Ecological Monographs, 67(3), 1997, pp. 345–366

SPECIES ASSEMBLAGES AND INDICATOR SPECIES: THE NEED FOR A FLEXIBLE ASYMMETRICAL APPROACH

MARC DUFRÊNE[1] AND PIERRE LEGENDRE[2]

[1]*Unité d'Écologie et de Biogéographie, Université catholique de Louvain, Croix du Sud, 5, B-1348 Louvain-la-Neuve, Belgium*
[2]*Département de Sciences Biologiques, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, Québec, Canada H3C 3J7*

Abstract. This paper presents a new and simple method to find indicator species and species assemblages characterizing groups of sites. The novelty of our approach lies in the way we combine a species relative abundance with its relative frequency of occurrence in the various groups of sites. This index is maximum when all individuals of a species are found in a single group of sites and when the species occurs in all sites of that group; it is a symmetric indicator. The statistical significance of the species indicator values is evaluated using a randomization procedure. Contrary to TWINSPAN, our indicator index for a given species is independent of the other species relative abundances, and there is no need to use pseudospecies.

The new method identifies indicator species for typologies of species relevés obtained by any hierarchical or nonhierarchical classification procedure; its use is independent of the classification method. Because indicator species give ecological meaning to groups of sites, this method provides criteria to compare typologies, to identify where to stop dividing clusters into subsets, and to point out the main levels in a hierarchical classification of sites.

Species can be grouped on the basis of their indicator values for each clustering level, the heterogeneous nature of species assemblages observed in any one site being well preserved. Such assemblages are usually a mixture of eurytopic (higher level) and stenotopic species (characteristic of lower level clusters). The species assemblage approach demonstrates the importance of the ''sampled patch size,'' i.e., the diversity of sampled ecological combinations, when we compare the frequencies of core and satellite species. A new way to present species–site tables, accounting for the hierarchical relationships among species, is proposed. A large data set of carabid beetle distributions in open habitats of Belgium is used as a case study to illustrate the new method.

Key words: Belgium; carabid beetles; core eurytopic species; ecological breadth; indicator species; multivariate analysis; patch size; satellite stenotopic species; species assemblages; TWINSPAN.

INTRODUCTION

The identification of characteristic or indicator species is a traditional activity in ecology and biogeography. Studies based on field work, describing sites or habitats, usually mention one or several species characterizing each habitat. So, it is surprising that several standardized and well-accepted methods to identify indicator species have not been developed. The only widespread method is TWINSPAN, proposed on heuristic grounds 18 yr ago (Hill 1979). Hill created a numerically based method to approach this ecological question and made available a computer program implementing it. TWINSPAN has several limitations that will be discussed below.

There is clearly a need for the identification of characteristic or indicator species in the fields of nature monitoring, conservation, and management. One of the most commonly used criteria for assessing the conservation value of a site is species richness, but this index is sensitive to several factors. For instance, He et al. (1994) have shown how some of the commonly used indices of diversity vary in a nonlinear fashion with the size of the elementary sampling units (quadrats, etc.) as well as with the extent of the study area. Prendergast et al. (1993*a*) have shown the role of sampling effort in artificially increasing the perceived biological richness of hotspots. On the other hand, high diversity does not insure that a site has a high ecological value (Dunn 1994). Even where the sampling effort has been equal across sites, species diversity is a questionable criterion when habitats with different productivity levels are compared, or when the number of rare species is large. Moreover, the protection of high-diversity sites does not guarantee the effective conservation of rare or spatially restricted organisms (Prendergast et al. 1993*b*). As proposed by Webb (1989) and Cousins (1991), representativeness, or representative diversity, would be a more satisfactory criterion. This implies a

Manuscript received 27 February 1995; revised 17 May 1996; accepted 21 May 1996; final version received 9 September 1996.

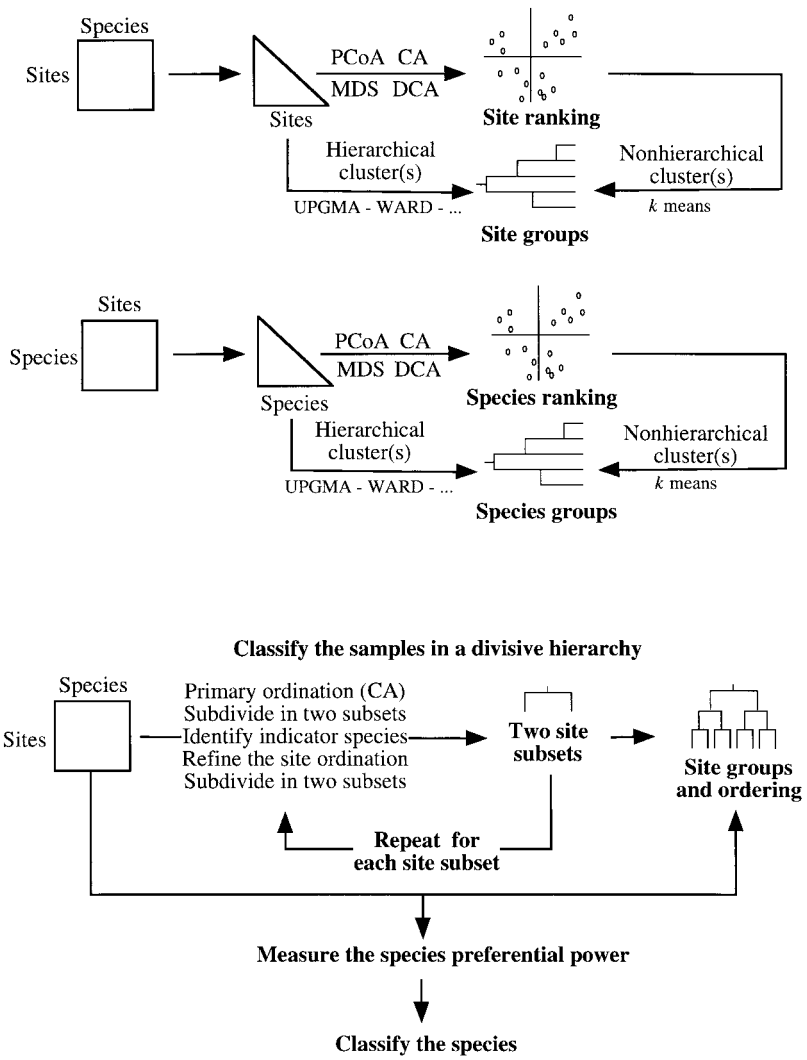

FIG. 1. Diagram of the analysis steps for the *Q*- and *R*-mode classical analyses and the TWINSPAN procedure. CA = Correspondence Analysis; DCA = Detrended Correspondence Analysis; MDS = Nonmetric Multidimensional Scaling; PCoA = Principal Coordinates Analysis; UPGMA = Unweighted Pair-Group Method using Arithmetic Average; WARD = Ward's clustering method.

list of the typical species assemblages for habitats or ecological factor combinations. Such species lists, or scores for habitats, would be very useful in evaluating the comparative richness of sites, or the effect of isolation or fragmentation. In monitoring programs designed to evaluate the impact of management practices, it will also be useful to weight the species abundances in a site in order to account more for the typical than the vagrant species. To summarize, representativity is a subjective notion that requires standardized methods for its efficient usage.

The characterization of habitats by species assemblages or indicator species is one component of the analysis of Structure–Activity Relationships (SAR, Lebreton et al. 1991). Its results only have predictive power at sites with habitats similar to those used to find the indicator species; SAR results are limited by how exhaustive and representative the sampling design has been. The analysis of Species–Environment Relationships (SER), which allows one to describe the environment using ecological variables rather than a typology of habitats, has greater calibrating power than SAR because it can be used to make predictions about unsampled habitats or ecological factor combinations. SER analysis, however, requires specific multivariate methods and procedures to account for the nonlinear and multiscalar relationships that explain species distributions (see Jongman et al. 1987). This paper presents a method for SAR analysis. Another paper (M. Dufrêne and P. Legendre, *unpublished manuscript*) will be devoted to SER analysis.

Generally, the SAR approach (Fig. 1) uses classical ordination methods such as Correspondence Analysis (CA) or Detrended Correspondence Analysis (DCA:

Hill and Gauch 1980), Principal Coordinates Analysis (PCoA), or Nonmetric Multidimensional Scaling (MDS), which can all produce ordinations (scalings) of sites and species. Ordination methods are traditionally used in conjunction with classical clustering methods (Gauch 1982, Legendre and Legendre 1983, 1984*a*, Austin 1985, Jongmann et al. 1987, etc.) to produce site or species clusters. When sites are compared, this approach is labeled *Q*-mode analysis. The same methods can be used to ordinate species or establish species groups (*R*-mode analysis). In that case, the SAR approach is symmetrical because site and species classifications are obtained independently with the same methods.

The TWINSPAN procedure proposed by Hill (1979) tries to tackle the problem of the characterization of habitats by indicator species using an asymmetric approach in which species are classified based on the results of a site typology. TWINSPAN is a two-way indicator species analysis that produces a tabular matrix arrangement approximating the results of a Braun-Blanquet table. The method first constructs a site typology and uses it to obtain a classification of the species based upon ecological preferences (fidelity of species to groups of sites). Sites are classified using a divisive hierarchical algorithm; they are first divided into two subsets according to their sign on the first axis of a correspondence analysis ordination (CA or DCA), whereby each subset is divided in two smaller subsets by repeating the same procedure, and so on (Fig. 1). At each step, each species receives an attribute describing its preference for one or the other side of the partition. These attributes are then used to produce a refined site ordination and to classify species into several groups. To model the concept of differential species (i.e., species with clear ecological preferences), which is mainly qualitative, TWINSPAN creates ''pseudospecies.'' Each species is subdivided into absence/presence vectors (0/1 dummy variables) for several different relative abundance levels. If the pseudospecies are defined as >0, >5, >26, >51, and $>76\%$, a relative abundance of 18% at a site will fill the first and the second dummy pseudospecies vectors with a presence; a site where the species is absent (0%) receives zero for every pseudospecies corresponding to that species. In this way, pseudospecies make it possible to use the species relative abundances as a measure of their indicator power.

When used to search for indicator species or to identify species assemblages, classical R-mode analyses (symmetrical approach), such as PCoA or R-mode clustering, do not account for several of the peculiarities of species distributions; this is particularly true when large differences in species ecological breadths occur. Rare and widespread species are compared on the same footing, although they do not respond to the same site clustering levels. In this type of analysis, there must be several widespread species typical of a high-level structure (e.g., all wet sites) to make the analysis recognize them as forming a homogeneous group, distinct from other species group(s) typical of lower level structure(s) (e.g., swamps only). Because the species assemblages that are generally created by such methods are characterized by several rare or satellite species associated with some widespread or core species, widespread species are placed in one or another of the specialized species groups, or else they are considered as outliers, instead of being placed in a ''widespread species'' cluster. Moreover, the allocation of widespread species to species groups often changes with the similarity index or the clustering method. Widespread species, however, are very important to the SAR because they are at the origin of many hierarchical structures observed when clustering sites. Site and species clusters are never completely superimposable. This problem indicates that an asymmetrical procedure should be used to analyze species relationships on the basis of clusters of sites, as in TWINSPAN.

The TWINSPAN algorithm has been criticized by Belbin and McDonald (1993) on two grounds: (1) TWINSPAN assumes the existence of a strong gradient dominating the data structure, and so it may fail to identify secondary gradients or other types of structure in data sets; and (2) the cutting points along the dominant axis are rather arbitrary; instead of selecting large gaps in the data, sites that may be very close in species composition may be separated.

There are other problems with TWINSPAN when it comes to identifying clusters of species or computing indicator values. (1) To identify species clusters, the program does not allow one to use a classification of sites, other than the one based on CA or DCA ordination and produced by TWINSPAN. (2) Secondly, the peusdospecies concept is based on relative species abundances, the relative abundance of a species depending on the absolute abundance of the other species in the quadrat. Such relative frequencies can be highly biased in the case of sampling or harvesting methods: with most sampling methods for living and mobile organisms, all species are not sampled in the same way because all species do not behave in the same way. There is always a distortion between the observed and the real species relative frequencies estimated at any given site (see *Materials and methods: Carabid beetle data set* below). The only appropriate comparisons are with other sample units of the same species. (3) Finally, whereas simple CA is well suited to study species distributions from several sample units (ter Braak 1985), the DCA algorithm has recently been criticized (Wartenberg et al. 1987, Jackson and Somers 1991, but see also Peet et al. 1988). For example, Jackson and Somers (1991) have shown that results obtained with DCA vary depending on the number of segments used to remove the arch effect. Therefore, several runs with different numbers of segments must be done to find stable factorial axes and interpretable results.

TABLE 1. Name of the selected habitats, number of sampled sites in each habitat, CORINE European Community standard habitat codes, phytosociological associations, and habitat descriptions are given.

Habitats	Number of sites	CORINE code	Phytosociological typology
Fringes of ponds	10	22.31, 54.4, 53.2, 22.11	*Littorelletalia, Caricetalia fuscae, Magnocaricion*; several transects along ponds with a variable water level
Swamps	12	54.4	*Caricetalia fuscae*; acidific fens and swamps, often with *Juncus acutiflorus* and *Carex* spp.
Raised mires	8	51.1	*Sphagnion magellanici*; near-natural raised bogs and mires
Peat moors	16	51.2, 31.1	*Ericion tetralicis*; purple moorgrass bogs, wet heath, moors, peat bogs
Sandy heathlands	20	31.1, 31.2	*Ericion tetralicis, Calluno-Vaccinion, Corynephoretum*; dry heathlands on sand (some of them are sometimes flooded during winter)
Gravelly-loamy heathlands	14	31.2, 35.13	*Calluno-Vaccinion, Deschampsia flexuosa* grasslands; dry heathlands on loamy and more or less gravelly soils
Alluvial meadows	11	37.1	*Filipendulion ulmariae*; meadowsweet stands and related alluvial communities
Chalky grasslands	22	34.32, 34.33	*Mesobromion, Seslerio-Xerobromion*; sub-Atlantic semi and very dry calcareous grasslands
Contaminated grasslands	10	34.2, 35.13	*Thlaspion calaminariae, Deschampsia flexuosa* grasslands; lowland heavy metal grasslands resembling to typical natural calaminarian grasslands

This paper presents an original and simple method to find indicator species and species assemblages characterizing groups of sites. While the new method is asymmetric like TWINSPAN, the first novelty is that it derives indicator species from any hierarchical or nonhierarchical site classification method. The second novelty of our approach is the way we measure the association between a species and a group of sites. The indicator value index is based only on within-species abundance and occurrence comparisons, without any comparison among species. The significance of each species indicator value is measured by a site randomization procedure.

We will compare the classical approaches mentioned above to the results obtained from this new method. The data set that will be used as an example concerns the structure of natural carabid beetle assemblages from open terrestrial habitats in southern Belgium.

MATERIALS AND METHODS

Carabid beetle data set

Many authors (Thiele 1977, Den Boer 1977, Desender 1986*a–d*, Eyre and Luff 1990*a*, *b*, Turin et al. 1991, Lindroth 1992, Luff et al. 1992) consider that carabid beetles, and invertebrates in general (Kremen et al. 1993), are sensitive and accurate indicators of the state of the environment. Among invertebrates, physical environmental factors seem to shape the species assemblages more than biological relationships such as competition, predation, and parasitism (Schoener 1986). Because they have been widely studied, carabid beetles are a good model group to reveal hierarchical interactions and structures at different spatial scales.

In southern Belgium, there is a strong need for information about the state of natural ecosystems, especially for invertebrates. Sixty-nine locations have been selected, representing a wide range of natural and subnatural open habitats. These 69 locations cover almost all combinations of a moisture and an alkalinity gradient. They represent nine habitat types defined and chosen a priori: fringes of ponds, swamps, raised mires, peat bogs, sandy heathlands, loamy-gravelly heathlands, alluvial meadows, xeric chalky grasslands, and heavy-metal-contaminated grasslands (Table 1). Most of the selected sites are generally considered to be the best and sometimes the last remnants of the corresponding natural and subnatural habitat types in southern Belgium: 80% of them have the status of Natural Reserves.

Pitfall traps were used to sample carabid beetle assemblages. According to Thiele (1977), it is the best technique available to assess species assemblages at any one site. Pitfall sampling carried out during one year, or at least along the whole activity season of the species, gives a reliable estimate of the densities of active adult specimens (Baars 1979) or is a good predictor of mean biomass and of the species quantitative importance (Loreau 1992). But, following Den Boer (*unpublished manuscript*) and Turin et al. (1991), pitfall abundances could only be used to compare the relative abundances of species in several sites (*Q* mode), not for comparisons among species (*R* mode). Because each species has characteristic catch parameters, such as activity pattern, life history traits, behavior, catchability, etc. (see Luff 1975), the dominance scale obtained from pitfall sampling is probably far from the real dominance scale in the wild (Halsall and Wratten 1988, Niemela and Spence 1994). For this reason, Den Boer advises against interpreting the equitability or the diversity of the species assemblages, or submitting the resulting abundance data to multivariate techniques of data analysis. Although we recognize potential problems in interpreting diversity values calculated from sampling data, we agree with Luff et al.

(1992) that multivariate analysis of the data from sampling programs that have been conducted under standardized conditions can lead to meaningful results. Indeed, we are interested in the information provided by species abundance differences between sites, and by the way in which these differences are reproducible, or not, when the ecological conditions are the same or different. Our method only uses within-species comparisons to construct the site typology, and the indicator value index for any given species is independent of the other species.

Ten pitfalls filled with 5–10% formaldehyde solution were placed at 54 sites during 2 yr (1986–1987), and during 1 yr at the remaining 15 sites (1991). The pitfall traps used were made of two cut-out plastic bottles, one 1.25 L, the other 1.5 L, fitted together and embedded in the soil to the rims. Pitfalls in a transect were 3–5 m from one another. Ten pitfalls per site, as used in this study, were sufficient to provide reliable information. Indeed, species rarefaction curves using 500 random permutations of these 10 pitfalls (see Ferry 1976) show that an 11th pitfall trap would likely have captured a new species at only three sites of our sampling program.

The 123 available year-catch cycles provide local distribution information about 39 984 specimens belonging to 189 species. This represents more than half the carabid species recorded in Belgium since 1950. All data were standardized to constant sampling intensity. The number of pitfall traps effectively active during the whole-year cycles were, in practice, <10, due to several kinds of uncontrollable reasons, such as floods in submerged habitats, fires in xeric grasslands, hard and long frosts in the high Ardennes area, and also game and human damage. In order to decrease the importance of vagrant specimens in the analysis, all abundances equal to one specimen were eliminated, but only when that one specimen represented $<5\%$ of that species sum of catches. This reduces the number of specimens in the analysis to 39 356 without eliminating any species. As the number of rare species is always large in carabid beetle assemblages, all species represented by <20 specimens were not taken into account in this analysis. Therefore, the studied data set contains 38 811 specimens belonging to 97 species in 123 year-catch cycles. This data set is characterized by 10 570 zeroes, corresponding to 88.6% of the records. This situation is characteristic of many species presence or abundance data sets studied by ecologists.

Taxonomy is based on Desender (1985). Specimens of the recently discovered *Calathus melanocephalus* species complex (Aukema 1990) have not been verified.

Classical methods

Ordinations by CA and DCA were computed using the CANOCO program of ter Braak (1988). For principal coordinate analysis (PCoA), we used a program written by P. Legendre implementing the Two-Way Weighted Summation algorithm described by ter Braak (1987). Ordination of the sites by PCoA was obtained from a similarity matrix computed using the Steinhaus coefficient, also called the Bray and Curtis coefficient (Legendre and Legendre 1983, 1984*a*), calculated on natural log-transformed data. Ordination of the species was carried out in the same way, also from log-transformed data.

To cluster species and sites, we used a nonhierarchical clustering procedure, *k* means. This method produces *k* groups, the value of *k* being decided by the user, after an iterative procedure of object reallocation whose aim is to maximize the among-group variability, as described by MacQueen (1967) and Späth (1980). The computation was repeated 250 times, after a random reallocation of sites to the initial groups used as the starting configuration of each run; this minimizes the risk of a local minimum (instead of the global minimum sought) of the overall sum of squares criterion, which is the sum of the within-group sums of squares. We applied the *k*-means method to the species or site coordinates on the first 20 PCoA axes of the Steinhaus similarity matrix; *k*-means clustering can be computed from a distance matrix, but since the program available (program KMEANS of Legendre and Vaudor 1991) does not allow it, we used PCoA species or site coordinates and let the *k*-means method minimize the within-group sums of squares of the derived ordination axes. This step also offers the advantage of filtering out the last few ordination axes, which represent mostly noise after $>90\%$ of the variance of the original data had been preserved in the first 20 ordination axes. Since the clusters so produced are not forced to be hierarchically nested, the actual presence of a hierarchical structure in the data can be verified.

The Indicator value method

The first step in our approach (Fig. 2) consists of obtaining a classification of sample units (localities, sites, times, etc.) using one of the classical methods of data analysis or any other appropriate technique, for instance, the ''Chronological clustering'' method of Legendre et al. (1985), when dealing with a time series or spatial transect of multispecies data. In our example, we obtained a typology from the *k*-means method, using the same procedure as described in the previous paragraph.

The second step is to identify indicator species corresponding to the various groups of the site typology. Indicator species are defined as the most characteristic species of each group, found mostly in a single group of the typology and present in the majority of the sites belonging to that group. This duality, which is of ecological interest, is rarely completely exploited in such analyses; often only the distribution of abundances in the various groups is used. In these cases, species occupying only one or two sites in one habitat group and

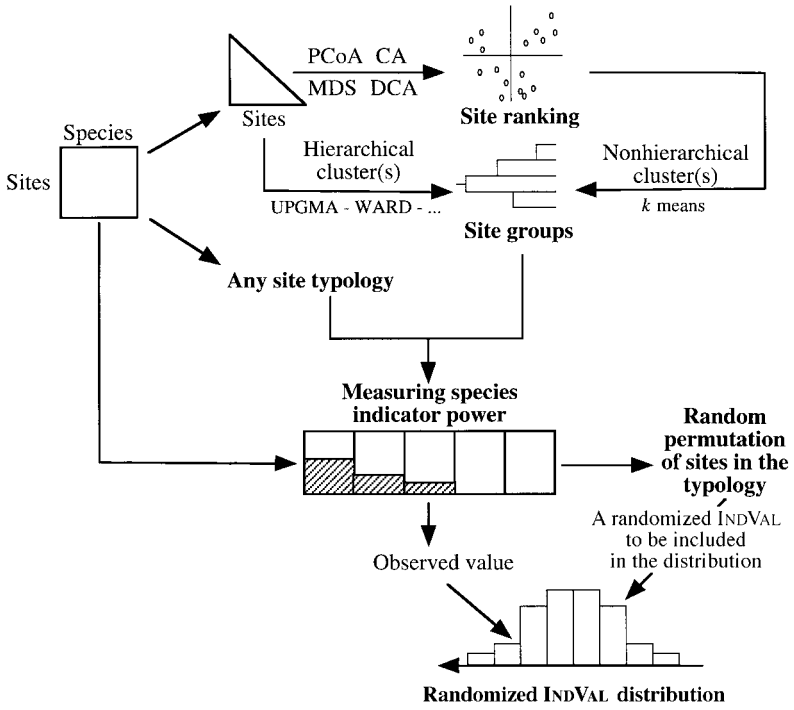

FIG. 2. Diagram of the analysis steps for the indicator value method. See Fig. 1 for explanation of abbreviations.

present only in that group (rare species) receive the same indicator value as species occupying all sites of that habitat group and found only in that group. There is an important difference between these two species, however. The first one is an asymmetrical indicator: its presence cannot be predicted in all sites of one habitat, but contributes to the habitat specificity. The second species, on the contrary, is a true symmetrical indicator: its presence contributes to the habitat specificity and one can predict its presence in all sites of the group.

To construct a symmetrical index of indicator value, we first selected two common measures of diversity that are commonly used to measure ecological niche breath. The first one is the Berger-Parker index (Hill 1973, Magurran 1988), which is simply the largest relative frequency of a species found over the various site clusters; one computes how many individuals of a species are found among the sites forming each group, expresses the results in relative frequencies, and takes the maximum of these relative frequencies across clusters. The second is the Shannon-Weaver exponential index exp(H′). With either index, a species that is only present in a few sites of a single-site cluster has the same indicator value as another species present in all sites of that same cluster. To take into account the second part of the symmetric indicator species definition in the expression of the indicator value, a new index was developed. For each species i in each site group j, we computed the product of A_{ij}, which is the mean abundance of species i in the sites of group j compared to all groups in the study, by B_{ij}, which is the relative frequency of occurrence of species i in the sites of group j, as follows:

$$A_{ij} = \text{Nindividuals}_{ij}/\text{Nindividuals}_{i.}$$

$$B_{ij} = \text{Nsites}_{ij}/\text{Nsites}_{.j}$$

$$\text{INDVAL}_{ij} = A_{ij} \times B_{ij} \times 100,$$

where INDVAL is the Indicator Value of species i in site cluster j. In the formula for A_{ij}, which is a measure of specificity, Nindividuals_{ij} is the mean number of individuals of species i across sites of group j, while $\text{Nindividuals}_{i.}$ is the sum of the mean numbers of individuals of species i over all groups. We use the mean number of individuals in each group, instead of summing the individuals, because this removes any effect of the number of sites in the various site groups, and of differences in abundance among sites belonging to the same group. A_{ij} is maximum when species i is only present in cluster j. In the formula for B, which is a measure of fidelity, Nsites_{ij} is the number of sites in cluster j where species i is present, while $\text{Nsites}_{.j}$ is the total number of sites in that cluster. B_{ij} is maximum when species i is present in all objects of cluster j. Quantities A and B must be combined by multiplication because they represent independent informations about the species distribution. Final multiplication by 100 produces percentages.

The indicator value of a species i for a typology of sites is the largest value of INDVAL_{ij} observed over all groups j of that typology:

TABLE 2. Test case example: abundance of three species in 25 sites divided into five clusters.

Species	Group 1					Group 2					Group 3					Group 4					Group 5				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	3	3	3	3	3	3	3	3	3	3
B	8	8	8	8	8	4	4	4	4	4	6	6	6	6	6	4	4	2	0	0	0	0	0	0	0
C	18	18	18	18	18	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$$\text{INDVAL}_i = \max[\text{INDVAL}_{ij}].$$

This index is maximum (=100%) when the individuals of species *i* are observed in all sites of only one site group. A numerical example is presented below to illustrate the calculations.

A random reallocation procedure of sites among site groups is used to test the significance of INDVAL_i (Edgington 1987). We used 250 permutations in the application below. The significance can first be evaluated by the difference between the observed value and the mean of those obtained from the random permutations, weighted by the standard deviation of the values obtained randomly, turning it into a regular *z* statistic, assuming approximate normality of the distribution of the permuted statistics. Another way is simply to note the rank of the observed value in the randomly generated distribution ordered in decreasing order; this produces a regular permutational probability.

This INDVAL index can be computed for any given site typology, or for all levels of a hierarchical classification of sites. This allows the more eurytopic species (i.e., species with large niche breadth that are present in a variety of habitats) to eventually react to the typology at their characteristic larger scale and thus take into account the hierarchical structure of the species distribution. Since species receive an indicator value for each level of the hierarchy, a distribution of these values can be plotted for a given species, in order to choose the most interesting clustering level. Generally, the level characterized by the largest value is chosen as the most appropriate for that species. If the clustering was also spatially constrained (i.e., sites belonging to a group are forced to be spatially contiguous: Legendre and Legendre 1984*b*, Legendre 1987, Dufrêne and Legendre 1997), the clustering level identifies the scale corresponding best to the spatial distribution of the species; if the ''best'' clustering level is the first (trivial partition where all sites belong to a single group) or the last (trivial partition where each site forms a distinct cluster), the spatial distribution of the given species probably corresponds to a larger or a smaller spatial scale than investigated by the study's sampling design. Sometimes, the value increases first abruptly and thereafter very slowly, the difference between successive values becoming <5%. Generally, large variations correspond to clustering steps where the group of sites of which the species is an indicator has been subdivided. The small variations observed, when no such subdivision occurs, are induced by variations in the mean abundances within site groups, used to compute the first part of INDVAL. We propose to use, as our ''best'' clustering level for a species, the level corresponding to the higher indicator value observed when the site group, of which the species is an indicator, is subdivided.

Table 2 and Fig. 3A present a simple test case describing the distribution of three species in 25 sites. The first species (A) is widespread in all sites; its indicator value (INDVAL) is maximum for the trivial partition where all sites belong to the same group; its INDVAL index decreases as more and more groups are formed (Fig. 3B). For the two-group clustering level, for example, the value is 60.9% (Table 3). For the first cluster composed of sites 1–15, INDVAL is computed as follows: A_{ij} = (70 individuals/15 sites) divided by [(70 individuals/15 sites) + (30 individuals/10 sites)] = 0.609; because this species is present in all 15 sites of the first group, B_{ij} = 1; so INDVAL(in percentage) = 0.609 × 1 × 100 = 60.9%. For the second cluster comprising sites 16–25, similar calculations produce a value of INDVAL equal to 39.1%. The value of INDVAL for this two-group partition is max(60.9%, 39.1%) = 60.9% for species A.

Species B reaches a maximum for the two-group level; 90% of its individuals belong to the first cluster, containing all sites of groups 1, 2, and 3. Afterwards, its value decreases slowly, to reach 75% at the three-group level, and more quickly thereafter. The slow decrease is explained by the use of mean abundances to compute A_{ij}: at the two-group level, the mean for the second group containing the last 10 sites is 10 individuals/10 sites = 1, so that the sum of the means is 6 + 1 = 7 for the denominator of A_{ij}; at the three-group level on the contrary, these 10 sites are split up and we have 10 individuals/5 sites = 2 for group 4, and 0 individuals/5 sites = 0 for group 5, so that the sum of the means is 6 + 2 + 0 = 8 for the denominator of A_{ij}. When the main group is subdivided, INDVAL decreases more quickly.

The last step is to present the results. A first way to display the indicator values is to write down, as in TWINSPAN, the species that are the best indicators on the nodes of the hierarchical dendrogram of sites. A second way is to show the hierarchical information for all species in a two-way table. The best solution is probably to flatten out the pyramid representing the hierarchy of site clustering levels in a site group-by-species two-way table (Fig. 4). Site groups are in col-

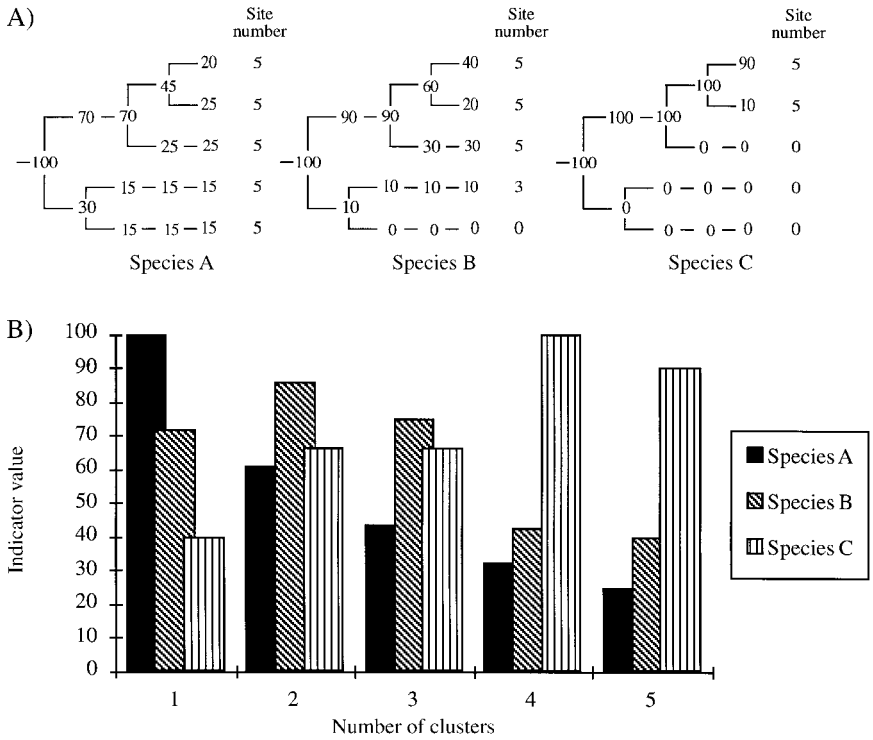

FIG. 3. Test case results. (A) Distribution of abundances of the three species in the five clustering levels. (B) Bar chart showing the decrease (species A) or increase (species C) of the indicator values when the site groups are subdivided.

umns, species in the rows. In the center rows of that table are found the species present in all site groups, as well as species that always have very low indicator values at all subsequent clustering levels; these last ones are the species that do not follow the hierarchical pattern. For the two-site group level, two groups of indicator species are created, one for each site group. The first group of species is placed in the table before the species present in all habitats, the other one after. In each of these groups, there are species that have their higher indicator values at this step and very low values subsequently. This operation is repeated for each clustering level. New species groups are always placed before and after the species group indicative of the previous node; this can produce very good species gradients when site groups (in columns) are ranked following the main ecological gradient. An example is provided with the results.

TABLE 3. Test case results: species indicator values and *z* statistics for the five clustering levels.

Species	Number of clusters				
	1	2	3	4	5
Indicator value					
A	100.0	60.9	43.8	32.3	25.0
B	72.0	85.7	75.0	42.9	40.0
C	40.0	66.7	66.7	100.0	90.0
z statistic					
A		7.17	5.70	4.25	2.88
B		6.20	5.60	2.19	2.84
C		4.21	3.14	6.59	6.39

RESULTS

Ordinations

CA.—The first four axes of the correspondence analysis, based upon 97 species, explain 23.2% of the total data set variance (Fig. 5). The first one (6.8%) clearly contrasts wet sites to drier ones along a more or less regular gradient. Moreover, two of the a priori habitat groups are subdivided in two additional groups, following the soil moisture level. For sandy heathlands, the wet sites are those that were temporarily flooded in the Campine area and, for mineral heathlands, those that are isolated in peat bogs. The second axis (6.1%) isolates the chalky grasslands, the third (5.3%) distinguishes contaminated grasslands from other drier sites, and the fourth (5.0%) isolates two peculiar fringes of ponds belonging to the *Littorellion* phytosociological association. The arch effect, expected when the first axis is a gradient, is only observed on the fifth axis.

DCA.—Fig. 6 presents the results of the DCA ordination based on 26 segments, a value recommended by Hill and Gauch (1980). The first four axes explain 19.2% of the total data set variance. The second axis (5.7%) is very similar to the second one of the CA ordination: a strong opposition between, on the one hand, chalky grasslands and, on the other hand, xeric heathlands and contaminated grasslands. This axis is

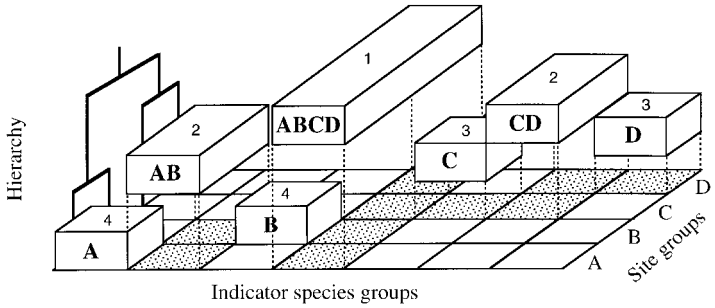

FIG. 4. Representation of a classical two-way table (species in rows, site groups in colums) showing the hierarchy of indicator species groups that determines the hierarchical clustering of sites.

stable in all analyses, performed with a wide range of numbers of segments. It is not the case for subsequent axes, which differ strongly from one value of number of segments to another. We did not observe high correlations between noncorresponding axes of different analyses, which could appear when several axes have nearly similar eigenvalues (Jackson and Somers 1991). These results suggest that the structure revealed by the third axis is not stable.

PCoA.—The first axis (25.2% of the variance) clearly separates the different habitats (Fig. 7) with, on the left, the wet sites such as the swamps and on the right, the drier sites (chalky and heavy metal contaminated grasslands, heathlands). The second axis (12.3%) is probably the result of an arch effect and should be interpreted like the first one. The upper panel shows the 80% confidence ellipses based on the a priori habitat groups, while the lower panel shows the same ordination scattergram with 80% confidence ellipses of site groups obtained with the *k*-means partitioning method.

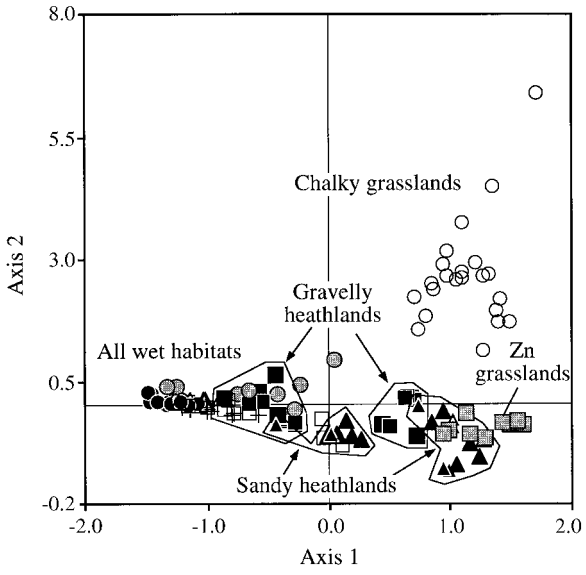

FIG. 5. Plot of the first two axes of the carabid beetle correspondence analysis. The symbols used are based on the a priori habitat typology (Table 1). All wet habitats are on the left; dry ones are on the right. Polygons surround two subgroups of gravelly and two subgroups of sandy heathlands showing differences based on the humidity level.

Clusterings

Q-mode analysis: site typology.—Results of the *k*-means procedure, for $k = 2$ to 10, show a very strong hierarchical structure that allows us to represent the site groups in the form of a hierarchical dendrogram (Fig. 8). Reallocations between the independent steps (numbers of clusters) are rare and concern only very peculiar sites. The most important reallocation concerns group 6, which we call ''temporary flooded heathlands'', which is associated with wet habitats for *k* values 2, 5, and 6. For other *k* values, this group is associated with drier habitats. Table 4 (see also Fig. 7) presents the correspondence between the a priori habitat typology (the ecologist's vision of these ecosystems) and the *k*-means typology for $k=10$ (the carabid beetles' vision).

The main structure ($k = 2$) contrasts dry (groups 1–6) and wet habitats (groups 7–10). Chalky grasslands

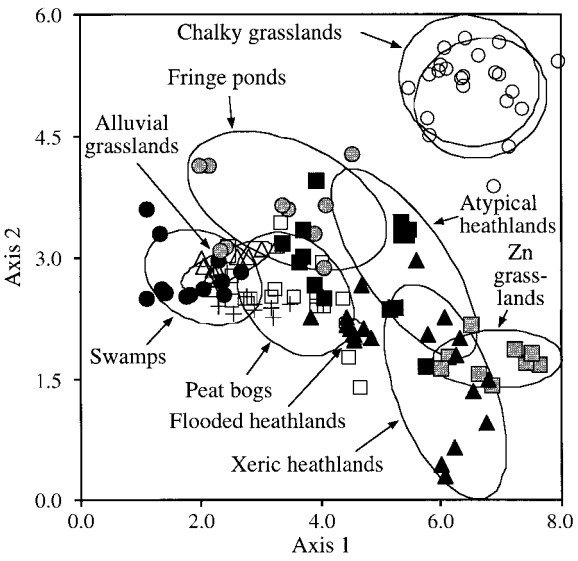

FIG. 6. Plot of the first two axes of the carabid beetle detrended correspondence analysis based on 26 segments. The symbols used are based on the a priori habitat typology, and 80% confidence ellipses represent the 10 site groups obtained with the *k*-means clustering method.

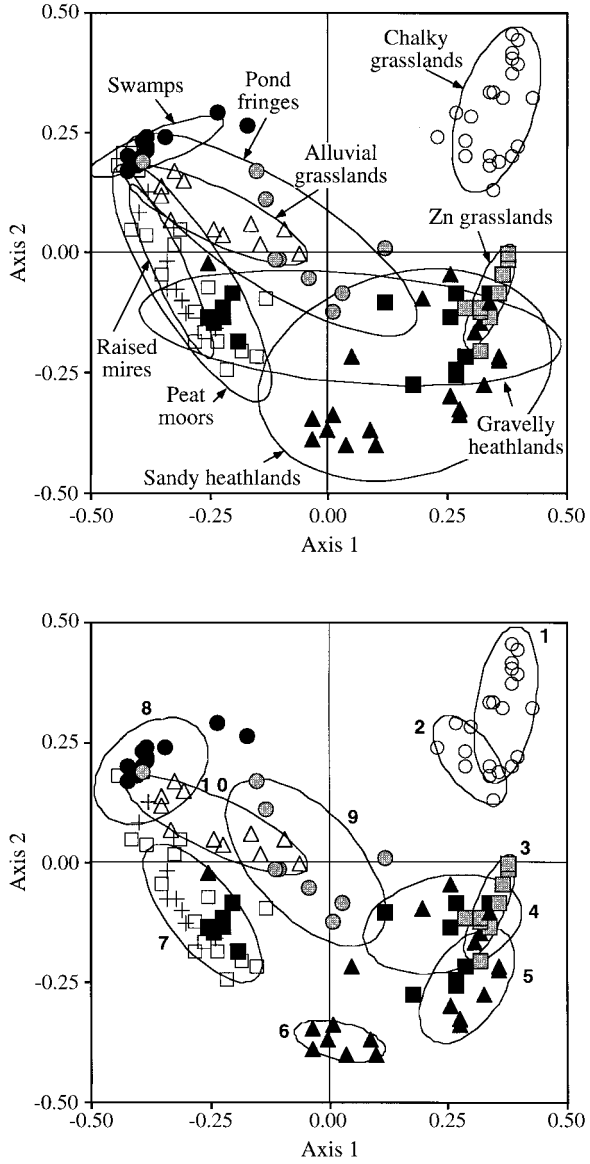

FIG. 7. Plots of the first two axes of a principal coordinate analysis based on a Steinhaus similarity index calculated on natural logarithms of abundances. In the upper plot, the symbols and 80% confidence ellipses are based on nine a priori habitat groups. In the lower one, the symbols are still based on the a priori habitat typology, but the 80% confidence ellipses are those of the 10 site groups obtained from the *k*-means clustering method.

(groups 1 and 2) are the most peculiar habitat at the three-group level. This group is subdivided in two, group 2 corresponding to six year-catch cycles coming from the lower Meuse basin, in an area near the Netherlands. Four of them are more mesic chalky grasslands, the last two are chalk scores isolated among grasslands. Contaminated grasslands are all clustered in group 3 and show more entomological affinities with heathlands than with chalky grasslands. Among heathlands (groups 4–6), the a priori typology mainly based on soil structure (sandy, mineral and peaty) is not realized. A first set of yearly catches (group 4) includes heathlands that are isolated in forest habitats or in which tree recolonization was important. In these sites, the herbaceous layer is at least as important as the *Calluna* stratum. All the year-catch cycles belonging to group 5 are xeric habitats, with sandy or loamy-graveled substratum. The *Calluna* stratum dominates, but there are also some patches devoid of any vegetation. The third set of heathlands (group 6) is very peculiar, because these seven year-catch cycles all come from the Campine region. These sites are occupied by either *Calluna vulgaris* or *Erica tetralix* vegetation and by patches without any vegetation, like other xeric heathlands, but they are prone to flooding during the whole winter season. These winter floodings explain why group 6 was sometimes allocated to the wet main group of habitats, for some values of *k.*

Wet sites fall into three year-catch cycle sets. The first set (group 7) includes most of the wet and oligotrophic sites: peat bogs and raised mires, but also six year-catch cycles on loamy-gravelled heathlands. This association is explained by the isolated nature of these six sites in large peat-bog areas dominated by herbaceous vegetation (mainly *Deschampsia flexuosa*) alternating with patches of *Vaccinium* spp. and *Calluna vulgaris.* Group 8 included all the year-catch cycles in swamps and other wet sites. The third set of wet sites comprises most of the wet and eutrophic sites, together with the pond fringes and the alluvial grasslands. These two habitats become completely separated when $k = 10$.

R-mode analysis: species typologies.—For low *k* values, the species groups are well separated but, at the seven-group step, a peculiar grouping places together species (e.g., *Agonum ericeti* or *Trechus rivularis* and *Callistus lunatus* or *Amara praetermissa*) that have nothing in common. The first two species are found in peat bogs, but never together; *C. lunatus* and *A. praetermissa* are present at only one site each, but not the same one (Table 5). The only feature common to these species and to others associated with them is their presence in a small number of sites. The computed relationships between species seem to be confounded by an undesirable structure linked with species frequencies. Indeed, the second axis is correlated with the number of occupied sites ($r_{\text{Pearson}} = 0.374$; $r_{\text{Pearson}} = 0.477$ with the logarithm of the number of occupied sites; $P = 0.001$; $n = 97$ species). Moreover, many species reallocations are observed from one step to the next, even though the number of random initial configurations has been increased to 500.

The six species groups can be associated with the habitats they dominate (Fig. 9). Differences between the *k* means and classical agglomerative clustering procedures concern species such as *Pterostichus lepidus, Cicindela campestris, Carabus nemoralis,* and *Amara aulica,* which are present in at least two habitat groups

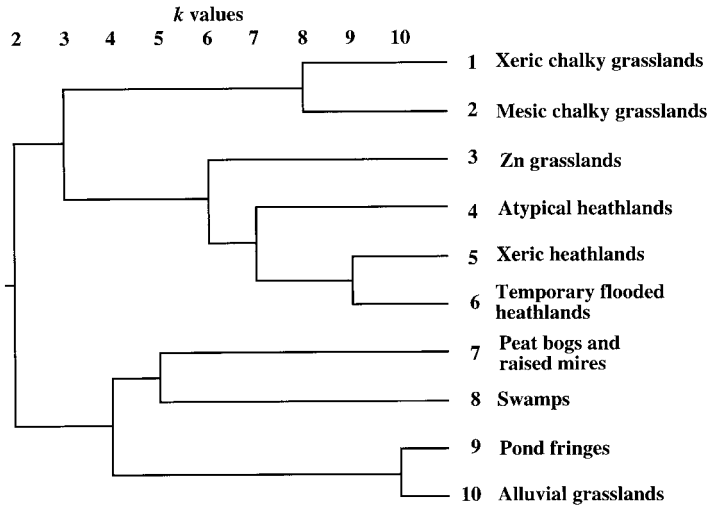

FIG. 8. Hierarchical dendrogram built with the results of the *k*-means reallocation clustering method. Reallocations are scarce and the main changes concern the ''temporary flooded heathland'' (group 6), which are allocated to wet habitats at the two-, five-, and six-group level and to dry habitats at the other clustering levels. Table 4 shows the correspondence between the a priori habitat and the *k*-means typology.

and are generally located at the cluster's border. These species are more eurytopic than the others but are nevertheless sensitive to ecological factors common to several habitats.

Indicator species

TWINSPAN.—This procedure was used with the cut-off levels of 0, 2, 5, 10, and 20% to construct pseudospecies, as proposed by Hill (1979). The results of year-catches classification with indicator species abundance levels are presented in Fig. 10. The first division separates dry from wet sites, as in previous analyses. The indicator species for wet sites are, in order of decreasing power, *Pterostichus diligens, P. rhaeticus, Dyschirius globosus, Agonum fuliginosum,* and *P. minor,* which are present in almost all these year-catches. For drier sites, only the presence of *Pterostichus cupreus* is indicative. The second division divides the wet sites into wet heathlands and other wet sites (fringes of ponds, alluvial grasslands, swamps, and raised mires). The third one separates chalky grasslands from dry heathlands and contaminated grasslands. In the next step, each of the four groups is subdivided in site subsets. The a priori habitat typology is largely preserved, but differences are observed. Some of the xeric sandy heathlands are associated with contaminated grasslands. For wet sites, fringes of ponds and alluvial grasslands on the one hand, and swamps and raised mires on the other, are not separated. Moreover, the indicator abundance levels are rather low. The complete table, not presented here because of its size, shows several small species groups in the center of the gradient. Extremes are thereby characterized by groups with many species.

The indicator value method.—The new method is computed for each step of the hierarchical structure of the *k*-means site clustering. For several species (Fig. 11a), the indicator value increases regularly or suddenly as the number of groups increases. These are species that are stenotopic (i.e., species with small niche breadth) and indicator of only one or two habitat groups. On the other hand (Fig. 11b, upper), there are species that are more eurytopic, typical of higher order groups; they show a more or less regular decrease of their indicator value, as the sites where they are abundant are split among different site groups (e.g., *Pterostichus diligens, P. rhaeticus*). For other species (Fig. 11b, lower), the patterns are uni- or bimodal, showing

TABLE 4. Confusion table comparing the a priori habitat typology to the *k*-means typology.

Habitats (a priori)	Habitats (*k*-means)										Sum
	1	2	3	4	5	6	7	8	9	10	
Chalky grasslands	**16**	**6**									22
Zn grasslands			**10**								10
Sandy heathlands				5	7	7	1				20
Other mineral heathlands				5	3		6				14
Peat bogs							**13**	3			16
Raised mires							6	2			8
Swamps								**12**			12
Pond fringes				1				2	**7**		10
Alluvial grasslands										**11**	11
Sum	16	6	10	11	10	7	26	19	7	11	123

Note: Values in boldface indicate good correspondence between the two classifications.

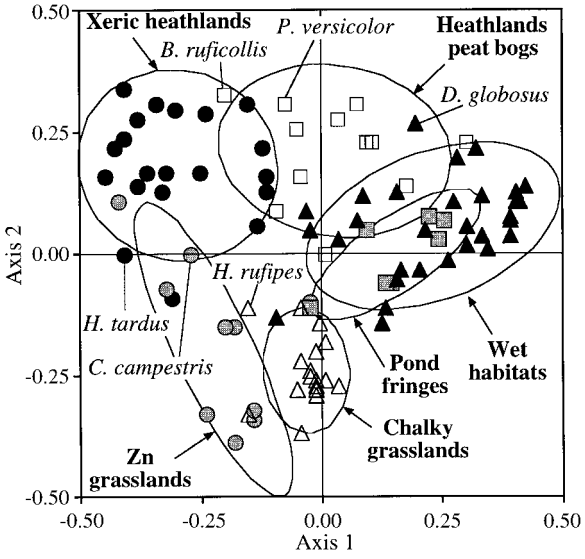

FIG. 9. Plots of the first two axes of a principal coordinate analysis based on Steinhaus similarity index calculated on natural logarithms of abundances. The 80% confidence ellipses and symbols are based on a *k*-means reallocation clustering of species. Six species groups are easily identifiable, but there are several reallocations between clustering levels. Some species belonging to more than one group are indicated.

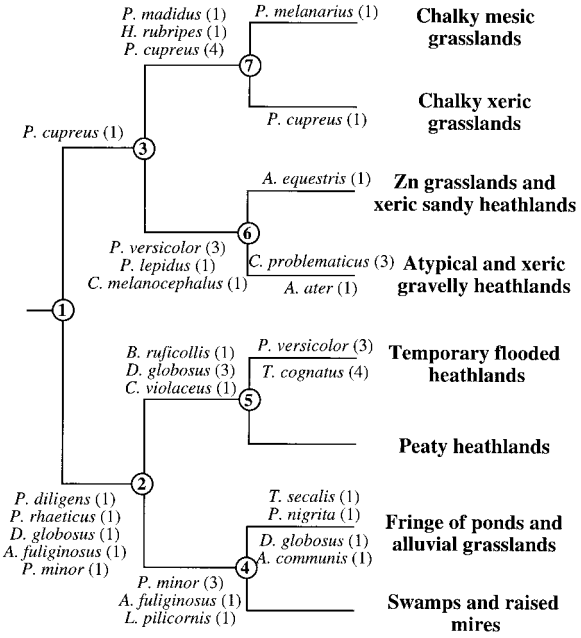

FIG. 10. Dendogram representing the TWINSPAN classification of the year-catch cycles. The indicator species relative abundance levels are expressed on an ordinal scale (1, 0–2%; 2, 2–5%; 3, 5–10%; 4, 10–20%; and 5, 20–100%).

that the species are characteristic of intermediate levels of the hierarchy (e.g., *Harpalus rubripes*, *Bembidion lampros*, *Pterostichus versicolor*).

These results allow us to determine the characteristic species for the different levels of the hierarchy. We arbitrarily chose a threshold level of 25% for the index, which supposes that a characteristic species is present in at least 50% of one site group and that its relative abundance in that group reaches at least 50%. If one of the two values reaches 100%, the other is always greater than or equal to 25%. Fig. 12 presents, on the dendrogram of sites, all the species that have an index value that is significant ($P \leq 0.01$) and >25%. Many species are typical of lower hierarchy levels, whereas others typify higher hierarchy levels.

With an index value of 97%, *Pterostichus rhaeticus* is one of the best examples of a characteristic species. Almost all the *P. rhaeticus* individuals are in one group (wet sites) and are present in all these sites. This species is associated with *Pterostichus diligens*, *Agonum fuliginosum*, and *Pterostichus minor*. The heterogeneity of dry sites is revealed by lower values, which increase when the chalky grasslands are isolated. *Pterostichus madidus* and *Harpalus species* are the two differential species of these chalky grasslands. All the contaminated grasslands are characterized by *Amara equestris* and several other species. For heathlands, the indicator value of all characteristic species grows regularly as the site group is subdivided and better characterized. Almost all the final groups are characterized by several indicator species. Only peat bogs and swamps are site groups with only one species possessing an indicator value >25%. All these sites are indeed dominated by species with a larger ecological amplitude, such as *Pterostichus rhaeticus*, *P. diligens*, and species that are indicators for other wet habitats such as *Agonum fuliginosum*, *Pterostichus minor*, and *Loricera pilicornis*. Detailed results are presented in Table 5. Fig. 13 presents the method of species group ordination, taking into account the hierarchy of site groups that has been followed to construct this two-way table.

Comparison of this species typology with clusters produced by *k* means and other classical clustering methods shows many similarities for the lower hierarchical levels, i.e., when the groups are numerous and well featured. For higher levels, i.e., for species present in several habitats, the clustering results were never as clear as the ones presented in Fig. 12.

DISCUSSION

Pitfall sampling

If pitfall sampling was not a reliable method, a significant amount of unexplained variation should appear in the site similarities. The strong habitat structure and the very great similarities between the two successive year-catches (96% of the year-catch cycle couples are in the same *k*-means group) support the potential of carabid beetles pitfall sampling to describe terrestrial habitats. Peculiar differences are revealed, but most of them can be explained by local features. Only the sampling of peculiar habitats like xeric chalky grasslands seems problematic, in which the number of specimens obtained is very low and a great heterogeneity in spe-

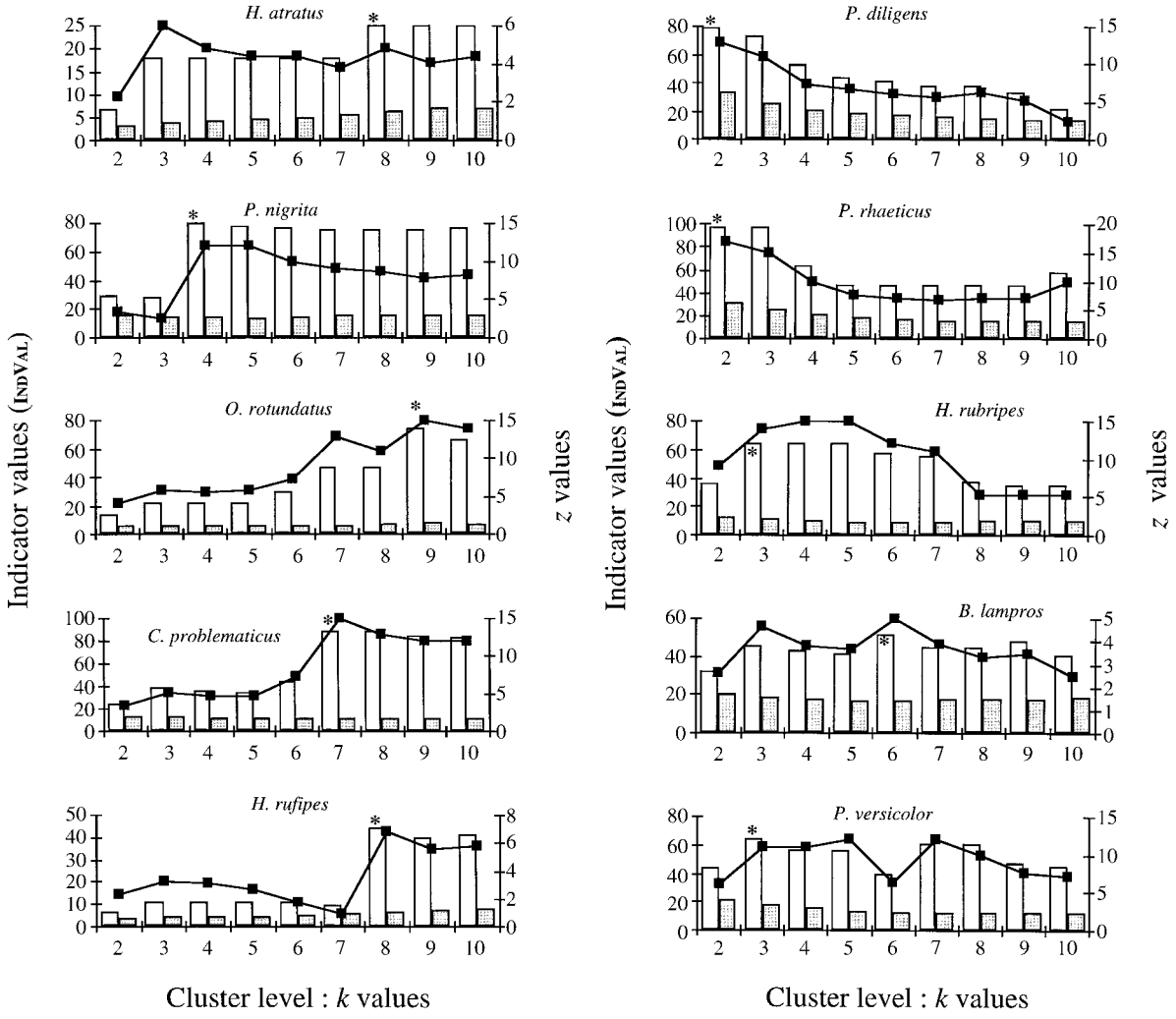

FIG. 11. Plot of the species indicator values obtained for the successive hierarchical clustering levels for selected species. Indicator values are represented by histograms (white bars for observed values and gray bars for the means after 250 random permutations of the sites). Differences (black squares) are expressed as *z* values. Assuming normal distributions for the randomized indicator values, any *z* value larger than 1.65 indicates significance at approximately the 0.05 level (one-tailed test). A high increase in observed values indicates that a species has a high indicator power for one group of sites at the corresponding clustering level (see Fig. 7). An asterisk identifies the highest INDVAL index observed among clustering levels.

cies local presence is observed. Rarefaction curves based on the number of pitfalls do not always reach an asymptote; for one relevé, one specimen only was caught in each one of the 10 pitfall traps and these specimens belong to 10 different species. As we do not yet have data from other sampling methods, we cannot establish whether this heterogeneous spatial distribution structure is due to the xeric chalky grassland species or is generated by a sampling method inadequate for this particular habitat.

Comparing ordination results

Correspondence analysis is very sensitive to outliers or outlier groups. This is not an artifact because it is precisely its goal, but in our case this analysis quickly reveals local peculiarities to the detriment of main structures. The variance explained by the four first correspondence analysis axes is very low (23.2%), in comparison with principal coordinate analysis where >50% of the data set variation was explained by the four first axes. Moreover, as species do not all have the same ecological amplitudes, the search for correspondences between species and sites is not the best way to detect relationships between species and site groups. This is illustrated by the differences in frequency distributions of marginal presence or abundance sums of the dataset. The frequency distribution of the number of species per site generally follows a truncated bell-shaped curve when the sampling is exhaustive, whereas the frequency distribution of number of sites per species (range size) is completely asymmetric (many rare species and few common spe-

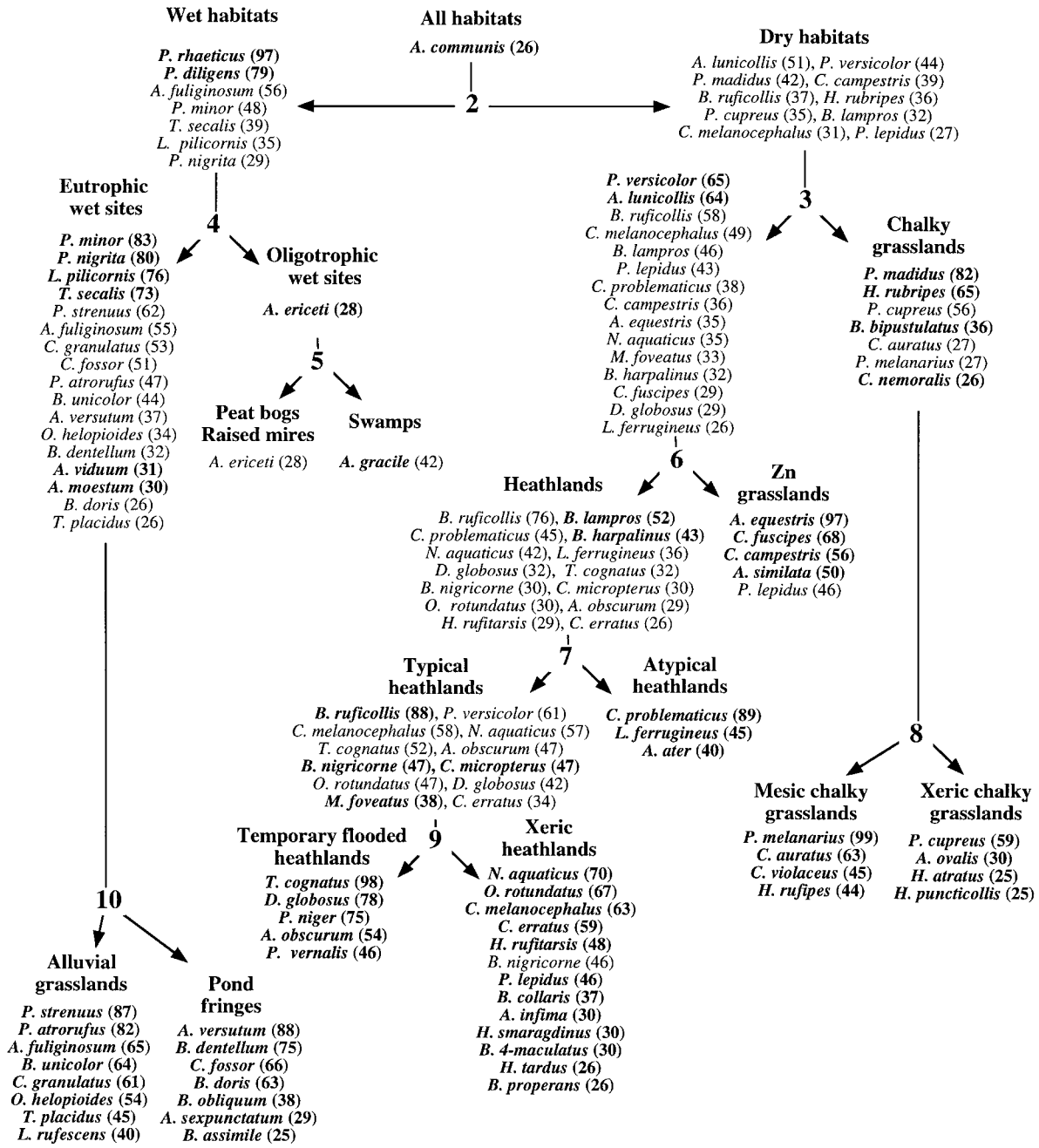

FIG. 12. Site clusters obtained with the *k*-means method (as in Fig. 7), but with the associated indicator species and indicator values in parentheses. All species with an indicator value >25% are mentioned for each site cluster where they are found, until they have reached their maximum indicator value. This maximum is indicated in bold.

cies; *J* shaped curves) or sometimes bimodal (see *Discussion*: *Regional distribution patterns and species assemblages* below). Also, since correspondence analysis weighs cell values by the marginal sums of abundances, such a pattern necessarily interferes with the search for species–site relationships.

An appropriate similarity index (the Steinhaus index in the present study) produces results that are theoretically less sensitive to these patterns, because similarities or differences among sites are the sums of species-by-species similarities or differences, without consideration of the similarities or differences among species or their weighted frequencies, as in correspondence analysis. Also, the importance of the arch effect is not clear; it certainly plays a role because it tends to bring together, along the first axis, sites that are very different. However, an advantage is that it separates sites that have affinities with the two extremes of the gradient

TABLE 5. Two-way indicator table showing the species indicator power for the site clustering hierarchy. The column headings correspond to those of Fig. 13, from alluvial grasslands (A) to xeric chalky grasslands (J). The species that have an indicator value >55% are symmetrical indicators. The other species must be considered accidental or anecdotal, although they have their maximum indicator value for this cluster level group (asymmetrical indicators). The INDVAL column indicates the species indicator value for the corresponding clustering level, which is the maximum indicator value observed in all the clustering hierarchy. Boldfaced numbers represent the main data set structure.

Species	IND-VAL (%)	A	B	C	D	E	F	G	H	I	J
Alluvial grasslands (A)											
P. strenuus	87**	**375/10**	7/2	2/1	2/1		3/1	2/1			
P. atrorufus	82**	**845/9**									
A. fuliginosus	65**	**435/11**	22/3	326/16	43/6						
B. unicolor	64**	**467/8**	2/1		26/3		40/2	2/1			
C. granulatus	61**	**668/7**	21/3								
O. helopioides	54**	**62/8**		36/2							
T. placidus	45**	**24/5**									
L. rufescens	40**	**58/6**		17/3	10/4			6/2			
T. rivularis	23	**39/3**			18/5						
T. laevicollis	15	**40/2**	2/1		2/1			3/1			
C. coriaceus	12	**22/3**			2/1			18/1			8/2
Eutrophic habitats (A+B)											
P. minor	83**	**541/9**	**376/8**	157/11	21/3			2/1			
P. nigrita	80**	**57/9**	**539/7**	35/3		3/1		3/1	5/1		9/2
L. pilicornis	76**	**212/10**	**26/6**	51/7		7/2				2/1	
T. secalis	73**	**1488/10**	**186/6**		454/11	12/5	112/3	11/1			
A. viduum	31**	**17/5**	**14/4**	40/2							
A. moestum	30**	**44/4**	**5/2**	5/1							
P. oblongopunctatus	22**	**33/3**	**40/2**		27/3	4/1	4/2				
A. muelleri	21**	**12/3**	**14/4**		2/1		23/2	12/4			
Pond fringes (B)											
A. versutum	88**		**285/7**								
B. dentellum	75**		**259/6**								
C. fossor	66**	52/3	**153/7**			9/2	2/1				
B. doris	63**		**76/5**								
B. obliquum	38**		**28/3**								
A. sexpunctatum	29**		**59/4**			6/3	4/2	40/1			
B. assimile	25		**39/2**								
A. binotatus	19		**53/3**			2/1		2/1	55/2		3/1
Wet habitats (A+B+C+D)											
P. rhaeticus	97**	**1400/11**	**39/8**	**807/19**	**1107/24**		2/1	5/2			
P. diligens	79**	**556/11**	**631/5**	**311/17**	**902/26**	278/7		80/3			
Swamps (C)											
A. gracile	42**	23/2		**211/9**	3/1						
Oligotrophic habitats (C+D)											
A. ericeti	28**			**93/4**	**269/11**	54/3					
C. caraboides	19**			**14/3**	**20/7**	2/1		2/1			
Peat bogs and raised mires (D)											
No indicator species											
All habitats											
A. communis	26	**126/6**	**51/5**		**130/4**	**2/1**	**14/4**	**616/3**	**5/2**	**5/2**	**23/5**
N. palustris	11	**13/3**		**14/2**	**12/3**		**9/3**	**15/2**			
Temporary flooded heathlands (E)											
T. cognatus	98**	2/1			8/1	**352/7**	5/2				
D. globosus	78**	81/8	258/8	2/1	695/25	**1712/7**	115/4				
P. niger	75**	14/2	8/2		5/1	**87/6**	4/1	130/3			
A. obscurum	54**		6/1			**457/4**	35/4				
P. vernalis	46**		52/4			**42/5**	6/2				
Typical heathlands (E+F)											
B. ruficollis	88**		3/1		181/14	**126/7**	**1479/10**	52/6		2/1	
C. micropterus	47**					**13/3**	**130/5**				
B. nigricornis	47**					**13/3**	**200/5**				
M. foveatus	38**					**80/3**	**138/6**	2/1	43/3	11/2	
C. arvensis	17	2/1	13/3		8/1	**22/1**	**27/3**				

TABLE 5. Continued.

Species	IND-VAL (%)	A	B	C	D	E	F	G	H	I	J
Xeric heathlands (F)											
N. aquaticus	70**					7/3	**158/8**	11/1	3/1		
O. rotundatus	67**					3/1	**88/7**				
C. melanocephalus	64**					97/3	**515/9**	7/2	70/4		
C. erratus	59**						**329/6**	7/1			
H. rufitarsis	48**						**398/5**	11/3	2/1		
P. lepidus	46**						**378/10**		440/6		
B. collaris	37**						**31/4**		3/1		
H. smaragdinus	30**						**81/3**				
A. infima	30**						**28/3**				
B. 4-maculatus	30**						**71/3**				
H. tardus	26**						**65/6**	2/1	3/1	31/2	48/3
B. properans	26		37/1				**116/3**				
H. latus	24		2/1				**24/4**	14/1			
T. obtusus	21	9/3					**11/3**				
H. anxius	18						**36/2**		3/1		2/1
Heathlands (E+F+G)											
B. lampros	52**	20/6	31/4		4/2	**124/3**	**600/6**	**41/7**	6/2	2/1	3/1
B. harpalinus	43**	2/1			2/1	**17/3**	**17/5**	**30/5**		2/1	
Atypical heathlands (G)											
C. problematicus	89**		14/2		13/5		29/4	**281/10**	3/1		2/1
L. ferrugineus	45**						25/5	**129/5**			2/1
A. ater	40**	2/1	130/2	2/1	16/3	15/3	20/2	**138/8**		3/1	87/8
Acid heathlands (E+F+G+H)											
P. versicolor	65**	32/4	178/6		10/3	**762/7**	**641/10**	**56/3**	**354/6**		6/3
A. lunicollis	64**	30/4	72/3	10/1	109/12	**64/5**	**54/8**	**524/9**	**383/7**	10/1	52/6
Zn grasslands (H)											
A. equestris	97**						23/3		**292/10**		
C. fuscipes	68**						48/2	18/2	**876/7**		9/2
C. campestris	56**	2/1			3/1		66/7	40/4	**350/7**	26/4	70/1
A. similata	50**		2/1						**682/5**		9/3
M. minutulus	19**						4/1		**80/2**		4/2
C. axillaris	19**								**18/2**		2/1
A. aenea	18	8/2					3/1	2/1	**106/2**	5/1	7/2
N. salina	18**						112/2		**304/2**		
T. 4-striatus	18**								**19/2**	4/2	
A. curta	17**								**146/2**		60/3
A. praetermissa	10								**25/1**		
Dry habitats (E+F+G+H+I+J)											
A. eurynota	5						2/1	3/1			19/1
Mesic chalky grasslands (I)											
P. melanarius	99**				2/1					**64/6**	
C. auratus	63**									**202/4**	27/2
C. violaceus	45**		2/1		108/12	14/3	12/3	17/1	2/1	**55/5**	2/1
H. rufipes	44**						6/1			**17/3**	
A. aulica	23	13/2								**9/2**	
Chalky grasslands (I+J)											
P. madidus	82**	8/2	29/4				23/3	31/4		**70/4**	**792/15**
H. rubripes	65**						32/3	2/1	22/1	**116/4**	**183/12**
B. bipustulatus	36**									**2/1**	**26/7**
C. nemoralis	26**	10/3	4/1				28/3			**2/1**	**53/7**
Xeric chalky grasslands (J)											
P. cupreus	59**	11/2	230/5		3/1		27/3	18/4	44/6		**467/16**
A. ovalis	30**		2/1								**61/5**
H. puncticollis	25**										**26/4**
A. atratus	25**										**23/4**
H. parallelus	19										**154/3**
M. piceus	19**										**29/3**
A. familiaris	17							2/1			**41/3**
N. germinyi	13										**38/2**
H. tenebrosus	12										**21/2**
C. lunatus	6										**28/1**
Number of sites		11	8	19	26	7	10	10	10	6	16

** Statistically significant ($P < 0.01$).

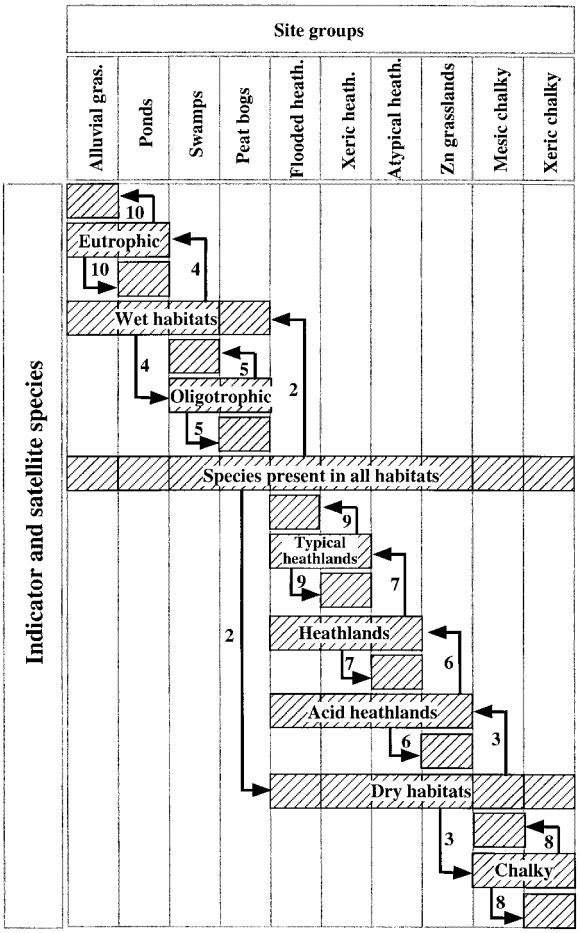

FIG. 13. Steps that are followed to build a two-way table from the hierarchical clusters indicator values. The first species group (center of figure) contains species that are common in all habitats (i.e., having their indicator value maximum when all sites are pooled in one group). At the next step, two species groups are created: one with species dominating in all wet habitats, and the other one with species that are common in all dry habitats. The procedure is repeated for each site cluster.

from the sites characterizing the gradient's center (Gauch 1982). This is not evidenced with DCA.

Comparing clustering results

The eight-group typology produced by TWINSPAN is similar to the *k*-means typology for the same level, although minor differences are observed. Contaminated grasslands, atypical heathlands, and peat bogs are well isolated in the *k*-means typology, whereas they are associated with other habitats in the TWINSPAN classification. The temporary flooded heathlands are associated with other wet habitats in TWINSPAN, while they are clustered with other sandy and dry heathlands in the *k*-means typology.

The divisive algorithm of TWINSPAN does not allow reallocations among site groups, and its division ranking is arbitrary. TWINSPAN always produces a hierarchical structure, even in cases where this structure is very subtle or nonexistent. As stated in the *Introduction,* the groups produced by a division at one step are not necessarily those that best explain the data set structure. This can lead to site groups that are more heterogeneous than others. In our case, the second division in TWINSPAN subdivides wet sites in two groups, while the *k*-means partitioning method as well as the ordinations contrast chalky grasslands to other dry sites at the same level. While the TWINSPAN algorithm is probably good in the case of a regular gradient, it is not the best for more heterogeneous situations. This has a significant impact on the species typology, because the species typology is based, in TWINSPAN as well as in our approach, on the site typology. In our approach, however, separating the search for the site typology from the search for the species typology and indicator values has several advantages. First, one can use a similarity coefficient that is appropriate to the data at hand. Then one can choose from among the clustering methods available in the literature, and eventually apply several well-suited methods to verify the stability of the clusters found, as was done in the present study. In all cases, some test of classifiability of sites should be performed at the outset of the analysis, at least visually by looking at an ordination plot (see Gordon 1994*a*, *b* for a review of methods). The Indicator Value method provides an alternative to which the results obtained by TWINSPAN can be compared. Austin (1985) recommended making comparisons of this type for all multivariate analysis results.

Q-mode and *R*-mode analyses show differences between site and species similarities. Site similarities, based on species abundances, allow one to establish clearly defined and hierarchically nested site groups. This pattern is not observed among species because there are large differences in species amplitude breadth. Several species are more or less eurytopic (e.g., *Pterostichus diligens, P. minor, P. madidus, P. rhaeticus, P. versicolor, Harpalus rubripes, Bembidion lampros*) and many others are restricted to one or two habitats. The more eurytopic species are responsible for similarities between habitats and for the nested hierarchical structure in the site typology. Without them, no hierarchical structure should emerge among the site groups. In such a case, a reallocation clustering method like *k* means is appropriate to verify the reality of the hierarchy and to measure the significance of the more eurytopic species.

Comparing indicator species results

It is not easy to strictly compare the TWINSPAN and INDVAL results because the site typology used as reference is not the same. TWINSPAN only identified, as indicators, very low cut-off level pseudospecies (class 1, indicating that a species has a frequency $>0\%$). This is not very useful for prediction; it simply indicates that a species is present in all sites of a group. None

of the higher abundance pseudospecies were identified by TWINSPAN.

Many species identified by TWINSPAN as indicators received a high indicator value with the INDVAL method, for the same or a closely related habitat class. The INDVAL method identified several other indicator species, with rather high indicator values, that also contribute to the specificity of the site groups. The INDVAL method is then more sensitive than TWINSPAN at identifying indicator species.

Other advantages of the IndVal index

Nonhierarchical patterns.—Although the main structure of our carabid date set is hierarchical, there are species that do not follow the general pattern. First, some species are not well characterized because of the sampling structure. *Pterostichus lepidus* and *Cicindela campestris* have their greatest indicator value for xeric heathlands and contaminated grasslands, respectively. These species were common in these two xeric habitats, and would have received a much greater indicator value if a group with only xeric and acid habitats had been created, excluding the atypical and flooded heathlands. This is also the case for *Pterostichus minor* and *Agonum fuliginosum,* which are abundant in alluvial grasslands and swamps. In such cases, one can always develop other site partitions, if they demonstrate ecological structures that are not revealed by a first site hierarchy. The INDVAL approach can be generalized to any classification of sites, based either on species distributions (as in the present study), or on a priori ecological variables or sampling structures.

Indicator value of species absence.—The mathematics of the INDVAL index lends itself to suggest an index describing the indicator value for the absence of a species. The philosophy of this absence indicator value index would be that it should be maximum for a site group when the species does not occur in any one of its sites but is present in all sites of the other groups. In such a case, the presence INDVAL index, described above, is rather low, yet the species may present an ecological interest as indicators of peculiar ecological conditions where the species is never present. Note that the absence indicator value index of species *i* is not simply symmetric to the presence INDVAL_i index described above. For species *i,* the relative mean realized abundances in group *j* (A_{ij}) should be replaced by the relative mean nonrealized abundance. One solution would be to compute the difference between the maximum, over all groups, of the mean number of specimens per group, and the mean number of specimens in group $_j$. Value A_{ij} for group $_j$ is obtained as the ratio of this difference, to the sum of these differences over all groups. The relative frequency of presence in a group (B_{ij}) should simply be replaced by the relative frequency of absences.

With such index, a species that has a low presence INDVAL index, because that species is absent from all the sites of a group but occupying many sites in different groups, should have a high absence INDVAL index, but only when the mean abundances are similar across the groups where the species is present.

Number of site groups.—When does one stop generating clusters with the *k*-means method, or accept those proposed by a hierarchical clustering method? This is a common question in cluster analysis because no single objective criterion receives general support (Milligan and Cooper 1985). Generally, one stops subdividing when the new subgroups cannot be explained, or when they are not supported by a gap in the ordination space. The INDVAL method provides the species indicator values as an a posteriori criterion to stop the subdivision process. When the indicator values of all species are decreasing, the clustering method does not explain anything more. So, we suggest using the sum of the species significant indicator values for each clustering level as a criterion to determine when to stop dividing, and to decide what are the most important clustering steps. Another criterion is to compute the frequency distribution of the number of the significant indicator species for each clustering step and to compare them.

Fig. 14 presents bar charts for the sums of the positive and negative differences in species INDVAL indices between successive clustering levels. The clustering levels from 2 to 10 groups are characterized by high positive sums, and a sum of all differences that is always positive. The four-group to five-group rearrangement is the only exception. At the five-group level, a large group of 45 wet and oligotrophic sites is subdivided into a swamp cluster and a peat bog and raised mire cluster. There are only two species more or less indicative of these two groups (Fig. 12), and many species of wet habitats show a decrease of their indicator value at this step. A stronger decrease is also observed for these species at the 10-group level (differences between alluvial grasslands and pond fringes); it is counterbalanced by a high INDVAL increase for several stenotopic species. Indeed, the eutrophic wet sites are characterized by a greater mean species richness than oligotrophic sites; this difference in species richness explains the small positive increase observed at the five-group level.

After the 10-group level, the sum patterns show clearly that the subsequent divisions do not explain an informative part of the species distributions. One exception is observed at the 13-group step, which subdivides the eight heavy-metal-contaminated grassland sites into two subgroups of four sites. This subdivision explains the distribution of several stenotopic xerophilic species.

The sum of all species indicator values does not decrease after the 20-group level, even if we take only the significant values into account. This can be explained by the high number of rare species in carabid

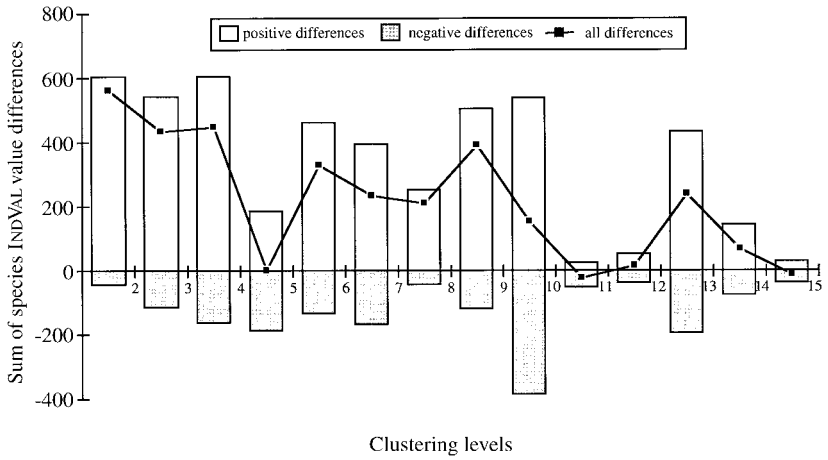

FIG. 14. The sum of the species indicator values for each clustering level is used as a criterion to identify the main steps of the clustering procedures. Bars represent the sums of positive and negative differences in the INDVAL index between successive clustering levels. The squares along the broken line represent the sum of all differences.

beetle assemblages, which characterize small differences between and within habitats.

Comparing typologies.—The sum of the species indicator values can be used as a criterion to compare typologies and choose the one that explains the species distributions best. Fig. 15 presents the comparison of the sum of all species indicator values for the clustering levels obtained with the *k*-means method and TWINSPAN. Although the two methods start at the same value, the *k*-means method has higher values for all the following clustering levels. The clusters from the two methods are not the same at any given level and, consequently, do not explain the same ecological information. It is clear, however, that the *k*-means classification explains the species distribution better than TWINSPAN.

Presence/absence data and pseudospecies.—In the case of presence/absence data, the first part of the INDVAL index can be modified to become the ratio of the number of species presences in a site group to the total number of species presences: $A_{ij} = (\text{Nsites}_{ij}/\text{Nsites}_{i.})$ where Nsites_{ij} is the number of sites in cluster *j* where species *i* is present, while $\text{Nsites}_{i.}$ is the total number of sites in cluster *i*. This index is maximum when the species occupies all the sites of a single group only. This is a very efficient way to identify strong indicator species of a site typology in the case of biomonitoring data, or of a geographic unit typology in the case of biogeographic studies.

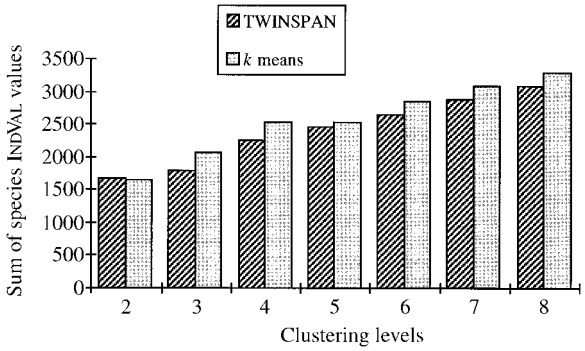

FIG. 15. Comparison of the pattern of the sum of all significant species indicator values at each clustering level for the TWINSPAN and the *k*-mean results.

We can also use the pseudospecies concept, defined as in TWINSPAN, to identify the species that reach a given threshold percentage value. As such, pseudospecies are presence/absence vectors of different levels of the species relative abundance across sites; the INDVAL method could then be used to derive the indicator potential of such species.

Carabid beetle assemblages

Although we have not sampled the complete range of southern Belgian habitats, the results show that the distribution of carabid beetles is structured by environmental conditions. Site clusters are easily delimited and, moreover, they are hierarchically nested. Using the Alatalo and Alatalo (1977) method, we found that >70% of the species diversity was conserved when the 123 year-catch cycles were clustered into 10 *k*-means site groups.

Several sites are from intermediate habitat types, however. In each case, changes in carabid beetle assemblages are observed. It is the case of the flooded heathland group, which, from a carabid beetle point of view, really lies between xeric heathlands and peat bogs. Temporary flooding allows the coexistence of species from wet and dry habitats. When heathland sites are isolated and perturbed by vegetation colonization, forest species from surrounding woodlands, like *Carabus problematicus* and *Abax ater,* are common, while many of the xeric and open heathland characteristic species have disappeared. In these two cases, differences in Carabid beetle assemblages were not ex-

pected because the vegetation was similar to other heathlands.

These results show the potential use of invertebrate data for local-scale assessment (Refseth 1980). When it is necessary to evaluate the local consequences of habitat fragmentation that are not disclosed by bird studies (their geographical scale is not appropriate because bird patch sizes are larger) or botanical relevés (the persistence of plant populations is too important), invertebrate inventories are the best alternative. Such studies on invertebrates are still rare (e.g., De Vries and Den Boer 1990) and should be developed, especially in countries where the impact of human activities is intense on most of the territory.

The flexible INDVAL method proposed here gives precise and accurate information on carabid species habitat preferences. Lindroth (1992) was the first to present a classification of a very large number of carabid beetle species into six groups. Recent works of Desender (1986*a–d*) and Turin et al. (1991) are more detailed in their descriptions, but some of their species groups, as the ''eurytopic species,'' could be more clearly defined. Many of these eurytopic species show habitat preferences at a higher hierarchical level and are responsible for the hierarchical similarities among habitats. The species classification obtained here could become a useful tool for biological assessment because species are subdivided in very precise assemblages, typical of combinations of several ecological factors, and also of several geographic scales. This typology could still be improved, however, by incorporation of other habitats, such as forests, or similar habitats in other biogeographical areas, to establish a detailed reference list of indicator species for biological conservation studies.

Regional distribution patterns and species assemblages

The study of regional distribution patterns is fundamental to understanding species distributions at several geographic scales. Several regional distribution models have been proposed in the literature (Levins 1970, Hanski 1982, Brown 1984). Our results clearly show that all species are not either rare (satellite) or common (core sensu Hanski 1982); these labels depend on the level of the sampling patch size (Nee et al. 1991). Species that are rare in higher hierarchical levels (where the sampling patch size is large) may be common at the lower ones (where the sampling patch size is small). The smaller the sampling patch size in an area, the higher the chances of finding generalist or core species. This confirms the hypothesis of Collins and Glenn (1991) who proposed that Brown's model is adequate for large sample patch sizes (>1 km^2), while Hanski's describes small sample patch sizes (<1 ha).

This fact is strongly connected with the Hutchinsonian multivariate ecological niche concept: the more we eliminate potential ecological factors of variation explaining species distributions from the sampling design (i.e., going from higher hierarchical clustering levels to lower ones), the greater are the chances of resampling the same species at several sites. It is likely that Brown's asymmetric distribution is a sum of Hanski's bimodal distributions in different small-scale ecological situations. Size frequency distributions of species range cannot be used to validate one or the other model. When the sampling patch size is smaller than several species patch sizes, bimodality appears without any relationship to an underlying distribution model.

Before evaluating the adequacy of species assemblages to a regional distribution model, it is important to clearly define them and to account for the sampling patch size induced by the sampling design structure. The flexible method proposed here seems to be a good way to define such precise species assemblages.

ACKNOWLEDGMENTS

We are thankful to K. Desender for advice in the identification and ecology of carabid beetles; to Henri André, Philippe Casgrain, Christopher Dunn, Michel Loreau, Bruce McCune, Jary Niemela, Michel Waeyenbergh, and an anonymous referee for comments on first drafts of this manuscript. Field work was partially supported by two grants (number 2.9008.89 and number 2.9014.91) from the Fond de la Recherche Fondamentale et Collective (Belgium). Marc Dufrêne's work was supported by a fellowship from I.R.S.I.A. (Belgium), a grant from Université Catholique de Louvain, and a grant from the Fonds National de la Recherche Scientifique (Belgium). This paper was written while he was a Fonds National de la Recherche Scientifique (Belgium) Senior Research Assistant, and his leave to Université de Montréal was supported by a NATO scientific research grant.

A program (INDVAL), written in FORTRAN, is available from M. Dufrêne for the computation of the indicator value index and the randomization testing procedure (http://www.biol.ucl.ac.be/ecol/html/Outils/Tools.IndVal.html). It was written for Macintosh System 7, but can easily be modified for other operating systems. The results can be imported into a spreadsheet to produce graphs and to sort the two-dimensional table.

LITERATURE CITED

Alatalo, R. V., and R. H. Alatalo. 1977. Components of diversity: multivariate analysis with interaction. Ecology **58**:900–906.

Aukema, B. 1990. Taxonomy, life history and distribution of three closely related species of the genus *Calathus* (Coleoptera, Carabidae). Tijdschrift voor Entomologie **133**: 121–141.

Austin, M. P. 1985. Continuum concept, ordination methods and niche theory. Annual Review of Ecology and Systematics **16**:39–61.

Baars, M. A. 1979. Catches in pitfall traps in relation to mean densities of carabid beetles. Oecologia (Berlin) **41**: 25–46.

Belbin, L., and C. McDonald. 1993. Comparing three classification strategies for use in ecology. Journal of Vegetation Science **4**:341–348.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. American Naturalist **124**:255–279.

Collins, S. L., and S. M. Glenn. 1991. Importance of spatial and temporal dynamics in species regional abundance and distribution. Ecology **72**:654–664.

Cousins, S. H. 1991. Species diversity measurement: choosing the right index. Trends in Ecology and Evolution **6**: 190–192.

De Vries, H. H., and P. J. Den Boer. 1990. Survival of population of *Agonum ericeti* PANZ. (Col., Carabidae) in relation to fragmentation of habitats. Netherlands Journal of Zoology **40**:484–498.

Den Boer, P. J. 1977. Dispersal power and survival carabids in a cultivated countryside. Miscellaneous papers. Landbouwhogeschool Wageningen **14**:1–192.

Desender, K. 1985. Naamlijst van de loopkevers en zandloopkevers van Belgie (Coleoptera, Carabidae). Studiedocumenten numéro 19, Koninklijk Belgisch Instituut voor Natuurwetenschappen, Brussels, Belgium.

———. 1986*a*. Distribution and ecology of carabid beetles in Belgium (Coleoptera, Carabidae). Part 1. Documents de travail numéro 26, Institut Royal des Sciences Naturelles de Belgique, Brussels, Belgium.

———. 1986*b*. Distribution and ecology of carabid beetles in Belgium (Coleoptera, Carabidae). Part 2. Documents de travail numéro 27, Institut Royal des Sciences Naturelles de Belgique, Brussels, Belgium.

———. 1986*c*. Distribution and ecology of carabid beetles in Belgium (Coleoptera, Carabidae). Part 3. Documents de travail numéro 30, Institut Royal des Sciences Naturelles de Belgique, Brussels, Belgium.

———. 1986*d*. Distribution and ecology of carabid beetles in Belgium (Coleoptera, Carabidae). Part 4. Documents de travail numéro 34, Institut Royal des Sciences Naturelles de Belgique, Brussels, Belgium.

Dufrêne, M., and P. Legendre. 1997. Biogeographical structure of carabid beetle distributions in Belgium. Journal of Biogeography, *in press.*

Dunn, Ch. 1994. Gaps in GAP. Plant Science Bulletin **40**: 120–121.

Edgington, E. S. 1987. Randomization tests. Marcel Dekker, New York, New York, USA.

Eyre, M. D., and M. L. Luff. 1990*a*. A preliminary classification of European grassland habitat using carabid beetles. Pages 217–226 *in* N. E. Stork, editor. Role of ground beetles in ecological and environmental studies. Intercept, Andover, New Hampshire, USA.

Eyre, M. D., and M. L. Luff. 1990*b*. The ground beetle (Coleoptera: Carabidae) assemblages of British grasslands. Entomologist's Gazette **41**:197–208.

Ferry, C. 1976. Un test facile pour savoir si la richesse mesurée d'un peuplement se rapproche de sa richesse réelle. Le Jean le Blanc **15**:21–28.

Gauch, H. G., Jr. 1982. Multivariate analysis in community analysis. Cambridge University Press, Cambridge, England.

Gordon, A. D. 1994. Clustering algorithms and cluster validation. Pages 497–512 *in* P. Dirschedl and R. Ostermann, editors. Computational statistics. Physica-Verlag, Berlin, Germany.

———. 1996. Hierarchical classification. Pages 65–121 *in* P. Arabie, L. J. Hubert, and G. De Soete, editors. Clustering and classification. World Scientific Publishing Company, Singapore, Malaysia.

Halsall, N. B., and S. D. Wratten. 1988. The efficiency of pitfall trapping for polyphagous predatory Carabidae. Ecological Entomology **13**:293–299.

Hanski, I. 1982. Dynamics of regional distribution: the core and satellite species hypothesis. Oikos **38**:210–221.

He, F., P. Legendre, C. Bellehumeur, and J. V. LaFrankie. 1994. Diversity pattern and spatial scale: a study of a tropical rain forest of Malaysia. Environmental and Ecological Statistics **1**:265–286.

Hill, M. O. 1973. Diversity and evenness: a unifying notation and its consequences. Ecology **54**:427–432.

———. 1979. TWINSPAN: a fortran program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. Cornell University, New York, New York, USA.

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

Jackson, D. A., and K. M. Somers. 1991. Putting things in order: the ups and downs of detrended correspondence analysis. American Naturalist **137**:704–712.

Jongmann, R. H. G., C. J. F. ter Braak, and O. F. R. van Tongeren. 1987. Data analysis in community and landscape ecology. PUDOC, Wageningen, The Netherlands.

Kremen, C., R. K. Colwell, T. L. Erwin, D. D. Murphy, R. F. Noss, and M. A. Sanjayan. 1993. Terrestrial arthropod assemblages: their use in conservation planning. Conservation Biology **7**:796–808.

Lebreton, J. D., R. Sabatier, G. Banco, and A. M. Bacou. 1991. Principal component and correspondence analyses with respect to instrumental variables: an overview of their role in studies of structure—activity and species—environment relationships. Pages 85–114 *in* J. Devillers and W. Karcher, editors. Multivariate analysis in SAR and Environmental studies. Kluwer, Dordrecht, The Netherlands.

Legendre, L., and P. Legendre. 1983. Numerical ecology: developments in environmental modelling, 3. Elsevier, Amsterdam, The Netherlands.

Legendre, L., and P. Legendre. 1984*a*. Écologie numérique. 2. La structure des données écologiques. Masson, Paris, France.

Legendre, P. 1987. Constrained clustering. Pages 289–307 *in* P. Legendre and L. Legendre, editors. Developments in numerical ecology. NATO ASI series. Volume G-14. Springer-Verlag, Berlin, Germany.

Legendre, P., S. Dallot, and L. Legendre. 1985. Succession of species within a community: chronological clustering, with applications to marine and freshwater zooplankton. American Naturalist **125**:257–288.

Legendre, P., and V. Legendre. 1984*b*. Postglacial dispersal of freshwater fishes in the Québec peninsula. Canadian Journal of Fisheries and Aquatic Sciences **41**:1781–1802.

Legendre, P., and A. Vaudor. 1991. The R package: multidimensional analysis, spatial analysis. Département de sciences biologiques, Université de Montréal, Montréal, Québec, Canada.

Levins, R. 1970. Extinction. Pages 77–107 *in* M. Gertenhaber, editor. Mathematical problems in biology. Lectures on mathematics in the life science. American Mathematical Society, Providence, Rhode Island, USA.

Lindroth, C. H. 1992. Ground beetles (Carabidae) of Fennoscandia. III. A zoogeographical study. Intercept, Andover, Massachusetts, USA.

Loreau, M. 1992. Species abundance patterns and the structure of ground-beetle communities. Annales Zoologici Fennici **28**:49–56.

Luff, M. L. 1975. Some features influencing the efficiency of pitfall traps. Oecologia (Berlin) **19**:345–357.

Luff, M., M. D. Eyre, and S. P. Rushton. 1992. Classification and prediction of grassland habitats using ground beetles (Coleoptera, Carabidae). Journal of Environmental Management **35**:301–315.

MacQueen, J. B. 1967. Some methods for classification and analysis of multivariate observations. Pages 281–297 *in* L. M. Le Cam and J. Neyman, editors. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume 1. University of California Press, Berkeley, California, USA.

Magurran, A. E. 1988. Ecological diversity and its measurements. Croom Helm, Ryde, Australia.

Milligan, G. W., and M. C. Cooper. 1985. An examination

of procedures for determining the number of clusters in a data set. Psychometrika **50**:159–179.

Nee, S., R. D. Gregory, and R. M. May. 1991. Core and satellite species: theory and artefacts. Oikos **62**:83–87.

Niemela, J. K., and J. R. Spence. 1994. Distribution of forest dwelling carabids (Coleoptera)—spatial scale and the concept of communities. Ecography **17**:166–175.

Peet, R. K., R. G. Knox, J. S. Case, and R. B. Allen. 1988. Putting things in order: the advantages of detrended correspondence analysis. American Naturalist **131**:924–934.

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

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

Refseth, D. 1980. Ecological analyses of carabid communities—potential use in biological classification for nature conservation. Biological Conservation **17**:131–141.

Schoener, T. W. 1986. Patterns in terrestrial vertebrate versus arthropod communities: do systematic differences in regularity exist? Pages 556–586 *in* J. Diamond and T. J. Case, editors. Ecology. Harper & Row, New York, New York, USA.

Späth, H. 1980. Cluster analysis algorithms. Ellis Horwood, Chichester, England.

ter Braak, C. J. F. 1985. Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. Biometrics **41**:859–873.

———. 1987. Ordination. Pages 91–173 *in* R. H. G. Jongman, C. J. F. ter Braak, and O. F. R. van Tongeren, editors. Data analysis in community and landscape ecology. PUDOC, Wageningen, The Netherlands.

———. 1988. CANOCO—a FORTRAN program for canonical community ordination by (partial) (detrended) (canonical) correspondence analysis, principal components and redundancy analysis (version 2.1). Updated in 1990 (version 3.1). Agricultural Mathematics Group, Ministry of Agriculture and Fisheries, Wageningen, The Netherlands.

Thiele, H.-U. 1977. Carabid beetles in their environments. Springer-Verlag, Berlin, Germany.

Turin, H., K. Alders, P. J. Den Boer, S. Van Essen, Th. Heijerman, W. Laane, and E. Penterman. 1991. Ecological characterization of carabid species (Coleoptera, Carabidae) in the Netherlands from thirty years of pitfall sampling. Tijdschrift voor Entomologie **134**:279–304.

Wartenberg, D., S. Ferson, and F. J. Rohlf. 1987. Putting things in order: a critique of detrended correspondence analysis. American Naturalist **129**:434–448.

Webb, N. R. 1989. Studies on the invertebrate fauna of fragmented heathland in Dorset, UK, and the implications for conservation. Biological Conservation **47**:153–165.