Ecology, 88(2), 2007, pp. 512–523

A NEW TECHNIQUE FOR ORDERING ASYMMETRICAL THREE-DIMENSIONAL DATA SETS IN ECOLOGY

SANDRINE PAVOINE,[1,5] JACQUES BLONDEL,[2] MICHEL BAGUETTE,[3,6] AND DANIEL CHESSEL[4]

[1]*Unité de Conservation des Espèces, Restauration et Suivi des Populations (UMR 5173), Muséum National d'Histoire Naturelle,*
55 Rue Buffon, 75005 Paris, France
[2]*CEFE/CNRS, 1919 Route de Mende, 34293 Montpellier Cedex 05, France*
[3]*University of Louvain, (UCL), Unité d'Ecologie et de Biogéographie, Croix du Sud 4, B-1348 Louvain-la-Neuve, Belgium*
[4]*Laboratoire de Biométrie et Biologie Évolutive (UMR 5558), CNRS, Univ. Lyon 1*
43 Bd. du 11 novembre 1918, F-69622 Villeurbanne Cedex, France

Abstract. The aim of this paper is to tackle the problem that arises from asymmetrical data cubes formed by two crossed factors fixed by the experimenter (factor A and factor B, e.g., sites and dates) and a factor which is not controlled for (the species). The entries of this cube are densities in species. We approach this kind of data by the comparison of patterns, that is to say by analyzing first the effect of factor B on the species-factor A pattern, and second the effect of factor A on the species-factor B pattern. The analysis of patterns instead of individual responses requires a correspondence analysis. We use a method we call Foucart's correspondence analysis to coordinate the correspondence analyses of several independent matrices of species × factor A (respectively B) type, corresponding to each modality of factor B (respectively A). Such coordination makes it possible to evaluate the effect of factor B (respectively A) on the species-factor A (respectively B) pattern. The results obtained by such a procedure are much more insightful than those resulting from a classical single correspondence analysis applied to the global matrix that is obtained by simply unrolling the data cube, juxtaposing for example the individual species × factor A matrices through modalities of factor B. This is because a single global correspondence analysis combines three effects of factors in a way that cannot be determined from factorial maps (factor A, factor B, and factor A × factor B interaction) whereas the applications of Foucart's correspondence analysis clearly discriminate two different issues. Using two data sets, we illustrate that this technique proves to be particularly powerful in the analyses of ecological convergence which include several distinct data sets and in the analyses of spatiotemporal variations of species distributions.

Key words: correspondence analysis; crossed-factor analysis; data cubes; multivariate analysis; ordination; partial triadic analysis; species abundances.

INTRODUCTION

The ordination of multi-arrays is a crucial question in ecology (Potvin and Travis 1993). For example, an experiment can be repeated several times in a year, leading to one array per sampling date. When an experiment is done by considering three crossed factors A × B × C (e.g., physicochemical variables × sites × observation dates), the collected data are organized to form a three-dimensional data set, often called "data cube." It consists of stacked two-dimensional data matrices as a function of a third coordinate. The rows and columns of each two-dimensional data matrix are given by two coordinates so that the data cube is defined by three coordinates called "modes." The analysis of data cubes covers a widespread field of research in which the three-mode analyses play a large part (Kroonenberg 1983, 1989*b*, Coppi and Bolasco 1989). These methods assume that the three modes represent three comparable factors that are symmetrically treated. However, ecologists face a very particular asymmetrical situation. Their data can include three very different modes: the species and two crossed factors (factor A and factor B) which define the zones and/or dates of experiment (Swaine and Greig-Smith 1980). The entries of the cube are the abundances or presences/absences of species. The two modes corresponding to the two crossed factors are fixed by the observer whereas the third mode is not controlled for because it is the list of the species included in the investigated ecosystems. We address the problems raised by this type of asymmetric data cube. There are two main objectives for analyzing these data.

The first objective concerns explanatory variables and is centered on the species: building the model of the effect factor A and factor B for each species (simple effect A or B, additive effect A + B, partial effect A given B, B given A, interaction between A and B, and so on)

Manuscript received 16 November 2005; revised 1 June 2006; accepted 13 June 2006. Corresponding Editor: N. C. Kenkel.

[5] E-mail: pavoine@mnhn.fr

[6] Present address: Muséum d'Histoire Naturelle, Département Ecologie et Gestion de la Biodiversité, CNRS-UMR 5176, 4 Avenue du Petit-Château, F-91800 Brunoy, France.

The question raised by this first objective is thus: which combination of factors explains the abundance distribution of each species? We will not tackle this first objective.

The second objective that we are addressing here is not centered on the species but on two patterns: the species-factor A pattern and the species-factor B pattern. Realizing an ordination of the species according to their response to the modalities of factor A and at the same time the ordination of the modalities of factor A according to their associated species distributions results in the revelation of the species-factor A pattern (Hill 1973). The extraction of this pattern is the objective of the correspondence analysis (CA; Greenacre 1984). For a species × factor A × factor B data cube, there are as many matrices of species × factor A type as there are modalities of factor B. Our second objective is therefore first to analyze the changes in the species-factor A pattern over the different modalities of factor B. As there are as many such matrices as there are modalities of factor B, there is an average pattern, which we call "reference pattern," and departures from this reference for each individual analysis induced by any individual modality of factor B. These departures constitute the effect of factor B on the jointed species-factor A pattern. Let us precise that factor A and factor B play a symmetric role, so one can be interested on the effect factor B on the species-factor A pattern and/or on the effect factor A on the species-factor B pattern. A data cube results thus in two separated analyses of patterns, which will be treated with the same method.

By its very nature, this second objective is explanatory, while constituting a first step toward the challenging question, which has not been solved so far, of the further development of models explaining patterns as functions of combination of factors. For this second objective, the species are not the central objects of the analyses, but constitute markers, as do the modalities of factor A or B. When the first objective is used to draw inferential conclusions, it requires assumptions about the shape of the abundance distribution of the species, and formulation of hypotheses. On the contrary, the second objective, which belongs to the family of explanatory multivariate analyses, does not rely on any assumption. We can quote Calenge and Dufour (2006) by saying that "one of the most positive properties of eigenanalyses is their wide applicability. Indeed these analyses do not rely on any distributional assumption, and do not make any hypothesis about the data." So the only requirement for the data we are dealing with is that they must be organized in a species × factor A × factor B type.

The purpose of this paper is to find a reference species-factor A (respectively B) pattern and to describe the departures from this reference according to each modality of factor B (respectively A). As a species-factor A (respectively B) pattern is revealed by a CA, and as there are as many individual CA as there are modalities of factor B (respectively A), for reaching this purpose, the main difficulty is to coordinate several CA. Similar situations have been investigated by Foucart (1978) in social sciences (for example, age group × region × year with population density as entry). The main difficulty in coordinating the CA of K matrices is due to the fact that each matrix is characterized by its own weightings of rows and columns (ter Braak and Schaffers 2003). The rationale of Foucart (1978) was to work first on an average matrix (which will be the "reference matrix") and then to look at the differences between the global pattern resulting from the reference matrix and the individual pattern given by each of the K constitutive matrices. We call this method "FCA" for Foucart's correspondence analysis. Foucart published it in the context of social sciences and it is not very well known presumably because the French journal where it was published ceased its activity several years ago. Although published in 1978, this method has not yet appeared in ecology. We present it and highlight its limitations and its promising properties for analyzing complex ecological data.

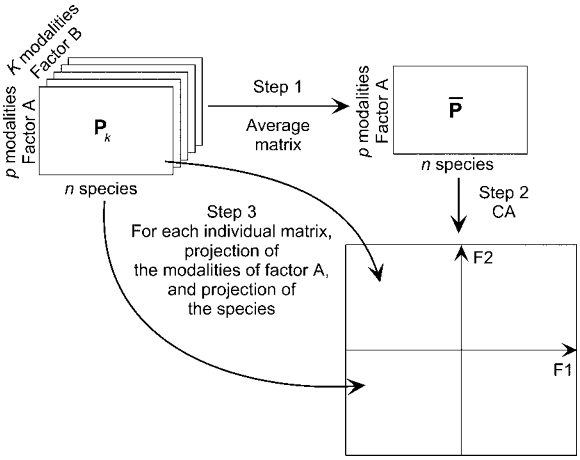

FIG. 1. Schema summing up Foucart's correspondence analysis. See *Methods: Foucart's correspondence analysis* for definitions of variables.

METHODS

All computations were carried out using R (Ihaka and Gentleman 1996), with personal routines available in the ade4 package (Chessel et al. 2004; see Appendix A and B).

Foucart's correspondence analysis

The procedure is summarized in Fig. 1. Consider an asymmetrical data cube formed by the three modes species × factor A × factor B. Consider that there are n species, p modalities for factor A, and K modalities for factor B. The FCA can be processed twice for analyzing first the changes in the species-factor A through the modalities of factor B and second the changes in the species-factor B pattern through the modalities of factor

A. We describe here the analysis of the changes in the species-factor A through the modalities of factor B, knowing that the second analysis will be treated with exactly the same process.

For the kth modality of factor B, there is a matrix $\mathbf{A}_{(k)} = [a_{ij(k)}]$ where $1 \leq i \leq n$, $1 \leq j \leq p$, and where $a_{ij(k)}$ denotes the abundance of a species. Each matrix $\mathbf{A}_{(k)}$, for $1 \leq k \leq K$, is first changed to be expressed as a percentage: $\mathbf{P}(k) = [a_{ij(k)}/a_{\cdot\cdot(k)}] = [p_{ij(k)}]$, where $a_{\cdot\cdot(k)}$ is the sum of the terms in $\mathbf{A}_{(k)}$. The FCA coordinates the separate analyses of the K matrices using an average matrix, which we call "reference matrix":

$$\overline{\mathbf{P}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{P}(k) = \left[\frac{1}{K}\sum_{k=1}^{K}p_{ij(k)}\right]_{\substack{1\leq i\leq n\\ 1\leq j\leq p}} = \left[\overline{p}_{ij}\right]_{\substack{1\leq i\leq n\\ 1\leq j\leq p}}.$$

The application of CA to $\overline{\mathbf{P}}$ provides a reference species × factor A pattern. The marginal weights in the CA of $\overline{\mathbf{P}}$ are defined by the diagonal matrices $\mathbf{D}_p = \mathrm{diag}(\overline{p}_{\cdot 1}, \ldots, \overline{p}_{\cdot j}, \ldots, \overline{p}_{\cdot p})$, and $\mathbf{D}_n = \mathrm{diag}(\overline{p}_{1\cdot}, \ldots, \overline{p}_{i\cdot}, \ldots, \overline{p}_{n\cdot})$, where $\overline{p}_{\cdot j} = \Sigma_{i=1}^{n}\, \overline{p}_{ij}$ and $\overline{p}_{i\cdot} = \Sigma_{j=1}^{p}\, \overline{p}_{ij}$. The core of the CA of $\overline{\mathbf{P}}$ is the following diagonalization:

$$\begin{aligned}\mathbf{D}_p^{1/2}(\mathbf{D}_p^{-1}\overline{\mathbf{P}}^{\,t}\mathbf{D}_n^{-1} - 1_{pn})\mathbf{D}_n(\mathbf{D}_n^{-1}\overline{\mathbf{P}}\mathbf{D}_p^{-1} - 1_{np})\mathbf{D}_p^{1/2}\\ = \mathbf{U}_m\Lambda_m\mathbf{D}_m^{t}.\end{aligned}$$

$\mathbf{\Lambda}_m$ contains m positive eigenvalues, and $\mathbf{U}_m$ contains m orthonormal eigenvectors. According to the reference pattern, the coordinates for the modalities of factor A are given in the rows of matrix $\mathbf{D}_p^{-1/2}\mathbf{U}_m\Lambda_m^{1/2}$, while the coordinates of the species are in the rows of matrix $\overline{\mathbf{P}}\mathbf{D}_p^{1/2}\mathbf{U}_m$.

The deviations between the reference species-factor A pattern and the pattern generated by the kth modality of factor B is obtained by projecting the rows and columns of matrix $\mathbf{P}_{(k)}$ as additional elements onto the reference analysis. Let us define the two matrices $\mathbf{P}_{(k)}^{\mathrm{col}} = [p_{ij(k)}/p_{\cdot j(k)}]$, and $\mathbf{P}_{(k)}^{\mathrm{row}} = [p_{ij(k)}/p_{i\cdot(k)}]$, where $p_{\cdot j(k)} = \Sigma_{i=1}^{n}\, p_{ij(k)}$ and $p_{i\cdot(k)} = \Sigma_{j=1}^{p}\, p_{ij(k)}$. They are expressed as percentages per column and per row, respectively. According to the kth modality of factor B, the coordinates for the modalities of factor A are then given in the rows of matrix $(\mathbf{P}_{(k)}^{\mathrm{col}})^t\,\overline{\mathbf{P}}\mathbf{D}_p^{1/2}\mathbf{U}_m\Lambda_m^{-1/2}$, and the coordinates for the species in the rows of matrix $\mathbf{P}_{(k)}^{\mathrm{row}}\mathbf{D}_p^{-1/2}\mathbf{U}_m$.

Comparison with existing methods for analyzing such asymmetrical data cubes

Global correspondence analysis.—The first approach used by Blondel and Farré (1988) for analyzing such asymmetrical data cubes consisted in applying CA to a single matrix which included all the data: by unrolling a cube, they obtained a single large matrix of species × relevé type, where the number of relevés is equal to the number of modalities for factor A multiplied by the number of modalities for factor B. We call this method "global CA." Unfortunately, by choosing this analysis, they failed to take into account a piece of information that they knew: the deliberate presence of two crossed factors in the sampling scheme. The consequence is that by mixing the two crossed factors, the global CA rendered the interpretation of its axes very difficult. They had therefore to arbitrarily ignore some factorial axes, focusing on others which were coherent with known ecological hypotheses, but yet expressed a weaker amount of variability (Blondel and Farré 1988).

The multi-array analyses.—The methods devoted to the analyses of several matrices, also called multi-array analyses can be differentiated according to two sets of criteria. The first one deals with the core analysis. Most of the time, two choices are possible: the principal component analysis (PCA), which deals with typologies of responses, and the correspondence analysis (CA), which concerns simultaneous ordinations of rows and columns. Other alternatives include the analyses of distances, the procrustes approaches (Gower 1975), and so on. The second set of criteria is the number of modes in common among the matrices. Three choices are possible: one mode (the rows), one mode (the columns), and two modes (the rows and the columns). As far as we now, Foucart's analysis is the only method that analyzes matrices sharing their two modes (same rows and same columns), which is the characteristic of data cubes, and takes the CA as its core analysis.

The method most related to the FCA is the partial triadic analysis (PTA; Tucker 1966, Thioulouse and Chessel 1987, Kroonenberg 1989*a*, Thioulouse et al. 2004). It aims at defining the common pattern given by K matrices, $X_1, \ldots, X_k, \ldots, X_K$, with same rows and same columns, but its core analysis is a PCA. Three points radically differ between FCA and PTA: (1) in PTA, all the matrices must have the same column and row weights, whereas each matrix has its own marginal weighting in FCA; (2) in PTA, these matrices must be normalized, whereas, in FCA, they are expressed as percentages; (3) in PTA, the reference matrix is obtained by a weighted mean, whereas FCA uses a standard mean with uniform weights. This third point is the first and main step of the PTA: the definition of an additional weight for each of the individual matrices. It aims at providing an average matrix $\mathbf{Y} = \Sigma_{k=1}^{K}\alpha_k\mathbf{X}_k$, which optimally captures the similarities among the individual matrices. The weights ($\alpha_1, \ldots, \alpha_k, \ldots, \alpha_K$) are chosen so that (1) $\Sigma_{k=1}^{K}\alpha_k^2 = 1$ and (2) they maximize the average covariance among pairs of matrices. Once they have been obtained, matrix $\mathbf{Y}$ is analyzed by principal component analysis (PCA) and the rows and columns of the individual matrices are projected onto the analysis as supplementary individuals and supplementary variables, respectively.

The essence of the differences between PTA and FCA is the difference between PCA, principal component analysis, and CA, correspondence analysis. In the PTA, the basic analysis is the PCA which works on an ordination of species due to their response to modalities of factor A. The ordination of each species is a response to the modalities of factor A relegated to a role of an indicator. This is not what we were looking for, because, as said in the introduction, our objective was to study the symmetric relationship between species and modalities of factor A, the so-called species-factor A pattern. Contrary to the PTA, the FCA is based on the CA, which results in a simultaneous ordination of species and modalities of factor A, therefore on the species-factor A pattern. Consequently, the FCA analyzes the variability of global patterns instead of the variability of individual responses induced by some factor acting as an indicator. The methodological difficulty resulting from the choice for CA instead of PCA is the non-uniform marginal weightings of the CA because they prevent us to use a method similar to the PTA which need similar weighting scheme for all individual matrices. Nevertheless, the solution chosen by Foucart, treating all the individual matrices with an equal weight, although more simple than PTA, appears to be an efficient alternative.

Another related ordination method is the multiple co-inertia analysis (MCoA; Chessel and Hanafi 1996, Dray et al. 2003). However, unlike the FCA and PTA, the MCoA is not devoted to the analysis of data cube. It aims to coordinate the separated analysis of *K* matrices sharing only one mode: the rows.

Case Studies

Convergence of bird assemblages in European and Mediterranean forests

Blondel and Farré (1988) suggested the existence of a regular convergence of bird assemblages in the last climactic forest stage along European habitat gradients. Although the reality of convergence is not disputable, the global CA used by Blondel and Farré (1988) hampered a detailed description of this convergence. We revisit this study using FCA because this new statistical procedure allows for a precise description of this convergence and brings new interesting results.

In their original paper, Blondel and Farré (1988) analyzed four local regions, but data from five local regions have been selected to carry out this study. Two of them were located in the Mediterranean region: Provence (southern France; Blondel 1979) and the island of Corsica (France; Blondel 1979). The others two were located in the medioeuropean region: Burgundy (central France; Ferry and Frochot 1970) and Poland (Glowacinski 1975). In order to widen the discussion in the context of the whole Mediterranean part of the western Palearctic, we added another region studied in Algeria by Benyacoub (1993) following exactly the same rationale and the same methods. In each local region, a habitat gradient has been conventionally divided into six seral stages in such a way that all five selected habitat gradients match one another in respect to the number and pattern of habitats. Their selection has been made using classical criteria of habitat pattern, especially the complexity and height of the vegetation (ranging from low bushy vegetation, height <1 m, stage 1, to forests with trees at least 20 m high, stage 6). This data set is given in the Supplement.

FCA allows us to study (1) the differences in the species-vegetation stage patterns among the five local regions; and (2) the differences in the species-local region patterns among the vegetation stages. The first point evaluates whether the ecological gradients defined within the five local regions are congruent. The second point gives an indication of whether the differences and similarities among the five local regions remain steady over the six vegetation stages. This second point also provides a description of the differences and similarities among the five local regions, therefore treating the question of ecological convergence. We considered both questions.

Temporal stability of impoverished calcareous grassland butterfly communities

This data set was collected in the Calestienne landscape straddling Northern France and Southern Belgium and characterized by a chain of chalky hills emerging from a clayish flood plain. In the beginning of the 20th century, those chalky hills were completely covered by calcareous grasslands, maintained by extensive agropastoral activities. Changes in agricultural practices, reforestation, and urbanization have led to the dramatic loss of calcareous grasslands. Nowadays, the remnants are small fragments less than 1.5 ha on average. Polus et al. (2006) showed previously that this dramatic habitat loss and fragmentation had a strong impact on butterfly communities, which are now significantly impoverished by the loss of specialist species. Here our aim is to test whether those impoverished butterfly communities remain stable from year to year. Theory predicts that small populations should incur higher extinction risks, which should generate, in turn, a high species turnover in local communities. We selected a total of 15 chalk grassland fragments located on different hills in a subset of the landscape. Although all these sites were calcareous grasslands, vegetation types vary from Xero- to Meso-Brometum because of variation in topography, soil depth, and sun exposure, both between and within sites.

Each site was visited once every two weeks from April to September in 2003, 2004, and 2005. However, we use here as basic data in the analyses the number of butterflies observed on each study site, summed over sampling periods for each year. In each study site, all butterflies were recorded within a 10 m wide stripe along transect routes in suitable weather conditions (see Pollard [1977] for line transect methods). Transect routes were designed a priori, aiming at (1) being representative of the site area (i.e., more routes or longer routes in larger sites) and (2) covering the heterogeneity of each site. Butterflies were identified in flight, and if this was not possible they were caught with a net and released immediately after identification. This procedure provides a reliable picture of species richness: exhaustiveness tests using rarefaction curves revealed that we had to sample at least twice as many sites to add only one species to the metacommunity (i.e., the total list of species recorded in the 15 sites; results not shown). The number of butterflies observed along transects also provides a reliable estimate of local population sizes, as shown by its significant correlation with daily abundances as estimated from capture–mark–recapture studies (results not shown). A grand total of 68 species has been collected, which is impressive in comparison to the 117 species constituting the total regional richness in southern Belgium. The grand total of individuals was double in the first year (~9800 individuals) in comparison to 2004 and 2005, reflecting the impact of the extreme hot and dry summer of 2003 in Europe.

FCA allows us to study the differences in the site–species pattern during the three years and the differences in the year–species pattern among the 15 sites. The first point evaluates whether the pattern of differences in species composition among sites remains steady with time. The second point gives an indication of whether the dynamic of species composition is similar in all the sites. We considered both questions.

RESULTS

Concerning the choice of a number of principal axes to be kept for the display of factorial maps, several rules have been defined (e.g., Peres-Neto et al. 2005). For our illustration purpose, we choose to retain all the axes when their number was less or equal to three, keeping then the whole information contained in the data. When the first axis expresses a very majority of the variability, we selected two axes. In addition, for each analysis, we calculated the absolute contribution of each species and each modality of factor A and B on the result of the analysis (Appendix B).

Convergence of bird assemblages in European and Mediterranean forests

FCA emphasizing the differences in the species–vegetation stage pattern between the local regions.—The habitat gradients of the distinct local regions reproduce the same pattern fairly well with the species point clouds having similar shapes (Fig. 2). Nevertheless, bushy habitats are closer to each other than on average in Algeria and conversely they differ more in Provence.

FCA examining the changes in the species–local region pattern along the vegetation stages.—F1 discriminates the medioeuropean from the Mediterranean region while F2 expresses differences between local regions within each of the two global regions. On the whole, the pattern of the differences between local regions remains unchanged in the six vegetation stages but the magnitude of differences gradually decreases as the complexity of the vegetation increases (Figs. 3 and 4). In forested habitats, the four local regions become very similar.

Temporal stability of impoverished calcareous grassland butterfly communities

FCA examining the changes in species–sites pattern along time.—A common pattern for the three years is revealed, with some discrepancies from this pattern (Fig. 5). The most stable pattern highlighted by this graph is the opposition between Fondry des chiens, Roche à lomme, Tienne aux pauquis3, and Tienne aux pauquis4 on the one hand and Tienne du lion, and Tienne du nord on the other hand.

FCA examining the changes species–year patterns among sites.—For the majority of the sites, 2003 was the most distinct year (Fig. 6). There remain, however, differences among sites: some are characterized by high temporal variations (e.g., Abannets1), whereas others were more stable than average (e.g., Fondry des chiens and Tienne Breumont).

DISCUSSION

Convergence of bird assemblages in European and Mediterranean forests

The first data set allows us to highlight two crucial points that are detailed in the following paragraphs. The first point is that, when data are arranged according to a strong pattern, illustrated here by a clear ecological convergence, this strong pattern will appear from a number of more or less appropriate analyses. However the precise description of this strong pattern requires a very suitable analysis, especially one eliminating confounding factors. The second point is that, in addition to this strong structure, other pieces of information requiring a more subtle analysis may be contained in the data set. This is also where the choice of a method suitable for the data and the objective of the study is crucial.

Toward a more precise description of the convergence.—Concerning the first point, Blondel and Farré's (1988) data set is structured by an ecological convergence, which is a strong emerging pattern: bird communities in the five local regions studied are similar in forested habitats, although they differ in open habitats. Blondel and Farré (1988) effectively obtained this pattern from a global CA, but they had to arbitrarily ignore some factorial axes because they did not provide an ordination of bird communities that could be explained from ecological grounds. A precise description of the convergence was therefore impossible.

On the other hand, the second FCA that we computed (Figs. 3 and 4) is directly devoted to the analysis of a potential convergence, because it does not rely on the confounding effects of the variations among vegetation stages which are not among but within local regions. Consequently, it highlights the convergence with all its factorial axes and provides a spectacular graphical description of this convergence (highlighted by Fig. 4).

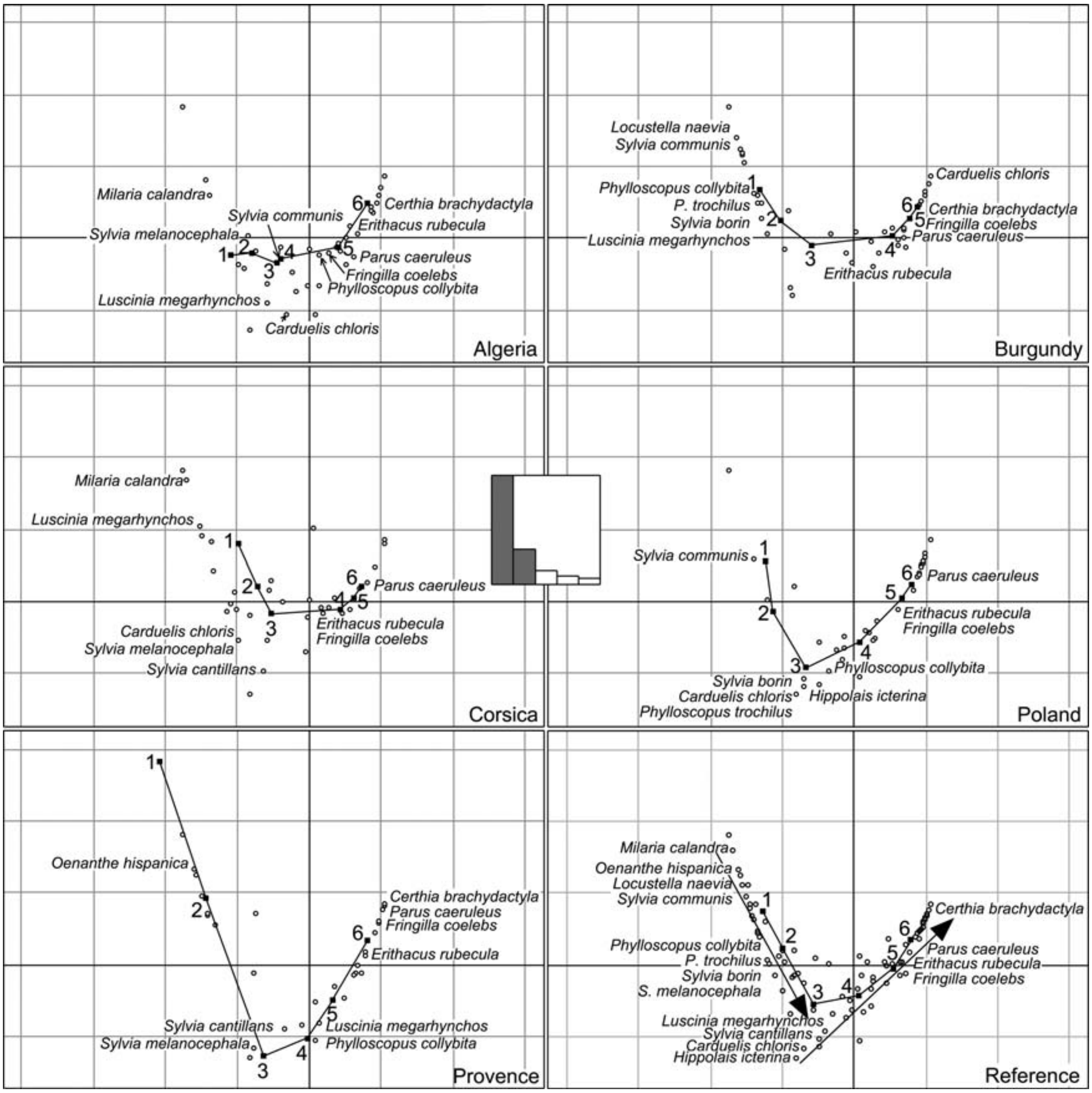

FIG. 2. Display of vegetation stages and species per local region on the bivariate space F1 × F2 of the FCA (Foucart's correspondence analysis) applied to five matrices (corresponding to the five local regions) with stages as columns and species as rows (F1, 63.5%; F2, 20.1% of the variation). Solid lines link the vegetation stages of an ecological gradient from bushy to pre-forested habitats. The species contributing the most to the reference analysis are indicated (see Appendix B). The last panel shows the reference. Arrows highlight the shape of species points, which is strongly conserved in all local regions. The eigenvalue bar plot is given in the middle of this figure. In each panel, a grid indicates the scale; the length of a square side is equal to 1.

The first new result provided by this analysis is that the pattern of differences between local regions is maintained from stage to stage while it is the amount of these differences which decreases along habitat gradients.

The convergence highlighted by these analyses is the consequence of the ecological trajectories of the bird communities across the five local regions which result in a homogenization of bird communities in the old forests, independently of the area considered. The influence of the climatic history of the Eurasian continent on the ecological characteristics and the geographical distribution of the fauna in Europe has been discussed by Steinbacher (1948), Moreau (1954), and Blondel and colleagues (Blondel 1985, 1986, Blondel et al. 1988, Blondel and Farré 1988). We can summarize the result of these discussions as follows. During the Pleistocene, while the Mediterranean shrublands subjected to local environmental conditions could have allowed the speciation processes and survival of endemic species, there has been very little speciation within the Mediterranean forests. Before human impact, the Mediterranean forested region was larger. The diversity of conditions of temperature and moisture and the diversity of the geotopography of the Mediterranean

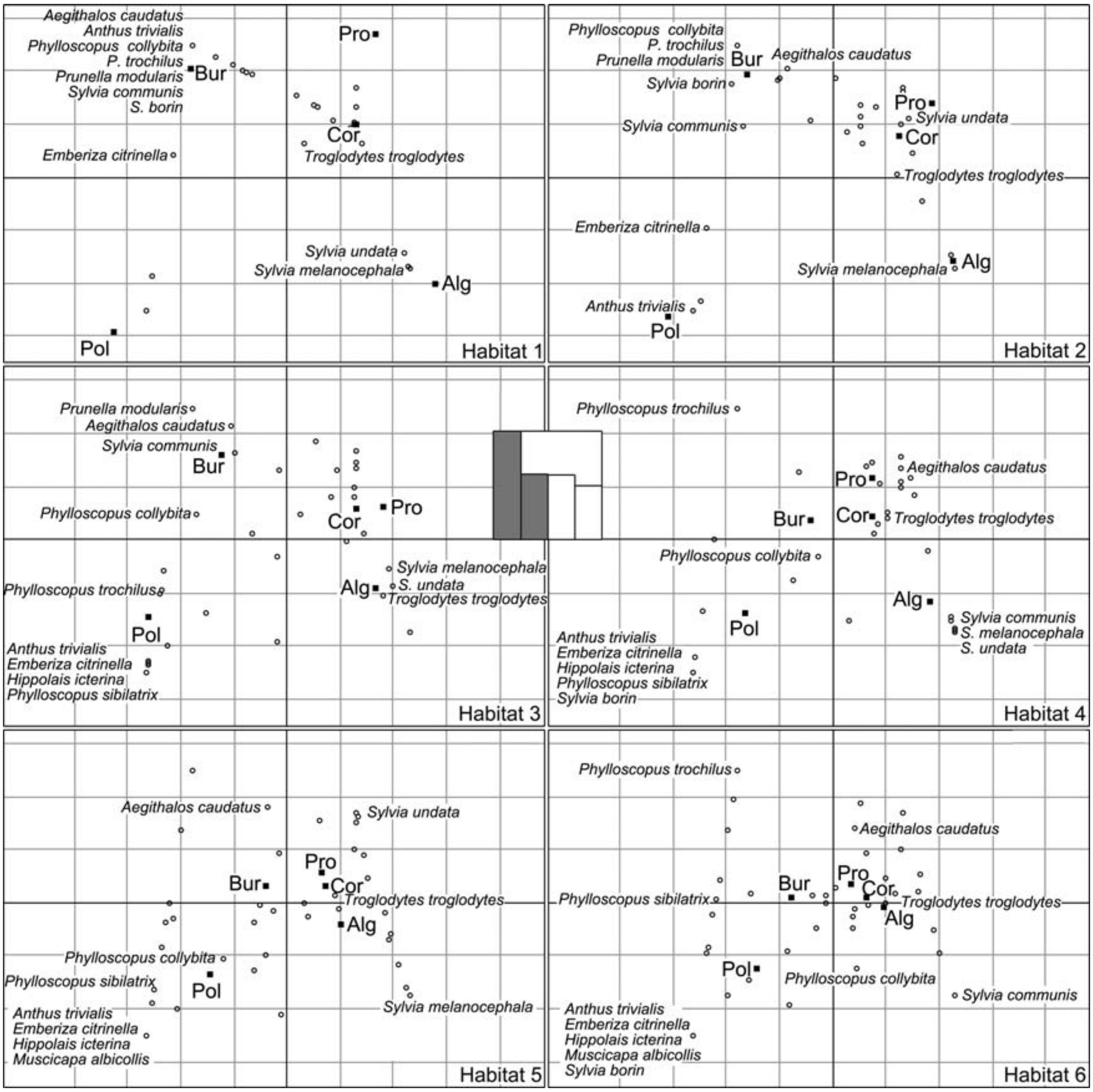

FIG. 3. Display of local regions and species per vegetation stage on the bivariate space F1 × F2 of the FCA applied to six matrices (corresponding to the six vegetation stages) with local regions as columns and species as rows (F1, 37%; F2, 22.4% of the variation). Labels for local regions are: Alg, Algeria; Bur, Burgundy; Cor, Corsica; Pol, Poland; and Pro, Provence. The species contributing the most to the reference analysis are indicated (see Appendix B). The eigenvalue barplot is given in the middle of this figure. In each panel, a grid indicates the scale; the length of a square side is equal to 0.5.

region allowed the coexistence, on a regional scale, of all the faunal types of Europe. Consequently, all the fauna of the forest type of Europe found refuge in the Mediterranean region during the most severe climatic phases. During interglaciation periods, the Mediterranean forested region was dominated by broad-leaved trees (not evergreen trees as today). As a result, after a severe climatic phase, the faunal elements expanded northward without leaving the Mediterranean region, which constituted a favorable habitat. This succession of changes during alternating expansions and contractions of glacial and arctic conditions led to similar bird fauna in central, northern, and Mediterranean forests.

Additional important patterns contained in the data set.—Concerning the second point, we analyzed by FCA first (Fig. 2) the species–vegetation stage patterns within local regions and, second (Figs. 3 and 4), the species–local region patterns within vegetation stages.

The first analysis revealed that, in open habitats, Provence has the most distinct species–stage pattern, with a rupture between stages 1 and 2 and stages 3 to 6. The maquis of Provence has very particular characteristics: 2–2.50 m high of very dense vegetation full of thorns (*Quercus coccifera*, *Ulex parviflorus*). Consequently, in this region, it proved very difficult to find

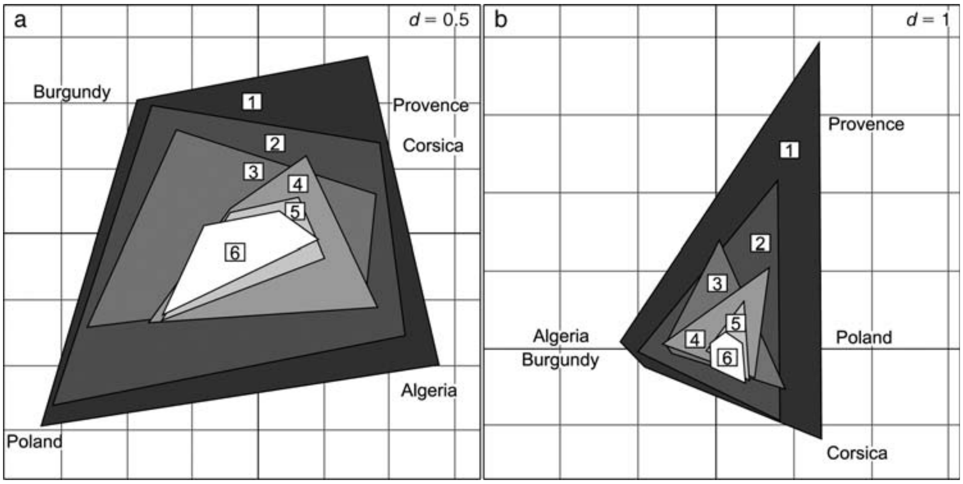

FIG. 4. Convex hull of the five local regions positioned per stage on the bivariate spaces (a) F1 × F2 and (b) F3 × F4 of the FCA applied to the six matrices (corresponding to the six vegetation stages) with local regions as columns and species as rows. Numbers indicate the stage (from 1 = bushy to 6 = forested). This figure shows how much Foucart's analysis brings an appropriate answer to the ecological question of convergence. It highlights the amplitude of the avifauna–local region pattern as a function of the habitat. We can now see how much the six individual matrices express the same pattern in different ways. These differences are mainly due to a variability of the degree of differences among local regions as a function of the habitat height. In each panel, a grid indicates the scale; the length of a square side is indicated by the *d* value.

an ecological gradient which perfectly matched the gradients of the other four regions.

With the second analysis, the FCA revealed that, whatever the vegetation stage, but especially in stage 1, the local region of Poland has the most distinct species composition. With only three scarce observed species, habitat 1 of Poland very much differs from its counterpart in Burgundy that includes 25 species, several of which are abundant (up to 99 breeding pairs/km^2). Compared to the other gradient, there is in Poland a shift of the species from earlier to older stages (1988), because, the Polish vegetation gradient keeps in the last forested stage some habitat patches of early stages.

To sum up, whereas the combination of the two crossed factors in global CA hampered a detailed description of the ecological convergence, the performance of the two complementary FCA clearly emphasized three main patterns in the data set: (1) the existence of a strong convergence of bird faunas in forested habitats; (2) the particular landscape of the maquis of Provence; and (3) the shift of the species in Poland from earlier to older stages.

Temporal stability of impoverished calcareous grassland butterfly communities

Concerning the second data set, the FCA (Fig. 5) revealed a stable site–species pattern corresponding to between-site differences in resource availability for caterpillars (axis 1) and adult butterflies (axis 2). The first axis opposes two groups of sites corresponding to the distinction between calcareous grasslands on dry, open slopes with superficial soil on the one hand (Fondry des chiens and related sites) and slightly wetter and shadier sites on plateaus with a deeper soil on the other hand (Tienne du lion and Tienne du nord). Those slight differences in ecological conditions had strong effects on the vegetation colonizing such sites, which was dominated mainly by small flowering thermophilous dicots in the first group, whereas graminaceaes dominated the second group. Vegetation of those two groups belonged therefore to two different phytosociological groups, *Xerobromion* and *Mesobromion* respectively (e.g., Rieley and Page 1990). According to differences in vegetation conditions, the composition of butterfly communities markedly differed in these two groups. Species with the highest contribution to the first axis were either those that fed at the larval stages on flowering dicots and that are therefore related to xerothermic calcareous grasslands, like *Colias alfacariensis*, *Cupido minimus*, and *Lysandra coridon*, or those whose host plants were graminaceaes like *Aphantopus hyperanthus*, *Coenonympha arcania*, or *Melanargia galathea*. Sites were distributed on axis 2 along a gradient of vegetation structure, from calcareous grasslands crisscrossed by bare soil to more forested areas. This distribution corresponds nicely to differences in adult butterfly thermal ecology, with on the one hand rapid flyers like *Lasiommata maera* or *L. megera* requiring sunny places of bare soils to bask and quickly warm up and on the other hand species with a slower flight, usually perching on forest edges and hedgerows like *Maniola jurtina* or *Pyronia tithonus*.

Accordingly, butterfly communities differ among sites due to difference in ecological conditions and site structures. Besides, this analysis showed differential community dynamics among sites (Fig. 6). First, the analysis of the reference situation comparing the three sampling years revealed that the exceptional dry and hot

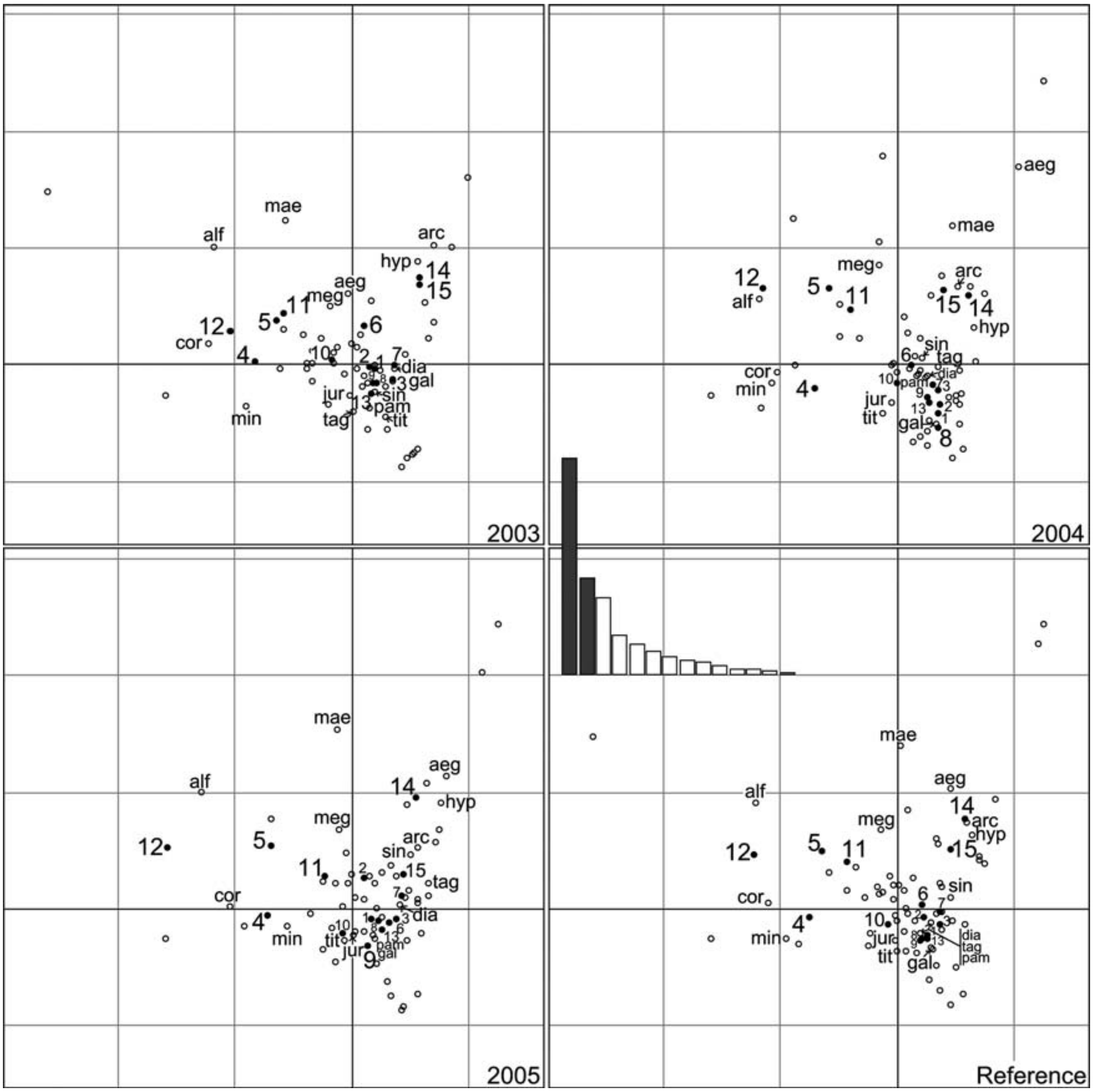

FIG. 5. Display of sites and species per year on the bivariate space F1 × F2 of the FCA applied to three matrices (corresponding to the three years) with sites as columns and species as rows (F1, 39%; F2, 17% of the variation). Sites are positioned by solid circles and numbers according to the following code: 1, Abannets1; 2, Abannets2; 3, Chalaine; 4, Fondry des chiens; 5, Roche a lomme; 6, Roche trouee; 7, Rosiere; 8, Tienne aux boulis; 9, Tienne aux pauquis1; 10, Tienne aux pauquis2; 11, Tienne aux pauquis3; 12, Tienne aux pauquis4; 13, Tienne breumont; 14, Tienne du lion; 15, Tienne du nord. The eigenvalue bar plot is given in the middle of this figure. The species contributing the most to the reference analysis (see Appendix B) are indicated by codes: aeg, *Pararge aegeria*; alf, *Colias alfacariensis*; arc, *Coenonympha arcania*; cor, *Lysandra coridon*; dia, *Clossiana dia*; gal, *Melanargia galathea*; hyp, *Aphantopus hyperanthus*; jur, *Maniola jurtina*; mae, *Lasiommata maera*; meg, *L. megera*; min, *Cupido minimus*; pam, *Coenonympha pamphilus*; sin, *Leptidea sinapis*; tag, *Erynnis tages*; tit, *Pyronia tithonus*. In each panel, a grid indicates the scale; the length of a square side is equal to 1.

2003 summer in Europe was extremely favourable to ectotherms like butterflies. Secondly, this analysis indicated that community dynamics were not correlated among sites, but contrarily to our expectation, the effect of site area was not evident. Two other different, nonexclusive, explanations may be suggested to account for this variation. Local populations of butterflies typically strongly fluctuate from one generation to the next, which might often be due to density dependence (Baguette and Schtickzelle 2006). Moreover, there is usually a weak synchronicity in dynamics among local populations of specialist species (e.g., Schtickzelle and Baguette 2004). Admittedly, the abundance of species in local communities is expected to change from year to year, which might generate the observed pattern. A striking example of this fluctuation in abundance is provided by two migrant species, *Aglais urticae* and *Vanessa cardui*, whose total numbers drop from 10 to 1

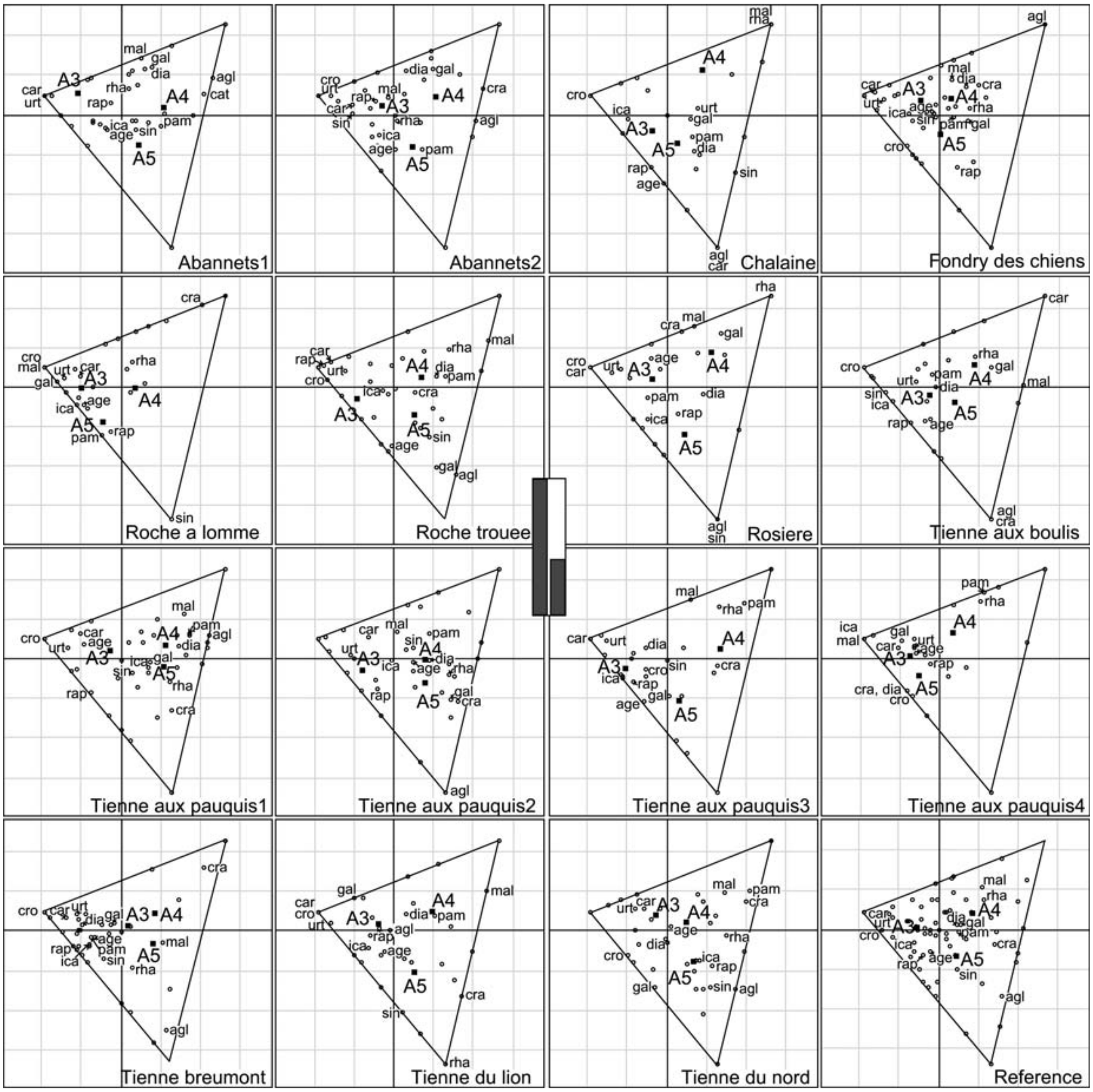

FIG. 6. Display of sites and species per year on the bivariate space F1 × F2 of the FCA applied to 15 matrices (corresponding to the 15 sites) with years as columns and species as rows (F1, 71%; F2, 29% of the variation). Because three years are under study, the whole information is concentrated on a two-dimensional map. For the same reason, the scatters of species points have a triangular shape. The apexes indicate species observed during one unique year (2003 on the left, 2004 at the top right, 2005 at the bottom right), while points on the triangle sides indicate species observed during two out of the three years, and points in the triangle indicates species observed during the three years. The eigenvalue bar plot is given in the middle of this figure. The species contributing the most to the reference analysis (see Appendix B) are indicated by codes: age, *Aricia agestis*; agl, *Argynnis aglaja*; car, *Cynthia cardui*; cra, *Aporia crataegi*; cro, *Colias crocea*; dia, *Clossiana dia*; gal, *Melanargia galathea*; ica, *Polyommatus icarus*; mal, *Pyrgus malvae*; pam, *Coenonympha pamphilus*; rap, *Pieris rapae*; rha, *Gonepteryx rhamni*; sin, *Leptidea sinapis*; urt, *Aglais urticae*. In each panel, a grid indicates the scale; the length of a square side is equal to 0.5.

between 2003 and 2004 for the former and from 20 to 1 between 2003 and 2005 for the latter.

Besides this stochastic variation, local communities may change in a more deterministic way along the ecological succession. Calcareous grasslands are the pioneer stage of a succession ending with the climacic forests. Historically blocked at this stage during centuries by sheep and goat grazing, they are now rapidly returning to shrubs and forests without management practices. Observed changes in butterfly community structure and composition might at least partly reflect this effect. For instance, changes in the Abannets1 community correspond to the decrease of grassland specialists and the increase of shrub- and forest-related species.

Properties of FCA regarding arch effects and rare species

Because FCA is based on the scheme of the correspondence analysis, it has same common properties as the correspondence analysis, for example concerning arch effects and the excessive influence of rare species.

Note that in Fig. 3 the clouds of points have a horseshoe shape. A fundamental criticism of the CA is that it bends gradients onto two dimensions (e.g., Guttman effect, Horseshoe or arch effect). These clouds of points which have a horseshoe shape are often taken for artificial products of the method, and some people eliminate them (detrending). These shapes can be artificial but they can also reveal some particular properties in the data set, which is the case in our study. In a study on bird communities Lebreton et al. (1988) demonstrated, in these European regions, the presence of two articulated sub-gradients: a subgradient in the open areas and a subgradient in the forested areas. Our results are congruent with those obtained by Lebreton et al. (1988: 279) in the same basic regions. The horseshoe shape is thus due to the presence of two sub-gradients.

A discussion about the influence of rare species in ordination methods can be found in Legendre and Gallagher (2001), who proposed solutions for adjusting the weights given to rare and non-rare species, by means of appropriate distance metrics. The distance metric used in CA is the chi-square metric, which gives high weight to the rarest species.

Therefore rare species in FCA can have a large influence on the analysis. If the data present a strong pattern, rare species won't prevent the determination of this pattern. If, on the contrary, the pattern is weak, rare species will have an excessive importance in the analysis. However, as Legendre and Gallagher highlighted, "for the analysis of community gradients, it does not matter that an analysis gives high weights to the rare species when the end-product is simply a reduced-space ordination diagram. The weights given to rare species do matter, however, when the end-product is a test of significance of the relationship between species composition and a set of explanatory variables." In addition, fortunately, if one of the matrices is largely composed of rare taxa, the prevalence of rare taxa in one matrix will not affect the analysis, because the ordination process is applied to the average matrix.

Brief conclusion

In conclusion, the Foucart's method is mathematically simple and proves to be particularly effective in observing the separated effects of two crossed factors on the species composition of a flora or fauna. It will particularly prove to be very useful and attractive for spatiotemporal analyses.

ACKNOWLEDGMENTS

The authors thank the three anonymous referees for their helpful comments. They all contributed to improve the presentation of this paper. The second data set was mainly collected by Emmanuelle Polus, in the frame of the BIOCORE project funded by BELSPO (Belgian Science Policy Office, federal government), grant PADDII EV10-26B to M. Baguette.

LITERATURE CITED

Baguette, M., and N. Schtickzelle. 2006. Negative relationship between dispersal distance and demography in butterfly metapopulations. Ecology 87:648–654.

Benyacoub, S. 1993. Ecologie de l'avifaune forestière nicheuse de la région d'El Kala (Nord-Est algérien). Dissertation. University of Burgundy, Dijon, France.

Blondel, J. 1979. Biogéographie et écologie. Masson, Paris, France.

Blondel, J. 1985. Historical and ecological evidence on the development of Mediterranean avifaunas. Pages 373–386 *in* Acta XVIII Congressus Internationalis Ornithologici. Moscow, Russia.

Blondel, J. 1986. Biogéographie évolutive. Masson, Paris, France.

Blondel, J., D. Chessel, and B. Frochot. 1988. Bird species impoverishment, niche expansion, and density inflation in Mediterranean island habitats. Ecology 69:1899–1917.

Blondel, J., and H. Farré. 1988. The convergent trajectories of bird communities along ecological successions in European forests. Œcologia (Berlin) 75:83–93.

Calenge, C., and A. B. Dufour. 2006. Eigenanalysis of selection ratios from animal radio-tracking data. Ecology 87:2349–2355.

Chessel, D., A. B. Dufour, and J. Thioulouse. 2004. The ade4 package. I. One-table methods. R News 4:5–10.

Chessel, D., and M. Hanafi. 1996. Analyses de la co-inertie de K nuages de points. Revue de Statistique Appliquée 44: 35–60.

Coppi, R., and S. E. Bolasco. 1989. Multiway data analysis. Elsevier Science Publishers B.V., North-Holland, The Netherlands.

Dray, S., D. Chessel, and J. Thioulouse. 2003. Co-inertia analysis and the linking of ecological data tables. Ecology 84: 3078–3089.

Ferry, C., and B. Frochot. 1970. L'avifaune nidificatrice d'une forêt de Chênes pédonculés en Bourgogne: étude de deux successions écologiques. La Terre et la Vie (Revue d'Ecologie) 24:153–250.

Foucart, T. 1978. Sur les suites de tableaux de contingence indexés par le temps. Statistique et Analyse des données 2: 67–84.

Glowacinski, Z. 1975. Succession of bird communities in the Niepolomice Forest (Southern Poland). Ekologia Polska 23: 231–263.

Gower, J. C. 1975. Generalized procustes analysis. Psychometrika 40:33–51.

Greenacre, M. J. 1984. Theory and applications of correspondence analysis. Academic Press, London, UK.

Hill, M. O. 1973. Reciprocal averaging: an eigenvector method of ordination. Journal of Ecology 61:237–249.

Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics 5:299–314.

Kroonenberg, P. M. 1983. Three-mode principal component analysis. DSWO Press, Leiden, The Netherlands.

Kroonenberg, P. M. 1989*a*. The analysis of multiple tables in factorial ecology. III. Three-mode principal component analysis: "analyse triadique complète." Acta Œcologica, Œcologia Generalis 10:245–256.

Kroonenberg, P. M. 1989*b*. Singular value decompositions in three-way contingency tables. Pages 169–184 *in* R. Coppi and S. Bolasco, editors. Multiway data analysis. Elsevier, Amsterdam, The Netherlands.

Lebreton, J. D., D. Chessel, R. Prodon, and N. Yoccoz. 1988. L'analyse des relations espèces-milieu par l'analyse canonique des correspondances. I. Variables de milieu quantitatives. Acta Œcologica, Œcologia Generalis 9:53–67.

Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129:271–280.

Moreau, R. E. 1954. The main vicissitudes of the European avifauna since the Pliocene. Ibis 96:411–431.

Peres-Neto, P. R., D. A. Jackson, and K. M. Somers. 2005. How many principal components? Stopping rules for determining the number of non-trivial axes revisited. Computational Statistics and Data Analysis 49:974–997.

Pollard, E. 1977. A method of assessing changes in the abundance of butterflies. Biological Conservation 12:115–134.

Polus, E., S. Vandewoestijne, J. Choutt, and M. Baguette. 2006. Tracking the effects of one century of habitat loss and fragmentation on calcareous grassland butterfly communities. Biodiversity and Conservation, *in press*.

Potvin, C., and J. Travis. 1993. Concluding remarks: a drop in the ocean. Ecology 74:1674–1676.

Rieley, J. O., and S. E. Page. 1990. Ecology of plant communities. Longman Scientific and Technical, Harlow, UK.

Schtickzelle, N., and M. Baguette. 2004. Metapopulation viability analysis of the bog fritillary butterfly using RAMAS/GIS. Oikos 104:277–290.

Steinbacher, G. 1948. Der Einfluss der Eiszeit auf die Europäische Vogelwelt. Biologische Zentralblatt 67:444–456.

Swaine, M. D., and P. Greig-Smith. 1980. An application of principal components analysis to vegetation changes in permanent plot. Journal of Ecology 68:33–41.

ter Braak, C. J. F., and A. P. Schaffers. 2003. Co-correspondence analysis: a new ordination method to relate two community compositions. Ecology 85:834–846.

Thioulouse, J., and D. Chessel. 1987. Les analyses multitableaux en écologie factorielle. I De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. Acta Œcologica, Œcologia Generalis 8:463–480.

Thioulouse, J., M. Simier, and D. Chessel. 2004. Simultaneous analysis of a sequence of paired ecological tables. Ecology 85: 272–283.

Tucker, L. R. 1966. Some mathematical notes on three-mode factor analysis. Psychometrika 31:279–311.

APPENDIX A

Instructions for resuming the calculations of this paper with R (*Ecological Archives* E088-031-A1).

APPENDIX B

Absolute contributions of species, modalities of factor A and B in the two case studies (*Ecological Archives* E088-031-A2).

SUPPLEMENT

The data set for the bird communities (*Ecological Archives* E088-031-S1).