RESEARCH ARTICLE

Landscape variables impact the structure and composition of butterfly assemblages along an urbanization gradient

Benjamin Bergerot · Benoit Fontaine ·
Romain Julliard · Michel Baguette

Abstract How urbanization affects the distribution patterns of butterflies is still poorly known. Here we investigated the structure and composition of butterfly assemblages along an urbanization gradient within the most urbanized and densely populated region in France (Île-de-France). Using a method issued from artificial neural networks, i.e. self-organizing maps (SOMs), we showed the existence of four typical assemblages ranging from urban-tolerant species to urban-avoider species. We identified indicator species of these assemblages: the peacock butterfly (*Inachis io*) in urbanized areas, the swallowtail (*Papilio machaon*) in sites with intermediate human pressure, or the meadow brown (*Maniola jurtina*), the small heath (*Coenonympha pamphilus*) and the gatekeeper (*Pyronia tithonus*) in meadows around Paris. A discriminant analysis showed that the four assemblages were mainly segregated by landscape elements, both by structural variables (habitat type, proportion of rural areas and artificial urban areas, patch surface) and functional variables (distance to the nearest wood, artificial area and park). Artificial neural networks and SOMs coupled stepwise discriminant analysis proved to be promising tools that should be added to the toolbox of community and spatial ecologists.

Keywords Community · Human pressure · Urban development · Landscape structure · Artificial neural networks

Introduction

Urbanization, i.e. human-driven land use change in and nearby an urbanizing zone, is usually considered as impacting biodiversity (Savard et al. 2000). Some basic empirical facts about urban systems and their development in France, Japan, and the USA over the last century or so (e.g. the increase of urban populations in artificial areas compared to rural population) have been pointed out and their histories are well-known (Eaton and Eckstein 1997; Dobkins and Ioannides 2001; Black and Henderson 2003; Ioannides and Overman 2003). Moreover, cities are generally growing in human population size over time. This growth is accompanied by the expansion of infrastructure in attractive territories (nearby existing cities for example) for many reasons of cultural, historical or strategic

B. Bergerot (✉) · B. Fontaine · R. Julliard
CNRS-MNHN-PARIS VI 'Conservation des Espèces,
Restauration et Suivi des Populations', CRBPO, 55 Rue
Buffon, CP 51, 75005 Paris, France
e-mail: bergerot@mnhn.fr

M. Baguette
CNRS—MNHN, UMR-7179 'Mécanismes adaptatifs: des
organismes aux communautés', 1 avenue du petit château,
91800 Brunoy, France

M. Baguette
CNRS USR 2936 'Station d'écologie expérimentale du
CNRS à Moulis', 09200 Moulis, France

order (Rubinoff and Powell 2004; Niell et al. 2007). So, current urban landscapes are the result of human activities and due to their structure, urban habitats generate a break in the ecological continuum of many living organisms, creating high spatial heterogeneity in habitats (Hardy and Dennis 1999). We predict that this heterogeneity will particularly impact species functioning in metapopulations (Steffan-Dewenter and Tscharntke 2000; Mennechez et al. 2003). Changes in habitat quality associated with urbanization have many different effects on biodiversity. They include a reduction in the size of habitat patches supporting indigenous fauna, the invasion and planting of alien flora, the increasing isolation of these remnant patches from other similarly "green" areas, an increase in the level of air pollutants due to road traffic and industry, and an increase in physical disturbance in and around the habitat fragments (Frankie and Ehler 1978; McIntyre and Barrett 1992). In urban areas, these disturbance effects often exist when a centre of concentrated development is surrounded by areas with decreasing levels of habitat alteration (McDonnell and Pickett 1990). These gradients of increasing habitat fragmentation and isolation have been associated with a reduction in the diversity of invertebrate species (Desender and Turin 1989) and may have dramatic consequences on species with poor dispersal abilities (Harrington and Stork 1995).

Butterflies are well-suited to investigate this issue (Blair and Launer 1997; New 1997; Leon-Cortés et al. 2000; Wood and Pullin 2002). They are sensitive indicators that react quickly to environmental changes (Erhardt 1985); sensitive indicators are crucial to detect subtle habitat changes and hence to understand new species assemblages (Boggs et al. 2003). Moreover, butterflies are a charismatic group, categorized in Red Lists (IUCN 2001) which make the results usable for public discussion and political actions (Fattorini 2006). The vast majority of butterflies have specific habitat requirements (Leon-Cortés et al. 2004); for these species, the world can be partitioned into hospitable or inhospitable habitats (Vanreusel and van Dyck 2007). From a butterfly's eyes, the urban context is a patchwork of many small (gardens, lawns) and several large (parks) suitable habitat patches embedded in a totally unsuitable urban matrix, which should hinder metapopulation dynamics when distances between suitable patches increase. The connectivity between patches varies according to the species

(e.g. two patches two patches could be considerate as connected if distances between them not exceed 5 km for *Pieris brassicae*, Feltwell 1981, or less than 1 km for *Pararge aegeria*, Merckx & Van Dyck 2007).

Like all over the world (Didham et al. 1996; Dumbrell and Hill 2005; Schmitt and Rakosy 2007), natural areas are decreasing in France, particularly in the Île-de-France region where urbanization associated with a high density of humans is the most important threat for plants and animals confronted with the fragmentation of their habitats. Although studies have been made in urban areas, to determine conservation priorities or to investigate the consequences of management plans (Nelson and Nelson 2001; Brown and Freitas 2002; Giuliano et al. 2004), a lack of knowledge exists at a higher level concerning the composition of species assemblages inside fragmented habitat patches and particularly on the impact of urban landscape elements on species communities (Blair and Launer 1997; Cole et al. 2005; Saarinen et al. 2005).

Some authors have shown the relevance of species assemblage studies and their usefulness for the assessment of the ecological habitat status in a heterogeneous landscape (Dufrêne and Legendre 1997; Hamer et al. 2003). New methodological approaches to investigate spatial assemblage patterns could provide useful tools for the understanding of the ecological continuum of species (Lasne et al. 2007). Methods such as artificial neural networks have shown that it was possible to successfully investigate the assemblage distribution patterns as well as individual species distribution patterns along complex gradients (Lek and Guegan 1999; Piscart et al. 2010). In this study, we investigated the effect of urbanization on butterfly assemblages. We monitored butterfly species along an urbanization gradient in the Île-de-France region and tested the role of landscape elements that could affect species distributions. Firstly, we classified butterfly assemblages using self-organizing maps (SOMs), a method issued from artificial neural networks. SOMs are designed to detect clusters of co-occurring species (Manel et al. 1999; Olden et al. 2006) and are considered to be more efficient than former analytical methods (Park et al. 2006). Then, by analyzing species assemblage with the Individual Value (IndVal) method (Dufrêne and Legendre 1997), we identified those indicator species that could be used to subsume each assemblage.

Fig. 1 a Île-de-France region network, ground cover types and sampling site locations. Ground cover types are represented by *various shades*. Sampling site locations correspond to transects and are identified by *white squares*. **b** The Île-de-France region network and sampling sites. Clusters of sites identified by the SOM procedure based on assemblage similarities are indicated by *various symbols and shades*
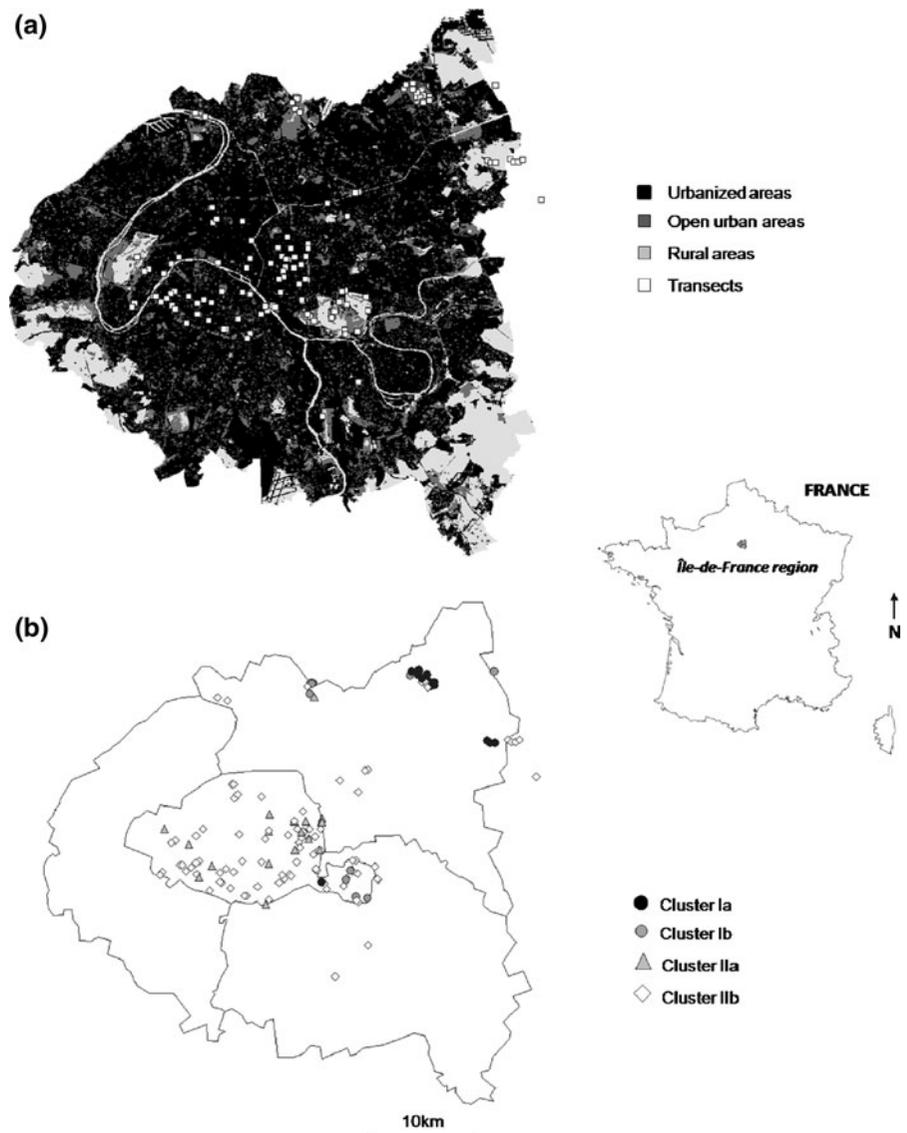
Finally we performed a stepwise discriminant analysis to sort out which landscape elements impacted the structure and composition of the butterfly assemblages.

Methods

Study area

With 12,001 km^2 and 11,490,500 habitants, Île-de-France is the most populated administrative region in France and one of the densest in the world (INSEE sources, French National Institute for Statistic and Economic Studies). For example, Paris (part of the Île-de-France) has an average density of 20.65 inhabitants per km^2 (INSEE sources). This high pressure generates a strong structuring contrast between patches at the landscape level. The centre of the Île-de-France region is located at 48°51′24.30″N and 2°20′34.46″E (Fig. 1).

Data sets

Butterfly data were collected using a standard method described by Pollard and Yates (1993). Fixed transect

routes were walked three times a year. Such reduced sampling effort required on average twice as many sites to achieve comparable precision to a weekly sampling effort scheme such as the United Kingdom Butterfly Monitoring Scheme (Roy et al. 2007). Data concerning 135 sampling sites were collected in 2008 and 2009 across the Île-de-France region (Fig. 1). Fixed dates were used to realize the three transects sessions: June 1st, July 1st and August 10th (±10 days). These dates correspond to the peaks of species richness and abundance in the Île-de-France region. The 10 days window allowed the realization of transect sessions under optimal weather conditions (wind speed less than 5 Beaufort, air temperature at least 17°C and at least 75% sunshine).

We investigated the distribution of 26 a priori selected butterfly species. The use of such a closed list could induce some loss of information, which is here traded-off against the risk of species misidentification. Indeed, the identification of species belonging to genera like *Pieris* spp. and *Polyommatus* spp. requires individual captures, which were not possible in every sampling sites. We chose to remove such genera from this study to avoid potential biases.

Environmental variables

A set of environmental variables was recorded in each sampling site (transect length varied between 50 and 450 m in parks ranging from 0.22 and 349 ha): (1) habitat type (6 categories: wasteland, lawn, urban park, meadow, vegetable garden and cemetery), (2) temperature (°C) and (3) cloud cover (five classes: 0, <5, 5–25, 25–50, and 50–75%). We scored these cloud cover classes from 1 to 5, respectively, and we calculated the mean value for each site over the three transect sessions.

The landscapes of Île-de-France are strongly contrasted with three broad types of ground covers according to the MOS classification (Soil Occupation Mode, ECOMOS 2003, Fig. 1a): (1) artificial urban surfaces (including 54 habitat classes such as buildings, parking lots or roads), (2) open urban areas (including 14 habitat classes such as private and public gardens) and (3) rural landscapes (including 15 habitat classes such as forest and crop fields). We assessed the urbanization degree surrounding each study plot by quantifying the surface occupied by each ground cover within a buffer around each transect.

Buffer radius was set to 5 km to exceed the maximal dispersal distance reported for the most sedentary butterfly species (Hanski and Kuussaari 1995). We calculated the area (m^2) of each habitat patch where transects were located. Then, for each transect, we also calculated the distances (m) to the nearest wood, open urban areas, agricultural areas and artificial areas according to the MOS database classification. These different landscape components were expected to impact butterfly dispersal abilities by acting either as barriers, corridors or stepping stones.

Statistical analysis

The study of ecological communities often results in large species * sampling unit matrices. Ordination methods are commonly used to detect relationships between species abundances and environmental gradients in such matrices. However, these methods presuppose that these relations are linear, which is an often unverified strong assumption (Kenkel and Orloci 1986). To tackle the problem of non-linearity, several methods have been developed in recent years.

Artificial neural networks have already been successfully been used in ecology (Lek and Guegan 1999; Kohonen 2001; Piscart et al. 2010) and are recognized as a powerful tool for describing species distributions and assemblages (Lek et al. 2005). The SOM (also called Kohonen map) used in this paper is a kind of ANN (Artificial Neural Network) method which is a robust method for clustering, classification, estimation, prediction, and data mining (e.g. Vesanto and Alhoniemi 2000; Giraudel and Lek 2001; Kohonen 2001). The predictive power of SOM has been validated in various studies (e.g. Somervuo and Kohonen 1999; Kohonen 2001; Ibarra et al. 2005). The SOM is a learning algorithm that was originally proposed by Kohonen (1982). SOM networks learn to cluster groups of similar input patterns using nonlinear methods from a multi-dimensional input space into a low dimensional (most commonly two-dimensional) composite, discrete lattice of neurons (Kohonen 2001). Neurons physically located close to each other in the output layer of the SOM have similar input patterns (combining clustering and ordering processes in SOM). Accordingly, SOM is both a similarity graph and a clustering diagram. The main advantage of the SOM algorithm is that it is a non-linear process, which preserves the topological structure of the data

(Corne et al. 1999). Indeed, when the target class of the patterns is unknown, unsupervised non-linear methods such as SOM are the only way to extract structure in the data. Simpler linear methods are often based on an analytical solution but they are outcompeted by nonlinear methods when the classification task consists in the transformation of a multi-dimensional space into a two (or three) dimensional composite space. The SOM algorithm usually clusters the samples or patterns into predefined classes (i.e. the number of neurons is selected by the modeller), and also orders the classes into meaningful maps.

To produce summaries (quantitative descriptions of data properties), groups of map units must be selected from the SOM. Here we used a classical hierarchical clustering analysis (Ward's Euclidean method, Ward 1963) to cluster the SOM.

To detect indicator species in each cluster, we applied the IndVal method (Dufrêne and Legendre 1997). IndVal selected indicator species on the basis of two criteria, fidelity and specificity. Fidelity was highest when the species was present in all the sites within a cluster. Specificity was highest when all the individuals of a species were found in the same cluster. The IndVal of species i in cluster type j is expressed as a percentage and is calculated as follows: $IndVal_{ij} = A_{ij} \times B_{ij} \times 100$, where A_{ij} (=Abundance$_{ij}$/Abundance$_i$) is a measure of the specificity of species i to the type j, and B_{ij} (=Nsites$_{ij}$/Nsites$_j$) is a measure of the fidelity of species i to type j. A randomization procedure was used to test the difference ($\alpha = 0.05$) of the IndVal of each species in the different clusters of a given hierarchical level. Only significant IndVal >25 were taken into account, because an IndVal >25 implies that the species is present in at least 50% of the sites of the cluster, and that this cluster contains at least 50% of the total abundance of the species. As suggested by Dufrêne and Legendre (1997), the level for which a species has its highest IndVal value should be considered as the best classification level for that indicator species. However, lower IndVal values may provide supplementary information on the distribution patterns of the species especially at lower hierarchical levels. Analyses of indicator values were performed using INDVAL 2.0.

We used the Butterfly Feeding Specialization Index (Bergerot et al. 2010) to estimate the degree of specialization of the butterfly community in each clusters identified by the SOMs process. This index considers the width of the feeding regime of adult butterflies, from monophagy to oligophagy to polyphagy. For each cluster, we calculated the mean of this feeding specialization index.

Mean abundance, species richness, butterfly specialization index and environmental variables in each cluster were compared using Kruskal–Wallis tests (Kruskal and Wallis 1952) and Dunn's post-hoc tests (Dunn 1964). Species richness was preferred to any other diversity index for several reasons. First, the small number of species recorded makes the interpretation of any diversity metric difficult. Second, the choice of a diversity metric is an arbitrary decision, giving different weightings to species number and species evenness; richness is less ambiguous and easier to interpret (Magurran 1988). The SOM and cluster analysis were computed with the SOM toolbox$^{\copyright}$ (Alhoniemi et al. 2000) within the Matlab environment (The Mathworks, Inc., Natick, MA, USA).

A backward stepwise discriminant analysis (performed with R2.7.0$^{\copyright}$) was used to determine whether clusters of sites derived from the SOM procedure and based on mean species abundance could be discriminated based on environmental variables. A random Monte-Carlo permutation test was used to assess if these variables efficiently segregated the clusters of butterfly communities.

Results

A total of 21 species (out of 26) among six families were sampled in the 135 sampling sites in 2008 and 2009 (Table 1). One species, *Vanessa cardui* (Nymphalidae), was very common (species mean occurrence >75%, Table 2). To a lesser extent, *Inachis io* (Nymphalidae), *Vanessa atalanta* (Nymphalidae) and *Pararge aegeria* (Satyridae) occurred in more than a quarter of all sites. In contrast, nine species were scarce (mean occurrence <5%, Table 1): *Issoria lathonia* (Nymphalidae), *Iphiclides podalirius* (Papilionidae), *Callophrys rubi* (Lycanidae), *Araschnia levana* (Nymphalidae), *Pyrgus malvae* (Hesperiidae), *Papilio machaon* (Papilionidae), *Cacyreus marshalli* (Lycaenidae), *Aglais urticae* (Nymphalidae) and *Celastrina argiolus* (Lycanidae).

The SOM procedure allowed us to ordinate samples in a two-dimensional map of 8 × 7 cells (Fig. 2a). Based on cell similarities, the clustering

Table 1 List of species classified according to their families (Tolman and Lewington 1997)

Scientific name	Common name	Mean overall abundance per 1,000 m² (SD)	Mean overall occurrence (SD)	KW	Mean abundance per 1,000 m² (SD) per cluster			
					Cluster 1	Cluster 2	Cluster 3	Cluster 4
Hesperiidae								
Pyrgus malvae (Linnaeus, 1758)	Grizzled skipper	0.007 (0.049)	0.022 (0.148)	**	0.004 (0.041) a	0.00 (0.00) a	0.00 (0.00) a	0.049 (0.121) b
Lycaenidae								
Cacyreus marshalli (Butler, 1898)	Geranium bronze	0.009 (0.073)	0.022 (0.148)	ns	0.01 (0.083)	0.00 (0.00)	0.022 (0.077)	0.00 (0.00)
Callophrys rubi (Linnaeus, 1758)	Green hairstreak	0.002 (0.02)	0.015 (0.121)	ns	0.001 (0.014)	0.011 (0.045)	0.00 (0.00)	0.00 (0.00)
Celastrina argiolus (Linnaeus, 1758)	Holly blue	0.01 (0.05)	0.044 (0.207)	ns	0.01 (0.051)	0.013 (0.052)	0.00 (0.00)	0.02 (0.067)
Nymphalidae								
Aglais urticae (Linnaeus, 1758)	Small tortoiseshell	0.006 (0.035)	0.03 (0.17)	ns	0.008 (0.042)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Araschnia levana (Linnaeus, 1758)	Map butterfly	0.003 (0.021)	0.015 (0.121)	ns	0.002 (0.016)	0.011 (0.045)	0.00 (0.00)	0.00 (0.00)
Inachis io (Linnaeus, 1758)	Peacock butterfly	0.113 (0.228)	0.341 (0.476)	***	0.077 (0.152) a	0.38 (0.446) b	0.064 (0.088) ab	0.058 (0.104) a
Issoria lathonia (Linnaeus, 1758)	Queen of Spain fritillary	0.003 (0.034)	0.007 (0.086)	ns	0.004 (0.041)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Polygonia c-album (Linnaeus, 1758)	Comma butterfly	0.035 (0.133)	0.104 (0.306)	ns	0.03 (0.127)	0.075 (0.203)	0.044 (0.118)	0.00 (0.00)
Vanessa atalanta (Linnaeus, 1758)	Red admiral	0.093 (0.272)	0.274 (0.448)	*	0.065 (0.214) a	0.099 (0.13) a	0.097 (0.158) a	0.319 (0.664) b
Vanessa cardui (Linnaeus, 1758)	Painted lady	1.296 (1.473)	0.881 (0.324)	***	0.814 (0.803) b	4.238 (1.625) b	0.76 (0.385) a	1.504 (1.142) a
Satyridae								
Coenonympha pamphilus (Linnaeus, 1758)	Small heath	0.134 (0.419)	0.163 (0.371)	***	0.012 (0.069) a	0.308 (0.641) a	0.091 (0.136) a	0.971 (0.816) b
Lasiommata megera (Linnaeus, 1767)	Wall brown	0.017 (0.08)	0.052 (0.223)	***	0.005 (0.034) a	0.016 (0.065) a	0.127 (0.216) b	0.00 (0.00) a
Maniola jurtina (Linnaeus, 1758)	Meadow brown	0.734 (2.081)	0.222 (0.417)	***	0.046 (0.22) a	0.658 (1.562) ab	1.535 (1.779) bc	5.917 (3.921) c
Melanargia galathea (Linnaeus, 1758)	Marbled white	0.536 (1.73)	0.2 (0.401)	***	0.071 (0.366) a	0.723 (1.874) a	1.489 (1.258) b	3.226 (4.509) b
Pararge aegeria (Linnaeus, 1758)	Speckled wood	0.198 (0.593)	0.267 (0.444)	ns	0.203 (0.653)	0.294 (0.539)	0.157 (0.386)	0.053 (0.175)
Pyronia tithonus (Linnaeus, 1771)	Gatekeeper	0.23 (1.088)	0.096 (0.296)	***	0.00 (0.00) a	0.329 (0.958) a	0.056 (0.137) a	2.248 (3.054) b
Papilionidae								
Iphiclides podalirius (Linnaeus, 1758)	Scarce swallowtail	0.002 (0.02)	0.015 (0.121)	ns	0.003 (0.024)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Papilio machaon (Linnaeus, 1758)	Swallowtail	0.009 (0.067)	0.022 (0.148)	***	0.00 (0.00) a	0.00 (0.00) a	0.096 (0.211) b	0.00 (0.00) a
Pieridae								
Colias crocea (Geoffroy, 1785)	Clouded yellow	0.068 (0.176)	0.2 (0.401)	***	0.019 (0.064) a	0.14 (0.221) b	0.309 (0.391) b	0.119 (0.156) b
Gonepteryx rhamni (Linnaeus, 1758)	Brimstone	0.027 (0.171)	0.052 (0.223)	ns	0.029 (0.198)	0.013 (0.052)	0.021 (0.073)	0.035 (0.117)

Mean overall abundances (SD in parentheses) per 1,000 m² transects (pooled over clusters) per 1,000 m² transects where the species was present), and mean abundance per 1,000 m² transects (pooled over clusters). KW: test for differences among clusters in mean abundances by a Kruskal–Wallis one-way analysis of variance. Mean values followed by the *same letter* are not statistically different ($\alpha = 0.05$) according to Dunn's post-hoc test when Kruskal–Wallis tests are significant ($\alpha = 0.05$; *** $\leq$0.001, ** $\leq$0.01, * $\leq$0.05, ns = not significant)

Table 2 Mean values (SD) of environmental variables in the four clusters

	P	KW	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Temperature (°C)	0.16	5.13 ns	24.54 (1.82)	24.83 (2.14)	24.14 (2.2)	25.76 (1.69)
Patch area (m²)	***	22.48 ***	154332.53 (403616.04) a	262224.21 (598750.23) a	181506.23 (195994.73) a	617822.55 (286139.16) b
Cloud cover	0.3	3.66 ns	4.37 (0.45)	4.37 (0.51)	4.44 (0.87)	4.39 (0.63)
Rural landscape (%)	***	39.9 ***	9 (8) a	12 (12) ab	22 (15) bc	28 (4) c
Open urban areas (%)	0.99	0.06 ns	13 (2)	13 (2)	13 (4)	13 (1)
Artificial urban landscape (%)	***	37.09***	78 (9) c	74 (11) bc	64 (12) ab	58 (3) a
Distance to the nearest artificial urban landscape (m)	***	20.07***	9.25 (33.88) a	18.61 (56.88) a	23.32 (29.63) a	56 (59.7) b
Distance to the nearest wood (m)	***	19.3***	1935.11 (1346.05) b	2050.33 (1184.46) b	372.66 (685.51) a	1358.59 (239.86) ab
Distance to the nearest agricultural area (m)	***	27.03***	3417.07 (2054.73) b	3352.07 (2402.2) b	630.36 (778.44) a	1227.55 (384.51) a
Distance to the nearest park (m)	0.13	5.58 ns	31.22 (85.21)	39.95 (71.25)	84.01 (122.38)	0.4 (1.34)

When Kruskal–Wallis (KW) values are significant ($\alpha = 0.05$; *** ≤ 0.001, ** ≤ 0.01, * ≤ 0.05, ns = not significant), mean value followed by the *same letter* are not statistically different (Dunn's post-hoc tests; $\alpha = 0.05$)

procedure identified four main clusters (Fig. 2a). On the artificial neural network representation, 95, 17, 12 and 11 sites composed respectively clusters 1, 2, 3 and 4 (Fig. 2a). Each cluster of the sites can be represented on the Île-de-France map (Fig. 1b). Thus on this map, sites were classified in four types which correspond to the four clusters. Sites ordered in cluster 1 were mainly located in urban areas, whereas cluster 4 sites were mainly rural sites (Fig. 1b, Table 2).

Some species (with a mean occurrence >5%, Table 2) were evenly distributed among clusters: *Polygonia c-album* (Nymphalidae), *Pararge aegeria* and *Gonepteryx rhamni* (Pieridae) (Table 1). The abundance of *Maniola jurtina* (Satyridae), *Melanargia galathea* (Satyridae) or *Pyronia tithonus* (Pieridae) was significantly different according to clusters (Table 1).

The IndVal method selected indicator species of clusters at the different levels of classification (Fig. 2b). The number of species with significant indicator values varied according to clusters. At the first hierarchical level, no indicator species was found in cluster 1, whereas eight species were indicator of the second assemblage (which integrates clusters 2, 3 and 4, Fig. 2a). At the lower level, from clusters 1 to 4, the number of indicator species was 0, 1, 2 and 3, respectively. Six indicator species (out of a total of 11 species with a mean occurrence >5%) had their highest IndVal values at the second classification level. These results suggested that clusters 1, 2, 3 and 4 had a sound ecological significance. The specific richness increased significantly from cluster 1 to cluster 4 (Fig. 3a). Moreover, butterfly species in clusters 3 and 4 had a mean specialization feeding index significantly higher than those in clusters 1 and 2 (Fig. 3b).

Data collected in each site and MOS (Soil Occupation Mode) showed high variations of environmental variables among sampling sites according to clusters (Table 2). Indeed, sampling sites were classified along a gradient of urbanization from cluster 1, which included mainly sites within artificial urban landscapes, to cluster 4, where sites were located in rural landscapes (mostly meadows).

To identify those most influential factors that segregated clusters, we used a backward stepwise discriminant analysis (Fig. 4) that selected only landscape variables. Landscape type, proportion of rural areas, proportion of artificial urban areas, patch area, distance to the nearest wood, to artificial areas
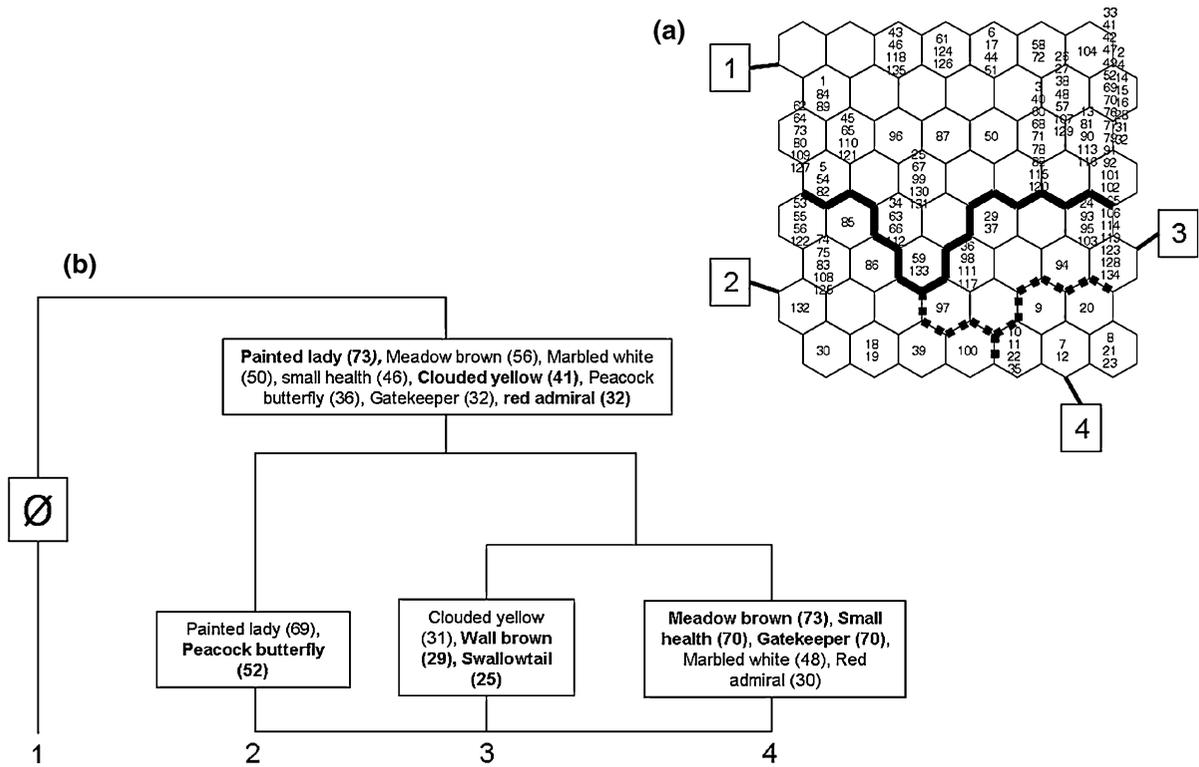
Fig. 2 **a** Distribution of the sampled sites on the SOM. Clusters of sites identified by the Ward's Euclidean method are indicated by a *full black* and *bold line* (higher hierarchical level) and *dotted black* and *bold lines* (lower hierarchical level). **b** Indicator species of the clusters of sites at the two levels. Indicator values (%) are given in parentheses, and *bold characters* indicate the highest indicator value for a given species
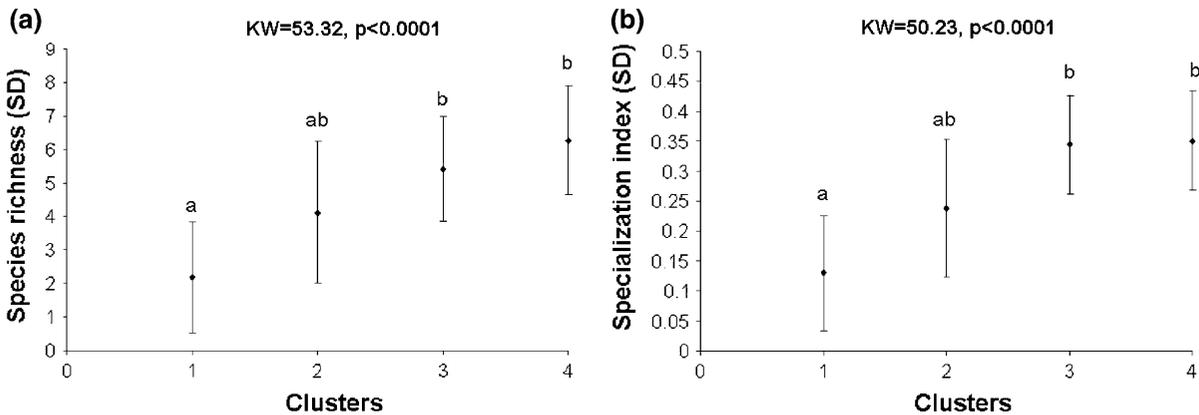

Fig. 3 Comparison of **a** the mean specific richness (±SD) and **b** mean specialization index of butterfly communities (±SD) between clusters. *Same letters above bars* shows not statistically different values (Dunn's post-hoc test; α = 0.05). KW: Kruskal–Wallis test value; *P*: Kruskal–Wallis *P*-value

and to parks had a significant effect on the organization of species in four clusters. Two significant discriminant functions were generated; a random Monte Carlo permutation tests showed that they were highly significant ($P < 0.001$). These functions (F1 and F2) accounted for 65.32 and 32.36% respectively of the global variability between clusters. F1 axis was only determined by landscape structural features, and

(a)

(b)
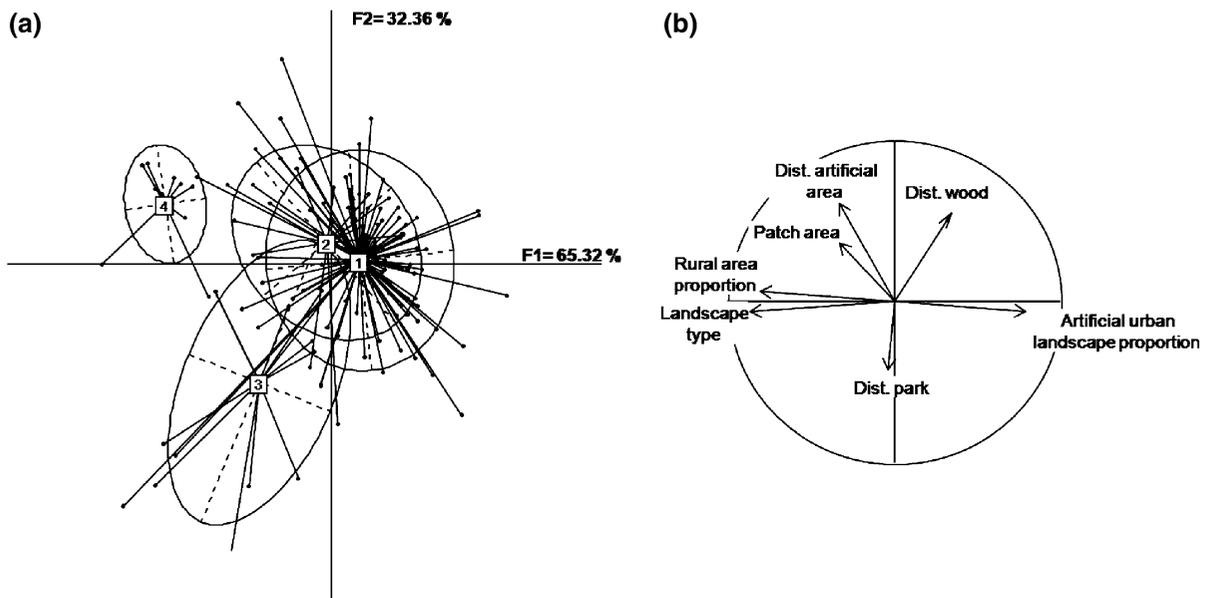
Fig. 4 a Backward stepwise discriminant analysis using environmental variables to predict the four clusters of sites. Axis 1 accounts for 65.32% and axis 2 for 32.36% of between group variability respectively. Each group of clusters is presented as an ellipsoid with the cluster name in the centre. **b** *Circle* showing the contribution of the variables to F1 and F2 (Dist. wood: distance to the nearest wood, Dist. Artificial area: distance to the nearest artificial area, Dist. park: distance to the nearest park)

more particularly by landscape type (cosine = −0.89), artificial area proportion (cosine = 0.86) and rural area proportion (cosine = −0.84). Sites in clusters 1 and 2 were mainly located in artificial areas and cluster 3 and 4 in more rural areas where habitat such as meadows, wastelands and parks predominated. F2 axis was determined by four landscape features more related to functional characteristics: patch area (cosine = 0.40), distance to the nearest wood (cosine = 0.48), distance to the nearest artificial area (cosine = 0.58) and distance to the nearest park (cosine = −0.44).

Discussion

We showed here that assemblages of butterfly species differed along an urbanization gradient. Four clusters of species were selected by SOMs at the lower hierarchical levels. The IndVal methods provided indicator species for three of them. Comparison of environmental variables by both non-parametric ANOVA (Analyses of Variance, Kruskal–Wallis test) and backward discriminant analyses provided evidences that sampling sites corresponding to these four

species clusters segregated according to a set of both structural and functional landscape attributes that covary along the urbanization gradient. In strongly urbanized sites (species clusters 1 and 2), both mean species richness and mean feeding specialization of adults were significantly lower than in less modified sites (species cluster 3 and 4). This result corresponds to the general pattern that generalists are more common than specialists in sites where urban pressure was higher (Blair and Launer 1997; Niell et al. 2007). Generalists could be more tolerant to urbanization due to particular life-history traits such as higher dispersal abilities and/or wider feeding regime. Accordingly, even small, low-quality habitat patches deeply embedded in the urban matrix are likely to be colonized by those generalist species that diffuse rather freely in the landscape, whereas species functioning in metapopulations should be more impacted by urbanization.

The first axis of the discriminant analysis clearly segregated butterfly species according to structural variables of the landscape. Clusters 1 and 2 correspond to those sampling sites that are located in the urban matrix, with a high proportion (ca. 75%) of artificial landscapes in their surroundings, and far from open or wooded natural areas. Accordingly,

clusters 1 and 2 had respectively 0 and 2 indicator species that were selected by the IndVal method. Both indicator species of cluster 2 are widespread Nymphalid butterflies (*Inachis io* and *Vanessa cardui*) with strong mobility and a wide adult feeding regime. Both species are characterized as having open population structures and non-permanent population structure (Bink 1992). Besides these species selected by the IndVal method, two other butterfly species were rather common in clusters 1 and 2: *Pararge aegeria* and *Celastrina argiolus*. These two species share some life history traits: males are territorial and defend small patches of woody habitats, there are at least two generations a year and adults have a wide feeding regime. Moreover, both gardening practices and park management increase the density of their larval food plants, which is a crucial factor for the persistence of butterflies in urban areas (Hardy and Dennis 1999). Finally, both species have a high vagrancy index (Cook et al. 2001), which means that adults have been often observed in areas where their host plant is absent. This vagrancy index is the most reliable estimate of a species' mobility (Stevens et al. 2010). Clusters 3 and 4 correspond to sampling sites with a lower urbanization rate (64 and 58% respectively) that are closer to open or wooded natural areas. The IndVal method selected three and five indicator species for cluster 3 and 4, respectively. The indicator species of cluster 3 (*Papilio machaon*, *Colias crocea* and *Lasiommata megera*) are all species with a rather open population structure (Bink 1992). However, the vagrancy index (Cook et al. 2001), available only for *Lasiommata megera*, is twice as low as those of *Pararge aegeria* and *Celastrina argiolus*. The five indicator species of cluster 4 are Satyrids (*Maniola jurtina*, *Pyronia tithonus*, *Melanargia galathea* and *Coenonympha pamphilus*) and the Nymphalid *Vanessa atalanta*. The four Satyrid species are typical meadow dwellers, where their larval food plants (*Poacae* sp.) grow. Sampling sites associated with cluster 4 have indeed a significantly higher proportion of rural landscape in their vicinity than the other sampling sites along the urbanization gradient. Given their lower mobility than other indicator species, their population structure is rather considered as closed (Bink 1992). A genetic study of three of these species in an urban context indeed confirmed the existence of a significant differentiation among populations in *P. tithonus* and

C. pamphilus (Wood and Pullin 2002). The Fst values of the third one (*M. jurtina*) were not significantly different from zero, but it is worth mentioning that the observed values (Fst = 0.044) in the urban context was three times as high as the values observed for populations of the same species in a non urban context (Fst = 0.015, Goulson 1993) on a comparable spatial scale (30 km).

The second axis of the discriminant analysis segregated those sampling sites that were associated with species cluster 3 from those associated with cluster 4. Species in cluster 3 were observed farther from parks than those observed in cluster 4. Species in cluster 4 were observed in significantly larger patches that were significantly farer from artificial urban landscapes and significantly farer from woods. We suggest that these landscape variables might have a functional role in the metapopulation dynamics at least for some species associated with cluster 4. Patch area and isolation are indeed used in metapopulation biology as surrogates of population density and connectivity, respectively (e.g. Hanski 1999; Baguette and Van Dyck 2007). Our results suggest that species in cluster 4 were thus observed in large meadows without interspersed urban structures or woods.

There are two methodological caveats to our study design. Firstly, samples were restricted to three periods in June, July and August, whereas the whole flying period for butterflies in the area ranged from March to October). Accordingly, we missed common species like *Antocharis cardamines* that flew in April–May. However, as we focussed on the detection of spatial structure in species assemblages, we chose a priori to trade-off the length of the flying period against the spatial dispersion of the sampling points along the urbanization gradient. Secondly, we used a close list of species to avoid misidentifications. Here again, we chose a priori a strategy aiming at maximizing the safety of the biological information at the expense of its exhaustiveness. Given the strong observed relationships between landscape variables and species' abundances, we are confident that our results are robust enough to be extended to a larger number of species.

To conclude, we showed here that butterfly species assemblages are strongly impacted by urbanization. As in previous studies, we showed that species diversity decreased along the urbanization gradient (Blair and Launer 1997; Dennis and Hardy 2001). We also confirmed that urban tolerant butterflies in highly

fragmented landscapes had a wider feeding regime at the adult stage than urban avoiders in more continuous landscapes, which seems a general rule along habitat fragmentation gradients (Steffan-Dewenter and Tscharntke 2000). Additionally, we provided new insights on the role of landscape variables on species assemblage segregation. We showed that butterfly species were so influenced by the urbanization gradient that it is possible to separate species assemblages according to significant differences in several landscape variables among sampling sites. We also related species' population structure, and hence mobility, to their ability to cope with urbanization. Species with metapopulation functioning are clearly urban avoiders, whereas urban tolerant species are mobile or highly mobile. Finally, artificial neural networks and self-organized maps coupled to discriminant stepwise analysis proved to be promising analytical tools that should be added to the toolbox of community and spatial ecologists.

References

Alhoniemi E, Himberg J, Parhankangas J, Vesanto J (2000) SOM toolbox. Laboratory of Computer and Information Science, Helsinki University of Technology, Helsinki. http://www.cis.hut.fi/projects/somtoolbox. Accessed September 2009

Baguette M, Van Dyck H (2007) Landscape connectivity and animal behavior: functional grain as a key determinant for dispersal. Landscape Ecol 22:1117–1129

Bergerot B, Fontaine B, Renard M, Cadi A, Julliard R (2010) Preferences for exotic flowers do not promote urban life in butterflies. Landsc Urban Plan (in press)

Bink BA (1992) Ecologische atlas van de Dagvlinders van Noordwest-Europa (Ecological atlas of the butterflies of NW Europe). Schuyt & Co, Haarlem

Black D, Henderson JV (2003) Urban evolution in the USA. J Econ Geogr 3:343–373

Blair RB, Launer AE (1997) Butterfly diversity and human land use: species assemblages along an urban gradient. Biol Conserv 80:113–125

Boggs CL, Watt WB, Ehrlich PR (2003) Butterflies: ecology and evolution taking flight. University of Chicago, Chicago

Brown JKS, Freitas AVL (2002) Butterfly communities of urban forest fragments in Campinas, São Paulo, Brazil: structure, instability, environmental correlates, and conservation. J Insect Conserv 6:217–231

Cole LJ, McCracken DI, Downie IS, Dennis P, Foster GN, Waterhouse T, Murphy KJ, Griffin AL, Kennedy MP (2005) Comparing the effects of farming practices on ground beetle (Coleoptera: Carabidae) and spider (Araneae) assemblages of Scottish farmland. Biodivers Conserv 14:441–460

Cook LM, Dennis RLH, Hardy PB (2001) Butterfly-hostplant fidelity, vagrancy and measuring mobility from distribution maps. Ecography 24:497–504

Corne S, Murray T, Openshaw S, See L, Turton I (1999) Using computational intelligence techniques to model subglacial water systems. J Geogr Sci 1:37–60

Dennis RLH, Hardy PB (2001) Loss rates of butterfly species with urban development. A test of atlas data and sampling artefacts at a fine scale. Biodivers Conserv 10:1831–1837

Desender K, Turin H (1989) Loss of habitats and changes in the composition of the ground and tiger beetle fauna in four West European Countries since 1950 (Coleoptera: Carabidae, Cicindelidae). Biol Conserv 48:277–294

Didham RK, Ghazoul J, Stork NE, Davis AJ (1996) Insects in fragmented forests: a functional approach. Trends Ecol Evol 11:255–260

Dobkins LH, Ioannides YM (2001) Spatial interactions among U.S. cities: 1900–1990. Reg Sci Urban Econ 31:701–732

Dufrêne M, Legendre P (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol Monogr 67:345–366

Dumbrell AJ, Hill JK (2005) Impacts of selective logging on canopy and ground assemblages of tropical forest butterflies: implications for sampling. Biol Conserv 125: 123–131

Dunn OJ (1964) Multiple contrasts using rank sums. Technometrics 5:241–252

Eaton J, Eckstein Z (1997) Cities and growth: evidence from France and Japan. Reg Sci Urban Econ 27:443–474

ECOMOS (2003) Ecological soil occupation mode. http://www.iau-idf.fr/lile-de-france/un-portrait-par-les-chiffres/occupation-du-sol.html. Accessed February 2010

Erhardt A (1985) Diurnal Lepidoptera: sensitive indicators of cultivated and abandoned grassland. J Appl Ecol 22: 849–861

Fattorini S (2006) A new method to identify important conservation areas applied to the butterflies of the Aegean Islands (Greece). Anim Conserv 9:75–83

Feltwell J (1981) Large white butterfly: the biology, biochemistry, and physiology of *Pieris Brassicae*. D.R.W. Junk Publishers, Boston

Frankie GW, Ehler LE (1978) Ecology of insects in urban environments. Annu Rev Entomol 23:367–387

Giraudel JL, Lek S (2001) A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. Ecol Model 146:329–339

Giuliano WM, Accamando AK, McAdams EJ (2004) Lepidoptera-habitat relationships in urban parks. Urban Ecosyst 7:361–370

Goulson D (1993) Allozyme variation in the butterfly, *Maniola jurtina* (Lepidoptera, Satyrinae) (L): evidence for selection. Heredity 71:386–393

Hamer KC, Hill JK, Benedick S, Mustaffa N, Sherratt TN, Maryati M, Chey VK (2003) Ecology of butterflies in

natural and selectively logged forests of northern Borneo: the importance of habitat heterogeneity. J Appl Ecol 40:150–162

Hanski I (1999) Metapopulation dynamics. Oxford University Press, Oxford

Hanski I, Kuussaari M (1995) Butterfly metapopulation dynamics. In: Cappuccino N, Price PW (eds) Population dynamics: new approaches and synthesis. Academic Press, London, pp 149–171

Hardy PB, Dennis RLH (1999) The impact of urban development on butterflies within a city region. Biodivers Conserv 8:1261–1279

Harrington R, Stork NE (1995) Insects in a changing environment. Academic Press, London

Ibarra AA, Park YS, Brosse S, Reyjol Y, Lim P, Lek S (2005) Nested patterns of spatial diversity revealed for fish assemblages in a west European river. Ecol Freshw Fish 14:233–242

Ioannides YM, Overman HG (2003) Zipf's law for cities: an empirical examination. Reg Sci Urban Econ 33:127–137

IUCN (2001) International Union for Conservation of Nature. http://www.iucn.org. Accessed March 2010

Kenkel NC, Orloci L (1986) Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology 64(4):919–928

Kohonen T (1982) Analysis of a simple self-organizing process. Biol Cybern 44:135–140

Kohonen T (2001) Self-organizing map. Springer, Heidelberg

Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47:583–621

Lasne E, Bergerot B, Lek S, Laffaille P (2007) Fish zonation and indicator species for the evaluation of the ecological status of rivers: example of the Loire basin (France). River Res Appl 23:1–14

Lek S, Guegan JF (1999) Artificial neural networks as a tool in ecological modelling, an introduction. Ecol Model 120:65–73

Lek S, Scardi M, Verdonschot PFM, Park YS, Descy JP (2005) Modelling community structure in freshwater ecosystems. Springer, New York

Leon-Cortés JL, Cowley MJR, Thomas CD (2000) The distribution and decline of a widespread butterfly *Lycaena phlaeas* in a pastoral landscape. Ecol Entomol 25:285–296

Leon-Cortés JL, Pérez-Espinoza F, Marin L, Molina-Martinez A (2004) Complex habitat requirements and conservation needs of the only extant Baroniinae swallowtail butterfly. Anim Conserv 7:241–250

Magurran AE (1988) Ecological diversity and its measurement. Croom Helm, London

Manel S, Dias JM, Ormerod SJ (1999) Comparing discriminant analysis, neural networks and logistic regressions for predicting species' distribution: a case study with an Himalayan river bird. Ecol Model 120:337–347

McDonnell MJ, Pickett STA (1990) Ecosystem structure and function along urban–rural gradients: an unexploited opportunity for ecology. Ecology 71:1232–1237

McIntyre S, Barrett GW (1992) Habitat variegation, an alternative to fragmentation. Conserv Biol 6:146–147

Mennechez G, Schtickzelle N, Baguette M (2003) Metapopulation dynamics of the bog fritillary butterfly: comparison of demographic parameters and dispersal between a

continuous and a highly fragmented landscape. Landscape Ecol 18:279–291

Merckx T, Van Dyck H (2007) Habitat fragmentation affects habitat-finding ability of the speckled wood butterfly, *Pararge aegeria* L. Anim Behav 74:1029–1037

Nelson GS, Nelson SM (2001) Bird and butterfly communities associated with two types of urban riparian areas. Urban Ecosyst 5:95–108

New TR (1997) Are Lepidoptera an effective 'umbrella group' for biodiversity conservation? J Insect Conserv 1:5–12

Niell RS, Brussard PF, Murphy DD (2007) Butterfly community composition and oak woodland vegetation response to rural residential development. Landsc Urban Plan 81:235–245

Olden JD, Joy MK, Death RG (2006) Rediscovering the species in community-wide predictive modeling. Ecol Appl 16:1449–1460

Park YS, Grenouillet G, Esperance B, Lek S (2006) Stream fish assemblages and basin land cover in a river network. Sci Total Environ 365:140–153

Piscart C, Bergerot B, Laffaille P, Marmonier P (2010) Are amphipod invaders a threat to regional biodiversity? Biol Invasions 12:853–863

Pollard E, Yates TJ (1993) Monitoring butterflies for ecology and conservation. The British Butterfly Monitoring Scheme. Conservation Biology Series 1

Roy DB, Rothery P, Brereton T (2007) Reduced-effort schemes for monitoring butterfly populations. J Appl Ecol 44:993–1000

Rubinoff D, Powell JA (2004) Conservation of fragmented small populations: endemic species persistence on California's smallest channel island. Biodivers Conserv 13:2537–2550

Saarinen K, Valtonen A, Jantunen J, Saarnio S (2005) Butterflies and diurnal moths along road verges: does road type affect diversity and abundance? Biol Conserv 123:403–412

Savard JPL, Clergeau P, Mennechez G (2000) Biodiversity concepts and urban ecosystems. Landsc Urban Plan 48(3–4):131–142

Schmitt T, Rakosy L (2007) Changes of traditional agrarian landscapes and their conservation implications: a case study of butterflies in Romania. Divers Distrib 13:855–862

Somervuo P, Kohonen T (1999) Self organizing map and learning vector quantization for feature sequences. Neural Process Lett 10:151–159

Steffan-Dewenter I, Tscharntke T (2000) Butterfly community structure in fragmented habitats. Ecol Lett 3:449–456

Stevens VM, Turlure C, Baguette M (2010) A meta-analysis of dispersal in butterflies. Biol Rev (in press)

Tolman T, Lewington R (1997) Butterflies of Britain and Europe. HarperCollins Ltd., London, 320 pp

Vanreusel W, van Dyck H (2007) When functional habitat does not match vegetation types: a resource-based approach to map butterfly habitat. Biol Conserv 135:202–211

Vesanto J, Alhoniemi E (2000) Clustering of the self-organizing map. IEEE Trans Neural Netw 11:586–600

Ward JH (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc 58:238–244

Wood BC, Pullin AS (2002) Persistence of species in a fragmented urban landscape: the importance of dispersal ability and habitat availability for grassland butterflies. Biodivers Conserv 11(8):1451–1468