendre, P. (1991) Geographic structure and potential ecological factors in Belgium. *Journal of Biogeography*, **18**, 257–266.

Dutilleul, P. (1993) Modifying the t-test for assessing the correlation between two spatial processes. *Biometrics*, **49**, 305–314.

EEA (2002) *Europe's biodiversity: biogeographical regions and seas*. European Environment Agency, Kopenhagen.

Fleishman, E., Austin, G.T. & Murphy, D.D. (2001a) Biogeography of Great Basin butterflies: revisiting patterns, paradigms, and climate change scenarios. *Biological Journal of the Linnean Society*, **74**, 501–515.

Fleishman, E., Mac Nally, R., Fay, J.P. & Murphy, D.D. (2001b) Modelling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology*, **15**, 1674–1685.

Goffart, P. & De Bast, B. (2000) *Atlas préliminaire des papillons de jour de Wallonie*. Groupe de Travail Lépidoptères, Marche.

Goffart, P., Baguette, M., Dufrene, M., Mousson, L., Neve, G., Sawchik, J., Weiserbs, A. & Lebrun, P. (2000) *Gestion des milieux semi-naturels et restauration de populations menacées de papillons de jour*. Direction Générale des Resources Naturelles et de l'Environnement, Division de la Nature et des Forêts – Direction de la Nature, Jambes.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **13**, 147–186.

Hardy, P.B. & Dennis, R.L.H. (1999) The impact of urban development on butterflies within a city region. *Biodiversity and Conservation*, **8**, 1261–1279.

Hawkins, B.A. & Porter, E.E. (2003) Does herbivore diversity depend on plant diversity? The case of Califorina butterflies. *American Naturalist*, **161**, 40–49.

Isaaks, E.H. & Srivastava, R.M. (1989) *An introduction to applied geostatistics*. Oxford University Press, Oxford, UK.

Kerr, J.T. (2001) Butterfly species richness patterns in Canada: energy, heterogeneity, and the potential consequences of climate change. *Conservation Ecology*, **5**, U131–U147.

Kerr, J.T., Vincent, R. & Currie, D.J. (1998) Lepidopteran richness patterns in North America. *Ecoscience*, **5**, 448–453.

Kerr, J.T., Southwood, T.R.E. & Cihlar, J. (2001) Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. *Proceedings of the National Academy of Science of the United States of America*, **98**, 11365–11370.

Kudrna, O. (2002) The distribution atlas of European butterflies. *Oedippus*, **20**, 1–342.

Lafranchis, T. (2000) *Les papillons de jour de France, Belgique, Luxembourg et leurs chenilles*. Collection Parthénope - éditions Biotope, Mèze.

Legendre, P. & Legendre, L. (1998) *Numerical ecology*. Elsevier, Amsterdam.

Lennon, J.L. (2000) Red-shifts and red herrings in geographical ecology. *Ecography*, **23**, 101–113.

Littell, R.C., Milliken, G.A., Stroup, W.W. & Wolfinger, R.D. (1996) *SAS system for mixed models*. SAS Institute, Inc., Cary, NC.

Lobo, J.M. & Martín-Piera, F. (2002) Searching for a predictive model for species richness of Iberian dung beetle based on spatial and environmental variables. *Conservation Biology*, **16**, 158–173.

Lobo, J.M., Lumaret, J.P. & JayRobert, P. (1997) Taxonomic databases as tools in spatial biodiversity research. *Annales de La Societe Entomologique de France*, **33**, 129–138.

Luoto, M., Toivonen, T. & Heikinnen, R.K. (2002) Prediction of total and rare plant species richness in agricultural landscapes from satellite images and topographic data. *Landscape Ecology*, **17**, 195–217.

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

Mac Nally, R. (2002) Multiple regression and inference in ecology and conservation biology: further comments on identifying important predictor variables. *Biodiversity and Conservation*, **11**, 1397–1401.

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

Maes, D. & Van Dyck, H. (1999) *Dagvlinders in Vlaanderen – Ecologie, verspreiding en behoud*. Stichting Leefmilieu i.s.m. Instituut voor Natuurbehoud en Vlaamse Vlinderwerkgroep, Antwerpen, Brussel.

Maes, D. & Van Dyck, H. (2001) Butterfly diversity loss in Flanders (north Belgium): Europe's worst case scenario? *Biological Conservation*, **99**, 263–276.

Magurran, A.E. (1988) *Ecological diversity and its measurement*. Princeton University Press, NJ.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., daFonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

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

OECD (1998) *Environmental performance reviews Belgium*. OECD Editions, Paris.

Pearson, D.L. & Carroll, S.S. (1999) The influence of spatial scale on cross-taxon congruence patterns and prediction accuracy of species richness. *Journal of Biogeography*, **26**, 1079–1090.

Prendergast, J.R., Quinn, R.M., Lawton, J.H., Eversham, B.C. & Gibbons, D.W. (1993a) Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature*, **365**, 335–337.

Prendergast, J.R., Wood, S.N., Lawton, J.H. & Eversham, B.C. (1993b) Correcting for variation in recording effort in analyses of diversity hotspots. *Biodiversity Letters*, **1**, 39–53.

Pullin, A.S. & Knight, T.M. (2001) Effectiveness in conservation practice: pointers from medicine and public health. *Conservation Biology*, **15**, 50–54.

Rossi, R.E., Mulla, D.J., Journel, A.G. & Franz, E.H. (1992) Geostatistical tools for modelling and interpreting ecological spatial dependence. *Ecological Monographs*, **62**, 277–314.

Settele, J., Feldmann, R. & Reinhardt, R. (1999) *Die Tagfalter Deutschlands*. Ulmer, Stuttgart.

Sparks, T.H., Dover, J.W., Warren, M.S. & Cox, R. (1995) How well can we model the distribution of butterflies at the landscape scale? *Landscape ecology: theory and applications* (ed. by G.H. Griffith), pp. 24–31. IALE (UK), Aberdeen.

Stoltze, M. (1996) *Danske dagsommerfugle*. Gyldendal, Copenhagen.

van Swaay, C.A.M. & Warren, M.S. (1999) *Red data book of European butterflies (Rhopalocera), nature and environment No. 99*. Council of Europe Publishing, Strasbourg.

Tax, M. (1989) *Atlas van de Nederlandse dagvlinders*. Vereniging tot behoud van Natuurmonumenten in Nederland/De Vlinderstichting, 's Graveland/Wageningen.

Van Goethem, J. (red.) (2001) *Second National Report of Belgium to the Convention on Biological Diversity*. Royal Belgian Institute of Natural Sciences (RBINS), Brussels.

Warren, M.S., Hill, J.K., Thomas, J.A., Asher, J., Fox, R., Huntley, B., Roy, D.B., Telfer, M.G., Jeffcoate, S.H.P., Jeffcoate, G., Willis, S.G., Greatorex-Davies, J.N., Moss, D. & Thomas, C.D. (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature*, **414**, 65–69.

Weibull, A.C., Bengtsson, J. & Nohlgren, E. (2000) Diversity of butterflies in the agricultural landscape: the role of farming system and landscape heterogeneity. *Ecography*, **23**, 743–750.

Williams, P.H. & Gaston, K.J. (1994) Measuring more of biodiversity: can higher-taxon richness predict wholesale species richness. *Biological Conservation*, **67**, 211–217.

BIOSKETCHES

Dirk Maes is a researcher at the Institute of Nature Conservation in Brussels. He works on the use of ecological and distributional data of insects, especially butterflies, in nature conservation.

Marius Gilbert is a post-doctoral fellow with the National Fund of Scientific Research (Belgium) working at the Free University of Brussels on spatial ecology of insect pest populations.

Nicolas Titeux is a PhD student at the Catholic University of Louvain-la-Neuve and works on landscape ecology, especially on the relationships between landscape structure and bird communities.

Philippe Goffart is a researcher at the Centre de Recherche de la Nature, des Forêts et du Bois of Gembloux, where he works on distribution, ecology and conservation of butterflies and dragonflies.

Roger Dennis is a geographer and anthropologist originally from the University of Durham. He is currently an honorary research associate of Manchester University (Museum Entomology) and Oxford Brookes University (Biological Sciences) researching into butterfly ecology and biogeography.