BOLLETTINO DI GEOFISICA TEORICA ED APPLICATA VOL. 40, N. 2, pp. 147-158; JUNE 1999

The use of principal component analysis in phytoplankton data: comparison of communities at one coastal station in the Gulf of Trieste

M. ZANGRANDI[1] and G. CRISPI[2]

[1]*Laboratorio di Biologia Marina, Trieste, Italy*
[2]*Osservatorio Geofisico Sperimentale, Trieste, Italy*

(Received June 26, 1998; accepted June 7, 1999)

Abstract. The phytoplankton succession in the Gulf of Trieste, during the autumn, winter and spring of the years between 1986-1988, is analysed by means of multivariate statistics to describe the temporal sequence of variation and the population groups. The principal component analysis gives the temporal dynamics of the data set considered and the cluster analysis identifies the groups of species dominant in each season. The typical diatoms-dinoflagellates succession is recovered, while no significant variation in the structure of the community in the water column in each-year data is identified. The between-year variation, which occurs in the species describing the temporal evolution in the community, is highlighted. This between-year species replacement suggests combining the classical principal component-cluster analysis with a rotation of the principal components. An orthogonal Procrustes transformation is applied both to sampling units and to species groups. It recovers the same underlying pattern as in the multi-annual succession of communities described at species level.

1. Introduction

The study of the phytoplankton community in the North Adriatic Sea dates back about one century. The peculiar characteristic of this basin is an annual cycle of the phytoplankton community structure with seasonal patterns of taxa-association which are probably among the strongest encountered in the Mediterranean region. Hence temporal series data from this area have a strong and somehow predictable correlation structure.

Nevertheless, because of the year to year change in the species of each seasonal taxa association, the analysis of the structure of the community does not recover any seasonal pattern if

Corresponding author: G. Crispi; Osservatorio Geofisico Sperimentale, Borgo Grotta Gigante 42c, 34010 Sgonico, Italy; phone: +39 040 2140205; fax: +39 040 327307; e-mail gcrispi@ogs.trieste.it

taxa are categorised at species level as well as from data spanning more than one year (Fonda, 1996).

In this paper, aspects which are relevant to the study of the community structure are addressed, starting from a presence and absence description of the species. The first objective of this work is to recognize how much the non-stratification conditions in the water column, documented by the chemical-physical parameters, can be seen in the phytoplankton community. The second one is to show how statistical methodology is capable of recovering the same underlying structure in the phytoplankton community, under interannual variation in the taxa composition for phytoplankton population and to recognize factors determining the seasonal biological change.

Multivariate techniques and, in particular, Principal Component Analysis (PCA) are used extensively in marine biology. PCA can be useful to highlight both plankton patterns (Cataletto et al., 1996) and fishery food trends (Orr and Bowering, 1997) and as well as to study catch data (Pech and Laloë, 1997).

An application of this methodology is shown for a coastal station in the Gulf of Trieste. The data sets and methods used are described in section 2. The third section tests the hypothesis that in late autumn, winter and early spring the water column can be considered homogeneous in recognizing factors determining the seasonal biological change for the phytoplankton population. To this aim a succession of Sampling Units is performed to yield the classes of the raw data and the PCA is applied to the binary biological data to obtain the groups of the species. A rotation of the components previously acquired is performed to interpret the succession of the present species in the different data sets and to recognize factors determining the seasonal biological change. Conclusions from applying these methods are outlined in the fourth section.

2. Data and methods

The data were collected between 1986-1988 at a station in the Gulf of Trieste, situated 200 m offshore from the Miramare Natural Reserve (Cabrini et al., 1988). The analysis of this multiannual sampling demonstrates the tight relation of freshwater input, variability of plankton and circulation (Fonda Umani et al., 1995).

The following physical-chemical parameters were measured: temperature, salinity, dissolved oxygen, saturated oxygen, Secchi disks, chlorophyll a, pheopygments. The sampling was fortnightly from March 1986 to December 1986 and monthly from February 1987 to September 1988. Water samples, of 500 ml each at four depths in the water column (0, 5, 10 and 15 m), were analysed by the Utermoehl method using subsamples of 50 ml giving the plankton population, expressed as cells/liter for every species (Milani et al., 1988a). The phytoplankton observed comprises 128 different species: 50 diatoms, 67 dinoflagellates, and 11 among Euglenophyceae, Crysophyceae and Prymnesiophyceae. The microzooplankton, analysed with similar methods, obtained 38 different species (Milani et al., 1988b).

In line with the objectives of this work, the autumn, winter and early spring measurements are selected. For the 1986/1987 period, hereafter denoted data set 1, the following dates are chosen for data measurements: 6 October, 18 October, 11 November, 3 December, 16 December for

1986; 10 February, 25 February, 25 March, 14 April, 20 May for 1987. For the 1987/1988 period, hereafter denoted as data set 2, the following samples are selected: 22 October, 27 November, 22 December for 1987; 9 January, 27 February, 24 April, 24 May for 1988.

The phytoplankton data are studied using the binary data of the presence-absence of the species because the unitary biomass is not available for all the species. All 49 species present, with occurrences of more than 5 % and less than 100 %, are included in the analysis (Table 1). Some taxa, e.g. microflagellates which have a seasonal change in biomass, are not considered in the analysis because they are always present and thus do not contribute to the total variance. The successive steps of the analysis have proven that the disregarded species, by not giving real informations about the biological gradients, are not characteristic of a bloom.

Hierarchical classification is performed on the phytoplankton data sets. The ecological distance between Sampling Units (SUs) is calculated with product-moment correlation (this function is suitable if there is more than 50% of blank data entries (Bakus, 1990)), while the complete linkage strategy is used. Two hypotheses are tested. The first hypothesis (H0) is that the sampling units are not clustered by depth, while in the second (H0') the SUs are not clustered by season. The test is performed using analysis, comparing expected occurrences (from null hypotheses) with observed occurrences of SUs within the cluster group.

The PCA is performed to study the ordination of the species and of the SUs to identify the essential components contributing to the total variance of the species, and finally to connect the different populations observed in the Gulf of Trieste during the period considered. The correlation coefficients are used and no prior transformation is performed on the raw data.

The use of PCA in ecological research requires the following assumptions: homoscedasticity of SUs, linearity of the response, implying short gradients, and statistical spatial-temporal independence of SUs. The first is not relevant because we use binary data. For the second the visual inspection of the data table shows the total replacement of species with time, and few species are present both in autumn and spring, thus the gradient is too long with respect to the linearity assumption. This will be discussed later in the PCA results. For the third the low sampling frequency (more or less monthly) suggests that the independence of SUs is a good approximation.

The violation of some of these hypotheses does not preclude the application of the PCA for descriptive purposes, provided that statistical inference is not required (Jolliffe, 1986). Here linear correlations between components of the PCA and physical quantities, in particular temperature and depth, are sought for.

In order to discuss the patterns in the community, the classes of the similarity analysis are reported on the plane of the first two components. The comparison of within-class versus between-class distance provides useful information on the relative magnitude of ecological gradients both in time and in depth.

To study the phytoplanktonic association, species with a significant correlation coefficient ($p < .05$) with the first two components, are considered for interpretation.

The temporal sequence of the data set 1 SUs is compared with those of data set 2. To compare the two series, data are reduced to the monthly data which are common for both years. The scores of the first 7 components of data set 1 (*X* data hereafter), recalculated on the reduced dataset (comprising the 18 October, 3 December, 16 December 1986, 25 February, 14 April and 20

Table 1 - Presence (1)-absence (0) table of the phytoplankton species for data set 1 and data set 2. The Sampling Unit number in each data set is progressive.

	Data set 1										Data set 2						
Sampling date Depth / 5 m	6 Oct 86 0 1 2 3	18 Oct 86 0 1 2 3	11 Nov 86 0 1 2 3	3 Dec 86 0 1 2 3	16 Dec 86 0 1 2 3	10 Feb 87 0 1 2 3	25 Feb 87 0 1 2 3	25 Mar 87 0 1 2 3	14 Apr 87 0 1 2 3	20 May 87 0 1 2 3	22 Oct 87 0 1 2 3	27 Nov 87 0 1 2 3	22 Dec 87 0 1 2 3	9 Jan 88 0 1 2 3	27 Feb 88 0 1 2 3	24 Apr 88 0 1 2 3	24 May 88 0 1 2 3
Amphidinium flagellans	0 0 1 0	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 0	0 1 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Ceratium furca	1 0 0 0	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 0	0 0 0 0	1 1 0 0	0 1 0 0
Ceratium fusus	0 1 0 0	0 0 0 0	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 0	1 1 0 0	0 0 0 0	0 1 0 0	0 1 1 0	0 1 1 0	0 0 0 0	0 1 0 1
Diplopsalis	0 0 1 0	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 0 0 1	0 1 0 0	0 1 0 0	0 1 0 1
Gymnodinium	0 0 0 0	0 1 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 0 0	0 0 0 0	0 0 1 1	1 1 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 0 0	0 0 0 0
Gyrodinium fusiforme	0 0 0 1	0 0 0 0	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 0 1	1 1 1 0	0 0 0 0	0 0 0 0	0 0 0 0	1 0 0 0	0 0 0 0	0 0 0 1
Gyrodinium lachrima	0 0 0 0	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 0 0 0	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Porocentrum micans	0 0 0 0	0 0 0 0	1 0 1 0	1 1 1 0	0 0 1 0	0 0 0 0	0 0 1 1	0 0 0 0	0 0 0 0	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 0 0	0 0 0 0	1 1 0 0	1 1 1 1
Porocentrum triestinum	1 1 0 0	0 0 0 1	0 0 0 0	1 0 0 0	0 0 0 0	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Protoperidinium bipes	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 1 0	0 0 0 1	1 0 0 0	1 1 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Chaetoceros steinii	0 0 0 0	0 0 0 0	0 0 0 0	1 1 0 1	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Chaetoceros affine	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 0 0	0 0 0 0	0 0 0 0	0 0 0 1	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Chaetoceros decipiens	1 0 0 1	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 0 0 1	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 1	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 1 1 1
Cerataulina pelagica	1 1 0 0	1 0 1 1	1 1 1 1	0 1 0 1	0 1 1 0	0 0 0 0	0 0 0 0	0 0 0 0	1 1 0 1	1 1 1 1	0 0 0 0	0 0 0 0	1 0 0 0	0 0 0 1	1 0 0 1	0 0 0 0	1 1 1 1
Guinardia	1 0 1 0	1 1 1 1	0 0 0 1	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 1	0 0 0 0	1 0 0 0	0 0 0 0	0 0 0 0	0 1 1 0	1 1 1 1	1 1 1 1
Licmophora	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 0 0 1	1 1 1 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Licmophora abbreviata	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 1 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Leptocylindrus danicus	1 1 1 1	1 1 1 1	1 1 1 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 1 1	1 0 0 1	1 0 1 1	1 1 1 1	1 1 1 0	1 0 0 0	0 0 0 0	1 0 0 0	0 0 0 1	1 1 1 1
Melosira moniliformis	1 0 0 1	0 0 0 0	0 0 1 0	1 0 0 0	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Melosira sulcata	1 0 1 1	0 0 0 1	0 0 0 0	0 0 0 0	0 1 0 1	1 1 0 0	0 0 1 1	0 0 0 0	0 0 1 0	0 1 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Nitzschia Pseudonitzschia	1 1 1 1	1 1 1 1	1 1 1 1	0 0 0 1	1 1 1 1	1 0 1 1	0 0 0 0	0 1 1 1	0 0 0 0	0 1 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 1
Nitzschia closterium	0 0 0 0	0 0 0 0	1 0 0 0	0 0 0 0	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 0	1 0 0 0	0 0 0 1	1 0 0 0	0 0 0 0	0 0 0 1
Nitzschia delicatissima	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 1 1 1	1 1 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Nitzschia longissima	1 0 1 1	0 0 0 0	1 1 1 1	1 0 0 0	0 0 0 0	1 1 1 1	1 1 0 0	0 1 0 0	1 0 1 0	0 0 0 0	0 0 1 1	1 1 0 0	1 0 0 1	0 0 0 0	0 0 0 0	1 1 1 1	0 1 1 0
Pleurosigma angulatum	0 1 0 1	0 0 0 0	0 1 0 1	0 1 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 1	0 0 0 1	1 0 0 0	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0
Rhizosolenia gracillima	1 0 0 1	1 1 1 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 0 0 1	0 0 1 0	0 0 1 1	0 0 0 0	0 0 1 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 1
Rhizosolenia indica	1 0 1 1	1 1 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Rhizosolenia delicatula	1 0 1 1	0 0 0 0	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 0	0 1 1 1	1 0 1 0	0 1 0 1	1 1 0 0	0 0 0 0	1 0 1 0	0 0 0 0	1 1 1 0
Rhizosolenia stolterfothii	0 0 1 1	1 1 1 1	1 1 1 1	0 1 0 0	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 0	1 1 1 0	0 1 1 1	0 1 1 0	0 0 1 0	0 0 0 0	0 0 0 0	1 1 1 0	0 0 0 1	0 0 0 0
Skeletonema costatum	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 0 1 1	1 0 0 1	1 1 0 1	1 1 1 1	0 1 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Thalassionema nitzschioides	0 0 0 1	0 0 0 0	1 0 0 1	1 0 0 0	1 1 0 0	0 0 0 0	0 0 1 0	0 0 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Thalassiosira decipiens	0 0 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 1	1 1 1 0	0 0 0 1	1 1 1 0	1 1 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Thalassiosira	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 1 0	0 0 0 0	1 1 1 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 1
Dictyocha fibula	0 1 1 0	1 1 1 1	0 0 0 0	1 1 0 1	0 0 0 0	1 1 0 1	1 1 1 1	1 0 0 0	0 0 0 0	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 0	0 0 1 0	0 0 1 0	0 0 0 0	0 0 0 0
Dinophysis sacculus	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 1 0 0	0 0 0 1	0 0 0 0	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Oblea	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 0	0 0 0 0	0 0 0 1	0 0 0 0	0 0 0 0	1 0 0 0	0 0 1 0
Protoperidinium diabolum	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 1 1	1 1 0 1	0 0 0 0	0 0 0 0	0 0 0 0
Protoperidinium steinii	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 1 0	1 1 1 1
Chaetoceros	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 1 1 1	1 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Chaetoceros curvisetum	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 0	0 0 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Chaetoceros lauderi	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 1 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 1
Hermiaulus	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 1 1 1	1 1 1 1	1 1 1 0	0 0 0 0	0 0 0 0
Navicula	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 1 1 0	0 0 0 0	0 0 0 0	0 1 1 0	0 0 0 0	0 0 0 0
Navicula crabro	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 1 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Nitzschia seriata	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 1 0 1	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
Rhizosolenia fragilissima	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 1 1
Dictyocha speculum	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 1	0 0 0 0	0 1 0 0	0 0 1 0	0 0 0 1	0 0 0 0
Dictyocha septenaria	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 1 1 1	0 0 0 0	0 0 0 0
Eutreptiella eupharyngea	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	1 0 0 1	1 1 0 0	0 0 0 0	0 0 0 0	1 0 0 0	0 0 0 0	0 1 0 0

May 1987 samples), undergo an orthogonal transformation (*H* hereafter) toward the reduced data set 2 scores (*y* data hereafter), chosen as target (data set 2 without 9 January 1988 sample). The factor analysis modelling, subject to different orthogonal and oblique transformations, is used extensively for treating behavioural and socio-economic data (Harman, 1976). This can also be a useful tool for recognizing factors in biological and ecological research.

The physical data chosen for a comparative analysis are the same as the samples of the biological data, at the four depths (0, 5, 10, 15 m), using the same sample units as the reduced data-set of the biological data. Thus for the physical-chemical parameters we consider 24 sampling units in the data set 1 and, because of one missing Sampling Unit (22 December 1987 at 10 meters), 23 in the data set 2.

The function to be minimized is the geometrical distance between *Y* and the rotated vector *XH*, Trace $[(Y\text{-}XH)^T (Y\text{-}XH)]$, i.e. the sum of all diagonal values of the resulting 7×7 squared residuals matrix.

This implies maximizing the expression Trace $(Y^T XH)$ subjected to the condition that *H* must be orthogonal.

Choosing the singular value decomposition

$$X\,Y^T X = U\Gamma V^T$$

and the associated transformation

$$H = VU^T,$$

we see that

$$\text{Trace}\,(Y^T XH) = \text{Trace}(U\Gamma V^T) = \text{Trace}(\Gamma).$$

This gives the required transformation, because all the diagonal values of Γ are positive.

To obtain the transformation for the PCA eigenvector, we start from the synthesis formula

$$X = AE^T = \sum_{k=1}^{7} a_{ki} e_{kj}$$

where *X* is written as product of the SUs score matrix, *A,* times the transposed component coefficient matrix, E^T, relative to the first seven PCs.

The rotated scores, A^R, are given by

$$A^R = AH,$$

and, keeping the presence-absence data constant, the new PCs are

$$E^R = EH.$$

3. Results and discussion

In this section the vertical structure of the phytoplankton structure, as well as its temporal evolution, is addressed through statistical techniques.

The binary data are used to connect the presence of the species with the total variance, while their quantitative individual contributions to the total biomass are taken into account during blooms periods.

For the Jaccard index (Legendre and Legendre, 1998) the value 0.43 is obtained, i.e with 49 species present in data set 1 and data set 2, only 21 are common.

The preceding result shows that the species in the two data sets are different. This agrees with the analysis of Cabrini et al. (1992) using the original phytoplankton data in cell/liter. This evidence is a motivation to analyze separately the data set 1 and data set 2 separately, and only then, to search for a common description.

During the two periods of the analysis the stratification is neglected according to the results of the preceding section. The average temperature in data set 1 is 12.07 °C , while in data set 2 it is higher at 13.85 °C . The average salinity of data set 1 is 37.22, decreasing to 36.42 during the data set 2 period.

These results are in agreement with the oceanographical analysis presented by Crisciani and Ferraro (1990), which gives a peak in daily mean rainfall and an increase of the mean temperature in the Gulf of Trieste during the winter of 1987-1988. Moreover, the comparison of the variances relative to the two data sets leads to a Fisher ratio of data set 1 versus data set 2 equal to 4.00. This rejects, at the 0.1% significance level, the hypothesis of homoscedasticity of the two samples. Thus during the two periods under analysis the stratification is neglected.

The similarity analysis of the Sample Units of data set 1 provides five classes: SUs No.15, 35 and 40 are not taken into account in the SUs classification because their similarity values are distant from any other class in the dendrogram.

The hypotheses tested provide the same results for all five classes: the first hypothesis (H0) is accepted at a 5 % significance level, while the second (H0') is rejected at the same significance level.

The classification into five classes, plotted according their temporal succession, is reported in Fig. 1.

The test on the hypothesis for data set 2 provides four main classes according to the seasonality (Fig. 2).

As long as the number of classes in the previous classifications is not constrained, we cannot compare the results of one year with respect to the other. In particular, the results presented in Fig. 1 do not show a higher species replacement than those in Fig. 2, because there are more classes at the same dissimilarity level.

The PCA performed on data set 1 gives a first, second and third component account for 15, 10 and 8% of the total variance. Fig. 3 reports the SUs scores with respect to the first two Principal Components (PC). The plot clearly shows an ecological gradient due to the seasonality. This gradient changes for the first component.

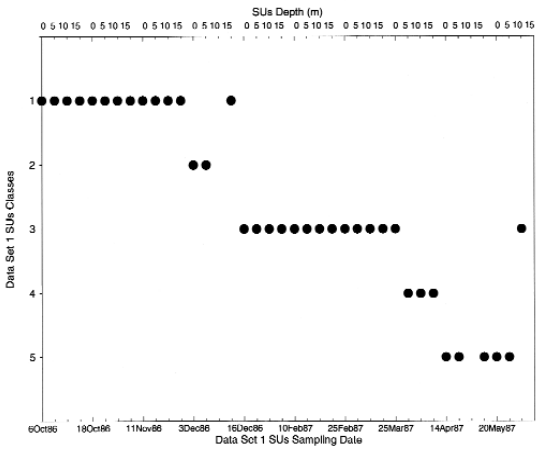

Fig. 1 - Sampling Units groups of data set 1 plotted according to the time succession. SUS 15, 35 and 40 are not shown because they represent single groups.

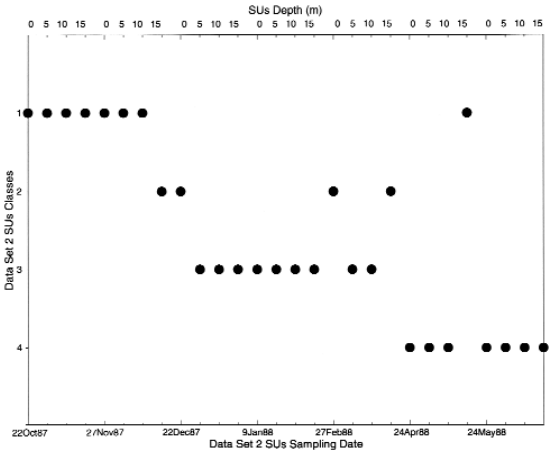

Fig. 2 - Sampling Units groups of the data set 2 plotted according to the time succession.

The horseshoe shape in Fig. 3 shows a non-linear pattern within the data set, which depends on the length of the gradient, in appearance showing a complete folding of the gradient. This does not imply that the fifth (last spring) group is similar to the first (autumn) group. In fact considering also the distance with respect to the third component (not shown), we note that the previous two groups are really very far from each other in the multidimensional space.

The first principal component is correlated with temperature: $r = 0.79$ at a significant probability of less than 0.05. None of the first three components are statistically correlated with depth. Species with a frequency of less than 15% (in the SUs data table) and those present several times during the year are not correlated ($p < .05$) with the first two components.

Fig. 4 suggests the definition of four aggregated groups of species, each denoted by the label given in the text, against the first and second principal components for data set 1.

The other species do not show any seasonal pattern and thus are not considered in the following analysis.

The first group present in spring (Group S) consists of dinoflagellates: *Protoperidinium bipes* (Pbi); diatoms: *Licmophora* sp. (Lsp), *Licmophora abbreviata* (Lab), *Skeletonema costatum* (Sco), *Thalassiosira decipiens* (Tde), *Thalassiosira* sp. (Tsp).

The autumn group (Group A) consists of diatoms: *Guinardia* sp. (Gsp), *Rhizosolenia gracillima* (Rgr), *Rhizosolenia delicatula* (Rde), *Rhizosolenia stolterfothii* (Rst), *Rhizosolenia indica* (Rin).

The autumn and spring group (Group M) consists of dinoflagellates: *Gymnodinium* sp. (Gsc); diatoms: *Leptocylindrus danicus* (Lda), *Cerataulina pelagica* (Cpe).

The winter group (Group W) consists of only one dinoflagellate: *Prorocentrum micans* (Pmi).

For the period analyzed the typical succession pattern (Aubert, 1988) is confirmed: first

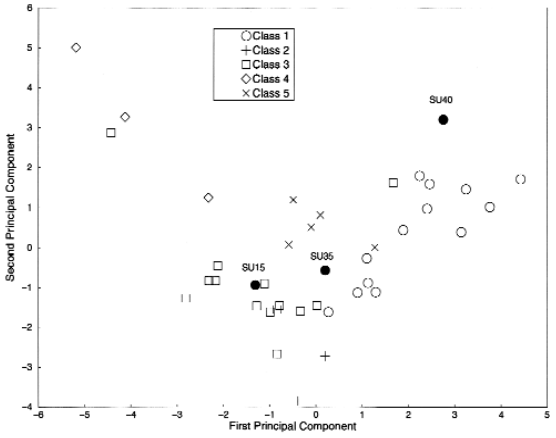

Fig. 3 - Data set 1 SUs Ordination in PC space. The classes are given by the Similarity Analysis (Fig.1). SUS 15, 35 and 40 are shown as single groups.

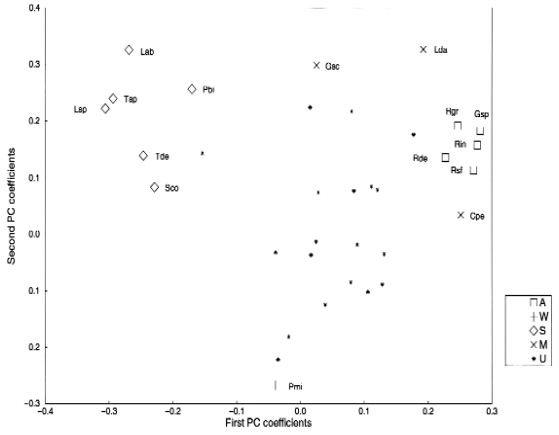

Fig. 4 - Data set 1 Species Component coefficients in PCA Space: A=autumn, W=winter, S=spring, M=autumn and spring, U=uncorrelated with seasons (see text for taxa labels).

there is an autumn bloom, mainly of diatoms, followed by dinoflagellate blooms.

The diatoms of groups A and M reach a value of about 200μgC/liter (Cataletto et al., 1993), while dinoflagellates present smaller values. The diatom biomass is mainly due to *Leptocylindrus danicus* and *Nitzschia* sp. (Fonda, 1992).

We note that the last genus is not put in evidence by the analysis because its contribution to the variance of the components in the presence-absence analysis is low.

The biomass of the species of group W, approximately 100μgC/liter, is dominated by the dinoflagellates and microflagellates (Cataletto et al., 1993).

Group S and part of group M are characterized by *Skeletonema costatum and Thalassiosira decipiens*. These are characteristic species of the Mediterranean environment and they have a biomass, dominated by *Skeletonema costatum*, of about $2 \cdot 10^6$cell/liter during spring bloom.

For data set 2 the first, second and third components account for 16, 14 and 9% of the total variance respectively. Fig. 5 reports the SUs scores with respect to the first two principal components.

The plot shows an ecological gradient according to the seasonality. The SUs of late autumn and winter are closer on the first two principal components plane than on those of the other periods. This gradient changes along the first and second components.

The autumn and spring groups are more scattered than the other two groups: this implies a reduced homogeneity of the phytoplanktonic population in the water column. This is well in accordance with other papers (Fonda et al., 1992) and also with the salinity and temperature data of the same samples when the stratification begins.

The third principal component is negatively correlated with temperature: r = -0.38 at a significant probability less than 0.05. None of the first three components is statistically correlated with depth.

With respect to the ordination of data set 2, the species with a frequency of less than 15% (in

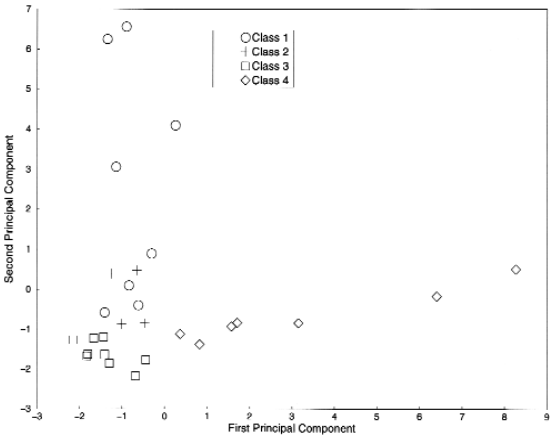

Fig. 5 - Data set 2 SUs Ordination in PC space. The classes are given by the similarity analysis (Fig.2).

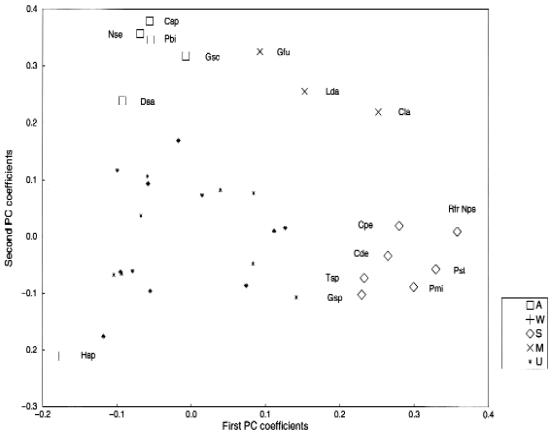

Fig. 6 - Data set 2 Species Component coefficients in PCA Space: A=autumn, W=winter, S=spring, M=autumn and spring, U=uncorrelated with seasons (see text for taxa labels).

the SUs data table), and those present several times during the year, are not correlated ($p < .05$) with the first two components.

Fig. 6 shows the four main groups of species (identified by the text labels) against first and second PCs, giving the seasonal pattern for this data as well.

The first group, present in autumn (Group A), consists of dinoflagellates: *Dinophysis sacculus* (Dsa), *Gymnodinium* sp. (Gsc), *Protoperidinium bipes* (Pbi); diatoms: *Nitzschia seriata* (Nse), *Chaetoceros* sp. (Csp).

The autumn and spring group (Group M) consists of dinoflagellates: *Gyrodinium fusiforme* (Gfu); diatoms: *Leptocylindrus danicus* (Lda), *Chaetoceros lauderi* (Cla).

The late spring group (Group S) consists of dinoflagellates: *Prorocentrum micans* (Pmi), *Protoperidinium steinii* (Pst), diatoms: *Chaetoceros decipiens* (Cde), *Guinardia* sp. (Gsp), *Nitzschia Pseudonitzschia* (Nps), *Rhizosolenia fragilissima* (Rfr), *Thalassiosira* sp. (Tsp), *Cerataulina pelagica* (Cpe).

Prorocentrum micans, Chaetoceros decipiens, Guinardia sp. and *Cerataulina pelagica* are present during the winter too.

The winter group (Group W) consists of only one diatom: *Hemiaulus* sp. (Hsp).

During autumn the diatom population is dominated by Chaetoceros sp. and *Leptocylindrus danicus* with about 50 - 100 $\cdot 10^3$ cell/liter, while *Nitzschia seriata* presents values one order lower.

During winter *Hemiaulus* sp., group W, has values of about 10^6cell/liter.

The S and M groups present species that are typical of the spring season. The diatoms are dominated by *Chaetoceros* sp. and *Leptocylindrus danicus* with about 50 $\cdot 10^6$cell/liter.

Hereafter we consider the species that clearly show a seasonal pattern, i.e. species correlated with the PCA axes. These provide information on the temporal dynamics of the phytoplankton. We refer to the species present during both periods. The following 7 species are common to data set 1 and data set 2 (Jaccard index 0.5):

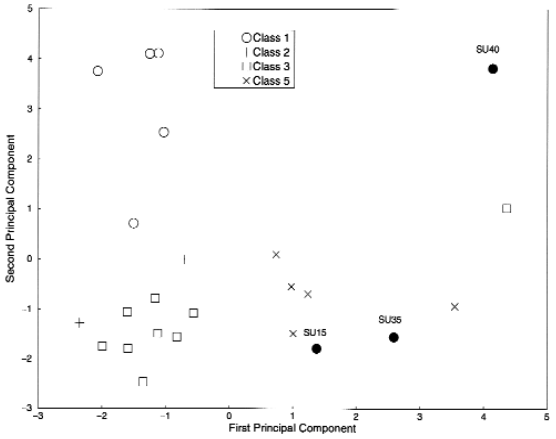

Fig. 7 - Reduced data set 1 SUS ordination after PC transformation (see further explanation in the text).

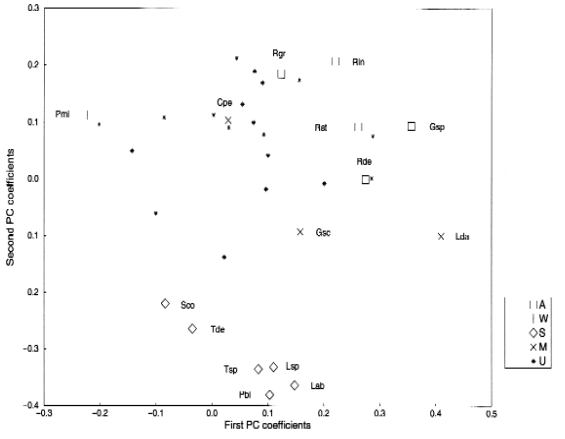

Fig. 8 - Reduced data set 1 Component Coefficients after PC transformation (see further explanation in the text).

Thalassiosira sp., *Cerataulina pelagica, Leptocylindrus danicus, Guinardia* sp. as diatoms; *Protoperidinium bipes, Gymnodinium* sp., *Prorocentrum micans* as dinoflagellates.

The winter groups (W) and autumn groups (A) have no common species; the spring groups (S) have only one common species, *Thalassiosira* sp.; the spring and autumn groups (M) have two common species, *Leptocylindrus danicus* and *Cerataulina pelagica*.

The results show that seasonal associations do not depend on the same species every year.

The interpretation of the first principal components does not provide the same model for both years. Assuming that for each year the same factors control the dynamics of the community, PCA orthogonal transformation is performed with the common features in the multidimensional space of the two years.

In Fig. 7, the succession, after the rotation, of the reduced data set 1 is shown. So far we have elucidated the same temporal structure of SUs during both years. In the following, we search for common characteristics of the ordination of the species in the two years. The orthogonal transformation performed assures that it is possible to match the seasonality of data set 1 to that of data set 2 and that the former shows a similar time ordering as to the latter.

Fig. 8 shows the component coefficients of all species of data set 1 with respect to the first and second PCs after rotation. Assuming that the following hypotheses are true, we can perform the transformation on the complete PCs.

Firstly, both the complete and reduced data sets must have a similar internal structure in terms of correlation between objects and association between variables. In this case the reduced data set is a "sample" of the complete data set.

Secondly, the temporal sequence of the complete data set is included within the temporal limits of the reduced one, because extrapolation is not allowed. Consequently, a higher number of species can be included in the transformation process.

Thus we obtain two main results for the time ordering of the SUs and the grouping of the phytoplankton. Firstly, although strong non-linear effect, as well as complete species replacement in time exists (Figs. 3 and 4), the chosen orthogonal transformation is capable of

recovering the same scores ordination.

Secondly, the first PC explains a low proportion of the total variance. This may be the result of our method which uses ordering with continuous values rather than binary data. This choice is linked to the fact that PCA reproduces the distance between the objects in the multidimensional space. The first eigenvalues are not very different in value and are not correlated with the same variables between the two periods being analysed here. After rotation a similar structure, in particular from the time ordering of the species, emerges.

A straightforward interpretation of the rotation gives a connection between the two data sets species succession. The first PC is influenced by the growth limitation in the water column, which enters its most favourable situation in early spring. The second PC is related to a physical factor instead, which in turn is related to the destruction of the thermocline and the increase of mixing processes.

4. Conclusions

The results show that a community having a fast replacement of species can be studied if taxa levels are quantified by their presence and absence: this allows a quick and less expensive description of the community (because taxa are not quantified by measuring them but only by identifying their presence). Nevertheless because of the year to year change in the species which are characteristic for each season, time series of more than one year are more difficult to describe, should categorisation at species level be adopted.

The problem may be overcome by adopting a suitable orthogonal transformation of one-year data on a target given by the other-years data. The SUs ordination (Fig. 7) indicates an increased accuracy of the description and the counterclockwise ordination of the species is obtained again after rotation of data set 1 (Fig. 8). Moreover, the different correlations with the factors in the different seasons can be treated in a qualitative way. For winter groups, for example, *Prorocentrum micans* is not related to mixing, while *Hemiaulus* sp. , which appears later in data set 2, is strongly connected to vertical processes.

Acknowledgments. This work was supported by the Italian Government, Programma di RIcerca e Sperimentazione per la salvaguardia del Mare Adriatico (PRISMA) fase 2, Contract No. 96.02131.PG03.

References

Aubert M.; 1988: *Théorie générale de l'eutrophisation.* In: Eutrophication in the Mediterranean Sea: receiving capacity and monitoring of long term effect, UNESCO reports in Marine Science, 49.

Bakus G. J.; 1990: *Quantitative Ecology and Marine Biology.* Balkema, Rotterdam.

Cabrini M., Milani L., Honsell G. and Fonda Umani S.; 1988: *The phytoplankton in a station in the Gulf of Trieste from March 1986 to September 1988: data report.* Nova Thalassia, **9**, 11-52.

Cabrini M., Cataletto B., Fonda Umani S., Milani L. and Pavesi C.; 1992: *Food webs in the Gulf of Trieste (Northern Adriatic Sea).* Rapp. Comm. int. Mer Médit., **33**, 367.

Cataletto B., Cabrini M., Fonda Umani S., Milani L. and Pavesi C.; 1993: *Variazioni del contenuto in C della biomassa fito-, microzoo- e mesozoo-planctonica nel Golfo di Trieste.* Biologia Marina, Suppl. al Notiziario S.I.B.M., **1**, 141-144.

Cataletto B., Feoli E., Fonda Umani S., Monti M. and Pecchiar I.; 1996: *Analyses of the relationship between mucous aggregates and phytoplankton communities in the Gulf of Trieste (Northern Adriatic Sea) by multivariate techniques.* Marine Ecology, **17**, 291-308.

Crisciani F. and Ferraro S.;1990: *Climatological aspects of the occurrence of "mare sporco" (dirty sea) episodes in the northern Adriatic Sea during the period 1841-1990.* Boll. Ocean. Teor. Appl., **8**, 289-298.

Fonda Umani S.; 1992: *Successioni fitoplanctoniche, micro e mesozooplanctoniche nell'Alto Adriatico.* In: Marchetti R. and Cotta Ramusino M. (eds), Atti V Congresso SITE, **15**, pp.221-248.

Fonda Umani S., Franco P., Ghirardelli E. and Malej A.; 1992: *Outline of oceanography and the plankton of the Adriatic Sea.* In: Colombo G., Ferrari I., Ceccherelli V. U. and Rossi R. (eds), Marine eutrophication and population dynamics, pp. 347-365.

Fonda Umani S., Sun C.-Y., Feoli E., Cataletto B., Cabrini M. and Milani L.; 1995: *Is it possible to identify any plankton succession in the Gulf of Trieste (Northern Adriatic Sea)?* In: Eleftheriou A., Ansell A. D. and Smith C. J. (eds), Biology and Ecology of Shallow Coastal Waters, pp. 59-65.

Fonda Umani S.; 1996: *Pelagic production and biomass in the Adriatic Sea.* Sci. Mar., **60** (Supp. 2), 65-77.

Harman H. H.; 1976: *Modern Factor Analysis.* University of Chicago Press, Chicago.

Jolliffe, I. T.; 1986: *Principal Component Analysis.* Springer, New York.

Legendre P. and Legendre L.; 1998: *Numerical Ecology*. Second English Edition, Elsevier, Amsterdam.

Milani L., Cabrini M., Fonda Umani S. and Honsell G.; 1988a: *The microzooplankton in a station in the Gulf of Trieste from March 1986 to September 1988: data report.* Nova Thalassia, **9**, 53-95.

Milani L., Cabrini M., Fonda Umani S. and Honsell G.; 1988b: *Environmental parameters (temperature, salinity, dissolved oxygen, chlorophyll a and pheopigments) in a station in the Gulf of Trieste from March 1986 to September 1988: data report.* Nova Thalassia, **9**, 97-145.

Orr D. C. and Bowering W. R.; 1997: *A multivariate analysis of food and feeding trends among Greenland halibut (Reinhardtius hippoglossoides) sampled in Davis Strait, during 1986.* ICES Journal of Marine Science, **54**, 819-829.

Pech N. and Laloë F.; 1997: *Use of Principal Component Analysis with Instrumental Variables (PCAIV) to analyse fisheries catch data.* ICES Journal of Marine Science, **54**, 32-47.