

# 'Phytoplankton events' in French coastal waters during 1987–1997

Benoît BELIAEFF<sup>a\*</sup>, Philippe GROS<sup>a</sup>, Catherine BELIN<sup>a</sup>, Bernard RAFFIN<sup>a</sup>, Isabelle GAILHARD<sup>b</sup>, Jean-Pierre DURBEC<sup>c</sup>

<sup>a</sup> Ifremer, BP 21105, 44311 Nantes cedex 3, France

<sup>b</sup> Ifremer, BP 70, 29280 Plouzané, France

<sup>c</sup> COM, Faculté des sciences de Luminy, case 901, 13288 Marseille cedex 09, France

**Abstract** – This study aims to propose a tool to describe the long-term (10 years) variability of phytoplanktonic assemblages monitored by Rephy (French monitoring programme for phytoplankton and phycotoxins) in the English Channel, the Bay of Biscay and the Mediterranean French coastal waters. According to the sampling strategy (systematic survey, with a time-step of 1 or 2 weeks, and a between-sampling station distance ranging from less than one to several kilometres), the information content of the data is mainly relevant to the characterization of seasonal variability of populations at the mesoscale. For any given area, and for each of the 56 taxonomic units considered here, the data are thus processed in order to recognize the temporal succession of 'phytoplankton events': an event is defined by retaining qualitative information only. It encompasses the phases of sudden growth, high level of abundance and decline of a population. Times at which an event begins or ends are detected by using a time-series segmentation method, which allows to summarize the whole data set as a multivariate set of event occurrences. Categorizing observations in such a way also provides an efficient tool for the identification of different patterns of variability over long-term time scales, for instance: 'recurrent events' (e.g. populations generating events in a periodic-like manner), or 'anomalies' (e.g. of climatic origin). On an univariate basis, an 'average' event can be defined for each taxum, characterized by its within-year position, its duration, its magnitude and the associated deviations. On a multivariate basis, graphical representation of event successions could also allow between-year comparisons. A simple multivariate analysis was also used to describe the seasonal pattern of some taxa. © 2001 Ifremer/CNRS/IRD/Éditions scientifiques et médicales Elsevier SAS

**Résumé** – Événements phytoplanctoniques dans les eaux côtières françaises de 1987 à 1997. L'objectif de cette étude est de proposer un outil pour la description de la variabilité à long-terme (10 ans) de communautés phytoplanctoniques observées par le Rephy (Réseau de surveillance du phytoplancton et des phycotoxines) dans la Manche, dans le golfe de Gascogne et dans les eaux côtières méditerranéennes françaises. Du fait de la stratégie d'échantillonnage (surveillance systématique avec un pas d'échantillonnage temporel de 1 à 2 semaines et une distance inter-stations allant d'environ un à plusieurs kilomètres), l'information contenue dans les données est principalement pertinente pour la caractérisation des variations saisonnières des populations à méso-échelle. Pour chaque site d'échantillonnage et pour chacune des 56 unités taxonomiques (genres) considérées ici, les données sont traitées afin de permettre l'identification d'une succession d'« événements phytoplanctoniques » : un événement est défini en conservant uniquement une information qualitative. Ainsi, l'événement comprend la phase de croissance subite, de niveaux élevés d'abondance, puis de déclin de la population. Les instants auxquels un événement débute ou finit (changements brusques d'abondance dans la population au cours du temps) sont détectés en utilisant une méthode de segmentation des séries chronologiques, qui permet de résumer les séries de données en un ensemble multivarié d'événements. Ce moyen de catégoriser les observations fournit un outil efficace pour l'identification des différents schémas de variabilité, par exemple des « événements récurrents » (c-a-d, des populations générant des événements de manière quasi-périodique),

\*Correspondence and reprints: fax: +33 2 40374073.

E-mail address: Benoit.Beliaeff@ifremer.fr (B. Beliaeff).

ou des « anomalies » (par exemple d'origine climatique). En mode univarié, un événement « moyen » peut être défini pour chaque taxon, caractérisé par son apparition dans l'année, sa durée, son amplitude et éventuellement les incertitudes attachées à ses estimations. En mode multivarié, la représentation graphique de successions d'événements autorise la comparaison visuelle entre années pour un site donné. Une application très simple d'analyse multivariée a finalement été utilisée pour décrire la variabilité saisonnière d'un groupe de taxons. © 2001 Ifremer/CNRS/IRD/Éditions scientifiques et médicales Elsevier SAS

**phytoplankton / French coasts / spatio-temporal variability / statistical analysis / time-series segmentation**

**phytoplankton / littoral français / variabilité spatio-temporelle / analyse statistique / segmentation de séries chronologiques**

## 1. INTRODUCTION

The large variability of phytoplankton abundance in coastal waters has been pointed out by several authors (e.g. Platt, 1972). This is true at every spatial and temporal scales and collecting representative samples in a given area and for a given period of time should take into account variability sources such as the influence of tidal and wind-induced currents, possible daily migrations in the water column, and the intrinsic dynamics of phytoplanktonic communities. These factors generate both temporal and spatial variability, and these two dimensions cannot be distinguished unless a fine spatial 3-D grid is sampled along time. Up to now we did not find any report of such sampling effort. This is especially the case for monitoring programmes basically constrained by their budget, as with the French programme for phytoplankton and phycotoxins monitoring (Rephy) managed by Ifremer. Although raw data generated by this programme exhibit a very large variability at small scales, the quantity of information of the whole data set should allow detection of large scale spatio-temporal patterns.

In this context, this paper aims to propose a way of processing data in order to better describe and characterize phytoplankton populations at large spatial and temporal scales. The key point here is to consider that a succession of phytoplankton observations at a given sampling station in the context of our monitoring programme could be seen as a single 'event', encompassing the phases of sudden growth, high level of abundance and decline of the observed taxum. This way a set of often erratic observations can be converted into a single event.

We used a time-series segmentation method to obtain phytoplankton events on an univariate basis. Statistics can then be computed on the event data base generated.

Eventually multivariate methods can then be applied to the multi-dimensional tables. A simple application using correspondence analysis on a genus  $\times$  month contingency table of event occurrences is given.

## 2. MATERIALS AND METHODS

### 2.1. Rephy data processing

Data, stored in the Ifremer Quadrige database, are concentrations of phytoplanktonic taxa in the water column measured since 1987 at sampling points scattered all over the French coastline, and grouped into 56 higher-scale areas called basins. Due to taxonomic identification difficulty, some species cannot be identified. In this case only the genus is reported in the database. Therefore, the taxonomic level chosen for this study is the genus. This gives a total number of 56 genera. For a given genus, the observed concentration provides an estimate of the number of cells per volume unit summed over all species identified within that genus (e.g. *Dinophysis* aggregates *D. acuminata*, *D. caudata*, *D. sacculus*, etc.).

In order to smooth the spatio-temporal variability at the sampling point scale, data are combined at a higher spatial scale, the basin scale, which corresponds somehow to an homogeneous spatial unit in terms of watershed and hydrodynamical characteristics. As a precautionary-type rule, in case of several observations in a basin for a given date, the maximum concentration value was retained.

## 2.2. Time-series segmentation

The method used for the time-series segmentation and thus for the event identification is the so-called Page-Hinkley Cusum (Page, 1954; Basseville et al., 1991), a method usually used in signal processing. This method detects a change of mean level on-line, the magnitude of which is unknown, in a piecewise stationary signal to which a white noise is added. Two detectors are activated in parallel to detect sudden increases in mean level (resp. decreases). The concept of a minimum jump  $\nu_m$  to be detected is introduced in the detector definition. The mean level is estimated on-line before the jump. A more formal description of the method is given below.

### 2.2.1. Model

For each taxonomic unit (i.e. genus), let us define the random variable  $X_t$  as  $X_t = \log_{10}[Q \text{ cells} \cdot (10 \text{ mL}^{-1}) + 1]$ , at  $t$  with  $Q$  as the genus abundance. For a given sampling point we assume that  $\{X_{t+k\Delta t}\}_{k=0,1,\dots} \equiv X_0, X_1, \dots, X_k, \dots$  is a purely random process. Thus  $X_0, X_1, \dots, X_k, \dots$  are mutually independent and identically distributed, with the following moments (for more theoretical details see, e.g. Chatfield, 1989):

$$\begin{cases} E(X_k) = \mu_r & \text{when } r \leq k < r' \\ E(X_k) = \mu_r + \nu & \text{when } r' \leq k < r'', \text{ and } \text{Var}(X_k) = \sigma^2, \end{cases}$$

where  $r, r', r''$  are temporal indices and  $\nu$  either a positive ( $\nu > 0$ ) or a negative jump ( $\nu < 0$ ).

In this context a ‘phytoplankton event’ will be defined as a set of  $X_k$  values bounded by two jumps, a positive jump ( $\nu > 0$ ) followed by a negative jump ( $\nu < 0$ ). The following Page-Hinkley Cusum algorithm allows to identify such jumps in the  $\{X_{t+k\Delta t}\}_{k=0,1,\dots}$  time series.

### 2.2.2. Algorithm

We first define  $\hat{\mu}_{k,h}$  as the signal ‘short-term’ memory, i.e. the average over the duration  $h\Delta t$  before a jump occurring at time  $k$ . We thus choose  $\hat{\mu}_{k,h}$  as:

$$\hat{\mu}_{k,h} = \frac{1}{h} \sum_{j=k-h-1}^{k-1} (X_j).$$

As for the sampling frequency, here  $\Delta t$  ranges from 1 to 2 weeks, by far the most frequent temporal lag between two consecutive observations.

Let  $\delta_k = X_k - \hat{\mu}_{k,h}$ , to be used in the following computations, and  $\lambda$  be a fixed threshold. A drop in mean level will be detected when  $g_n = M_n - T_n \geq \lambda$ , with

$$T_0 = 0, T_n = \sum_{k=1}^n \left( \delta_k + \frac{\nu_m}{2} \right) \text{ and } M_n = \max_{0 \leq k \leq n} T_k.$$

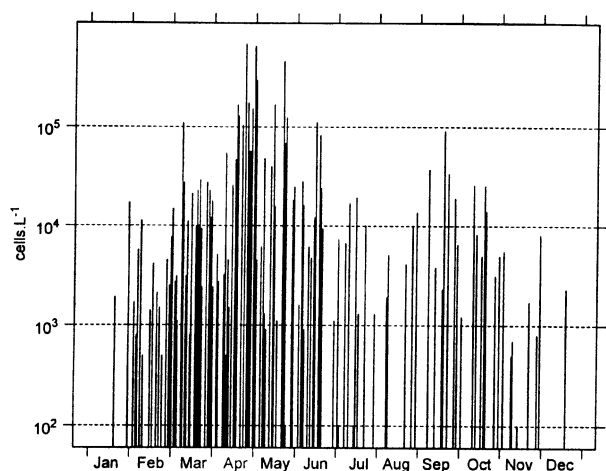
Conversely, an increase in mean level will be detected when  $g_n = U_n - m_n \geq \lambda$ , with

$$U_0 = 0, U_n = \sum_{k=1}^n \left( \delta_k - \frac{\nu_m}{2} \right), \text{ and } M_n = \min_{0 \leq k \leq n} U_k.$$

An alarm is triggered when one of the two detectors exceeds  $\lambda$ . Theoretically, the jump instant estimator corresponds to the last time the current maximum ( $M_n$ ), and respectively minimum ( $U_n$ ), was reached. Due to a massive number of zeroes in phytoplankton time series we defined starting and ending dates of a given event as respectively the first and the last non-zero values bounding the theoretical event period.

## 2.3. Multivariate analysis

Factorial analysis is used as a description tool in order to identify hidden structures in multi-dimensional tables. These techniques have been widely used in ecology, and more specifically in marine phytoplankton studies. In particular, correspondence analysis (for theoretical details, see e.g. Escofier and Pagès, 1990) proved to be an adequate method for describing spatio-temporal patterns of phytoplankton community (e.g. Matthews et al., 1991). In correspondence analysis, the chi-square distance between points is appropriate to analyse either a species presence/absence or abundance table (e.g. Ryckaert et al., 1983), or a contingency table comparing two qualitative (categorical) variables. We used the latter approach, where our points are number of event occurrences for a given value of the first categorical variable (e.g. genus) crossed with the value of the second categorical variable (e.g. month of event beginning). Other categorical variables could also be used such as the geographical location or the year.



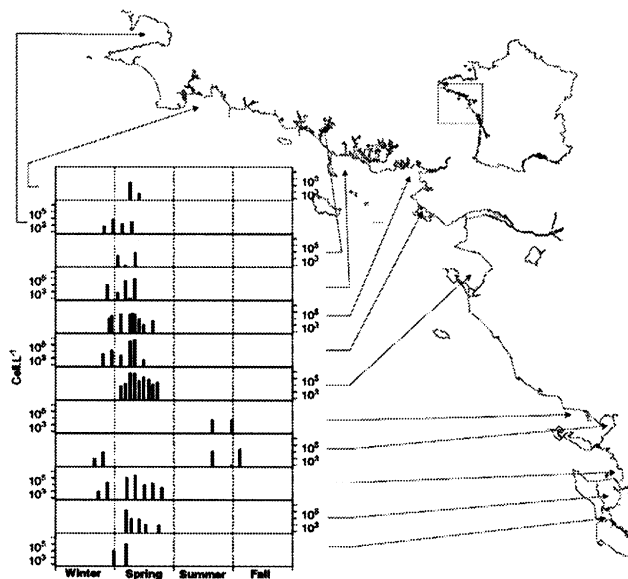
**Figure 1.** Global pattern of seasonal variation of *Asterionella* in the Bay of Biscay for the 1987–1997 period.

### 3. RESULTS

#### 3.1. Event database

After several attempts,  $\lambda = 0.5$  and  $\nu_m = 2.5$  were the chosen values for the algorithm tuning parameters. Those values were calibrated by examining the generated events: they proved to be satisfying whichever taxa was considered thanks to the logarithmic transformation. Namely, it was successful for bloom-forming taxa s.a. *Alexandrium*, reaching  $10^8$  cells·L<sup>-1</sup>, as well as for *Dinophysis*, maxima of which are in the order of  $10^4$  cells·L<sup>-1</sup>. Single non-zero values were identified and in this case the minimum value to be detected was  $2 \cdot 10^3$  cells·L<sup>-1</sup>.

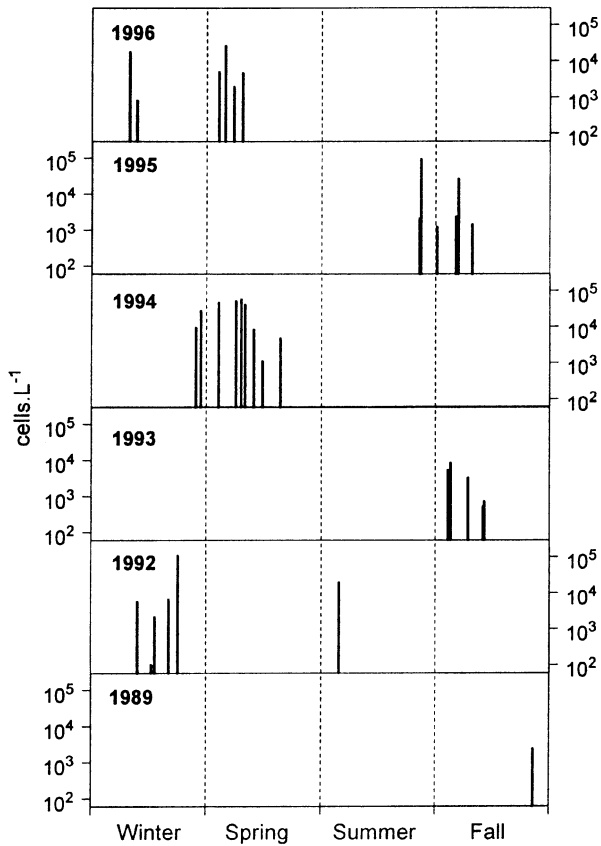
Using only observations within identified events, *figure 1* illustrates the whole pattern of seasonal variations of the genus *Asterionella* in the Bay of Biscay for the 1987–1997 period. Maxima, close to  $10^6$  cells·L<sup>-1</sup>, are reached in April–May. Consecutive fall blooms show lower maxima in September and are seemingly less frequent. *Figure 2* shows the spatial distribution of the corresponding 1994 *Asterionella* time series. For this particular year, *Asterionella* seasonal pattern looks remarkably stable with a spring bloom except for two sampling sites, where mid-summer and start-of-fall events were identified. Eventually, high between-year variability, presumably due to changes in meteorological conditions, can be visualized in *figure 3*, where *Asterionella* was not observed in years 1987, 1988, 1990, 1991



**Figure 2.** Spatial distribution of 1994 *Asterionella* events in the Bay of Biscay.

and 1997 and where winter blooms were observed in 1993 and 1995 and fall blooms in 1992 and 1996.

The Cusum method allowed us to generate a new database, which reduces the original database from around 80 000 raw data records to 4 882 event records. In this new base, each event is identified by its geographical location (basin), its starting and ending dates as estimated by the algorithm, its duration in days and the number of observations. Empirical distributions for these parameters can be drawn from this database. For example various event duration frequency histograms can be obtained (*figure 4*). More than 60 % of 42 *Phaeocystis* events correspond to unique values pointing out the narrow temporal window where blooms for this genus can be captured by the sampling process. Two-thirds of 432 *Pseudonitzschia* events correspond to a duration of less than 3 weeks whereas *Dinophysis*, with 68 events, shows a more symmetric distribution with a month-long event mode. The following statistics are given in *table 1* by genus: number of events, median number of observations per event, median event duration in days, median month of event start, median maximum concentration and maximum maximum concentrations. Early genera s.a. *Scenedesmus* can be opposed to late genera s.a. *Odontella*, omnipresent diatom *Pseudonitzschia* to rare

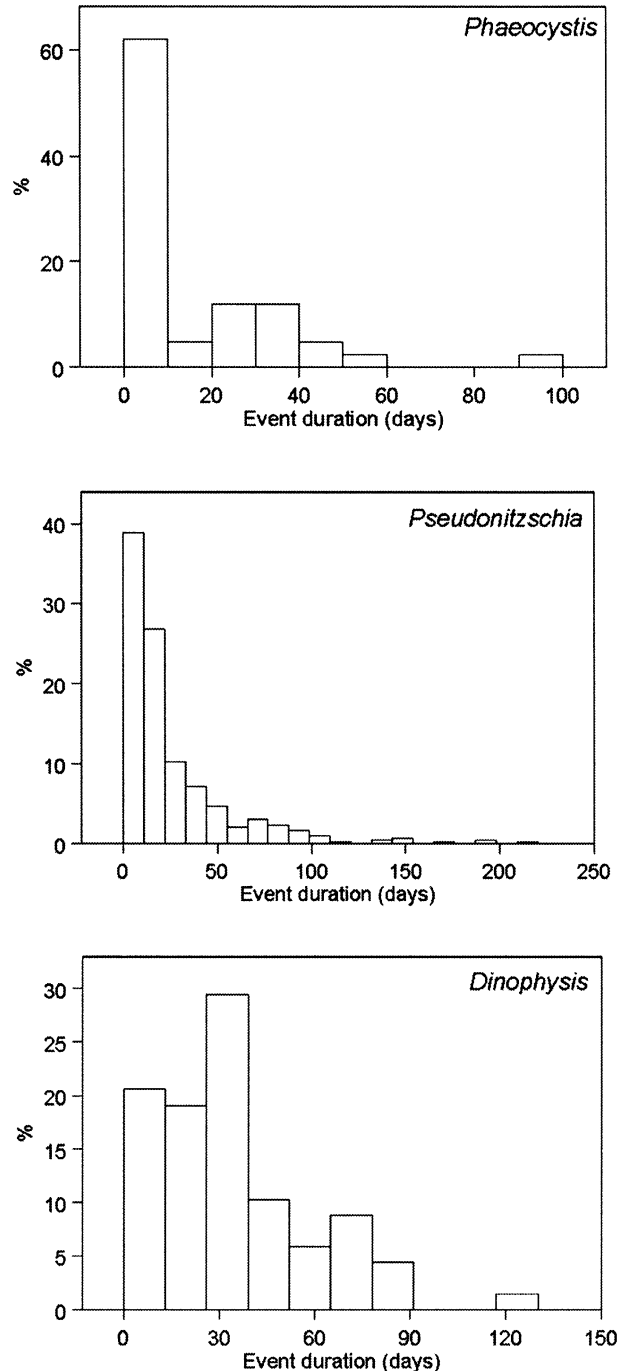


**Figure 3.** *Asterionella* event between-year variation (1987–1997) in the Bay of Vilaine.

*Protoperidinium*, ever-lasting *Navicula* to fleeting *Bacillaria* or *Phaeocystis* in terms of duration or number of observations, a ‘red-tide’ forming *Chaetoceros* to a toxic but ‘invisible’ dinoflagellate *Dinophysis* with a four-logarithmic unit difference in maxima concentrations.

### 3.2. Multivariate analysis

A correspondence analysis was performed on a genus × month (event start) contingency table, the cells of which contain event occurrences for the period 1988–1997 in the Bay of Biscay. In this area, for example, there were 19 *Chaetoceros* events starting in April over the whole period. The year 1987 was not used as it was the starting year of the Repty monitoring programme and taxonomic identification problems were encountered a posteriori. Moreover not all taxa but thirty



**Figure 4.** Frequency distributions of event duration (in days) for the genera *Phaeocystis*, *Pseudonitzschia* and *Dinophysis*.

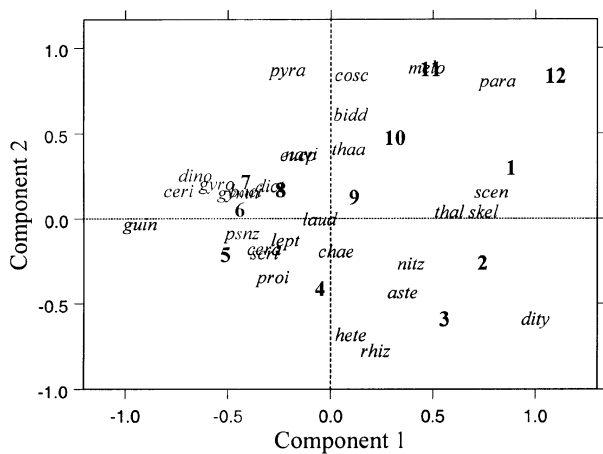
among them (*table II*) were considered to give more clarity in the analysis graphical results. The first two

**Table I.** Event statistics for the 1987–1997 period; for explanation see text. Genus Quadriga codes are given in parentheses.

Genus	Number of events	Median event start month	Median number of observations	Median duration (d)	Median maximum concentration (cells·L <sup>-1</sup> )	Maximum concentration maximum (cells·L <sup>-1</sup> )
<i>Achnanthes</i> ( <i>achn</i> )	11	3	1	0	4 300	4.32E+04
<i>Alexandrium</i> ( <i>alex</i> )	68	7	3	13	12 150	1.80E+08
<i>Asterionella</i> ( <i>aste</i> )	161	4	3	32	21 000	7.77E+07
<i>Bacillaria</i> ( <i>baci</i> )	35	5	1	0	4 000	4.00E+04
<i>Bacteriastrium</i> ( <i>bact</i> )	38	7	1	0	5 700	3.96E+05
<i>Biddulphia</i> ( <i>bidd</i> )	35	8	3	28	6 500	1.34E+05
<i>Cerataulina</i> ( <i>cera</i> )	127	5	2	14	13 600	7.04E+05
<i>Ceratium</i> ( <i>ceri</i> )	35	6	5	40	7 100	8.62E+04
<i>Chaetoceros</i> ( <i>chae</i> )	391	5	6	64	165 000	4.59E+08
<i>Chrysochromulina</i> ( <i>chru</i> )	14	5	3	14	88 500	1.10E+07
<i>Cocconeis</i> ( <i>cocc</i> )	24	7	4	32	8 500	3.15E+05
<i>Coscinodiscus</i> ( <i>cosc</i> )	37	7	5	57	11 000	4.52E+05
<i>Cylindrotheca</i> ( <i>cyli</i> )	26	6	3	25	9 850	1.50E+07
<i>Detonula</i> ( <i>deto</i> )	33	6	1	0	8 400	6.74E+05
<i>Dichyoca</i> ( <i>dict</i> )	30	6	4	38	7 300	1.62E+05
<i>Dinophysis</i> ( <i>dino</i> )	68	7	6	28	6 750	1.58E+05
<i>Diplopsalis</i> ( <i>dipo</i> )	21	7	2	7	3 000	1.95E+04
<i>Ditylum</i> ( <i>dity</i> )	62	3	4	36	6 600	7.30E+04
<i>Ebria</i> ( <i>ebra</i> )	14	8	2	18	4 100	4.96E+04
<i>Eucampia</i> ( <i>eucp</i> )	95	7	2	12	7 500	3.05E+05
<i>Fragilaria</i> ( <i>frag</i> )	25	4	2	16	6 000	3.03E+07
<i>Gonyaulax</i> ( <i>gony</i> )	27	6	3	27	9 000	7.30E+06
<i>Grammatophora</i> ( <i>gram</i> )	33	7	3	29	6 800	1.76E+05
<i>Guinardia</i> ( <i>guin</i> )	64	5	3	27	10 400	1.10E+05
<i>Gymnodinium</i> ( <i>gymn</i> )	193	6	6	56	75 800	6.00E+07
<i>Gyrodinium</i> ( <i>gyro</i> )	82	6	5	41	7 700	4.20E+06
<i>Hemiaulus</i> ( <i>hemi</i> )	17	8	3	27	5 200	1.57E+05
<i>Heterocapsa</i> ( <i>hete</i> )	57	5	3	27	8 300	1.00E+07
<i>Heterosigma</i> ( <i>hetg</i> )	11	9	3	15	15 600	1.33E+08
<i>Katodinium</i> ( <i>kato</i> )	52	5	2	13	10 100	7.64E+05
<i>Kryptoperidinium</i> ( <i>kryp</i> )	19	7	2	5	205 000	4.14E+06
<i>Lauderia</i> ( <i>laud</i> )	175	5	2	13	10 800	1.07E+06
<i>Leptocylindricus</i> ( <i>lept</i> )	332	6	5	48	77 850	8.00E+06
<i>Licmophora</i> ( <i>licm</i> )	60	6	4	32	8 250	1.92E+05
<i>Lithodesmium</i> ( <i>lith</i> )	19	8	4	29	6 000	1.27E+05
<i>Melosira</i> ( <i>melo</i> )	92	5	3	31	12 500	1.86E+06
<i>Navicula</i> ( <i>navi</i> )	68	5	8	85	23 350	5.43E+06
<i>Nitzschia</i> ( <i>nitz</i> )	162	4	7	84	58 100	3.10E+07
<i>Odontella</i> ( <i>odon</i> )	25	9	3	14	7 200	1.45E+05
<i>Oxyrrhis</i> ( <i>oxyr</i> )	11	4	1	0	29 600	1.95E+05
<i>Paralia</i> ( <i>para</i> )	92	5	3	28	7 900	4.00E+05
<i>Phaeocystis</i> ( <i>phae</i> )	42	4	1	0	216 500	2.00E+08
<i>Plagiogramma</i> ( <i>plag</i> )	57	3	3	21	10 000	2.88E+08
<i>Pleurosigma</i> ( <i>ples</i> )	30	7	6	52	7 900	6.92E+04
<i>Polykrikos</i> ( <i>poly</i> )	19	7	2	16	5 200	6.00E+04
<i>Proboscia</i> ( <i>prob</i> )	15	7	3	25	9 400	2.10E+05
<i>Protoperidinium</i> ( <i>prot</i> )	99	5	7	76	15 800	1.97E+06
<i>Protoceratium</i> ( <i>prot</i> )	4	4	1	0	6 000	5.24E+05
<i>Pseudonitzschia</i> ( <i>psnz</i> )	432	6	3	14	32 150	1.00E+07
<i>Pyramimonas</i> ( <i>pyra</i> )	30	7	3	14	7 100	2.15E+05
<i>Rhizosolenia</i> ( <i>rhiz</i> )	278	4	8	91	93 800	3.14E+06
<i>Scenedesmus</i> ( <i>scen</i> )	31	2	2	14	6 800	6.58E+06
<i>Scropsiella</i> ( <i>scri</i> )	121	5	7	63	19 000	2.53E+06
<i>Skeletonema</i> ( <i>skel</i> )	349	6	4	32	40 600	8.30E+07
<i>Thalassionema</i> ( <i>thaa</i> )	178	7	5	56	18 200	4.26E+06
<i>Thalassiosira</i> ( <i>thal</i> )	256	4	5	56	31 850	1.39E+07

**Table II.** Genus × month (event start) contingency table; number in cells are event occurrences.

Genus	Month											
	1	2	3	4	5	6	7	8	9	10	11	12
<i>aste</i>	1	5	14	15	5	5	2	2	6	3	1	2
<i>baci</i>	0	1	0	3	5	2	6	2	1	0	1	1
<i>bidd</i>	0	1	0	1	2	2	1	2	5	3	3	0
<i>cera</i>	0	0	1	18	13	6	3	2	7	3	3	0
<i>ceri</i>	0	0	0	0	7	7	6	0	4	1	0	0
<i>chae</i>	8	14	23	19	21	31	13	11	12	6	2	1
<i>cosc</i>	0	0	0	1	0	1	4	1	1	6	0	1
<i>dict</i>	0	0	1	2	0	2	4	3	3	1	0	0
<i>dino</i>	0	0	0	1	7	12	3	5	1	3	1	0
<i>dity</i>	0	7	17	3	0	1	0	0	2	5	2	0
<i>eucp</i>	1	0	1	5	4	5	9	4	10	7	2	0
<i>guin</i>	0	0	0	2	16	18	5	1	0	0	1	0
<i>gymn</i>	1	2	7	3	13	28	26	9	16	5	0	0
<i>gyro</i>	0	1	1	1	10	6	7	10	1	3	1	0
<i>hete</i>	0	1	9	5	6	1	1	3	1	1	0	0
<i>laud</i>	1	0	6	21	6	12	6	5	9	6	4	2
<i>lept</i>	1	6	16	21	21	23	19	15	12	8	2	0
<i>melo</i>	1	1	3	3	3	7	1	3	3	11	11	4
<i>navi</i>	1	1	2	0	5	6	5	4	2	5	1	1
<i>nitz</i>	4	6	14	3	5	5	6	1	11	1	1	0
<i>para</i>	7	4	1	4	2	2	2	3	3	7	7	5
<i>proi</i>	1	4	6	7	13	10	4	4	1	1	0	0
<i>psnz</i>	1	1	8	36	36	35	21	15	15	11	3	0
<i>pyra</i>	0	0	0	0	0	5	6	1	2	4	3	0
<i>rhiz</i>	0	6	33	28	16	6	3	3	6	4	0	0
<i>scen</i>	0	2	5	2	0	3	1	0	0	3	1	3
<i>scri</i>	0	1	10	5	18	14	5	3	4	5	1	0
<i>skel</i>	10	13	29	6	2	7	8	3	17	15	5	5
<i>thaa</i>	0	2	3	3	4	4	4	7	13	9	4	0
<i>thal</i>	7	16	19	8	5	8	4	10	17	12	4	4



**Figure 5.** Graphical results of correspondence analysis on the genus × month contingency table: simultaneous representation of genera and months in the first factorial plan.

components accounted for ca. 60 % of the total variability. The plot of the first factorial plan (*figure 5*) shows a remarkable temporal connectivity between consecutive months, and the genera are located along this temporal path presumably according to their within-year ecological preferenda. For example, *Rhizosolenia* appears to be an end-of-winter/spring blooming species, *Pseudonitzschia* comes later around June, *Dinophysis* more of a summer genus and *Paralia* a cold-water taxum, as was also found in the Channel by Ryckaert et al. (1983).

#### 4. DISCUSSION AND CONCLUSION

What is a phytoplankton event? This seems an embarrassing question to address to scientists working in this

field of study. Some concentration value thresholds are sometimes given in the literature for particular events such as blooms or 'red tides'. In our case, in a crude way, we define an event as what will be given by our algorithm, eventually. Of course, a validity check was made consisting of observing that what we wanted to be detected as an event was retained in the event database and what seemed to us as insignificant noise was not. The results showed that the way the Cusum algorithm was tuned led to the identification of events with as few as one non-zero observation with concentration as low as  $1\,900\text{ cells}\cdot\text{L}^{-1}$ . The underlying reasoning which could justify such skinny events is that single non-zero values are the only testimonies of some broader dynamics in time and space, from which our sampling strategy could only catch one isolated non-zero observation. As determined by the algorithm tuning,  $2\,10^3$  is arbitrary but seems to correspond to some 'rule of thumb' critical value from which people involved in the monitoring have the feeling that 'something happened'. Justification of considering events instead of a suite of abundance measurements also comes from the hypothesis that what is finally important is that a taxum appeared at a given location and at a given time. Thus we considered that what triggered this occurrence, a punctual concept, is at least as information-rich, if not more, as the suite of observations constituting the event. Once started the suite is modulated by environmental factors, and this is summarized by statistics related to the event: magnitude, duration, etc. Other ways of identifying events could have been used (e.g. Poisson process), but Cusum proved to be efficient as soon as data were log-transformed, and as the Cusum algorithm was adequately tuned, i.e. convenient  $\lambda$  and  $\nu_m$  values were determined.

On one hand an algorithm seems to be an objective way of identifying events, on the other hand it contains some arbitrariness and as mentioned the choice of  $\lambda$  and  $\nu_m$ . Other arbitrary procedures were used in the data processing. First, species were aggregated at the genus level. Belin et al. (1995) already used this level to classify French coast 'homogeneous areas' based on dominant phytoplankton assemblages. This seems to be the best compromise between a higher level with a too important loss by reduction of taxonomic resolution and species identification subject to significant within- and between-laboratory biases. The higher the taxonomic level and the less sensible the aggregation in terms of ecological significance. Such procedure might namely cluster spe-

cies from, e.g., the same genus with different physiological and ecological characteristics. This would lead to a non-realistic description of spatio-temporal patterns for the so-built groups. Ecological preferences of the considered species should be carefully checked before aggregating them in view of an exhaustive study on the subject, which this paper does not intend to be. Another arbitrary choice was to retain the concentration maximum value at the basin scale at a given date. The underlying idea is to approach more closely the actual maximum concentration value in a given geographical area, broader than a sampling station in order to reduce part of the spatial variability. Moreover, as there are often several sampling points within each basin, considering a lower spatial scale would lead to more events, in the same body of water with roughly the same environmental conditions prevailing at the day of sampling, and would thus bias the large scale analysis.

The exploitation of the event database provides a better way of characterizing spatio-temporal structures on an univariate basis, and allows to define a typological partition of the counted taxa. *Figure 1* might be seen as an estimate of *Asterionella* seasonal fingerprint in the Bay of Biscay over the years of observation. However, splitting *figure 1* in time and space, leads to the conclusion that it is obviously not easy to distinguish site-specific seasonal variations (*figure 2*) from high between-year variations (*figure 3*). Fall blooms in *figure 2* cannot be directly related to some specific spatial conditions unless some more thorough studies can be conducted on this matter. Although univariate analysis can provide a deep insight in the description of the 'average' event of a particular taxum (*figure 4*), it does not take into account the inter-specific relations in the interpretation, contrary to multivariate analysis. In most previous studies applying correspondence analysis to plankton, the used multi-dimensional tables contain abundance, or phytoplankton count, per unit volume of each taxum (e.g. Ryckaert et al., 1983). Here thanks to the massive RePHY information contained in the Quadrigé database, we could build contingency tables and in light of a very simple application show that event occurrences could provide a sensible description of the seasonal pattern of the phytoplankton communities for the 1988–1997 period in the Bay of Biscay. Monitoring programmes, and the RePHY in particular, are not designed to provide sensible quantitative information to correlate to some other factors at small scales, but their



extension in time and space allowed to catch large scale variability patterns, which we aim to confront to some other patterns, such as those contained in hydrodynamical and weather information.

Finally a set of poor quantitative information at small scales can be converted into an informative set of event occurrences, providing a concept and a tool possibly allowing us to reach the RePHY main objective, i.e. to adequately describe large scale spatio-temporal patterns. In view of achieving an exhaustive description, perspectives of work are (i) to fully validate the phytoplankton RePHY data in the Quadriga database, (ii) to find out with the help of experts the most sensible taxonomic aggregation based on ecological and/or physiological criteria and (iii) to refine the multivariate analysis, using multi-table techniques and/or making use of environmental – mainly hydro-climatic – factors, in view of finding causative agents. Beside factorial analyses, another innovative way of dealing with phytoplankton should be explored: it consists in using hydrodynamic modelling to recreate and summarize physical conditions which are well known to be the main explicative factors for phytoplankton dynamics (Legendre, 1981; Nõmmann and Kaasik, 1992). The model exists for the Bay of Biscay (Lazure and Jégou, 1998). The main difficulty here will be to express model results in a suitable way and at scales relevant to the observed biological phenomena.

### Acknowledgements

The authors are thankful to the twelve Ifremer coastal laboratories responsible for water sample collection and for phytoplankton identification. This study was supported by the French National Programme for Coastal Environment (PNEC).

### REFERENCES

- Basseville, M., Benhallam, A., Doncarli, C., Lucas, M.-F., Rix, H., Thierry, E., De Brucq, D., Colot, O., Kauffman, F., 1991. Segmentation de signaux : fiches descriptives d'algorithmes, GDR 134 TdSI – GT 2 : Reconnaissances et Ruptures. CNRS.
- Belin, C., Beliaeff, B., Raffin, B., Rabia, M., Ibanez, F., 1995. Phytoplankton time-series data of the French phytoplankton monitoring network: toxic and dominant species. In: Lassus, P., Arzul, G., Erard, E., Gentien, P., Marcaillou, C. (Eds.), *Harmful Marine Algal Blooms. Technique et Documentation – Lavoisier, Intercept Ltd*, pp. 771–776.
- Chatfield, C., 1989. *The Analysis of Time Series – An Introduction*. Chapman and Hall, London.
- Escofier, B., Pagès, J., 1990. *Analyses factorielles simples et multiples – Objectifs, méthodes et interprétation*. Dunod, Bordas, Paris.
- Lazure, P., Jégou, A.M., 1998. 3D Modelling of seasonal evolution of Loire and Gironde plumes in the Bay of Biscay continental shelf. *Oceanol. Acta* 21, 165–177.
- Legendre, L., 1981. Le contrôle physique de la production phytoplanktonique à échelle courte et intermédiaire. *Oceanis* 7, 119–129.
- Matthews, R.A., Matthews, G.B., Ehinger, W.J., 1991. Classification and ordination of limnological data: a comparison of analytical tools. *Ecol. Model.* 53, 167–187.
- Nõmmann, S., Kaasik, E., 1992. Hydrodynamical control of phytoplankton succession during the vernal light-limited phase in the Baltic Sea. *Mar. Ecol. Prog. Ser.* 84, 279–292.
- Page, E.S., 1954. Continuous inspection schemes. *Biometrika* 41, 100–115.
- Platt, T., 1972. Local phytoplankton abundance and turbulence. *Deep-Sea Res.* 19, 183–187.
- Ryckaert, M., Gros, P., Erard-Le Denn, E., 1983. Seasonal succession of coastal phytoplankton populations in the Channel. *Oceanol. Acta. Proceedings 17th European Marine Biology Symposium*, Brest, France, September-1 October, 1982, 171–175.