ELSEVIER

Chemometrics and Intelligent Laboratory Systems 47 (1999) 289–297

Chemometrics and intelligent laboratory systems

Oceanographic characterization of northern São Paulo Coast: a chemometric study

Márcia M.C. Ferreira [a,*], Cristiano Gomes Faria [a], Eduardo T. Paes [b]

[a] Instituto de Química, Universidade Estadual de Campinas-UNICAMP, Campinas SP, 13081-970, Brazil
[b] Departamento de Biologia Marinha, Universidade Federal Fluminense-UFF, Niterói, RJ, 24220-970, Brazil

Abstract

A study of the northern São Paulo Brazilian coast was carried out by four survey cruises during the summers of 1985/1986 and winters of 1986/1987. In this work, the seasonal variations of physical parameters and chemical composition of marine waters collected during these four cruises are investigated through the multivariate approaches of principal component (PCA) and hierarchical cluster analysis (HCA), besides the application of the traditional thermohaline analysis. The 'island effect', coastal and shelf break upwelling phenomena and its influence in the phytoplankton community are clearly visualized through the spatial interpolation of PCA scores and of chlorophyll concentration for the column water profiles. The intense penetration of tropical water mass (TW) over the continental shelf towards the coast observed during the winter 1987 cruise, can also be visualized through the scores contour plots. Complementing the usual thermohaline analysis, chemometrics has shown that other elements such as NO_3^-, PO_4^{3-} and O_2 should be taken into account especially when biological processes are important.

Keywords: Principal component analysis (PCA); Hierarchical cluster analysis (HCA); Chemometrics; Oceanography; Water masses; Upwelling

1. Introduction

The northern coast of São Paulo State is a rich environment, which has attracted attention from geologists and oceanographers. In this paper, a data analysis of the water samples collected during four survey cruises in this region is carried out. As a whole, the phenomena happening in this region are very interesting and peculiar, especially the intense upwelling observed during the summertime. In order to have a better understanding of these data sets, a brief introduction about the nature and general circulation of water masses at the Southern Atlantic Ocean will be presented.

The surface circulation of the oceans is intimately tied to the prevailing wind circulation of the atmosphere. The horizontal movement of surface waters in the southern Atlantic is counterclockwise and opposite to its movement in the Northern Hemisphere [1]. The composition of those waters and also their density, which is basically a function of temperature and salinity, are subjected to different environmental conditions such as temperature, sunlight radiation, pluvial influxes and melting of polar ices among other factors. Such differences in density promote the strat-

* Corresponding author. E-mail: marcia@iqm.unicamp.br

0169-7439/99/$ - see front matter
PII: S0169-7439(98)00216-0

ification of water layers giving origin to a coexistence of waters with different characteristics. This process make the whole dynamic system very complex [2].

On the oriental border of Atlantic Ocean, the Benguela current flows northward roughly parallel to the African coast and deviates west near Angola, in the direction of the South American continent. This current carries low temperature waters with high content of nutrient salts (nitrates, nitrites, phosphates and silicates). As this water mass flows to the west, the temperature and salinity increases due to intense evaporation and sunlight radiation which is always present east from Bahia's coast [3]. The current then splits into two streams, most of it flowing to the northern hemisphere, while the remaining deviates southward originating the Brazilian current, a warm and saline mass of tropical water (TW). This current is characterized as being poor of nutrients and having temperatures above 20°C and salinity above 36‰ [4]. As it flows southward, this TW gradually cools down becoming denser and finally mixing at the Plata River latitude with the cold and eutrophic waters originated in subantarctic region called Malvinas current. The meeting of these two masses of waters in the region of subtropical convergence originates new mass named South Atlantic Central water (SACW). Due to its high density, part of this SACW sinks and returns as a counter current northward in deeper ocean layers under the Brazilian current.

Coastal waters (CW) of low salinity are present in regions close to the coast where currents are induced by winds and tide. Their extension and volume varies with respect to the pluviometric conditions and are influenced by the proximity of river estuaries and the intensity to which they interconnect with higher salinity waters coming from the Continental shelf [5].

Vertically, the studied region is divided into the euphotic zone, at depths lower than 60 m (where the sunlight reaches), and the aphotic zone (without sunlight). The euphotic region is populated by phytoplanktons, which are microscopic single-celled marine algae, and the zooplanktons, the floating animals that rely on the phytoplanktons as food source.

Upwelling is a very interesting phenomenon by which the cold dense bottom waters from aphotic zone rich in nutrients rise towards the euphotic zone. Their occurrences are important because they affect directly the ocean fertility causing a new chlorophyll production due to the phytoplankton boom. Such phenomenon is driven by classical ocean–atmosphere interaction (Ekman transport)—coastal upwelling [6,7] or by meanders and eddies of Brazilian current—shelf break upwelling [8].

In this work, the multivariate methods of principal component analysis (PCA) and hierarchical cluster analysis (HCA) are used to visualize and analyze both upwellings mentioned above besides other interesting phenomena which may occur in this region.

2. Experimental

Water samples were collected in Nansen bottles during four 1-week oceanographic cruises during the summers of 1985/1986 and winters of 1986/1987, at several depths in 60 different marine stations, distributed in six parallel profiles perpendicular to the coastline of the northern São Paulo State's coast (Fig. 1), in the region of Ubatuba between 23 and 24°S latitude [9]. The chemical analysis included temperature, salinity, dissolved oxygen, ammonia, nitrate, nitrite, silicate, phosphate and chlorophyll concentrations. The laboratory analysis was performed using classical methods [10,11]. The Oceanographic Institute from University of São Paulo kindly provided the data set used in this work.

The chemometric methods [12,13] used for data analysis are described below and was carried out using Pirouette Software [14]. Scores and chlorophyll contour maps representations were accomplished through the use of a krigging method [15] using the software Surfer [16].

3. Methods

One of the most important multivariate methods of data analysis is PCA [12,13], based on the correlation between variables. It aims to group these correlated variables, generating a new set of variables called principal components, PCs, onto which the data is projected. These PCs are built as linear combination of original variables and have the important property of being completely uncorrelated. The first new axis, PC1, is chosen in such a direction that it maximizes the variance along that axis, the second

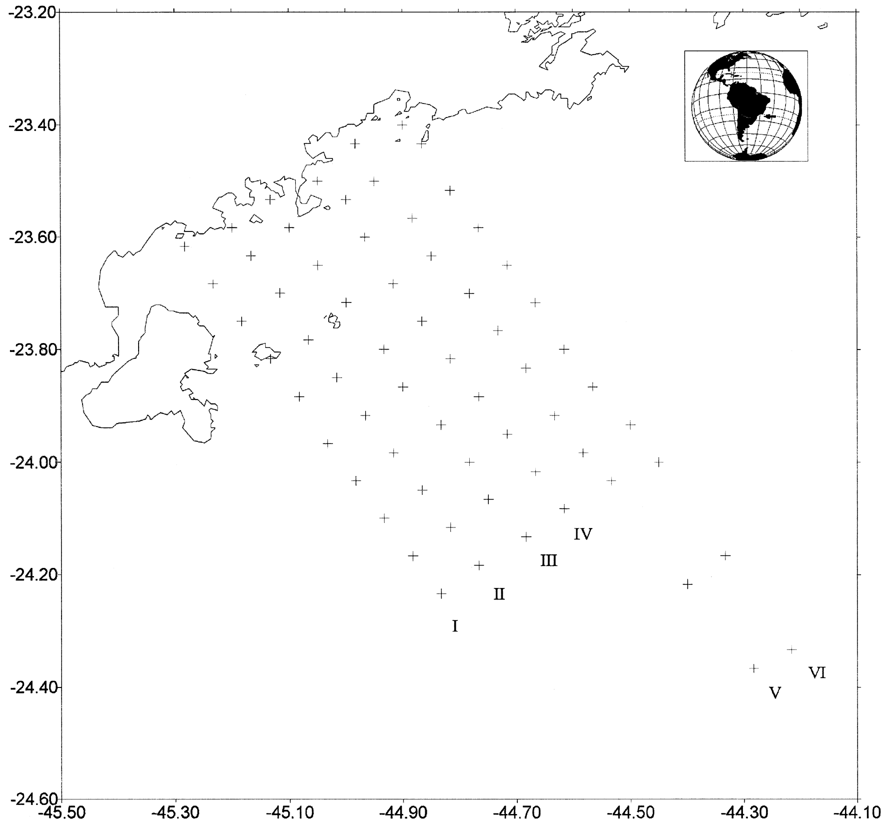

Fig. 1. Brazilian coastline with geographical location of the 60 marine stations in six parallel profiles (I–VI).

axis must be chosen orthogonal to the first one and in the direction to describe as much variance left as possible and so on.

The raw data matrix, represented by **X**, is decomposed into two matrices, **T** and **V** where

$$\mathbf{X} = \mathbf{T}\mathbf{V}^{\mathrm{T}}. \quad (1)$$

The matrix **T**, known as 'score' matrix, represents the position of the samples in the new coordinate system. The second matrix, **V**, is the 'loading' matrix and describes how the new axis, i.e., the PC's, are built from the old axes.

HCA [12] is another important multivariate method of data analysis. Its primary purpose is to display the data in such a way as to emphasize its natural clusters and patterns in a two-dimensional space. The results presented in the form of dendograms, make it possible to visualize the relationships between samples or variables. In HCA, the distances between samples or variables are calculated and compared through the similarity scale which ranges from zero, i.e., no similarity and large distance among samples, to one, for identical samples. The incremental clustering technique, especially indicated when the clusters are not well separated (two classes abut one another) is used in this work. It employs a sum of squares approach to calculate the nearest cluster, weighted by the cluster population.

4. Results and discussion

4.1. Summer I (December 1985)

During the summer season, 107 samples were collected at the stations shown in Fig. 1. The thermohaline scatter plot (temperature and salinity distribution) revealed (Fig. 2) a triangular conformation indicating the coexistence of three masses of water (m_1 = SACW, m_2 = TW and m_3 = CW) in the studied region. Each one of those masses is characterized by a vertex of a triangle defined by a pair of coordinates (Tm_i, Sm_i) which is called the respective thermohaline coefficient (see Table 1). The inferior (at low temperatures and salinity) and superior vertices (high temperatures and low salinity), correspond to thermohaline coefficients for SACW and CW, respectively, while the TW's coefficient is characterized by high temperature and high salinity. The triangle's sides in this scatter plot show the mixture of the water masses, less intense among SACW and CW during this summer.

The Shtokman method [17], well known in the oceanographic community, has been used to classify these samples into classes. This method takes the measured temperature (T) and salinity (S) of each sample as a convex linear combination of the thermohaline coefficients

$$\sum x_i = 1 \qquad i = 1,2,3$$

$$\sum x_i(Tm_i, Sm_i) = (T,S) \tag{2}$$

and solves for the xs which represent the fractions of respective water masses SACW, TW and CW. A given sample is classified as belonging to a certain class if its fraction is greater than 0.5 (> 50%), otherwise, the sample is classified as M (mixture). The samples shown in Fig. 2 have been classified according this method.

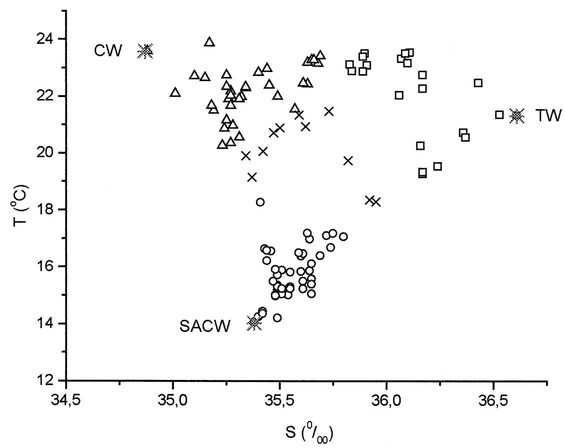

Fig. 2. Temperature–salinity scatter plot. Summer I cruise. △, CW; ○, SACW; □, TW; ×, M; ∗, thermohaline coefficient.

Table 1
Thermohaline coefficients

Water masses	CW		TW		SACW	
	T (°C)	S (‰)	T (°C)	S (‰)	T (°C)	S (‰)
Summer I	23.58	34.87	21.32	36.61	14.03	35.38
Winter I	21.79	35.09	23.37	37.07	14.46	35.43
Summer II	26.49	34.17	23.13	36.95	14.71	35.43
Winter II	21.58	33.20	22.81	36.80	9.99	35.17

Fig. 3 shows the HCA results obtained for the autoscaled data (107 samples). As expected, due to the coexistence of the masses a sharp grouping is not possible. The samples are labeled according to Shtokman criteria described above. There is a separated group (G V) consisting of samples (SACW) originated from deeper waters, i.e., depths beyond 65 m. The other SACW samples are grouped with some samples from CW in G I and are characterized by being less than 15 m of depth. There are also some SACW samples in G II, containing a mixture of CW and samples (M), which do not have a definite characteristic. That is an indication that as deeper waters rise toward the surface, they lose some of their identity, becoming more similar to the CW. The CW in G I have a high concentration of chlorophyll and the fact that they have a similar profile of the eutrophic waters, suggests that large quantities of chlorophyll were generated by the phytoplankton boom due to the water enrichment caused by the coastal upwelling phenomenon. The group G III consists of TW samples only and the G IV group consists of a mixture of TW and CW, the later, being offshore samples which acquired some TW characteristics.

The HCA of variables indicates that there is a good correlation between chlorophyll and the following nutrients: phosphate, nitrate, silicate and nitrite. Again, that is a proof of the coastal upwelling phenomenon, which provides nutrients to the surface layers of the continental shelf.

The PCA scores for the first two PCs, which explain 55.6% of the total variance, are shown in Fig. 4. The first PC, accounting for 42.2% of the total variance in the data, describes a depth gradient separating the SACW samples on the right-handed side, with positive scores. From the loadings results for

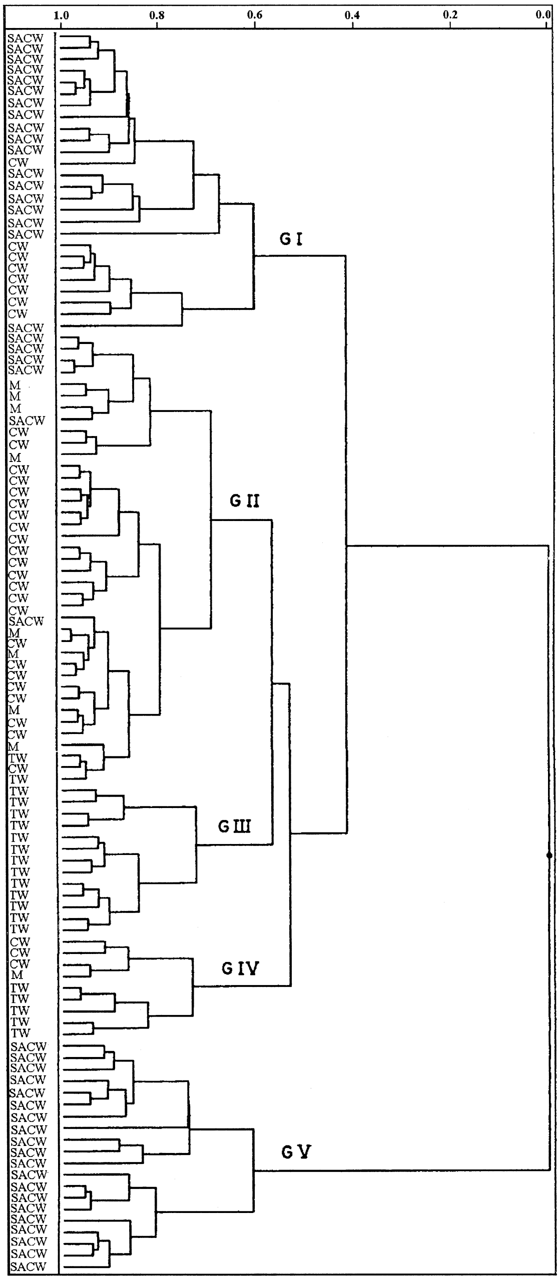

Fig. 3. HCA results on water samples. Summer I cruise.

PC1 at Table 2, it is evident that these samples have a high content of nutrients and low dissolved oxygen at a rather low temperature, as expected. On the other hand, the CW and TW samples with negative scores, are at higher temperature and have dissolved oxygen, being poor in nutrients (see Table 2). It is interesting to note that the mixture of CW with SACW at score values around zero in the first PC corresponds to those samples in G I and upper side of G II groups in the dendogram shown in Fig. 3. These SACW samples shows a higher temperature and dissolved oxygen and chlorophyll concentration as indicated by the loadings. That is an evidence of cold deep-water movement towards the surface and as it mixes with the CW, acquires some of its characteristics (high temperature and dissolved oxygen concentration). In the second PC (13.36% of total variance), TW have positive scores, while CW have negative values. That separation is mostly due to the differences in salinity and chlorophyll concentration (see Table 2). The samples classified as mixture (M) in Fig. 3 have scores near zero on PC2 and are mostly related with the mixing of CW and TW. Farther from the coast, the TW and CW become more similar.

In order to have a better idea of the spatial distribution of SACW, the PCA scores can be regionalized using a krigging method [15], where the values of spatially distributed variables are estimated by interpolation. The plot shown in Fig. 5 is especially useful to elucidate the water dynamics. The scores for PC1 are plotted as a function of the distance from the coast vs. depth, for profile VI (see Fig. 1). It is remarkable to see the presence of SACW in the euphotic zone, at depths lower than 15 m. The superposition of chlorophyll concentration to the scores plot generate contour plots with very similar profiles, evi-

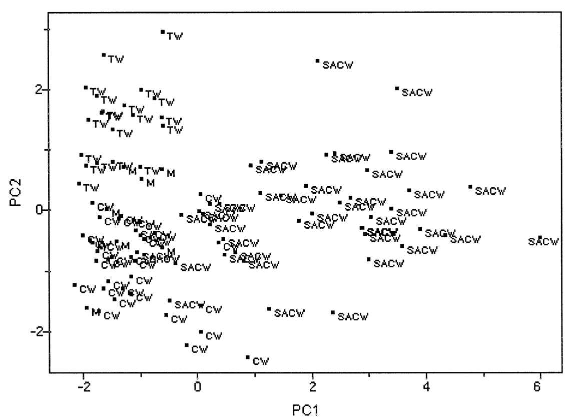

Fig. 4. PCA scores plot for first two PCs (PC1 and PC2). Summer I cruise.

Table 2
Loadings and total percent (%) variance for PC1 and PC2; Summer I cruise

	Variance (%)	Clf	NO_3^-	NO_2^-	NH_3	PO_4^{3-}	SiO_2	O_2	T	S
PC1	40.21	0.28	0.42	0.37	0.01	0.46	0.22	−0.39	−0.44	−0.10
PC2	13.36	0.43	0.04	0.11	0.39	0.02	0.24	−0.21	0.02	0.74

dencing its influence in the production of new chlorophyll. The same coastal upwelling phenomenon has been observed in this summer season for all profiles shown in Fig. 1.

4.2. Winter I (July 1986)

During the second cruise, 165 water samples were collected and analyzed as described before. Table 1 has the thermohaline coefficients for this season. Note that only the temperature of CW decreases slightly with respect to the previous summer season. That is explained by the extremely high thermal capacity of the oceans when compared to the atmosphere, which causes its temperatures to fluctuate seasonally, much less than the atmospheric temperature. In contrast to the previous cruise, no samples between the vertices CW and SACW have been observed in the thermohaline scatter plot, indicating that in this situation the

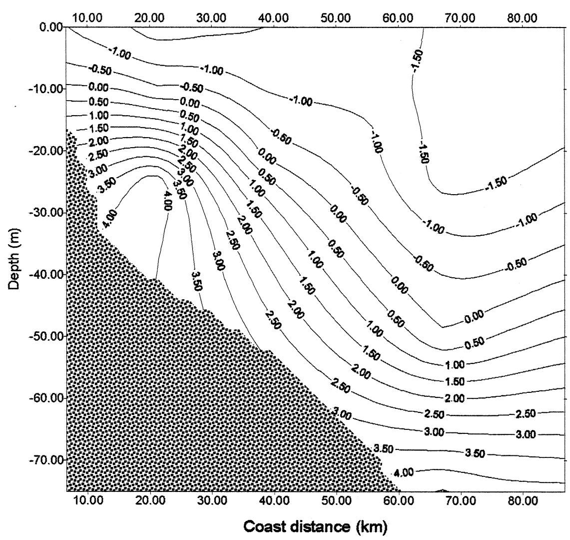

Fig. 5. PCA scores contour plot for PC1 as a function of the distance from the coast vs. depth for profile VI, evidencing the coastal upwelling. Summer I cruise.

CW and SACW do not mix with each other (similar to Fig. 8, for Winter II). Note also, that in this case there are no deep eutrophic water samples near the surface having only two SACW samples about 30 m deep. The HCA for this situation has a much more definite grouping as expected. There are three main clusters. The SACW are grouped in a single cluster, with a few samples classified as mixture (M). The other cluster, containing the CW, has a definite subgroup of CW samples, which are close to the coast. The last cluster consists basically of three groups. One, of CW farther from the coast, the other with TW samples and a third, with CW, M and TW samples describing the mixing of these two water masses. As the CW becomes farther from the coast, it loses gradually its peculiar characteristics, acquiring TW characteristics.

The PCA reflects the tendencies observed in HCA. The first PC, describing 37.9% of total variance in the data, is again responsible for the separation of SACW, which in this case have negative scores.

The PCA scores isolines for PC1 from profile II (Fig. 6), indicate the presence of a cold water dome, rich in nutrients. However, the nucleus of the dome is not so close to the coast when compared to the previous case. The upwelling in this case has a distinct nature, being independent of the wind's regime. In this case, it is observed that the shelf break upwelling driven by the meandering of Brazilian current. The superposition of chlorophyll concentration contour plots and PCA scores isolines in Fig. 6 show how the eutrophication of euphotic zone alters the phytoplankton production, causing an increase of the chlorophyll concentration in the dome's nuclei and its neighborhood.

In the case of profile I, the superimposed PCA scores (PC1) and chlorophyll concentration contour plots (Fig. 7), show that the phenomenon in this case is directly related to the 'island effect' originally described by Doty and Ogury [18], where they ob-

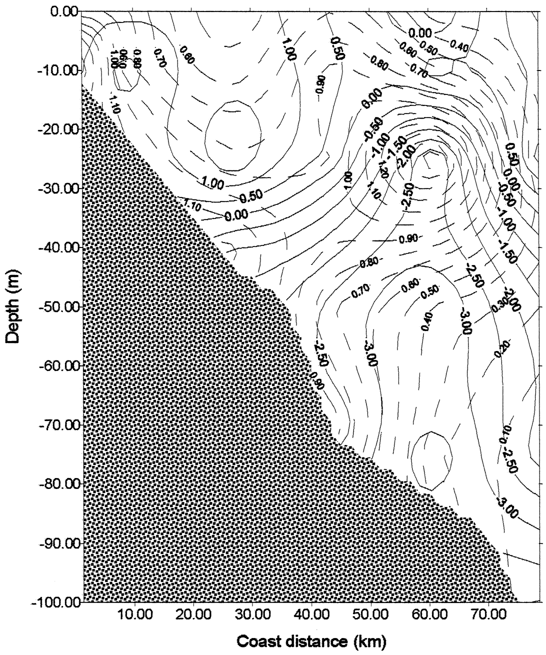

Fig. 6. Overlay contour plots of Chlorophyll (— —) concentration and PC1 scores (—) as a function of the distance from the coast and depth for profile II. Winter I cruise.

served a higher organic production in the neighborhood of islands and oceanic banks, caused by perturbation in the stratification of water column.

4.3. Summer II (December 1986)

In this second summer season cruise, 184 samples have been used for the data analysis. The chemical analysis included one extra measurement, which is the active chlorophyll concentration. The present thermohaline coefficients (Table 1) indicate some small variations with respect to the previous values. CW has a higher temperature and lower salinity. TW has higher salinity and temperature when compared with the previous year but lower than the winter season. The SACW has basically the same salinity, and just a small variation in the temperature, which is a bit higher in the present study. The temperature–salinity diagram is not much different from the previous cruise (Winter I), indicating a small interaction between CW–SACW suggesting a less intense coastal upwelling. The dendogram shows four distinct classes: one containing samples from all different origins (CW, TW SACW and M) and the others having basically one type of water mass each.

The first two PCs from PCA analysis describe 62.7% of total variance and account for the separation of the three water masses with similar patterns as that in Fig. 1. SACW samples have positive scores on PC1. CW and TW have positive and negative scores on PC2, respectively.

For this cruise a less intense coastal upwelling has been observed especially at the northern profile, and no break shelf upwelling.

4.4. Winter II (July 1987)

In this last survey cruise, some different phenomena took place, making it very interesting. The temperature–salinity scatter plot (Fig. 8) does not show any mixture of CW–SACW, indicating that no coastal upwelling occurred, but at the same time, an extensive mixture of CW and TW can be observed. The PCA confirms this fact (Fig. 9). As in all other cases presented, PC1 reveals a depth gradient, but contrary to the previous cases TW and CW samples are roughly separated by PC1.

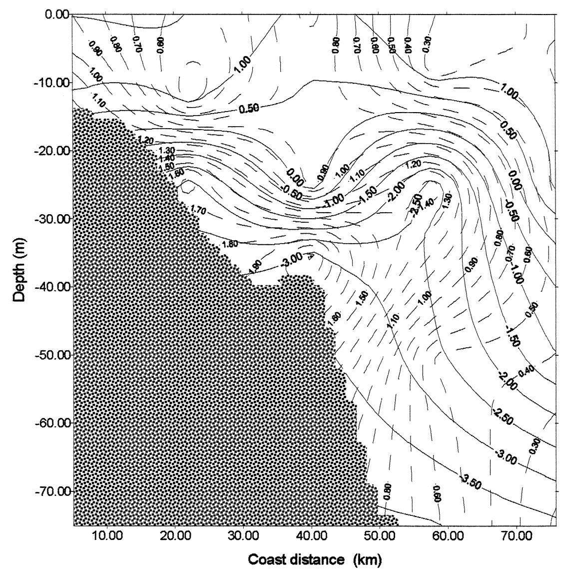

Fig. 7. Overlay contour plots of chlorophyll (— —) concentration and PC1 scores (—) as a function of the distance from the coast and depth for profile I. Winter I cruise.

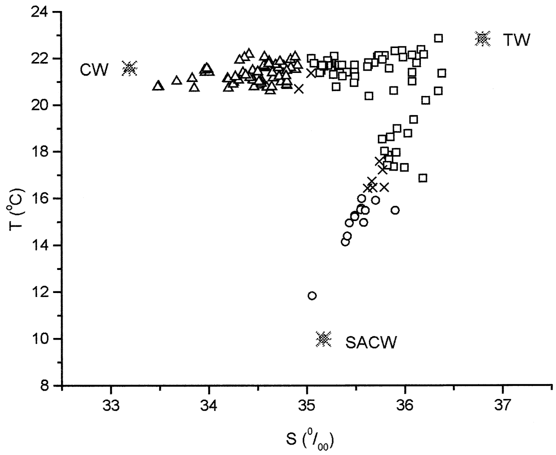

Fig. 8. Temperature–salinity scatter plot. Winter II cruise. △, CW; ○, SACW; □, TW; ×, M; ∗, thermohaline coefficient.

The HCA of the present variables shows a strong correlation among chlorophyll and nitrite concentration, contrary to the other cases where the chlorophyll was correlated with the nutrients as a whole. Since the nitrite is a result of phytoplankton excretion, the above conclusion suggests a high metabolic activity of the phytoplankton community, which is an interesting biological information.

The grids for all six profiles from Fig. 1 indicate an intense penetration of TW mass towards the coast and under the CW mass. That can be seen in the grid for the profile I (Fig. 10) where a homogeneous tongue of TW water, i.e., large region with small variation on scores values equal to 1.0 is present. The SACW with scores equal or lower than −2.0, are confined to deeper layers indicating no upwelling phenomenon in this case.

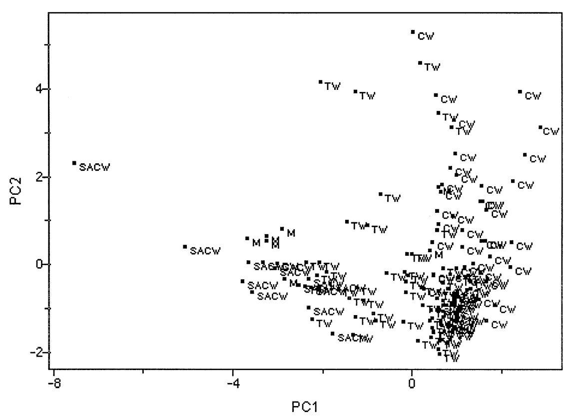

Fig. 9. PCA scores plot for the first two PCs (PC1 and PC2). Winter II cruise.

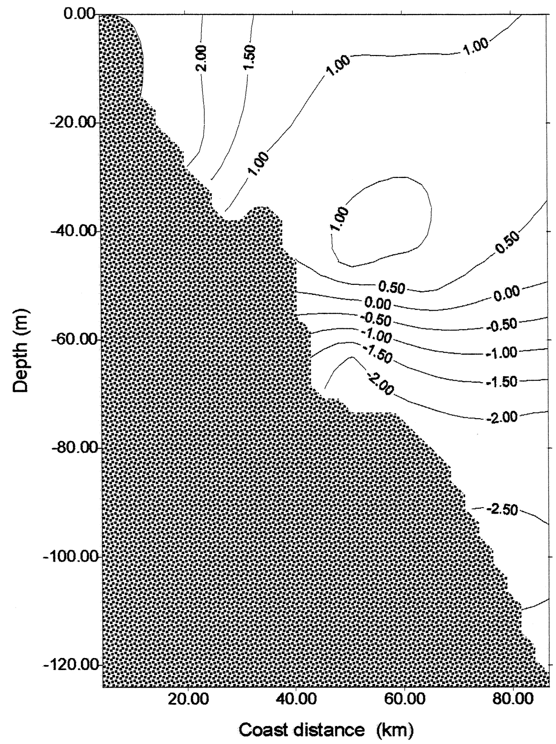

Fig. 10. PCA scores plot for PC1 as a function of the distance from the coast vs. depth for profile I. Winter II cruise.

5. Conclusions

It is shown in this paper that the multivariate approach is an efficient instrument for the characterization of ocean water masses. The spatial interpolation of PCA scores superimposed to the chlorophyll concentration are an effective procedure for the visualization of interesting phenomena such as the coastal upwelling observed during the Summer I, the ‘island effect’ and shelf break upwelling during Winter I cruises. The intense penetration of the TW mass over the continental shelf towards the coast can also be visualized through the scores isolines. The correlation among chlorophyll and nitrite that appeared in the last cruise suggests a phytoplankton community in high metabolic activity. It is also shown that other elements besides temperature and salinity such as NO_3^-, O_2 and PO_4^{3-} might be important and should be used when biological activities are present. No anthropogenic factors are considered in these studies because samples were collected farther than 10 km from the coast.

Acknowledgements

The authors thank CNPQ, FAEP and FAPESP (Brazilian Agencies) for financial support. The authors also thank Dr. Elizabeth Aydar from Oceanographic Institute, University of São Paulo for kindly providing the data for this work.

References

[1] H.U. Sverdrup, M.W. Johnson, R.H. Fleming, The Oceans, Their Physics, Chemistry and General Biology, Prentice-Hall, Englewood Cliffs, NJ, 1954.
[2] B.M. Castro Filho, L.B. Miranda, S. Miayo, Bol. Inst. Oceanogr. Univ. São Paulo 35 (1987) 135–151.
[3] I. Emílsson, Ciênc. Cult. São Paulo 11 (1959) 44–54.
[4] L.B. Miranda, Bol. Inst. Oceanogr. Univ. São Paulo 33 (1985) 105–119.
[5] I. Emílsson, Bol. Inst. Oceanogr. Univ. São. Paulo 11 (1961) 101–112.
[6] G.T. Csanady, Circulation in the Coastal Ocean, Reidel, Dordrecht, 1982.
[7] K. Yoshida, Jpn. J. Geophys. 4 (1967) 1–75.
[8] E.J.D. Campos, J.E. Gonçalves, Y. Ikeda, J. Geophys. Res. 100 (1995) 18537–18550.
[9] E. Aydar, S.A. Gaeta, S.M.F. Gianesella-Galvão, M.B.B. Kutner, C. Teixeira, Publ. Esp. Inst. Oceanogr. Univ. São. Paulo 10 (1993) 9–43.
[10] K. Grasshoff, M. Ehrhardt, K. Kremling, Methods of Sea Water Analysis, 2nd revised and extended edn., Verlag Chemie, Wienhein, 1983.
[11] J.D.H. Strickland, T.R. Parsons, A practical handbook of the sea water analysis, Bull. Fish. Res. Bd. Can. 167 (1968) 1–311.
[12] M.A. Sharaf, D.L. Illman, B.R. Kowalski, Chemometrics, Wiley, New York, 1986, pp. 153–295.
[13] E.R. Malinowski, Factor Analysis in Chemistry, Wiley, New York, 1991.
[14] Pirouette Multivariate Data Analysis for IBM PC Systems, Version 2.02, Infometrix, Seattle, WA, 1997.
[15] T.C. Bailey, A.C. Gatrell, Interactive Spatial Data Analysis, Wiley, New York, 1995.
[16] Surfer for Windows, Version 6.04.27, Golden Software, Golden, CO, 1995.
[17] W.B. Shtokman, J. Mar. Res. 6 (1946) 1–24.
[18] M.S. Doty, M. Ogury, J. Cons. Int. Explor. Mer. 22 (1956) 33–37.