Survey of environmental complex systems: pattern recognition of physicochemical data describing coastal water quality in the Gulf of Trieste

Pierluigi Barbieri,[a] Gianpiero Adami,[a] Sergio Predonzani,[b] Edoardo Reisenhofer[a] and Desiré Luc Massart[c]

[a]Department of Chemical Sciences, University of Trieste, via L. Giorgieri 1, I-34127, Trieste, Italy
[b]Laboratory of Marine Biology, Strada Costiera 336, I-34010 Santa Croce-Trieste, Italy
[c]ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium

Received 28th September 1998, Accepted 1st December 1998

A data set reporting temperature, salinity, dissolved oxygen, nitrogen as ammonia, nitrite and nitrate, silicate, chlorophyll *a* and phaeopigment values, determined in seawaters sampled during two years with a monthly frequency in 16 stations in the Gulf of Trieste, and at different depths of the water column, has been studied.

In order to find synthetic descriptors useful for following the spatial and temporal variations of biogeochemical phenomena occurring in the considered ecosystem, the data set has been factorized using principal component analysis. A graphical display of scores, by means of boxplots and biplots, helped in the interpretation of the data set. The first factor conditioning the system is related to the input of freshwater from the estuary of the Isonzo River and to the stratification of the seawater (thermohaline discontinuity), while the second and third components describe interactions between biological activity, nutrients and physicochemical parameters; typical spring and autumn phytoplankton blooms were identified, in addition to an exceptional winter bloom conditioned by anomalous meteorological/climatic conditions. The fourth principal component explains the reducing activity of seawaters, which often increases when the decomposition of organic matter is relevant. The simple linear model proposed, and the related graphs, are shown to be useful tools for monitoring the main features of such a complex dynamic environmental system. The outlined approach to the considered complex data structure presents in a cognitive easy way (graphical outputs) the significant variations of the data, and allows for a detailed interpretation of the results of the monitoring campaign. Temporal and spatial effects are outlined, as well as those related to the depth in the water column.

Introduction

To obtain a better comprehension of phenomena occurring in the environment and to understand how human activities affect natural cycles, environmental authorities have promoted a growing number of monitoring campaigns in which data on many variables are collected over wide areas and for long periods. Huge data sets are obtained from these research programmes, and their rationalization requires the application of multivariate statistical methods.

This paper discusses the results obtained within such a broad research programme aimed to define the quality of water in the Adriatic Sea;[1] our attention was focused on the specific complex environment of the Gulf of Trieste. The shallow depth of the sea and the strong winds from inland provoke mixing of the whole water column, so that this area has been the object of much attention from oceanographers.[2] Many factors contribute to increase the complexity of this ecosystem: (i) the presence of the estuary of the Isonzo River, draining to the Gulf nutrients from agricultural soils, in addition to other chemicals, such as mercury from the Idrija (Slovenia) mines; (ii) several submarine freshwater springs—related to the Timavo River—along the coast;[3] (iii) the urban and industrial settlements of Trieste and Monfalcone, constantly discharging sewage through collectors.[4–6]

In the last ten years, anomalous algal and mucilage blooms, with effects on the production of mussel hatcheries and on tourism, have been reported along the whole northern Adriatic coast, giving further impetus to studies on the interactions between physicochemical parameters and biological activities in these waters.

Here, we aim to detect the spatial and temporal variations of the physicochemical pattern of these waters, in order to be able to recognize in the future if critical values of some of the measured parameters can have a 'trigger effect' on further anomalous algal blooms. In order to detect the pattern of the data array, we make use of well-known tools of exploratory data analysis, such as principal component analysis,[7] biplot[8,9] and boxplot graphs.[10]

Experimental

The physicochemical measures are: temperature (TEMP), salinity (SAL), dissolved oxygen (O_2), nitrogen as ammonia (NH_3), nitrite (NO_2) and nitrate (NO_3), silicate [$Si(OH)_4$], chlorophyll *a* (Chlo.a) and phaeopigments (Phaeop.). These were sampled at 16 stations (see Fig. 1) at different levels of the water column during 24 months. As a rule, three days were necessary to complete sampling operations.

Sampling design

Samples were collected with a monthly frequency from April 1995 to March 1997 by the oceanographic ship *Effevigi* of the

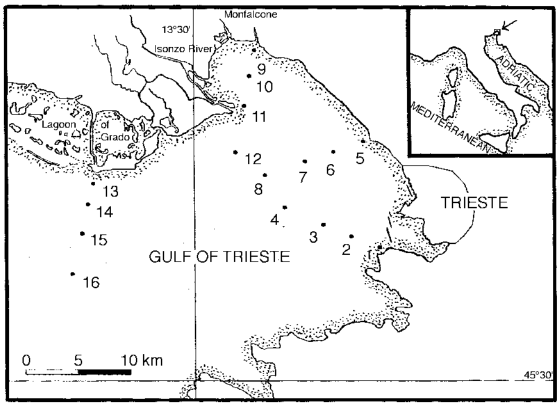

Fig. 1 Map of the sampling area.

Laboratory of Marine Biology of Trieste. The position of the 16 stations was defined by means of the Global Positioning System. Aliquots of water for chemical analysis were sampled by means of Niskin bottles throughout the water column: at the surface, at 5 and 10 m and on the sea bottom (wherever these depths have physical meaning).

Experimental data acquisition

Temperature and salinity were determined *in situ* with an IDRONAUT 401 probe, and dissolved oxygen by a modified Winkler method,[11] on the day of sample collection. Ammonia, nitrite, nitrate and silicate were determined by segmented continuous flow analysis,[12] with an Alliance Integral Autoanalyser. Two litres of seawater for chlorophyll *a* and phaeopigment analysis were filtered on GF/F Whatman (0.8 μm) filters, and extracted with acetone 90%; the acetone phase was then measured by spectrofluorimetry,[13] with a Perkin-Elmer Model LS50B spectrofluorimeter, and the concentrations were computed as in ref. 14. Calibration and quality control of the methods were periodically checked by interlaboratory calibration.[1]

Statistical methods

The data were organized in a matrix with chemical parameters as columns and samples as rows. Elimination of rows containing missing values led to the identification of 909 valid samples. The columns were autoscaled to obtain values that could be meaningfully processed together. Principal component analysis (PCA) using the Kernel algorithm[15] was performed on the matrix obtained. A biplot[8] was used to display relationships between scores and loadings. Matlab 5.0 software[16] was used for the calculations, while the graphs were obtained by SPSS 6.1.3 facilities.[10]

Results and discussion

Basic statistics for the complete data set are reported in Table 1.

The examination of the correlation matrix (see Table 2) extracted from the (909 × 9) data matrix reveals that linear relationships exist between some parameters; it is therefore preferable to obtain latent factors and plot these.

Models with different numbers of principal components (PCs) were evaluated, and different rotations of components

Table 1 Basic statistics for the 909 valid cases

	TEMP/ °C	SAL/ practical salinity units	O_2/ μmol dm^{-3}	N-NH_3/ μmol dm^{-3}	N-NO_2/ μmol dm^{-3}	N-NO_3/ μmol dm^{-3}	$Si(OH)_4$/ μmol dm^{-3}	Chlo.a/ μg dm^{-3}	Phaeop./ μg dm^{-3}
Mean	14.74	36.06	243.24	0.90	0.28	4.67	3.35	0.47	0.47
s	5.25	2.74	34.11	1.10	0.31	9.30	4.41	0.42	0.64
Min	5.85	17.58	18.30	0.05	0.02	0.02	0.02	0.01	0.01
Max	26.99	37.90	305.80	13.73	2.60	125.38	44.65	2.90	9.72

Table 2 Correlation matrix for 909 cases

	TEMP	SAL	O_2	N-NH_3	N-NO_2	N-NO_3	$Si(OH)_4$	Chlo.a	Phaeop.
TEMP	1.0000								
SAL	−0.2541	1.0000							
O_2	−0.5302	−0.1108	1.0000						
N-NH_3	0.1851	−0.2283	−0.3033	1.0000					
N-NO_2	−0.3291	−0.1139	0.1382	0.1261	1.0000				
N-NO_3	0.1295	−0.7243	0.0999	0.2437	0.1002	1.0000			
$Si(OH)_4$	0.0612	−0.4569	−0.1968	0.3029	0.1040	0.4550	1.0000		
Chlo.a	0.0024	−0.1350	0.0102	−0.0615	−0.0653	0.0400	−0.0386	1.0000	
Phaeop.	−0.1547	−0.0765	0.0800	−0.0487	0.1134	0.0138	0.0024	0.6798	1.0000

Table 3 Principal component loadings

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
TEMP	0.2593	−0.4223	0.3845	−0.2120	0.3702	−0.2890	−0.3373	−0.1450	0.4541
SAL	−0.5342	−0.1878	0.0626	0.2967	−0.2104	0.1045	0.3079	−0.0827	0.6562
O_2	−0.1571	0.4480	−0.3877	−0.3838	0.1414	0.3172	−0.4364	−0.0388	0.4055
N-NH_3	0.3579	−0.1608	0.0039	0.5087	0.2649	0.7099	−0.1079	0.0090	0.0387
N-NO_2	0.0713	0.2864	−0.3590	0.5701	0.3700	−0.5326	−0.0466	0.1096	0.1555
N-NO_3	0.5193	0.1678	−0.1573	−0.2755	0.0923	0.0324	0.7013	−0.1267	0.2899
$Si(OH)_4$	0.4670	0.0160	−0.1037	0.1529	−0.7619	−0.1163	−0.3004	0.0485	0.2468
Chlo.a	0.0393	0.4298	0.5728	0.0045	0.0235	0.0525	0.0485	0.6741	0.1598
Phaeop.	0.0101	0.5144	0.4531	0.1915	−0.0454	−0.0077	−0.0523	−0.6970	−0.0509
% Experimental variance	26.3722	20.8379	17.2645	12.2250	7.4553	6.8704	3.5602	3.2305	2.1841

were tried. We discuss the model with four unrotated components, which explains 76.6% of the considered variance. Such a model was chosen since it explains a relatively high percentage of the total variance of the data with few components, and its loadings are interpretable on the basis of our knowledge of environmental phenomena. Table 3 displays the loadings of each of the four PCs and the explained variances for each variable.

The first PC [26.37% experimental variance (e.v.)] has the highest loadings for salinity (negative sign), ammonia, nitrate and silicate. The presence of nitrogen simultaneously in the lowest and highest oxidation states does not make the interpretation of this component straightforward. Boxplots for the PC1 scores of each layer, in the 16 sampling stations (see Fig. 2), show that surface samples in sites labelled from 9 to 12 have clearly higher scores than all the others. These sites are situated in proximity to the Isonzo River mouth, providing evidence of a strong contribution to this component from freshwater input in the Gulf, bringing nutrients and silicate. Moreover, there is an almost systematic variation according to depth: scores decrease from the surface to the two intermediate layers, and increase again in the deep sea bottom waters (>20 m).

Outliers with high values are detected in the surface layers in June 1995 and 1996, July 1995, October and November 1995 and 1996, and December 1996. October 1995 and 1996, and November 1996 are outliers also for the sea bottom samples. To help in the interpretation of these observations, the total dispersion of temperature values for each month is plotted in Fig. 3. This graph shows the presence of temperature gradients, and therefore gives information about density gradients and water column mixing.[17] We note that both June–July and October–November boxplots display extreme situations, the former representing the maximal spread of values in the Gulf (due to the presence of a thermocline in the water column characterized by a steep temperature gradient, separating in this case the warmer layer above from the colder layer below) and the latter representing minimal spread (conditions most homogeneous and favourable for mixing of the water column).

The presence of a thermocline in June–July, which minimizes water mixing, means that surface waters (with a composition similar to that of the river input, *i.e.* low salinity, high nitrate and silicate) move over the deeper layers, so generating the summer outliers reported in Fig. 2. The increase in the PC1 scores at the sea bottom may be interpreted by recalling the reducing environment caused by the decomposition of organic debris, consuming oxygen and delivering nitrogen in reduced form. Relatively high ammonia values occur in October, when the warm water is less oxygenated and the organic matter from phytoplankton is abundant. In particular, the deep sea bottom waters (>20 m) are, to a far smaller extent, involved in the effects of water column mixing, so that the contrast between the lower layer and the remaining water body is high. The surface waters of October 1995 present, anomalously, the highest values of ammonia of the whole campaign. We attempt to interpret the October 1995 outliers, as well as those of December 1996, later in the text. On the whole, we can see PC1 as describing the effects of the gradients of temperature and density (thermohaline discontinuity) on the chemical composition of these coastal waters influenced by the Isonzo estuary.

The examination of the loadings of the second (20.84% e.v.) and third (17.24% e.v.) PCs reveals that both have high absolute values of the same variables, namely temperature, dissolved oxygen, nitrite, chlorophyll *a* and phaeopigments, which are degradation products of chlorophylls. Chlorophyll *a* and phaeopigments have the same sign in both components. In PC2, O_2 and NO_2 have the same sign as the pigments, while temperature has the opposite sign; all these signs are inverted in PC3. The scores of samples and loadings of variables are drawn together in the biplot reported in Fig. 4. Samples are plotted so that, in the graph, the distances between them are approximated by a least square fit of rank two, while the distances between the variables are the square root of the so-called Mahalanobis distance. The small angle between the vectors identifying objects (samples and/or variables) in the biplot indicates the strong association between them.

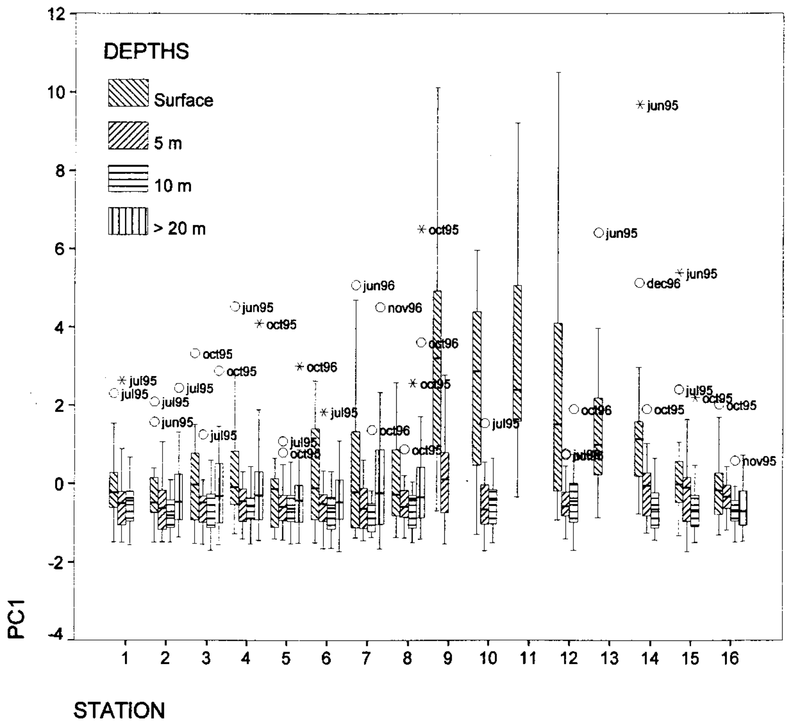

Fig. 2 PC1 scores plotted as boxplots of sampling depth clustered according to the different stations (○, outliers; *, extreme values; labelled by months).

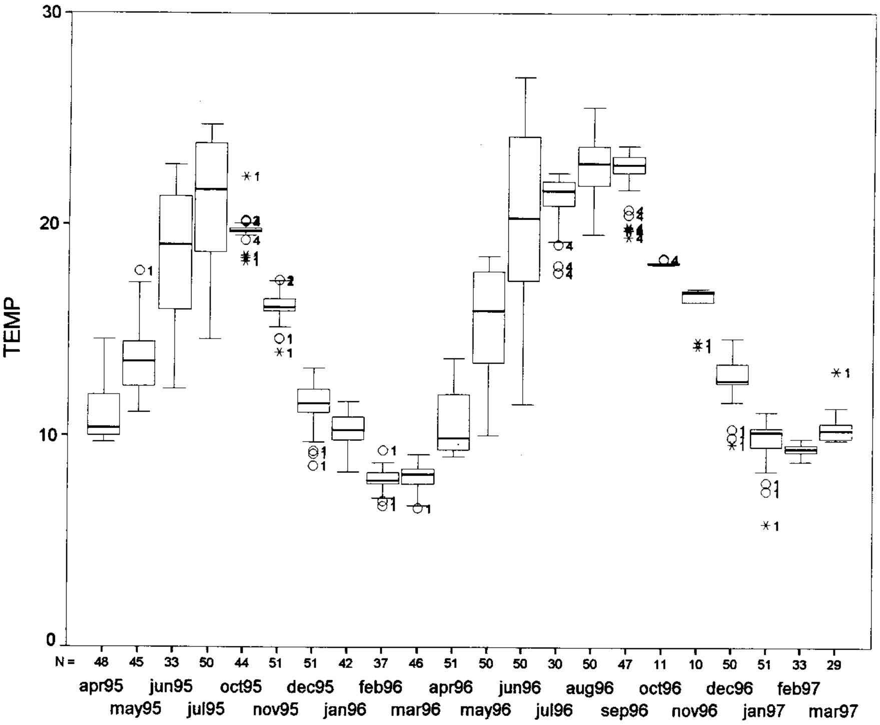

Fig. 3 Boxplots of water temperature (TEMP) for each considered month (○, outliers; *, extreme values; labelled by stations); *N* is the number of available samples.

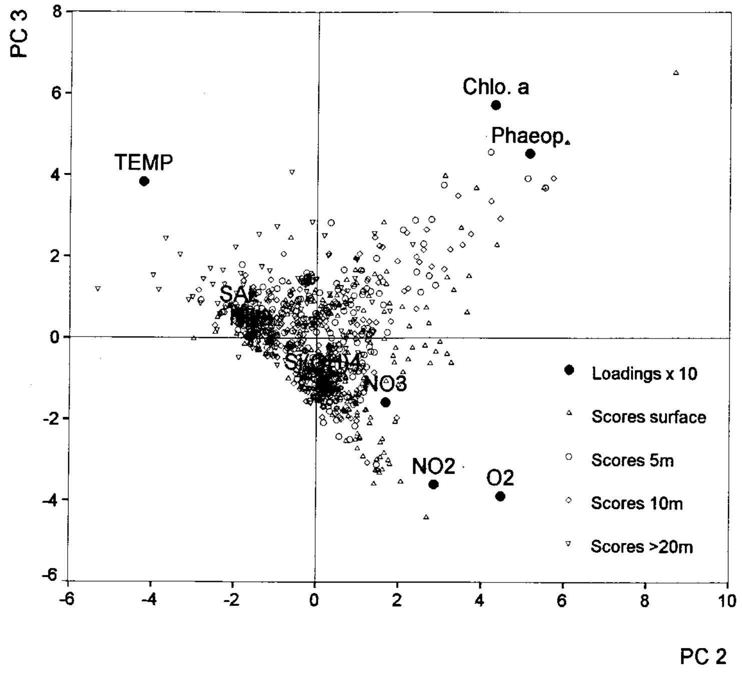

Fig. 4 Biplot of PC2 and PC3.

Chlorophyll *a* and phaeopigments are situated close to the bisectrix of the first quadrant of the cartesian plane defined by PC2 and PC3; temperature (second quadrant) and O_2 (fourth quadrant) are situated on an axis orthogonal to this bisectrix. NO_2 loadings are only slightly different from those of O_2, showing how O_2 measurements reflect redox tendencies within the water environment. Phaeopigment loadings are similar to those of chlorophyll *a*.

The scores for each of the four considered depths are plotted in Fig. 4 with different symbols, in order to recognize the eventual depth effects on the scores. A group of samples collected close to the deep sea bottom (>20 m) is found in the direction of temperature. Two clusters are found on the line opposing temperature and O_2. Towards the temperature loading, we note the presence of a cluster consisting of summer–autumn samples (from July to November 1995 and 1996), while towards O_2 we detect samples from winter and early spring (December 1995, January–April 1996, February–March 1997). Positive signs of both components are scored by samples from April–May 1995, but the highest are those of December 1996 (the warmest December in the last 30 years).[17] To elucidate this temporal evolution of the components, we plot in Fig. 5 the means of the scores for each month, linking them by spline interpolation.[10] We see that the system oscillates

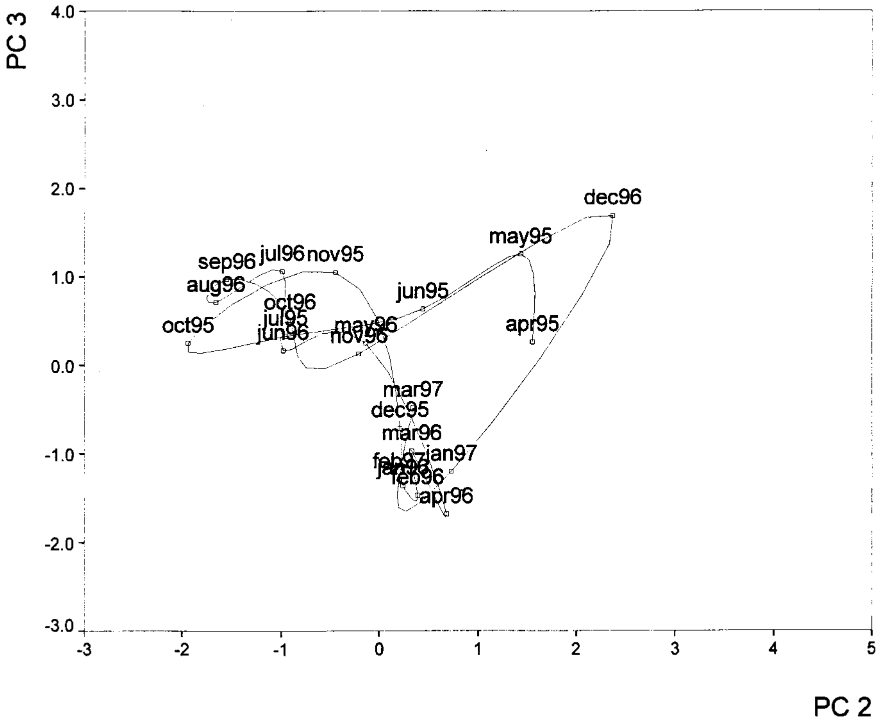

Fig. 5 Spline interpolation linking the mean values of scores for each sampling month in the PC2–PC3 plane.

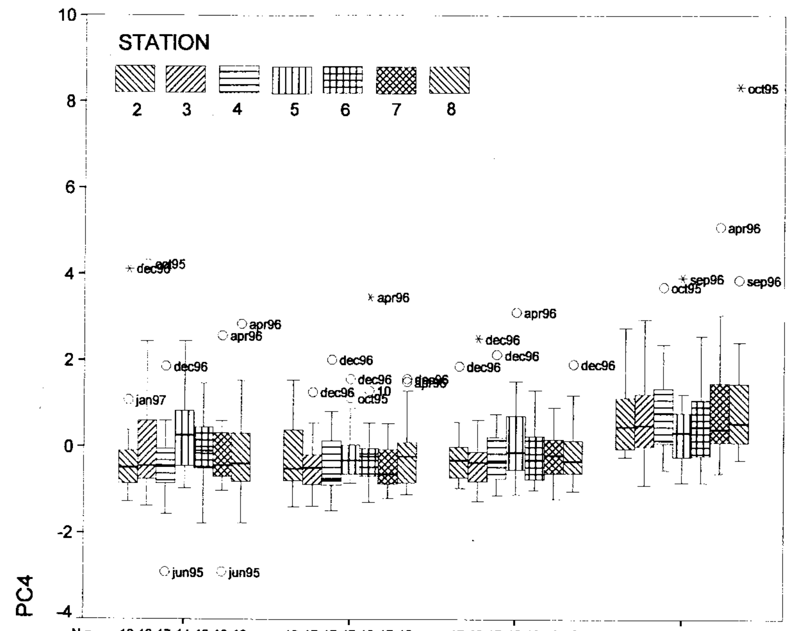

Fig. 6 PC4 scores plotted as boxplots of 7 stations clustered according to the different sampling depths (1, surface; 2, 5 m; 3, 10 m; 4, >20 m); *N* indicates the number of cases for each boxplot (○, outliers; *, extreme values; labelled by months).

between the second and fourth quadrants, with exceptional incursions (December 1996) into the former.

The second and third components can be interpreted mainly as describing the yearly evolution of chlorophyll production in the Gulf, and indirectly the primary production of organic carbon by phytoplankton, *i.e.*, the basic level of the trophic chain in a marine ecosystem. In particular, PC2 describes the early spring phytoplanktonic bloom, reported in the literature as the most important during the whole year (in our case PC2 has 20.84% e.v., larger than PC3).[18] This bloom (positive loading for chlorophyll *a*) occurs in cold and oxygenated waters. The loading of phaeopigments is even higher than that of chlorophyll *a*, witnessing the huge production and degradation of photosynthetic molecules. PC3 describes the summer–autumn bloom, occurring in warmer but less oxygenated waters.[18] Mild winter or early-spring conditions can result in exceptional chlorophyll production. Probably a critical combination of dissolved oxygen, temperature and light penetration into the sea triggers this effect: further studies are required in order to clarify this issue.

The loadings of the fourth PC (12.23% e.v.) with highest absolute values are dissolved oxygen (negative sign) and the two reduced forms of nitrogen, ammonia and nitrite. The examination of the PC4 scores in different sites shows that the four sites most influenced by the river (labelled from 9 to 12) present the lowest values. We can reasonably conclude that this is due to the input of Isonzo estuarine waters, which are more oxygenated than marine ones. In order to achieve a deeper insight into the meaning of this PC, according to the distribution of its scores in space and time, we focused our attention on seven sampling stations (labelled from 2 to 8) of the central part of the Gulf, all having four sampling depths. Scores clustered by depth, reported as different boxplots for each of the sites, are displayed in Fig. 6; in this figure, we see that the values are the highest at the sea bottom, where, as we have already mentioned, dead marine organisms decay and decompose, with the consumption of oxygen and the formation of reduced nitrogen. The examination of distribution in time of the upper outliers reveals how the observations for PC4 are linked with those of the previous two PCs. A total of 83.3% of these outliers correspond to the vertices of the time triangle of chlorophyll production shown in Fig. 5, *i.e.*, December 1996, April 1996 and October 1995. Therefore, we interpret PC4 as being related to the decomposition of organic matter, which is not limited to the sea bottom, but can also be relevant in the water column, especially during algal blooms. As we stated concerning PC1, there is evidence that, during October 1995, an important mixing of the water column was in progress, possibly bringing up to the surface most of the ammonia generated from organic matter decomposition in deeper layers. Unfortunately, we lack information about chlorophyll production in the two previous months. The lower outliers were registered at the surface layer at sites 4 and 7 in June 1995. Clearly, a plume of low density and oxygenated estuarine waters reached these sites during the rainy spring.

Conclusions

The proposed model, based on principal component analysis (PCA), gives a reasonable explanation of many important features of the considered data set, showing how physicochemical parameters and macroscale physical phenomena can be correlated in a complex dynamic system such as that considered.

The gradients of temperature and salinity—and of correlated chemical parameters—in the water body were explained by the first principal component, and they are due essentially to both the input of freshwater from the Isonzo River and to summer stratification. The second and third PCs are related to chlorophyll production, to spring and autumn phytoplankton blooms, respectively, as the biplot effectively displays. The fourth PC describes the reducing nature of the seawater, highlighting situations in which this factor is more relevant than average, as at the sea bottom, or in the water column where there is a high production and decomposition of organic matter.

This model can be considered as the starting frame on which more sophisticated models will be developed, using tools such as multi-way PCA,[19] repeated measures and mixed-design ANOVA,[20] aimed at providing the local authorities with instruments adequate for performing an effective statistical control of coastal environmental quality.

Acknowledgements

The authors thank the Italian National Research Council (C.N.R.) and the Ministry for University and Scientific Research for permitting publication of the results and Dr. Massimo Celio for *in situ* measurements.

References

1 C.N.R., M.U.R.S.T., PRISMA: Programma di Ricerca per la Salvaguardia del Mare Adriatico; http:// adria.irpem.an.cnr.it/.
2 A. Artegiani, R. Azzolini and E. Salusti, *Oceanologica Acta*, 1989, **12**, 151.
3 G. Adami, A. Favretto and E. Reisenhofer, *Fresenius' J. Anal. Chem.*, 1996, **354**, 729.
4 L. Favretto, B. Campisi, E. Reisenhofer and G. Adami, *Anal. Chim. Acta*, 1997, **344**, 251.
5 G. Adami, F. Aleffi, P. Barbieri, A. Favretto, S. Predonzani and E. Reisenhofer, *Water, Air and Soil Pollution,* 1997, **99**, 615.
6 P. Barbieri, G. Adami, A. Favretto and E. Reisenhofer, *Fresenius' J. Anal. Chem.*, 1998, **361**, 349.
7 K. V. Mardia, J. T. Kent and J. M. Bibby, *Multivariate Analysis*, Academic Press, London, 1994.
8 K. R. Gabriel, *Biometrika*, 1971, **58**, 453.
9 J. C. Gower and D. J. Hand, *Biplots*, Chapman & Hall, London, 1996.
10 M. J. Norusis, *SPSS for Windows* 6.1.3 *User's Guide*, SPSS Inc., Chicago, 1996.
11 J. H. Carpenter, *Limnol. Oceanogr.*, 1965, **10**, 141.
12 K. Grasshoff, M. Ehrhardt and K. Krenling, *Methods of Seawater Analysis*, Verlag Chemie, Weinheim, 1983.
13 T. R. Parsons, Y. Maita and C. M. Lalli, *Manual of Chemical and Biological Seawater Analysis*, Pergamon Press, Oxford, 1985.
14 G. Magazzu', *Metodi per lo Studio del Plancton e della Produzione Primaria*, G.M., Roma, 1978.
15 W. Wu, D. L. Massart and S. de Jong, *Chemom. Intell. Lab. Syst.*, 1997, **36**, 165.
16 The MathWorks, *Matlab 5.0*, Natick, MA, 1994.
17 F. Stravisi, Chair of Meteorology, University of Trieste, personal communication, 1997.
18 J. P. Riley and G. Skirrow, *Chemical Oceanography*, Academic Press, London, 1965, pp. 576–577.
19 R. Henrion, *Chemom. Intell. Lab. Syst.*, 1994, **25**, 1.
20 P. J. Diggle, K. Y. Liang and S. L. Zeger, *Analysis of Longitudinal Data*, Oxford University Press, Oxford, 1994.

Paper 8/07528J