Vol. 120: 243-250, 1995 | MARINE ECOLOGY PROGRESS SERIES
Mar. Ecol. Prog. Ser. | Published April 20

Statistical method for selecting representative species in multivariate analysis of long-term changes of marine communities. Applications to a macrobenthic community from the Bay of Morlaix

Claude Manté[1], Jean-Claude Dauvin[2], Jean-Pierre Durbec[1]

[1] Centre d'Océanologie de Marseille (M.A.I. & Gr.3) URA CNRS 041, Campus de Luminy case 901, F-13288 Marseille Cedex 09, France
[2] M.N.H.N., Laboratoire de Biologie des Invertébrés Marins, URA CNRS 699, 57 rue Cuvier, F-75231 Paris Cedex 05, France

ABSTRACT: A new method is presented for analysing biological time series with only a few observations of many species. It operates by selecting the more informative species and describing the temporal changes of a community using a Principal Components Analysis method based on the relative proportions of selected species. The method was tested on a macrobenthic time series data from the Bay of Morlaix (western English Channel).

KEY WORDS: Principal components · Species selection · Time series

INTRODUCTION

Systematic sampling of communities over time results in the recording of the abundance of all species present at given sampling locations. Unfortunately, many rare species are unlikely to be detected by sampling and, even when detected, the estimated abundances of such species are unreliable. Therefore, in relation to statistical methods used for analysing data, it might be more efficient to discard rare species from subsequent analyses. In quantitative analysis, according to the multivariate method used, numbers of rare taxa cannot usually be employed in Principal Components Analysis (PCA) or in Correspondence Analysis. In PCA analysis, rare species are not important in calculation of distances between samples; on the contrary, because of the large number of species, they are inappropriate in the interpretation of Principal Components. In the case of Correspondence Analysis, some rare species will have high loadings without any ecological or statistical significance (Stephenson & Cook 1980, Greenacre 1984). On the other hand, in qualitative analysis, only the presence or absence of species in samples is taken into account; the 'rare' ones can be very informative on the state of the community, and the dominant ones can be less informative (Gray et al. 1990). In the present work, we essentially focused on Metric Multidimensional Scaling (MMDS), i.e. PCA from interdistance tables. This approach is well suited to the exploratory analysis of populations and leads to easily interpretable displays of distances between samples (Besse et al. 1986). The selection of 'representative species' enables us to discard rare species, which would not significantly alter the interdistances, to obtain legible displays with a reduced number of descriptors. This new approach is tested on a macrobenthic series (1977 to 1991) from an infralittoral muddy fine-sand community of the Bay of Morlaix (English Channel; northern Brittany, France).

MATERIAL AND METHODS

Study site. To test the method, we chose an available macrobenthic time series having a large number of collected species and à relatively large number of observations through time. It is an *Abra alba-Hyalinoecia bilineata* macrobenthic community from the Bay of

Morlaix which was regularly sampled from April 1977 to December 1991. The main characteristics of the sampling of the abiotic and biotic conditions were described in detail elsewhere (Dauvin 1984, 1991, Dauvin & Ibanez 1986) and are only summarized in this paper. The Pierre Noire station was located in the eastern part of the Bay of Morlaix (48° 42′ 30″ N, 3° 51′ 58″ W, 17 m depth). The sediment was fine sand (median particle size, 148 to 184 µm). The bottom-water temperature varied from 8°C in March to 15.5°C in September and salinity between 34.5 in winter and 35.3 at the beginning of autumn. During the 15 yr survey there were 112 sampling events, i.e. 5 observations for April 1977 to December 1977, 11 observations in 1978, 12 observations per year from 1979 to 1981, 7 observations in 1982, 5 in 1983, 6 in 1894, 10 in 1985, 7 in 1986, and 5 observations per year from 1987 to 1991. Samples were collected with a Hamon grab for April and August 1977 and 4 replicate samples were taken each time, covering a total area of 1.2 m^2. All other times, a Smith McIntyre grab was used and 10 samples were taken each time, for a total area of 1.2 m^2. After collection, the sediment was sieved (1 mm circular mesh) and the retained material was fixed with 10% neutral formalin and sorted twice. The second sorting was done afer staining with 10% Rose Bengal. Species were identified and counted for each grab, then their densities were pooled together and expressed in ind. m^{-2}

The Pierre Noire station was impacted by the hydrocarbons of the 'Amoco Cadiz' oil spill in spring 1978, which caused disappearance of the dominant *Ampelisca* spp. populations (Dauvin 1984, Dauvin & Ibanez 1986). Subsequently, the station was recolonised by these species whose populations had been locally eliminated by the oil spill (Dauvin 1987, 1991).

Characteristics of the data. Pierre Noire station showed high diversity: the total cumulative number of species collected was 421 (112 m^2 sampled). The number of species m^{-2} generally varied from a minimum of 90 in March to 130 in October (see Fig. 1b).

Method. After identification and counting of the individuals, samples were directly described by raw frequencies and identified with independent observations of multinomial random vectors. In consequence, the dissimilarities between samples were quantified through distances suited to multinomial laws. There are few usual distances on discrete probability law space; among them, the Hellinger and Bhattacharrya distances are easily computable and clearly interpretable in a geometrical framework (Rao 1982, Qannari 1983, Kass 1989, Amari 1990).

Selection of representative species. At any given time, each sample of a series drawn from the population under study is composed of N individuals belonging to some random number of species (Q). The true number of species actually present in the community is $Q' \geq$ Q. Methods for estimating Q' are beyond the scope of this paper and are extensively discussed in Chao & Lee (1992). Each sample can be considered as some value of a multinomial random vector with parameters $\pi_1, \pi_2, ..., \pi_Q$ where π_i ($i = 1, 2, ..., Q'$) is the theoretical proportion of the ith species in the sample. Consequently, the variate 'number of individuals belonging to the ith species' (K_i) follows a binomial distribution of parameters π_i and N. Due to sampling fluctuations, only Q species are detected in a sample and the presence in a particular sample of some 'rare species' in the community can be attributed to chance.

Let α_0 equal the probability a species being absent from a sample when its true proportion is π_0 and the sample size is N. Under the independence hypothesis for sampling, the probability of absence is

$$\alpha_0 = (1 - \pi_0)^N.$$

By definition, a species will be said to be 'rare at the α_0 level' if its true proportion π_i ($i = 1, 2, ..., Q'$) in the macrobenthic community is less than or equal to π_0. Clearly π_i is unknown and we need to decide whether the species is 'rare' or not from an estimate $\hat{\pi}_i$. We used the unrestricted maximum likehood estimate k_i/N, where k_i is the observed abundance of the ith species. The probability $P_{\pi_i}(k_0)$ for the random number K_i being less than or equal to k_0, a given integer $0 \leq k_0 \leq N$, is equal to:

$$P_{\pi_i}(k_0) = P_{\pi_i}(K_i \leq k_0) = \sum_{m=0}^{k_0} \binom{N}{m} \pi_i^m (1-\pi_i)^{N-m}.$$

This is a decreasing function of π_i, equal to 1 for $\pi_i = 0$ and 0 for $\pi_i = 1$.

The upper limit of a unilateral confidence interval for π_i at the η level may be obtained by determining the value of π_i such that (Van der Waerden 1967)

$$P_{\pi_i}(k_i) = \eta$$

where, as above, k_i is the observed number of individuals of the ith species. Now if we assume that η is fixed at a given value η_0, and if we suppose that $\pi_i = \pi_0$, we can determine the greatest integer $k_{\max}$ such that

$$P_{\pi_0}(k_{\max}) = \eta_0.$$

$k_{\max}$ is the greatest observable number of individuals compatible with the α_0 level rarity of a species when the size of the population is N. In other words, π_0 would be the estimated upper limit of a unilateral confidence interval for π_i at the α_0 level, when the observed number is $k_{\max}$. Practically, the determination of $k_{\max}$ is based on the expression of $P_{\pi_0}(k_{\max})$ as a function of the incomplete beta distribution $\beta(\pi, m, n)$

(Van der Waerden 1967):

$$P_{\pi_0}(k_{\max}) = 1 - \beta(\pi_0, k_{\max}, N - k_{\max} - 1).$$

One obtains $k_{\max}$ by inverting $P_{\pi_0}(k_{\max}) = \eta_0$ using, for instance, the algorithm of Majumder & Bhattacharjee (1985). After determining $k_{\max}$, all species for which $k_i \leq k_{\max}$ are discarded from the sample; all the rare species are pooled and constitute a special group named 'background noise'. This variate reflects some kind of biodiversity in the sample. Lastly, the species discarded from all the processed samples are eliminated. The number of selected species is an increasing function of α_0 and η_0 (see Table 1).

PCA on multinomial data. The comparison of the distributions of retained species at different dates was carried out by a weighted PCA (Rao 1964, Jolliffe 1986) based on 2 metrics (Hellinger and Bhattacharrya distances) suited to the particular nature of the data. The weight of a sample is the total number of individuals belonging to 'active' (as opposed to supplementary; see Greenacre 1984) species collected, such that each sample remains represented by a multinomial law when some species are removed to a supplementary list.

Hellinger distance. This distance between 2 multinomial laws is

$$d_H^2(\pi^d, \pi^{d'}) = \sum_{i=1}^{Q} \left(\sqrt{\pi_i^d} - \sqrt{\pi_i^{d'}}\right)^2$$

where π^d and $\pi^{d'}$ are associated with dates *d* and *d'* respectively. Because in the multinomial case the Hellinger distance is the usual euclidian distance in R^Q after square-root transformation, this analysis can be performed using a conventional weighted-PCA program (e.g. SAS 1989). This metric should not be confused with Orloci (1978) chord distance, which is defined as an empirical correlation coefficient between 2 samples (Pielou 1984).

Bhattacharrya distance. This distance was originally defined on the space of multinomial laws (Bhattacharrya 1946); it is the geodesic distance on this space when the Riemannian metric is defined from Fisher's information (Kass 1989, Amari 1990). It can be expressed as a function of Hellinger distance according to

$$d_B(\pi^d, \pi^{d'}) = 2 \cdot \arccos\left(1 - d_H^2(\pi^d, \pi^{d'})/2\right).$$

When distances are 'small', Bhattacharrya's and Hellinger's metrics are essentially the same, with both of them approximately identical to the square of the Kullback-Leibler divergence, which is nearly identical to the chi-squared distance (Benzécri 1967, Kass 1989). So, all of these distance notions coalesce when samples are close enough together. In contrast to Hellinger's distance, it is necessary to perform metric or non-metric MDS from the interdistance table. The relationships between the species and the principal components are no longer straightforward. To overcome, in part, this drawback, the linear correlation coefficients between species and components are to be computed as a usual assistance to interpretate the analysis.

RESULTS

The number of collected individuals for some of the species was low and the procedure described above was performed with several threshold values for the probabilities α_0 and η_0. Table 1 shows the results of the procedure.

PCA is a method for decomposing the total variance of a set of samples. In consequence, its results are not affected by 'rare species' (weak variance), at least not in the range of the considered thresholds; so, we used only the 124 species corresponding to $\alpha_0 = 0.01$ and $\eta_0 = 0.01$ (Table 1). The total number of individuals showed strong temporal changes that ranged between approximately 2000 ind. m^{-2} and values greater than 60 000 ind. m^{-2} (Fig. 1a). The first peak was observed in September 1977, which was before the 'Amoco Cadiz' oil spill, and corresponded to high abundance of *Ampelisca* species; a second peak occured in August 1982 and corresponded to a proliferation of the opportunistic polychaete *Polydora pulchra*. During the last 4 annual cycles, we observed seasonal variations with a summer maximum, a winter minimum, and a trend of increasing density due to the re-establishment of *Ampelisca* species. Temporal changes of total species richness and those of selected species were quite similar and approximately synchronous (Fig. 1b). Number of species was affected by seasonal variations, with a maximum in summer and a minimum in winter. We also observed a decrease in species richness in 1978 and 1979 linked to the disappearance of a great number of species, especially amphipods, just after the oil spill (Dauvin 1984, 1987). Using Hellinger and Bhattacharrya's distances, 2 weighted-PCA analyses were performed. As the results were very similar, we choose to report only those related to Bhattacharrya's interdistances table, processed using the ADDAD (1992)

Table 1. Number of selected species as a function of the thresholds α_0 and η_0

α_0	η_0	
	0.01	0.05
0.01	124	144
0.05	141	163
0.10	154	177

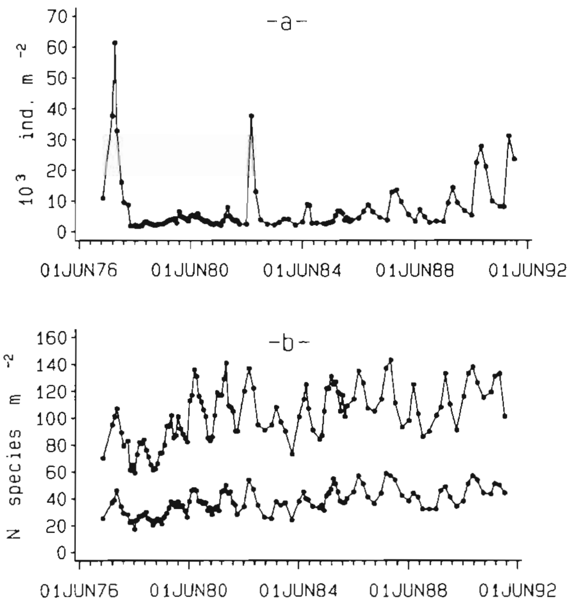

Fig. 1 (a) Density (ind. m^{-2}) versus sampling date, and (b) number of observed (upper line) and selected (lower line) species from April 1977 to December 1991

MMDS program. After the distances between samples were computed, the first 2 principal components were determined. They accounted for 71.6% of total variance. The temporal changes of the first component (Fig. 2a) showed a large increase in spring of 1978 linked to the decrease of *Ampelisca* spp. abundance after the 'Amoco Cadiz' oil spill. Thereafter, we noted a 'linear' trend towards the initial values of 1977 with the re-establishment of *Ampelisca* populations. Thus, the first factor was strongly correlated with the temporal changes of *Ampelisca* spp. (Table 2), e.g. *Ampelisca armoricana* (r = –0.69; see Fig. 3a). The other species representative of this factor were *Ampelisca sarsi* (r = –0.75), *Ampelisca tenuicornis* (r = –0.69), *Ampharete acutifrons* (r = –0.71), and *Scoloplos armiger* (r = +0.60). The increase identified in this analysis for the first component, just after the 'Amoco Cadiz' oil spill, was similar to the increase of the Shannon diversity observed by Warwick & Clarke (1993) on the same Pierre Noire data set. The second component (Fig. 2b) showed essentially seasonal behaviour related to annual recruitment cycles that occurred from the end of the spring to the beginning of autumn. It displayed an increasing trend in seasonal changes from 1981 to 1987, with low values in summer and high values in winter. The minimum was in August 1982 when *P. pulchra* showed very large densities. During the 4 first and the 4 last annual cycles, seasonal variations remained moderate. This factor was mainly due to *P. pulchra* (r = –0.76, Fig. 3b) which presented very high densities in 1982 and 1984. Other species representative of this factor (Table 2) were *Leucothoe incisa* (r = –0.79, Fig. 3c), *Phtisica marina* (r = –0.75), *Leptonereis glauca* (r = +0.72), *Gnathia oxyuracea* (r = –0.69), *Pisidia longicornis* (r = –0.67), and *Tharyx marioni* (r = +0.66).

The first analysis on the data table (112 samples × 124 species) identified *Ampelisca* species and *Polydora*

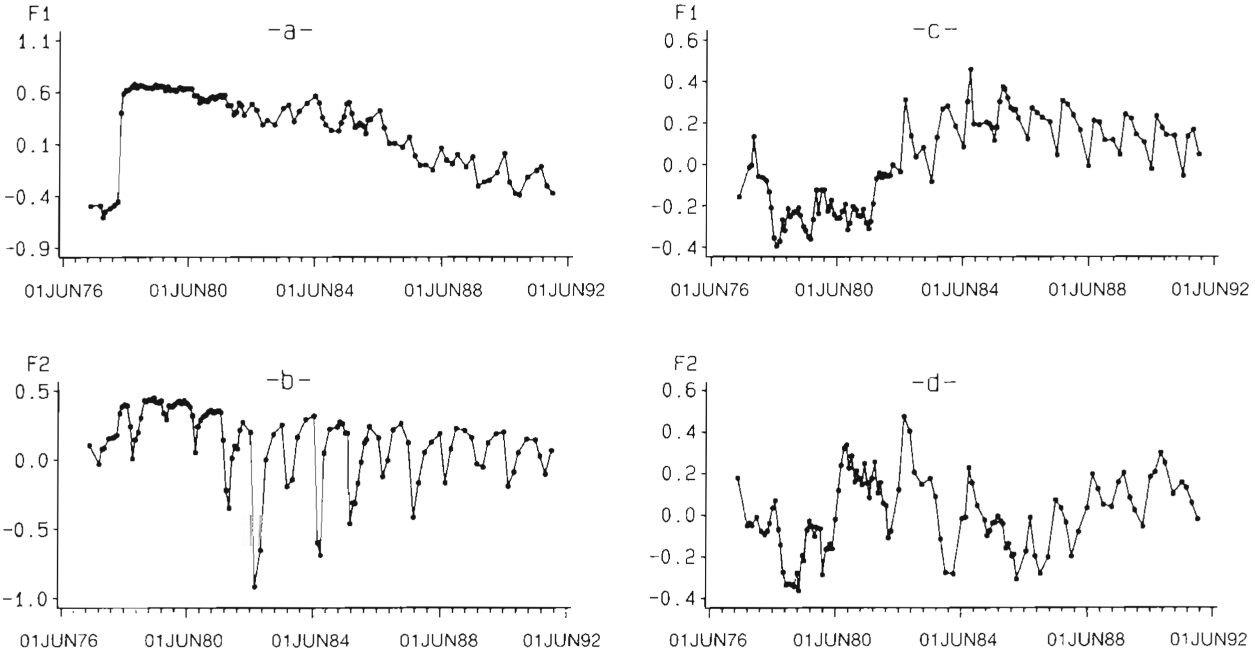

Fig. 2. Principal components versus sampling date. (a, b) PCA with 124 species: (a) F1 (first component) and (b) F2 (second component). (c, d) PCA without *Ampelisca* species and *Polydora pulchra*: (c) F1 and (d) F2

Table 2. Values of R^2 (multiple correlation coefficient between the first 7 principal components and each species) and linear correlation coefficient between species and the first 2 components (F1 and F2) in the PCA with 124 species × 112 samples. BGN: background noise

Species	R^2	F1	F2
Abra alba	65.29	0.123	0.315
Ampelisca armoricana	68.94	−0.692	0.071
Ampelisca brevicornis	56.16	−0.335	−0.484
Ampelisca sarsi	77.30	−0.754	−0.050
Ampelisca spinipes	54.31	−0.553	0.063
Ampelisca tenuicornis	65.05	−0.688	0.074
Ampelisca typica	66.13	−0.050	−0.491
Ampharete acutifrons	65.48	−0.712	−0.120
Aricidea cerrutii	64.70	−0.414	0.358
Aricidea fragilis	52.54	−0.439	0.367
Aricidea minuta	52.21	−0.269	0.184
Bathyporeia elegans	60.08	0.416	0.221
Bathyporeia tenuipes	56.26	0.200	0.087
Chaetozone setosa	74.81	0.206	0.341
Euclymene oerstedi	53.73	−0.189	−0.403
Cultellus pellucidus	56.48	0.399	0.192
Exogone gemmifera	61.93	0.232	−0.623
Gnathia oxyuraea	54.60	0.128	−0.690
Heterocirrus alatus	50.49	0.199	0.139
Lanice conchilega	67.79	0.230	−0.648
Leiochone clypeata	52.09	−0.340	−0.076
Leptonereis glauca	77.84	0.258	−0.724
Leucothoe incisa	68.59	0.148	−0.790
Magelona mirabilis	58.66	0.274	−0.117
Marphysa bellii	52.26	−0.147	0.251
Nephtys hombergii	73.70	0.026	−0.836
Paradoneis armata	52.04	−0.021	0.503
Photis longicaudata	52.60	−0.643	0.011
Phtisica marina	67.31	0.235	−0.748
Pisidia longicornis	57.82	0.202	−0.669
Polydora pulchra	78.34	0.233	0.761
Scoloplos armiger	78.82	0.602	0.292
Spio decoratus	74.65	−0.134	0.149
Spiophanes bombyx	50.76	−0.483	−0.106
Tharyx marioni	65.93	0.218	−0.658
Thyasira flexuosa	78.50	0.357	0.335
Urothoe pulchella	73.88	0.538	−0.344
Venus ovata	68.53	0.353	0.231
BGN	50.10	−0.032	−0.599

pulchra as important for explaining both first factors. So, in a second step we discarded the 8 species of amphipod *Ampelisca* and the polychaete *Polydora pulchra* to define the temporal changes of the other main species. The second analysis was performed with a reduced table (112 samples × 115 species). The first component accounted for 22% of total variance and showed the seasonal variations of the community with a maximum in summer and a minimum in winter (Fig. 2c). Two main periods could be distinguished: (1) from the beginning of sampling (April 1977) to December 1981, i.e. the annual cycle before the spill and the first 3 cycles afterwards, when the community remained perturbed and presented low total densities (Dauvin & Ibanez 1986) and (2) from 1982 to the end of the observations in 1991, when this restricted community reached a steady state with a significant seasonal pattern. The temporal changes of the subcommunity were also affected by the oil spill, but the impact of hydrocarbons on the selected species was not as dramatic as in the first analysis. After the initial stress, this subpopulation rapidly reached a steady state (see Fig. 2c). The species which had the highest correlation with the first component (Table 3) were *Leucothoe incisa* (r = 0.62), *Nephtys hombergii* (r = 0.66), *Spiophanes bombyx* (r = 0.67) and *Paradoneis armata* (r = −0.78).

The second component (Fig. 2d) (15.5 % of total variance) seemed to correlate with species richness (Fig. 1b). It displayed correlation with abundance of marginal species ('background noise': r = 0.59) and with temporal changes of *Lanice conchilega* (r = 0.60) and *Spio decoratus* (r = −0.59). Maximal values from 1980 to 1983 corresponded to increases in population of several main species as a secondary effect of the

Table 3. Values of R^2 (multiple correlation coefficient between the first 7 principal components and each species) and linear correlation coefficient between species and the first 2 components (F1 and F2) in the PCA without *Ampelisca* spp. and *Polydora pulchra* wih 115 species × 112 samples. BGN: background noise

Species	R^2	F1	F2
Abra alba	81.30	−0.577	0.272
Abra prismatica	63.93	−0.327	0.098
Ampharete acutifrons	73.81	0.474	0.192
Aricidea fragilis	60.35	−0.202	−0.349
Aricidea cerrutii	56.25	−0.336	−0.214
Bathyporeia elegans	58.59	0.132	−0.416
Chaetozone setosa	80.83	−0.573	0.072
Cultellus pellucidus	56.03	−0.457	0.445
Euclymene oerstedi	59.83	0.437	0.479
Exogone hebes	60.36	0.324	−0.286
Heterocirrus alatus	64.47	−0.274	−0.363
Leiochone clypeata	65.92	0.352	0.321
Leucothoe incisa	66.62	0.619	0.200
Magelona filiformis	55.99	0.613	−0.185
Magelona mirabilis	57.97	0.572	−0.263
Marphysa bellii	58.11	−0.498	0.383
Nephtys hombergii	76.12	0.657	0.492
Notomastus latericeus	58.17	0.212	0.371
Ophiura albida	53.21	−0.120	0.469
Paradoneis armata	78.42	−0.776	−0.163
Photis longicaudata	58.44	0.275	0.090
Phtisica marina	56.75	0.243	0.437
Scoloplos armiger	78.17	−0.388	0.013
Spio decoratus	79.62	0.220	−0.589
Spiophanes bombyx	79.20	0.670	0.143
Thyasira flexuosa	75.11	−0.525	0.084
Urothoe pulchella	82.95	0.429	0.053
Venus ovata	68.42	−0.460	0.202
BGN	61.64	0.424	0.590

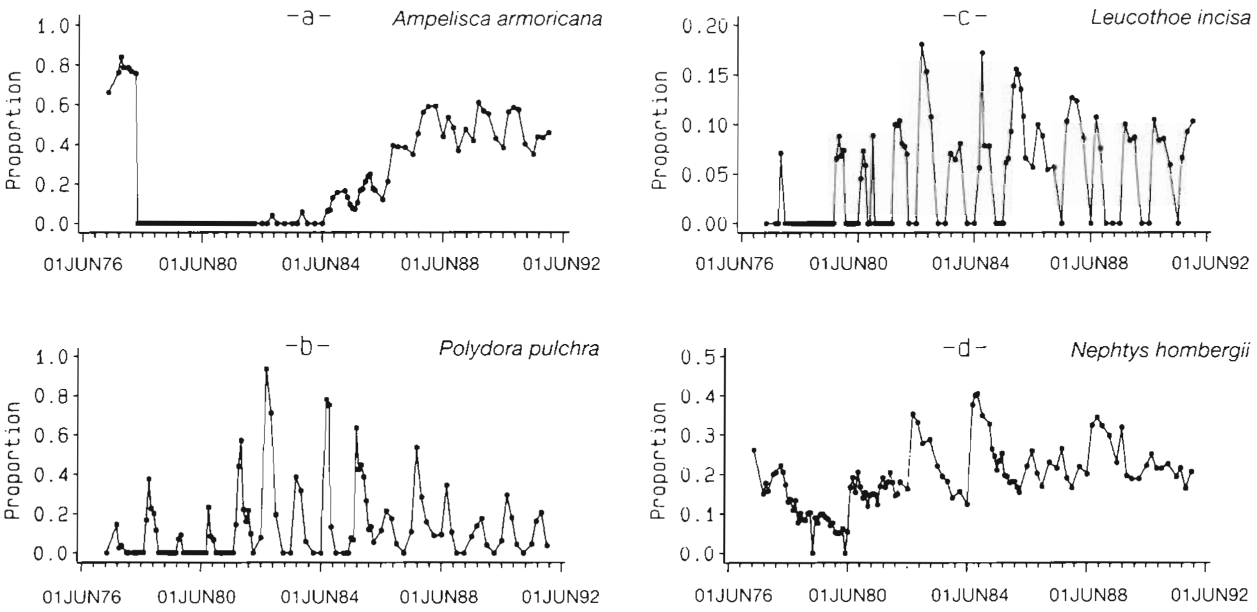

Fig. 3. Changes of relative abundance of (a) *Ampelisca armoricana*, (b) *Polydora pulchra*, (c) *Leucothoe incisa*, and (d) *Nephtys hombergii*

perturbation of the community after the initial impact of oil pollution (Dauvin & Ibanez 1986). The elimination of *Polydora pulchra* induced the disappearance of its associated factor (seasonality), probably because the other associated species were too sparse.

To test the robustness of our method of selection of the main species, we plotted the 6216 Bhattacharrya interdistances after selection species (D1: PCA with 124 species; D2: PCA with 115 species) against the corresponding interdistances without selection (D: PCA with 421 species). Fig. 4 shows the relationships between D and D1, and between D and D2. In the first case, the selection method does not alter the interdistances since the cloud of points is very close to the line $x = y$. Plotting D vs D2 (Fig. 4b), we obtain a scattered cloud of points showing that the elimination of some dominant species dramatically changes the interdistance structure between samples. In summary, both PCAs on a subset of the species in the community identified the general trend and the seasonal

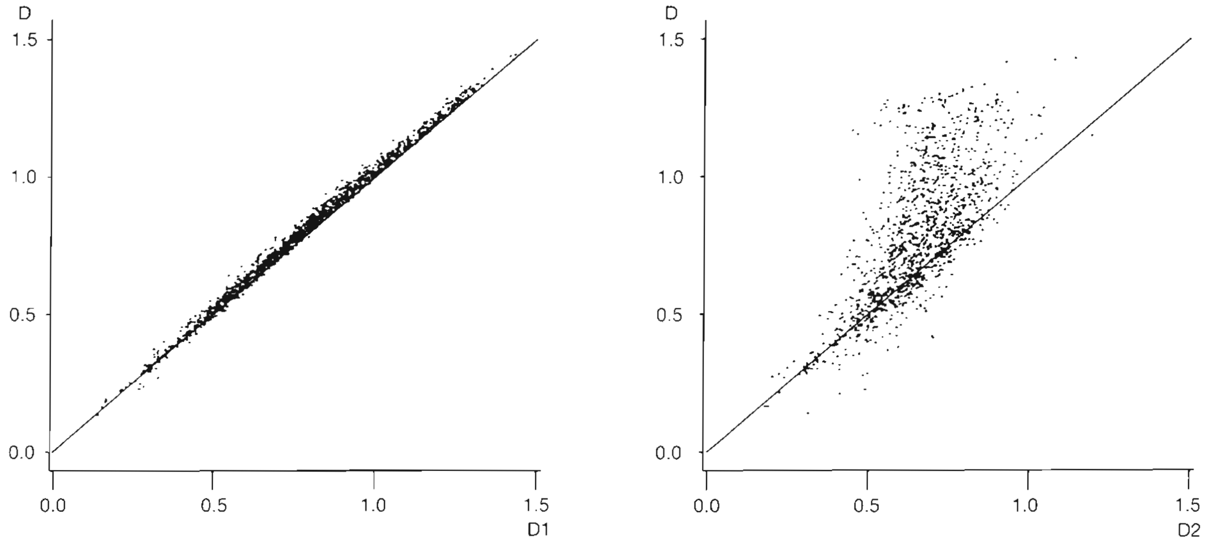

Fig. 4. Relationships between Bhattacharrya interdistances. (a) D (421 species) versus D1 (124 selected species). (b) D versus D2 (115 remaining species)

changes of the assemblages, along with successional periods following the oil spill including destruction of the dominant *Ampelisca* species and the progressive re-establishment of the community to nearly its prespill state.

DISCUSSION

There is a large body of literature on various types of multivariate analysis (see Kenkel & Orloci 1986, Minchin 1987, Clarke & Green 1988, Palmer 1993). The choice of an analysis method essentially depends on the aims of the study and the nature of data, but also on personal preference, availability of programs and validity in terms of statistical assumptions (Heip et al. 1988). In the present work, we developed a method selecting the more informative species from a quantitative point of view. The analysis of temporal changes in the community was performed using a particular PCA method based on the relative proportions of selected species. Another way used to reduce the number of biological descriptors consists of aggregating species according to higher taxonomical levels (e.g. genus, family, and phyla) as proposed by Heip et al. (1988) and Gray et al. (1990). The major groups identified by multivariate analysis with aggregation to higher taxa, even to the level of phyla, remain distinguished. Thus, with the Pierre Noire data set Warwick & Clark (1993), using ordinations of phylum, showed the impact of the 'Amoco Cadiz' oil spill on a macrobenthic community. To our knowledge, the selection of representative species before data analysis has been studied only by Stephenson & Cook (1980) and, before them, by Dale & Williams (1978). The method of Stephenson & Cook was very different from ours since it was based on studying the influence of removing species on the stability of the dissimilarity structure between samples. As a consequence, the retained species depend on the chosen dissimilarity index. The usual technique for excluding species is based on the examination of proportions; for instance, species whose proportion is $\leq 5\%$ are excluded from the analysis. Our opinion is that since the observation of 'rare species' is relatively improbable, some are collected, while others are not. So it is quite natural to identify such species and pool them in a residual category. The method used in this paper for selecting species for PCA assumes that each sample is assimilated to a multinomial vector where the components are the numbers of individuals of each species. The main interest is thus focused on proportions. In the comparison of observations by PCA, the size of each sample is taken into account as a weight associated with that sample. Since in PCA the observations have less influence when weights are low, smaller samples play a minor role. In the same way, if a species is 'rare', i.e. the probability of a random individual belonging to this species is low, it may be discarded from the analysis without affecting the overall results. The critical values chosen in the determination of 'rare species' have been fixed between 0.1 and 0.01. The first value, α_0, is the fixed probability that a species is not observed in N samples; from α_0 and N, a maximum proportion π_0 can easily be computed such that, if the true proportion of some species, π_s, is less than π_0, the probability of this species being absent in N samples is greater than α_0. Such a species is 'rare at the α_0 level'. As we don't know the true proportion π_s, we must determine, afterwards, the probability of the event for each species s ($\pi_s \leq \pi_0$) when the observed effective is K_s and eliminate rare species for which this probability is greater than the second critical value η_0. Another way to choose the dominant species would be to perform a similar analysis on the whole frequency table. Rare species should be processed by methods suited to the analysis of presence/absence data, such as (1) Gower's maximal predictive classification (1967) or Govaert's variant of the dynamical clusters method for binary data (1990) for determining homogeneous groups of samples, (2) multiple correspondence analysis (Greenacre 1984, Goodman 1985, Escofier & Pagès 1990) for displaying proximities between samples through chi-square distance, or (3) tree analysis (Buneman construction) for displaying the samples as leafs or nodes on a tree (Barthélemy & Guénoche 1988). Multiple Factorial Analysis (Escofier & Pagès 1990) could be a synthetic tool for exploring eventual links between both groups of species (rare and abundant ones).

Our method allows the relationships between species over time to be investigated and allows a comparison of temporal changes of several communities at a local or mesoscale (Gray & Christie 1983, Souprayen et al. 1991, Ibanez et al. 1993). Used on macrobenthic data from Pierre Noire station in the Bay of Morlaix, our method was shown useful to time series analysis of data with many species of low abundance.

Acknowledgements. This study was carried out under the program (P.N.O.C.) IFREMER/INSU topic 'Long term Series'. It is also a part of the program 'GDR Manche du CNRS' sponsored by IFREMER annual contracts since 1990. The authors thank the crew of N.O. 'Mysis' and technicians from the 'Station Biologique de Roscoff' for their valuable assistance with field work and D. Bernardo, L. Dauvin, and M. Joncourt for their help in sorting the samples.

LITERATURE CITED

ADDAD (1992). Users guide. ADDAD Institute, Paris

Amari, S. I. (1990). Differential geometrical methods in statistics, 2nd edn Springer Verlag, Berlin, New York

Barthélemy, J.-P., Guénoche, A. (1988). Les arbres et les représentations des proximités. Masson, Paris

Benzécri, J.-P. (1976). L'analyse des données. Dunod, Paris

Besse, P., Caussinus, H., Ferré, L., Fine, J. (1986). Some guidelines for principal components analysis. In: Compstat 1986. Physica Verlag, Heidelberg, p. 23–29

Bhattacharrya, A. (1946). A measure of divergence between two multinomial populations. Sankhya 7: 401

Chao, A., Lee, S. M. (1992). Estimating the number of classes via sample coverage. J. Am. statist. Ass. 87: 210–217

Clarke, K. R., Green, R. H. (1988). Statistical design and analysis for a 'biological effect' study. Mar. Ecol. Prog. Ser. 46: 213–226

Dale, M. B., Williams W. T (1978). A new method of species reduction for ecological data. Aust. J. Ecol. 3: 1-5

Dauvin, J.-C. (1984). Dynamique d'écosystèmes macrobenthiques des fonds sédimentaires de la baie de Morlaix et leur perturbation par les hydrocarbures de l'Amoco Cadiz. Thèse de Doct. Etat, Sci. nat., Univ. Pierre & Marie Curie, Paris

Dauvin, J.-C. (1987). Evolution à long terme (1977–1986) des populations d'Amphipodes des sables fins de la Pierre Noire (Baie de Morlaix, Manche Occidentale) après la catastrophe de l'Amoco Cadiz. Mar. environ. Res. 21: 247–273

Dauvin, J.-C. (1991). Effets à long terme de la pollution de l'Amoco Cadiz sur la production de deux peuplements subtidaux de sédiments fins de la Baie de Morlaix (Manche Occidentale). In: Elliot, M., Ducrotoy, J. P. (eds.) Estuaries and coasts: spatial and temporal intercomparisons. International Symposium series, ECSA 19 Symposium. Olsen & Olsen, Fredensborg, p. 349–358

Dauvin, J.-C., Ibanez, F. (1986). Variations à long terme (1977–1985) du peuplement des sables fins de la Pierre Noire (Baie de Morlaix, Manche Occidentale): analyse statistique de l'évolution structurale. Hydrobiologia 142: 171–186

Escofier, B., Pagès, J. (1990). Analyses factorielles simples et multiples. Objectifs, méthodes et interprétations. Dunod, Paris

Goodman, L. (1985). The cross classified data having ordered and unordered categories: Association Models, Correlation Models and Asymetry Models for contingency tables with or without missing entries. Ann. Statist. 13: 10–69

Govaert, G. (1990). Classification binaire et modèles. Rev. Statist. appl. 38: 67–81

Gower, J. C. (1967). Maximal predictive classification. Biometrics 30: 643–654

Gray, J. S., Christie, H. (1983). Predicting long-term changes in marine benthic communities. Mar. Ecol. Prog. Ser. 13: 87–94

Gray, J. S., Clarke, K. R., Warwick, R. M., Hobbs, G. (1990). Detection of initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North Sea. Mar. Ecol. Prog. Ser. 66: 285–299

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

Heip, C., Warwick, R. M., Carr, M. R., Herman, P. M. J., Huys, R., Smol, N., Van Holsbeke, K. (1988). Analysis of community attributes of the benthic meiofauna of Frierfjord/Langesundfjord. Mar. Ecol. Prog. Ser. 46: 171–180

Ibanez, F., Dauvin, J.-C., Etienne, M. (1993). Comparaison des évolutions à long terme (1977–1990) de deux peuplements macrobenthiques de la baie de Morlaix (Manche Occidentale). Relations avec les facteurs hydroclimatiques. J. exp. mar. Biol. Ecol. 169: 181–214

Jolliffe, I. T (1986). Principal component analysis. Springer Verlag, New York

Kass, R. E. (1989). The geometry of asymptotic inference. Stat. Sci. 4(3): 188–234

Kenkel, N. C., Orloci, L. (1986). Applying metric and nonmetric multidimensional scaling to ecological studies: some new results. Ecology 67: 919–928

Majumder, K. L., Bhattacharjee, G. P. (1985). Inverse of the incomplete beta function ratio. In: Griffith, P., Hill, I. D. (eds.) Applied statistics algorithms. Ellis Horwood Ltd/Royal Statistical Society, London, p. 117–120

Minchin, P. R. (1987). An evaluation of the relative robustness of techniques for ecological ordination. Vegetatio 69: 89–107

Orloci, L. (1978). Multivariate analysis in vegetation research, 2nd edn. Junk, The Hague

Palmer, M. W. (1993). Putting things in even better order: the advantages of canonical correspondence analysis. Ecology 74: 2215–2230

Pielou, E. C. (1984). The interpretation of ecological data. John Wiley & Sons, New York

Qannari, E. M. (1983). Analyses factorielles de mesures. Thèse de 3ème cycle, Univ. Paul Sabatier, Toulouse

Rao, R. C. (1964). The use and interpretation of principal components analysis in applied research. Sankhya A 26: 329–358

Rao, R. C. (1982). Diversity and dissimilarity coefficients: a unified approach. Theor. Populat. Biol. 21: 24–43

SAS (1989). User's guide Statistics Ver. 6. SAS Institute Inc., Cary, NC

Souprayen, J., Dauvin, J.-C., Ibanez, F., Lopez-Jamar, E., O'Connor, B., Pearson, T H. (1991). Long-term trends of macrobenthic communities: numerical analysis of four north-western european sites. In: Keegan, B. F. (ed.) Space and time series data analysis in coastal benthic ecology. An analytical exercise organized within the framework of the COST 647, Project on Coastal Benthic Ecology. Commission of the European Communities, Bruxelles, p. 265–438

Stephenson, W., Cook, S. D. (1980). Elimination of species before cluster analysis. Aust. J. Ecol. 5: 263–273

Van der Waerden, B. L. (1967). Statistique mathématique. Dunod, Paris

Warwick, R. M., Clarke, K. R. (1993). Comparing the severity of disturbance: a meta-analysis of marine macrobenthic community data. Mar. Ecol. Prog. Ser. 92: 221–231

This article was submitted to the editor

Manuscript first received: February 2, 1994
Revised version accepted: December 6, 1994