Geographic Structure and Potential Ecological Factors in Belgium

Marc Dufrene; Pierre Legendre

Journal of Biogeography, Vol. 18, No. 3 (May, 1991), 257-266.

Journal of Biogeography (1991) **18**, 257–266

Geographic structure and potential ecological factors in Belgium

MARC DUFRÊNE and PIERRE LEGENDRE* *Unité d'Ecologie et de Biogéographie, 5 Place Croix du Sud, B-1348 Louvain-la-Neuve, Belgium, and *Département de Sciences Biologiques, Université de Montréal, C.P. 6128, Succursale A, Montréal, Québec, Canada H3C 3J7*

Abstract. The available potential ecological factors have been scored in the form of presence/absence in U.T.M. squares in Belgium. A correspondence analysis shows a strong underlying gradient in the data set which induces an extraordinary horseshoe effect. This gradient follows closely the altitude component. Applying the *k*-means clustering method on U.T.M. squares produces geographically compact groups which are largely hierarchically nested. This indicates strong regional trends in the ecological data set. As homogeneous groups may also be artefacts created by the clustering algorithms on a continuous gradient, the relevance of the borders between homogeneous areas is tested. In general, *k*-means borders correspond to the main breaking lines between adjacent U.T.M. squares. They can be referred to as natural borders.

Key words. Ecological factors, geographic structure, multivariate analysis, homogeneous geographical areas, geographical borders, Belgium.

Résumé. Les facteurs écologiques disponibles en Belgique ont été encodés en absence/présence par carrés U.T.M. Une analyse des correspondances montre un important gradient sous-jacent dans le jeu de données qui se traduit par un extraordinaire effet d'arche. Ce gradient est bien expliqué par l'altitude. Le groupement des carrés U.T.M. par la méthode *k*-means produit des aires homogènes qui sont aussi hiérarchiques. Cela indique une forte structure régionale dans le jeu de données écologiques. Comme des groupes homogènes peuvent aussi être créés sur un gradient continu par l'algorithme de groupement, la pertinence des frontières entre les régions est testée. En général, les frontières produites par la méthode de groupement *k*-means correspondent à des zones de transition entre carrés U.T.M. Elles peuvent donc être considérées comme de véritables frontières géographiques.

Mots clés. Facteurs écologiques, structure géographique, analyses multivariables, territoires géographiques homogènes, frontières géographiques, Belgique.

INTRODUCTION

The determination of homogeneous biogeographic areas and their interpretation through environmental variables have always been an important topic of biogeographic studies. Although it is often clear that some distribution patterns are shared by several taxa, biogeographers may have difficulties in delimiting biogeographically homogeneous zones when the distributions are more diffuse, when the data available are spatially incomplete, or when the number of taxa is large (Birks, 1987). Another problem is the fact that possible causal factors and knowledge of biogeographic structures based on other data sets such as geological or climatical may have been overlooked in any given study. Recent developments in quantitative methods of ecological data analysis allow to overcome the problems caused by large numbers of taxa and the consideration of other geographical structures (Legendre, 1990).

The primary purpose of this paper is to describe the ecological geographic structure of Belgium, as a support for the interpretation phase of the faunal and floral biogeographic studies now under progress in that country. After a description of the dataset, the geographical structure of Belgium will be apprehended through the study of the covariation of several ecological factors, searching for areas characterized by similar values for the ecological variables.

ECOLOGICAL DATA SET

The investigated area comprises the 380 U.T.M. (Universal Transverse Mercator projection system) squares of 10 km side covering Belgium. This projection system has been chosen because it is generally used in Europe to score and represent species distributions. The U.T.M. squares represent the Operational Geographical Units (*sensu* Crovello, 1981) of the study, in short the OGUs. For each OGU, we scored the eleven available potential ecological factors

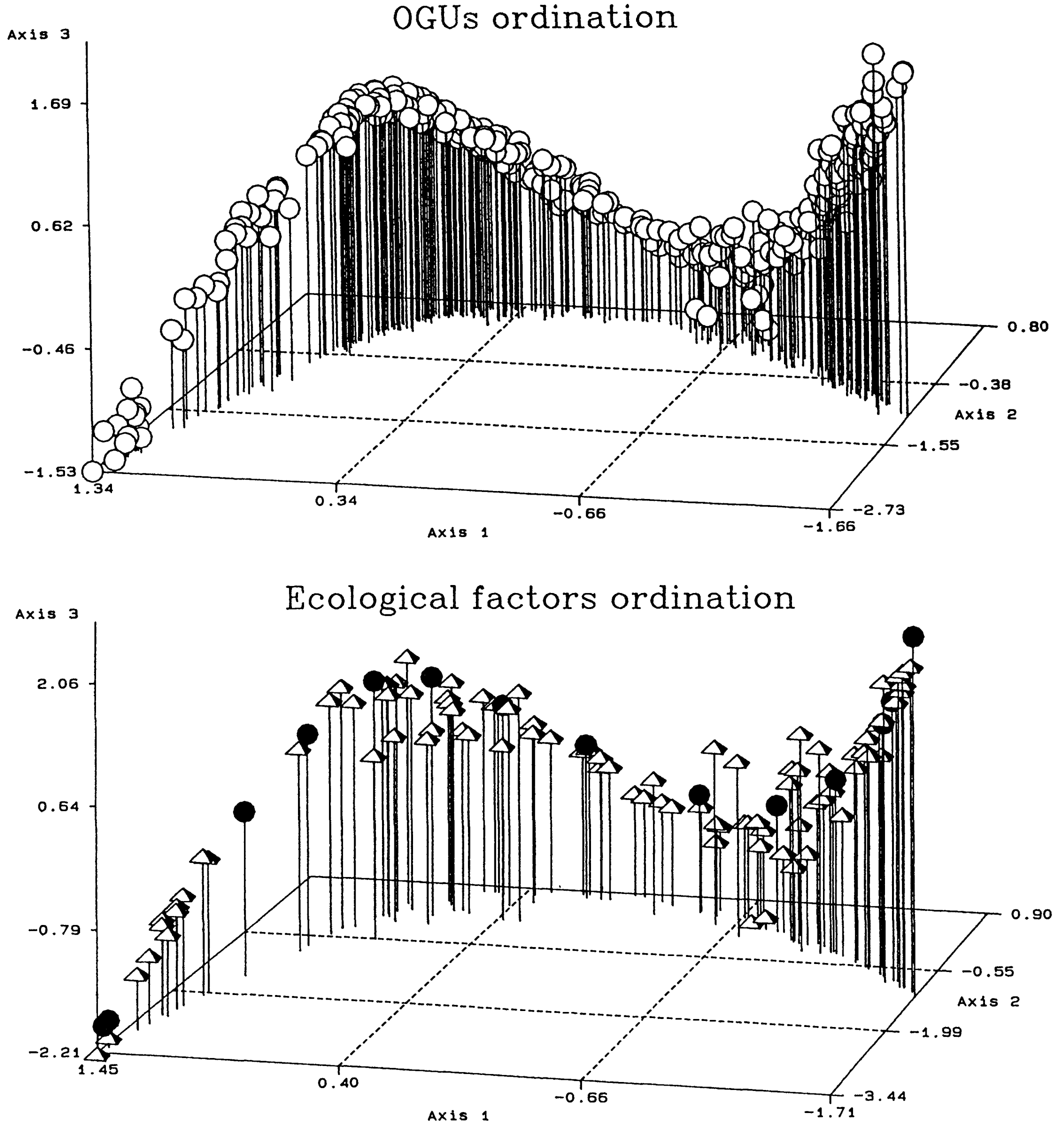


FIG. 1. Tridimensional plots of the first three correspondence analysis axes for the OGUs (upper) and the descriptor states (lower). Altitude classes are represented by black circles.

from the Atlas of Belgium (Anonymous, 1950–72) in the form of presence/absence of the various qualitative (nominal) states listed in Table 1. Presence/absence are used because the data sets are scored from maps of qualitative states for each variable. The factors included in this study are the number of days with and without frost, the first and last day of frost, average temperature, average rainfall, number of rainy days, lithology, geology, pedology, and altitude, for a total of 380 Belgian OGUs and 109 factor states. All these ecological factors, which have a spatial component, represent potential explanations for ecological distribution patterns.

RELATIONSHIPS AMONG POTENTIAL ECOLOGICAL FACTORS

Correspondence analysis has been used (program ACOBI of Lebart, Morineau & Tabard, 1977) to describe the interactions among characters and to bring out the main directions of covariability. This ordination method is adapted to absence/presence data. Surprisingly, the first axis accounts for 17.6% of the total variance; this is a very high value for a large presence/absence data table. The following axes display an extraordinary horseshoe effect (Kendall, 1971) (Fig. 1). Horseshoes are produced by multivariate ordination analysis methods when axes are related to each other in a non-linear way. They reveal that there is a strong underlying gradient in the data set (Hill & Gauch, 1980; Wartenberg *et al.*, 1987). In our case, this gradient is very important because at least the first five following axes are non-linear functions of the first one.

This NW–SE gradient clearly follows the altitude component, as shown by the succession of altitude classes along axis one (Fig. 1, lower panel). The lower altitude classes are on the left, together with the states describing the presence of dunes and polders along the coast. On the right, the highest altitude classes are associated with descriptor states characterizing the high Ardennes, like the presence of peaty soil and abundant precipitation.

To prevent an undue influence of the altitude character states, we repeated the correspondence analysis without them. The OGU position as well as the character state scores remain the same as in the first correspondence analysis, and this on each axis (Pearson *r* correlation comparing the first four axes of the two correspondence analyses, for state scores as well as for OGUs, is 0.99). Another correspondence analysis, computed on edaphic factors only (excluding all climatic and altitude characters states), shows an identical first axis (Pearson *r* for character state scores, comparing the first four axes of the two correspondence analyses, is 0.99; for OGU positions, 0.96), while the second axis is rather different (see below). The third axis of this analysis is correlated with the second axis of the global analysis. As can be expected, most of the character states are correlated with each other. The lower altitude areas, near the coast, are characterized by high temperatures (mean annual temperature = 9.5°C), fewer rainy days (160 days or 800 mm) and recent Quaternary deposits. At the opposite, in Ardennes, temperatures are lower (mean annual temperature = 7.5°C), precipitations are more important (more than 180 days of precipitation, or more than 1100 mm), and the geological substrate is the old Primary shield. Thus, almost all the covariation among character states is correlated with the underlying altitude gradient.

Another way of representing the altitude gradient is to compute a spatial autocorrelogram (program AUTOCOR from 'The R package for multivariate data analysis', referred to in Legendre & Fortin, 1989) on mean altitude. For each 100 km^2 OGU, the dominant altitude class was found over the 1 km^2 subunits. Fig. 2 shows the evolution of Moran's *I* index of spatial autocorrelation for different distance classes among OGUs. The shape of the correlogram is typical for a gradient, be it linear or in the form of a step function (Legendre & Fortin, 1989).

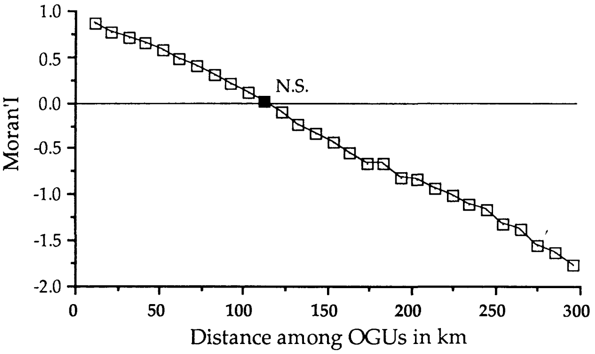


FIG. 2. Spatial autocorrelation (Moran *I*) on OGUs mean altitude. White squares are statistically significant values ($P \leq 0.05$).

Searching for other potentially interesting ecological factors, we applied a detrended correspondence analysis to our data (program CANOCO of ter Braak, 1988). Detrended methods were developed by several authors (Swan, 1970; Williamson, 1978; Hill, 1979) for unbending the horseshoe effect so as to obtain linear representations for linear gradients. The most widely used method is that of Hill (1979), despite criticisms by Wartenberg *et*

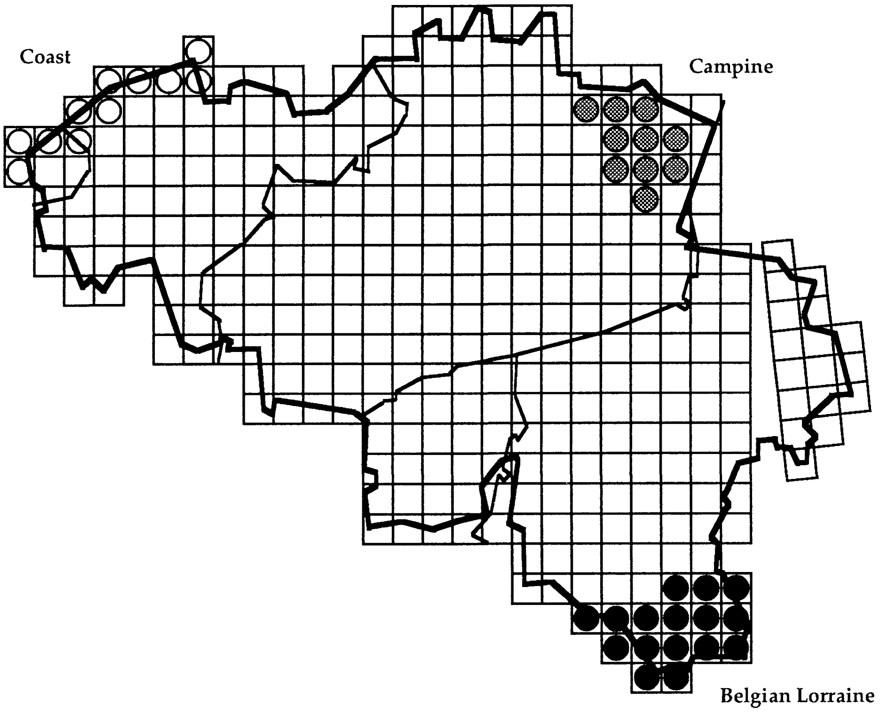


FIG. 3. Location of the OGUs isolated on axes two (black), three (grey) and four (white circles) by the correspondence analysis detrended by segments. Each square is 10 × 10 km.

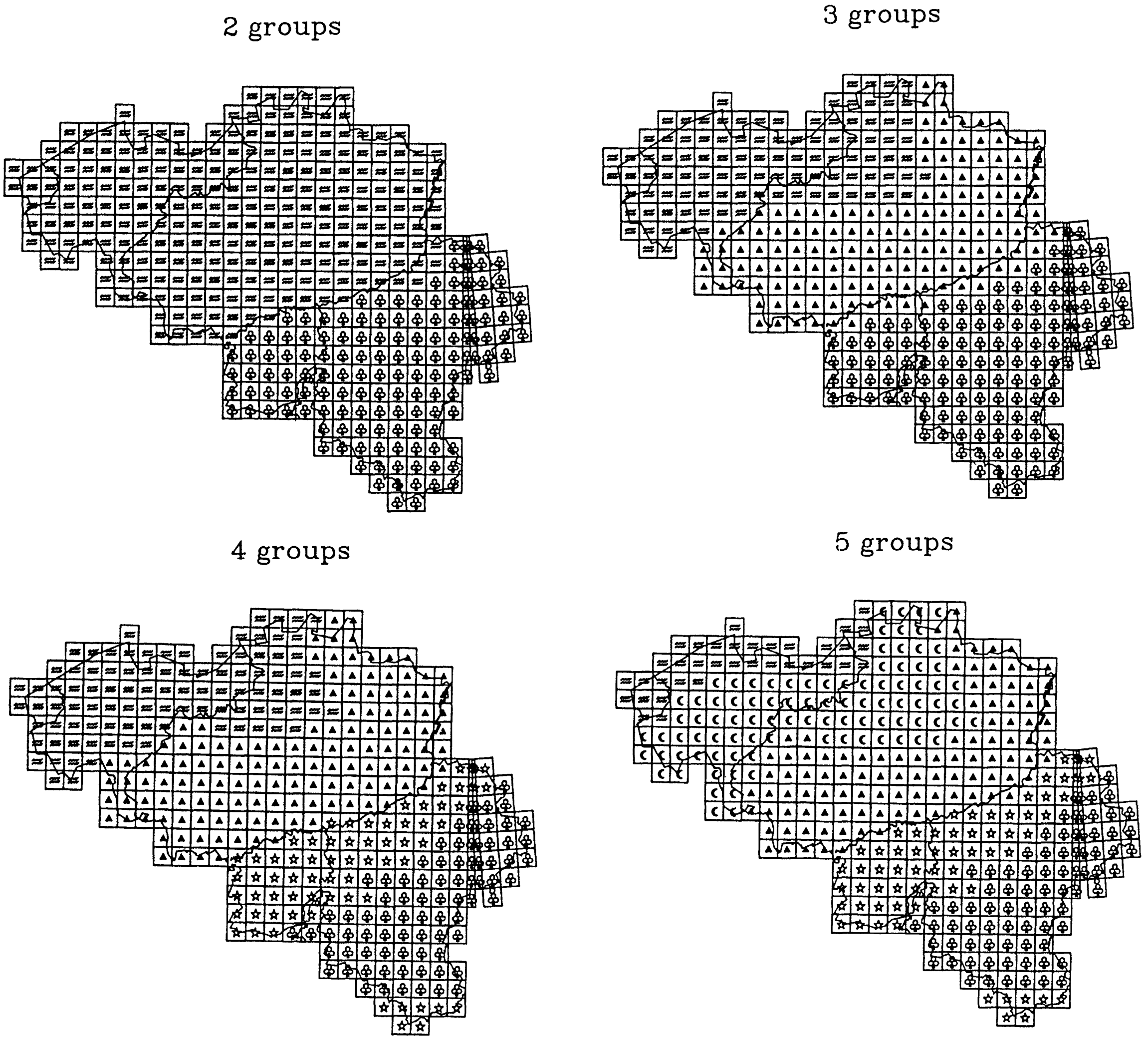


FIG. 4(a). Mapping of the clusters produced by *k*-means runs for $k = 2$–5 groups. Each group is represented by a different symbol. Each square is 10 × 10 km.

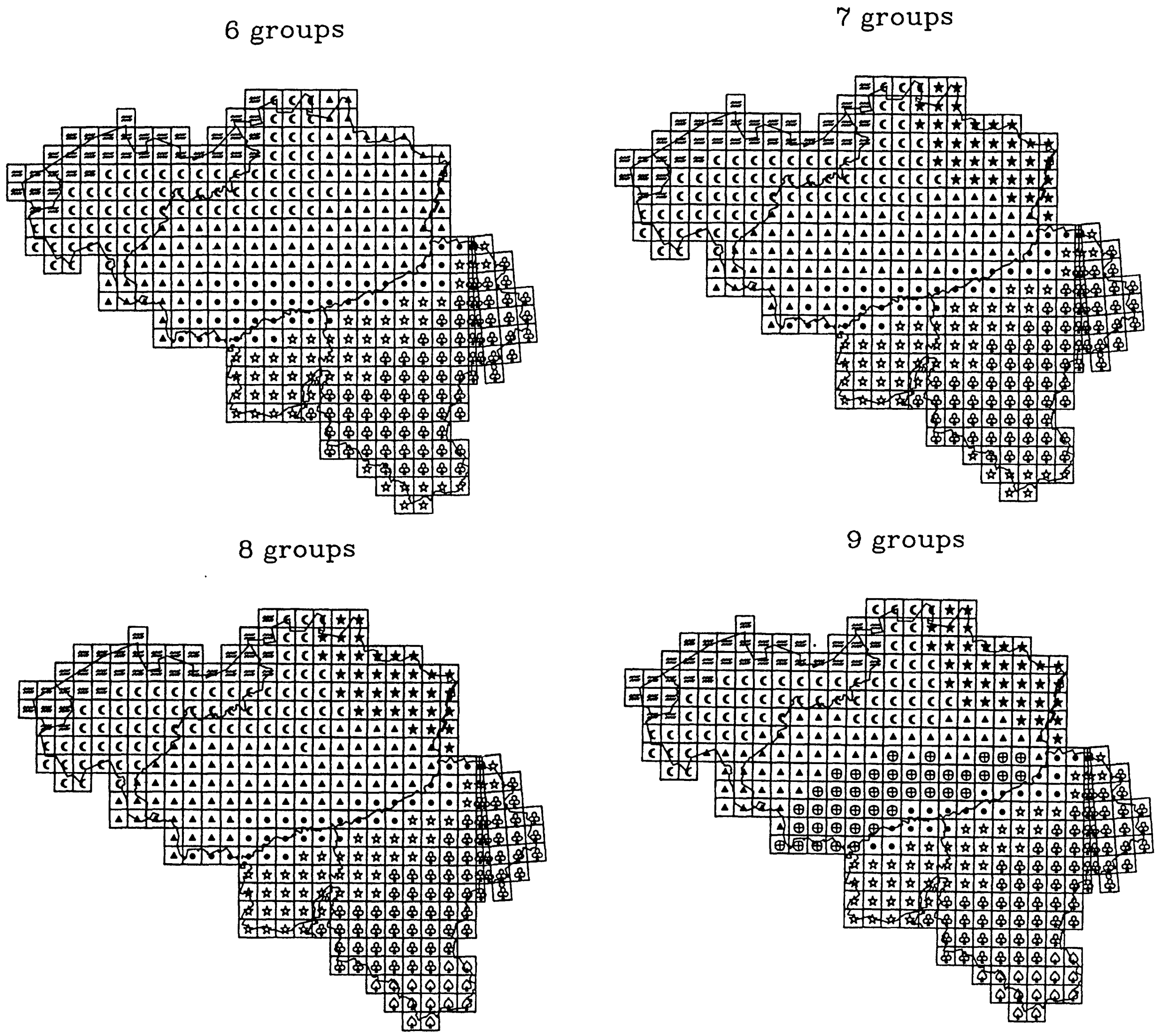
al. (1987). A recent paper (Peet *et al.*, 1988) seems to indicate that it is nevertheless the most powerful detrending method, so that we used it on our data, as well as the detrending-by-polynomials methods proposed by ter Braak (1988). These last attempts did not succeed in completely unbending our horseshoes because it affected so many axes; after a second order polynomial detrending, axis 2 still showed a third order horseshoe effect, while the third and fourth axes displayed non-linearities of other orders.

Detrending-by-segments (Hill & Gauch, 1980) produces a first axis that linearly follows the altitude classes, as in our first correspondence analysis. The next axes isolate groups of descriptor states and OGUs. The set of characters states isolated by the second axis comprises nine edaphic states that are to a large extent typical of the Belgian Lorraine region of southeastern Belgium. The corresponding OGUs are mapped in Fig. 3 (black circles). These OGUs and associated character states had already been pointed out by a non-linear regression analysis of the third axis of the global correspondence analysis, by the detrending-by-polynomials analysis, as well as by the correspondence analysis limited to the edaphic descriptor states. Thus, we may conclude that this area differs from the other parts of Belgium by a potential ecological factor distribution that does not follow altitude. It is likely to be induced by the presence of the chalky Parisian Basin expansion into southeastern Belgium.

The third and fourth axes respectively isolate character states and OGUs that are located in the Campine area (grey circles) and on the coast (white circles in Fig. 3). These regions had not been pointed out by any other analyses, although they are known to possess geographical and biological peculiarities, compared to the rest of Belgium.

GEOGRAPHIC REGIONS

The delineation of homogeneous areas or ‘provinces’, based on present-day floras and faunas, is one of the typical problems in ecological biogeography. Clustering methods represent one of the interesting approaches to this problem



6 groups

7 groups

8 groups

9 groups

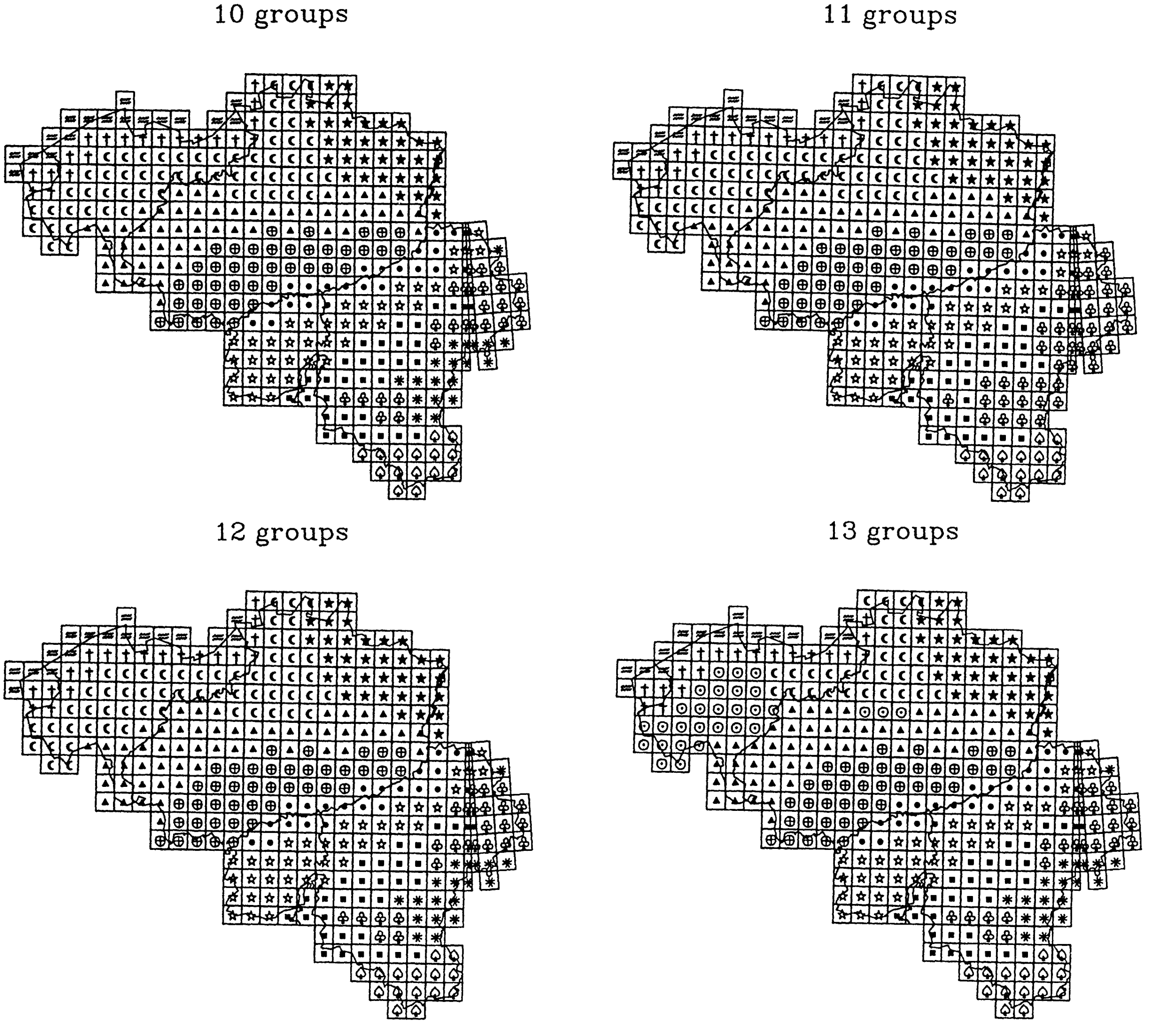
FIG. 4(b). Mapping of the clusters produced by *k*-means runs for *k* = 6–9 groups. Each group is represented by a different symbol. Each square is 10 × 10 km.

(Birks, 1987; Legendre, 1990). These methods can also be applied on our dataset to search for geographically homogeneous areas. When using resemblance-based clustering methods, the first step is to choose a resemblance measure among OGUs which is appropriate to the problem at hand. The Jaccard similarity coefficient was chosen for the present study because it does not take into account the number of joint absences. This coefficient is simply the ratio of the number of descriptor states present in the two areas, to the total number of descriptor states present in at least one area.

The next step is the choice of a clustering method. First, we used a proportional-link linkage agglomerative clustering algorithm (Sneath, 1966) with constraint of spatial contiguity (Legendre & Legendre, 1984; Legendre, 1987). The program that we used (BIOGEO) outputs a series of maps corresponding to the various clustering levels. Secondly, in order to avoid small distortions in the clustering structure such as might have been generated by the agglomerative clustering algorithm, the *k*-means method (MacQueen, 1967; Späth, 1980) was used to permit a reallocation of OGUs among clusters, at each level. The *k*-means method produces *k* groups (the value of *k* is decided by the user) after an iterative procedure of object reallocation; the procedure stops when the overall sum of squares, which is the sum of the within-group sums of squares, has reached a minimum. The clusters obtained by proportional-link linkage agglomerative clustering were used as initial configurations to the *k*-means algorithm; one hundred random allocations of points were also used as starting configurations, in order to make sure that the minimum obtained from the agglomerative clustering initial configuration was not a local minimum. The choice of the initial agglomerative clustering method was thus without important consequence.

Computations of the similarities between the 380 OGUs and of the clustering methods were done using ‘The R package for multivariate data analysis’ already mentioned above.

Figs. 4(a), 4(b) and 4(c) present the results obtained by



10 groups

11 groups

12 groups

13 groups

FIG. 4(c). Mapping of the clusters produced by k-means runs for k = 10–13 groups. Each group is represented by a different symbol. Each square is 10 × 10 km.

k-means clustering for values of k (the number of clusters) from 2 to 13. Geographically compact groups were obtained, although the k-means clustering runs were done without applying any constraint of geographic contiguity. This indicates strong regional trends in the ecological data set subjected to the analysis. The general structure consists of oblique strips of land, each largely homogeneous in altitude. For values of k larger than thirteen groups, many different local minima are found with very similar values of the sum of within-group sums of squares criterion, so that no single clustering structure emerges; this is why we stopped the analysis at thirteen groups.

The first split in Belgium (in two groups) is centered on the Meuse and Sambre valley and marks the difference between northern and southern Belgium. The next step divides the northern region into two sub-regions, with the first border moving back a little: eight OGUs move from the southern region into the middle one. The next grouping brings together two regions (Condroz-Fagne-Famenne and Belgian Lorraine, respectively in the median and the meridional groups) which are not contiguous but nevertheless similar. These two chalky regions are very different from the remainder of the southern region, which is the Ardennes plateau. The fifth group is created by the split of the northern one in two parts: coastal (sandy Flanders) *versus* inland (sandy-loamy Flanders). At the six-group level, the Meuse and Sambre valley steps out as a homogeneous area. The Campine area, restricted to the old Meuse alluvial cone, is the seventh group. The eight-cluster level separates the fourth OGU group in its two components: the Condroz-Fagne-Famenne and the Belgian Lorraine regions. Subsequent levels are represented on the maps.

The succession of maps (Fig. 4) shows that a natural hierarchy exists in the Belgian landscape, although the k-means clustering method does not impose a hierarchical structure to the results. This indicates that limits between areas must be largely stable, although some OGUs change groups at different steps.

TABLE 1. List of the available character states, as they appear in the Atlas, which have been scored for each one of the 380 U.T.M. squares covering Belgium.

1. Number of days without frost

1.1 more than 220 days
1.2 from 210 to 220 days
1.3 from 200 to 210 days
1.4 from 190 to 200 days
1.5 from 180 to 190 days
1.6 from 170 to 180 days
1.7 from 160 to 170 days
1.8 from 150 to 160 days
1.9 from 140 to 150 days
1.10 less than 140 days

2. Date of the first frosts

2.1 after 5/11
2.2 from 30/10 to 5/11
2.3 from 20/10 to 30/10
2.4 from 10/10 to 20/10
2.5 before 10/10

3. Date of the last frosts

3.1 before 10/4
3.2 from 10/4 to 20/4
3.3 from 20/4 to 30/4
3.4 from 30/4 to 10/5
3.5 from 10/5 to 20/5
3.6 after 20/5

4. Number of days with frost

4.1 less than 50 days
4.2 from 50 to 60 days
4.3 from 60 to 70 days
4.4 from 70 to 80 days
4.5 from 80 to 90 days
4.6 from 90 to 100 days
4.7 from 100 to 110 days
4.8 from 110 to 120 days
4.9 more than 120 days

5. Mean temperature

5.1 less than 7 C°
5.2 from 7 to 8 C°
5.3 from 8 to 9 C°
5.4 from 9 to 10 C°
5.5 more than 10 C°

6. Mean annual precipitation

6.1 less than 800 mm
6.2 from 800 to 900 mm
6.3 from 900 to 1000 mm
6.4 from 1000 to 1100 mm
6.5 from 1100 to 1200 mm
6.6 from 1200 to 1300 mm
6.7 from 1300 to 1400 mm
6.8 more than 1400 mm

7. Number of rainy days

7.1 less than 160 days
7.2 from 160 to 180 days
7.3 from 180 to 200 days
7.4 more than 200 days

8. Pedology

8.1 dunes
8.2 young polders
8.3 middle polders
8.4 old polders
8.5 moeres
8.6 covered Pleistocene
8.7 sandy soils
8.8 loamy soils
8.9 loamy-stony soils
8.10 peaty soils
8.11 peaty soils inclusions
8.12 loamy-sandy soils
8.13 steeply slopes
8.14 clayey soils

9. Geology

9.1 Holocene
9.2 Pleistocene
9.3 Pliocene
9.4 Miocene
9.5 Oligocene
9.6 Eocene
9.7 Cretaceous
9.8 Oolithic
9.9 Liasic
9.10 Triasique
9.11 Permian
9.12 Carboniferous
9.13 Dinantian
9.14 Neo-meso Devonian
9.15 Eodevonian
9.16 Silurian
9.17 Cambrian

10. Lithology

10.1 recent alluviums
10.2 Meuse's alluvial cone
10.3 Cenozoic sands
10.4 Mesozoic sands and sandstones
10.5 Malmedy's conglomerate
10.6 Neo-Devonian sandstones
10.7 polders clays
10.8 Cenozoic clays
10.9 Mesosoic clays and schists
10.10 Devonian schists
10.11 Silurian schists
10.12 Ceno- and Mesozoic schists
10.13 gravellous chalks and limestones
10.14 Jurassic limestones
10.15 Paleozoic limestones
10.16 Carboniferous schists and sandstones
10.17 Cambrian and eodevonian schists

11. Altitude

11.1 more than 600 m
11.2 from 500 to 600 m
11.3 from 400 to 500 m
11.4 from 300 to 400 m
11.5 from 200 to 300 m
11.6 from 150 to 200 m
11.7 from 100 to 150 m
11.8 from 50 to 100 m
11.9 from 20 to 50 m
11.10 from 10 to 20 m
11.11 from 5 to 10 m
11.12 from 0 to 5 m
11.13 tidal area
11.14 sea

GEOGRAPHIC BORDERS

Ordinations and the *k*-means clustering method seem to lead to conflicting results. On the one hand, the ordinations point out the existence of a strong continuous gradient that follows the altitude classes; on the other hand, the *k*-means clustering method brings out discontinuous areas, which are largely hierarchically nested. This is not surprising *per se*, since ordination techniques are designed to bring out the most important axes of variation of the data structure, while clustering methods look for finer structures within the data set and are designed to recognize groups in all cases, even when these do not correspond to reality (Legendre & Legendre, 1983). In descriptive multivariate data analysis, one is always faced with the problem of assessing the reality of the clusters one has identified; in other words, since the homogeneous areas are isolated from one another by borders, one must question their relevance. Are these real

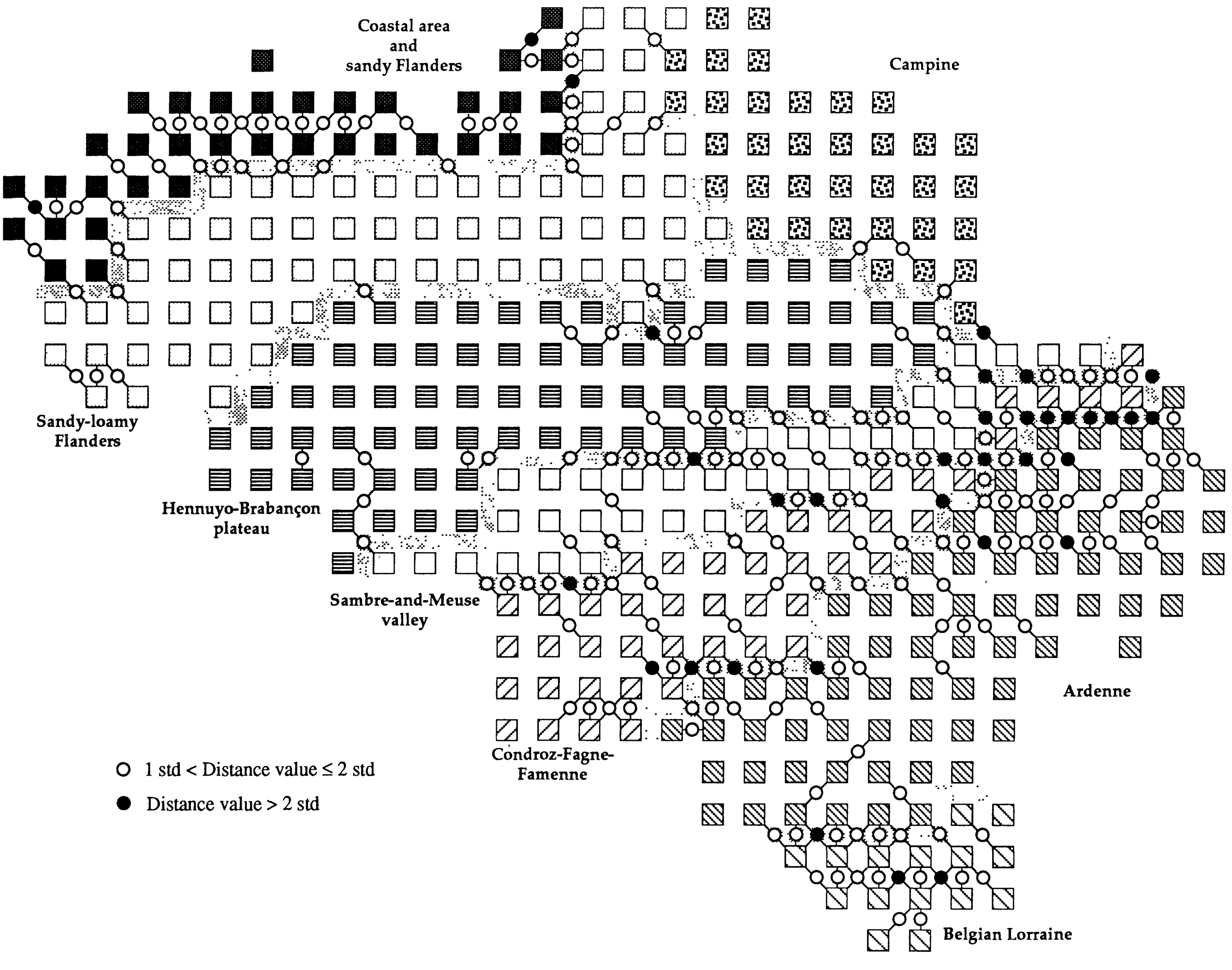


FIG. 5. Location of the standardized distance values between adjoining OGUs that are larger than one standard deviation. OGUs are represented by squares, distance values by small circles and *k*-means borders (for eight groups) by grey strips in the space between squares. Small lines behind circles indicate the directions of maximum change. Each square is 10 × 10 km.

geographic borders, or are they simply artefacts created by the clustering algorithms on a continuous gradient?

The difficulty comes from the fact that borders cannot be statistically tested through analysis of variance, since the data one would like to test with are the very same that gave birth to the borders to be tested; Perruchet (1983) presents a review of this question. Our approach here will be to look again at the data and try to identify the regions of abrupt change of the variables on the map, without considering the clustering results. If these natural borders correspond to those identified by the *k*-means method, we will have more confidence in their being natural discontinuities.

This will be accomplished by computing and mapping the distance measures between adjoining OGUs on the map. The Euclidean distance was computed, over the 109 binary descriptors of Table 1, among those OGUs that are adjacent by their sides and corners (king's chess movement). Values in this distance matrix were standardized (i.e. centered on the mean and divided by the standard deviation of the distances), and the distance values larger than one standard deviation were then plotted on a map; the limits between *k*-means regions for eight groups were also plotted for comparison (Fig. 5). This method is simpler than the Wombling method recently advocated by Barbujani *et al.* (1989), but it should lead to largely the same results.

The borders found by *k*-means clustering have a tendency to follow preferentially the breaking lines between neighbouring OGUs. Although the largest distances among neighbours and the *k*-means borders have been computed on the same data, a chi-square test can be computed on an indicative basis, under the null model that the *k*-means method might have created groups with imprecise borders, that do not necessarily correspond to those found with the very local method used here to detect points of maximum change in the data structure. The chi-square test computed between the presence/absence of a *k*-means border and the presence/absence of a two-standard-deviations distance is significant at $P<0.0001$; 77.1% of the distance values larger than two standard deviations above the mean are crossed by a *k*-means borders. For the distance values between one and two standard deviations above the mean, the proportion decreases to 42.6%. The proportion is only 11.5% for distance values that are below the mean. So, in general, *k*-means

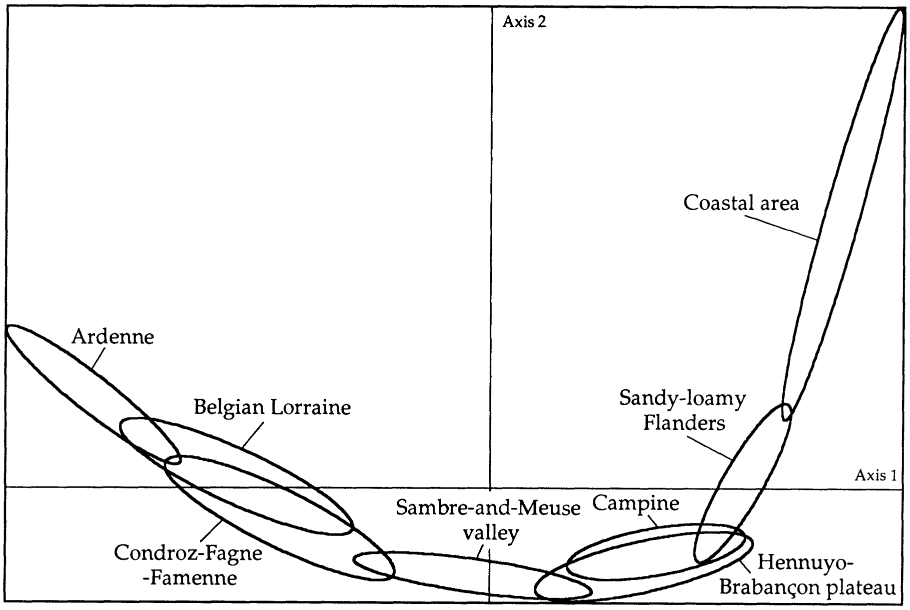


FIG. 6. Plot of *k*-means groups on CA coordinates with 90% dispersion ellipses (for eight groups).

borders correspond to the main breaking lines between adjacent OGUs. In the coastal and Belgian Lorraine regions, two straight breaking lines are not crossed by *k*-means borders at the eight-group level. However, the borders formed at eleven-group level for the coastal region run over the breaking line (Fig. 4). For the Belgian Lorraine regions, the southern breaking line was already crossed by the *k*-means borders at the four-group level (Fig. 4) and will be crossed again at the level 15 (not represented).

This kind of representation shows also that with the exception of the coast, the northern area is very homogeneous compared to the south of Belgium.

DISCUSSION

The analysis reported in this paper includes all the potential ecological factors presently available and covering the whole territory of Belgium. Ordination methods show that a strong continuous gradient is underlying the potential ecological factors included in this study. This gradient is highly correlated with the altitude variable; although in Belgium, altitudes do not exceed 700 m, this factor explains almost all the geographic structure because many other factors are correlated with it. Meanwhile, peculiar edaphic factors isolate the Gaume from the other Belgian regions.

In spite of a strong continuous gradient, the *k*-means clustering method generates several homogeneous regions which are isolated from one another by stable borders, which in turn largely correspond to the points of steepest gradient in the set of variables. This indicates that the delineated areas are not artefacts generated by a continuous gradient structure. The discontinuities found across the gradient are sufficiently important to isolate stable areas. These discontinuities can be referred to as natural geographical borders.

The horseshoe effect shows a strong gradient structure although cluster analysis suggests strong stable borders. This may appear as a contradiction. On the one hand, the *k*-means clustering method always produces the number of groups that it is asked to produce, irrespective of the fact that the steps between groups are large or small. On the other hand, the divisions between groups will always correspond to an area of larger differences. In the present case, what we find are a series of groups, well-defined by their border areas, that are ordered along the overall gradient as shown by Fig. 6.

The comparison of our results with the only biogeographic studies available over Belgium (Massart, 1910; Tournay, 1968; De Langhe *et al.*, 1978, all on vegetation) reveals much similarity in the areas identified by the two approaches. At the seven- or eight-group level, the areas are the same, except for the region centered on the Meuse and Sambre valley, which is not found on the phytogeographers' maps; this area is certainly an important transitional zone because the first Belgian partition in our study occurs precisely at the Meuse and Sambre valley. Our results show also a hierarchical arrangement of the partition levels between areas, some divisions being more important than others, corresponding more or less to the phytogeographers' hierarchy (domain, sector, district, ...).

The great similarity between our geographic maps and those of the phytogeographers indicates a good relationship between our method of analysis and the perception of the phytogeographers. Using biological data, such as those of the phytogeographers, hypotheses concerning the role of physical geographic borders, as identified in the present study, can now be tested, and the role of the climatic and edaphic factors in determining the distribution of the fauna and flora can be evaluated. Methods to do so have recently been proposed by Sokal, Oden & Thomson (1988), Barbujani *et al.* (1989) and Legendre *et al.* (1990).

Meanwhile, results from a preliminary analysis on the distribution of two insect families in Belgium (Dufrêne & Rasmont, 1989) show that the biogeographic structure is more confused than expected. This study will be completed and reported elsewhere. It would be very interesting to enlarge this kind of studies to other European areas where geological and climatic factors seem to have more homogeneous or more confused distributions and where 'potential biogeographical regions' are, at a first sight, difficult to establish or seem to be very patchily distributed.

ACKNOWLEDGMENTS

We are grateful to C. C. Downer, M. Baguette, J. De Sloover, Ph. Lebrun, P. Rasmont and G. Neve for their comments on earlier drafts of this manuscript. This work was supported by research grants from the Fonds de Développement Scientifique (Université Catholique de Louvain) and the Communauté française de Belgique, as well as by the cooperation programme between Québec and the Communauté française de Belgique.

REFERENCES

Anonymous (1950–72) *Atlas de Belgique*. Comité national de Géographie, Commission de l'Atlas, Palais des Académies, Bruzelles.

Barbujani, G., Oden, M.L. & Sokal, R.R. (1989) Detecting regions of abrupt change in maps of biological variables. *Syst. Zool.* **38**, 376–389.

Birks, H.J.B. (1987) Recent methodological developments in quantitative descriptive biogeography. *Ann. Zool. Fennici*, **24**, 165–178.

Crovello, T.J. (1981) Quantitative biogeography: an overview. *Taxon*, **30**, 563–575.

De Langhe, J.-E., Delvosalle, L., Duvigneaud, J., Lambinon, J. & Vanden Berghen, C. (1978) *Nouvelle Flore de la Belgique, du Grand Duché de Luxembourg, du Nord de la France et des régions voisines*. Patrimoine du Jardin Botanique national de Belgique, Meise.

Dufrêne, M. & Rasmont, P. (1989) Recherche d'une structure zoogéographique de la Belgique par l'étude des Carabidae (Coleoptera) et des Apoidea (Hymenoptera). Comptes-rendus du Symposium 'Invertébrés de Belgique', pp. 415–418. Institut Royal des Sciences Naturelles de Belgique, Bruxelles.

Hill, M.O. (1979) DECORANA—A FORTRAN program for detrended correspondence analysis and reciprocal averaging. Section of Ecology and Systematics, Cornell University, Ithaca, New York.

Hill, M.O. & Gauch, H.G., Jr (1980) Detrended correspondence analysis: an improved ordination technique. *Vegetatio*, **42**, 47–58.

Kendall, D.G. (1971) Seriation from abundance matrices. *Mathematics in archaeological and historical sciences* (ed. by C. R. Hodson, D. G. Kendall and P. Tautu), pp. 215–252. Edinburgh University Press.

Lebart, L., Morineau, A. & Tabard, N. (1977) *Techniques de la description statistique*. Dunod, Paris.

Legendre, L. & Legendre, P. (1983) *Numerical ecology*. Developments in Environmental Modelling, 3. Elsevier, Amsterdam.

Legendre, P. (1987) Constrained clustering. *Developments in numerical ecology* (ed. by P. Legendre and L. Legendre), pp. 289–307. NATO ASI Series, Vol. G14. Springer, Berlin.

Legendre, P. (1990) Quantitative methods and biogeographic analysis. *Evolutionary biogeography of the marine algae of the North Atlantic* (ed. by D. J. Garbary and G. R. South), pp. 9–34. NATO ASI Series, Vol. G22. Springer, Berlin.

Legendre, P. & Fortin, M.-J. (1989) Spatial pattern and ecological analysis. *Vegetatio*, **80**, 107–138.

Legendre, P. & Legendre, V. (1984) Postglacial dispersal of freshwater fishes in the Québec peninsula. *Can. J. Fish. Aquat. Sci.* **41**, 1781–1802.

Legendre, P., Oden, N.L., Sokal, R.R., Vaudor, A. & Kim, J. (1990) Approximate analysis of variance of spatially autocorrelated regional data. *J. Class.* **7**, 53–75.

MacQueen, J.B. (1967) Some methods for classification and analysis of multivariate observations. *Proc. fifth Berkeley symp. math. stat. probab.*, Vol. 1 (ed. by L. M. Le Cam and J. Neyman), pp. 281–297. University of California Press, Berkeley.

Massart, J. (1910) Esquisse de la géographie botanique de la Belgique. *Rec. Inst. Bot. Léo Errera* (*Univ. Bruxelles*), 332p.

Peet, R.K., Knox, R.G., Case, J.S. & Allen, R.B. (1988) Putting things in order: the advantages of detrended correspondence analysis. *Am. Nat.* **131**, 924–934.

Perruchet, C. (1983) Significance tests for clusters: overview and comments. *Numerical taxonomy* (ed. by J. Felsenstein), pp. 199–208. NATO ASI Series, Vol. G1. Springer, Berlin.

Sokal, R.R., Oden, N.L. & Thomson, B.A. (1988) Genetic changes across language boundaries in Europe. *Amer. J. Phys. Anthrop.* 76, 337–361.

Späth, H. (1980) *Cluster analysis algorithms*. Ellis Horwood, Chichester.

Sneath, P.H.A. (1966) A comparison of different clustering methods as applied to randomly-spaced points. *Classification Soc. Bull.* **1**, 2–18.

Sokal, R.R., Oden, N.L. & Thomson, B.A. (1988) Genetic changes across language boundaries in Europe. *Amer. J. Phys. Anthrop.* 76, 337–361.

Swan, J.M.A. (1970) An examination of some ordination problems by use of simulated vegetational data. *Ecology*, **51**, 89–102.

ter Braak, C.J.F. (1988) CANOCO—An extension of DECORANA to analyse species–environment relationships. *Vegetatio*, **75**, 159–160.

Tournay, R. (1968) Les territoires géobotaniques de la Belgique. *Bull. Jard. Bot. Nat. Belg.* **38**, 277–294.

Wartenberg, D., Ferson, S. & Rohlf, F.J. (1987) Putting things in order: a critique of detrended correspondence analysis. *Am. Nat.* **129**, 434–448.

Williamson, M.H. (1978) The ordination of incidence data. *J. Ecol.* **66**, 911–920.