ELSEVIER

Analytica Chimica Acta 398 (1999) 227–235

ANALYTICA CHIMICA ACTA

www.elsevier.com/locate/aca

Modeling bio-geochemical interactions in the surface waters of the Gulf of Trieste by three-way principal component analysis (PCA)

P. Barbieri [a,*], C.A. Andersson [b], D.L. Massart [c], S. Predonzani [d], G. Adami [a], E. Reisenhofer [a]

[a] *Department of Chemical Sciences, University of Trieste, Via Giorgieri 1, I-34127 Trieste, Italy*

[b] *Chemometrics Group, Food Technology, Department of Dairy and Food Science, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark*

[c] *ChemoAC, Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, B-1090 Brussels, Belgium*

[d] *Laboratory of Marine Biology, Via Piccard 54, I-34010 S.Croce-Trieste, Italy*

Received 20 November 1998; received in revised form 6 May 1999; accepted 10 May 1999

Abstract

Data of temperature, salinity, dissolved oxygen, nutrients and chlorophyll measured on samples of surface seawater and collected monthly during 2 years in different sites of the Gulf of Trieste are modeled by means of three-way principal component analysis (PCA). Missing values are handled using an expectation maximization algorithm, regression or substitution with random numbers, depending on their origin. Physicochemical parameters are described by three different components that explain the effect of the river input on the seawater pattern, the effect of temperature, and metabolic–catabolic activity of the phytoplankton, respectively. One spatial component accounts for the gradient of influence of the estuarine waters in the Gulf, and three temporal components characterize three main seasonal conditions. Anomalous situations, generated by meteoclimatic events, are highlighted.

Keywords: Three-way principal component analysis (PCA); Missing values; Environmental monitoring; Estuary; Micronutrients; Chlorophyll a; Gulf of Trieste

1. Introduction

Environmental monitoring programs often produce huge data sets describing spatial and temporal variations of many physical, chemical and biological parameters. Data reduction methods are required to extract useful and interpretable information which is able to explain physicochemical parameter patterns as well as spatial and temporal ones. Among these methods, three-way, and generally *N*-way principal component analysis (PCA) methods [1] are generating growing interest [2–5]. Beside methodological works aimed essentially at proving the applicability of the *N*-way approach to environmental systems, there is a lack of applications to real case studies in this field. The main purpose of this paper is to present a three-way model built from a three-dimensional data array produced during a survey, covering 2 years, on the quality of surface seawaters sampled with monthly frequency at eight stations in the Gulf of Trieste, near the Isonzo river estuary. Temperature, salinity, dissolved oxygen, micronutrients and chlorophylls were the measured variables. As it often happens in real field surveys, the data set contained several cases of missing values.

* Corresponding author. Tel.: +39-040-676-3915; fax: +39-040-676-3903

E-mail address: barbieri@dsch.univ.trieste.it (P. Barbieri)

0003-2670/99/$ – see front matter
PII: S0003-2670(99)00418-3

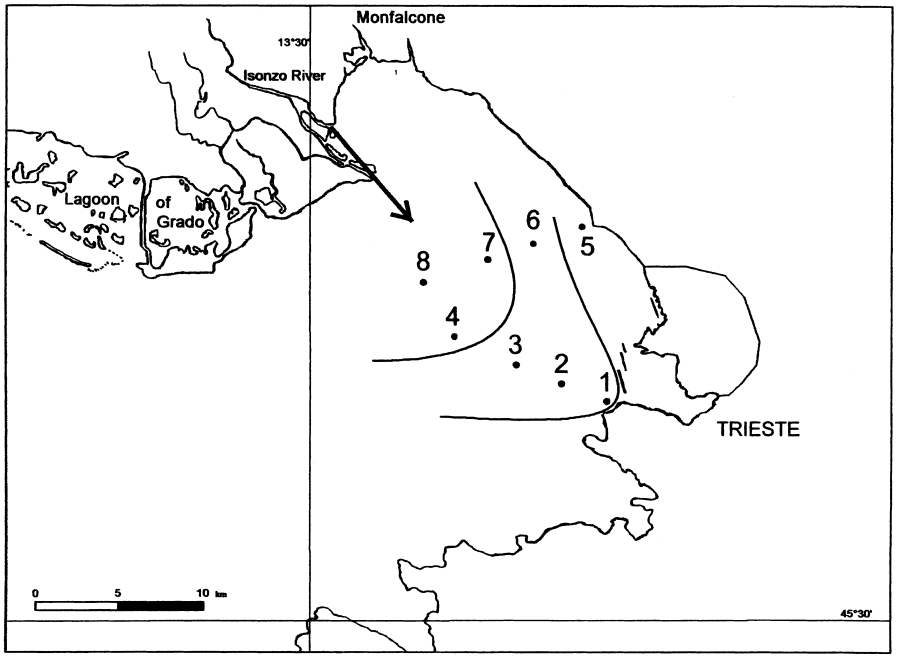

Fig. 1. Map showing the position of sampling sites.

2. Experimental

2.1. Physicochemical parameters

Surface seawater samples were collected with monthly frequency from April 1995 to March 1997 by the oceanographic ship Effevigi of the Laboratory of Marine Biology of Trieste; the positions of the eight sampling stations are indicated in Fig. 1. The 11 physicochemical parameters are: temperature (Temp), salinity, dissolved oxygen (DO), nitrogen as ammonia (NH_4), nitrite (NO_2) and nitrate (NO_3), ortho-phosphate (PO_4), total phosphorous (P_{tot}), dissolved silica (SiO_2), chlorophyll-a (Chlo.a) and phaeopigments (Phaeop). It typically lasted 3 days to complete the collection of samples at all eight sites.

Temperature and salinity were determined in situ with a multiparametric probe, an Idronaut CTD 401 profiler. Dissolved oxygen was measured by a modified Winkler method [6] in the laboratory later on the day of sampling. Ammonium, nitrite, nitrate, ortho-phosphate, total phosphorous and dissolved silica were determined by segmented continuous flow analysis (SCFA) as in [7], using an Autoanalizer Alliance Integral instrument. Chlorophyll-a and pheopigments were extracted with acetone from the filtered seawater, and measured by spectrofluorimetry [8] using a Perkin–Elmer Model LS50 fluorimeter. Their concentrations were evaluated as in [9]. Calibration and quality control of the analytical methods were checked periodically by interlaboratory calibrations within a monitoring program concerning the whole Adriatic Sea [10].

2.2. Mathematical methods: multi-way analysis

The data were collected in an array that can be visualized as a parallelepiped of dimensions: 11 (parameters) × 8 (sites) × 24 (months). Physicochemical parameters, sampling sites and sampling months constitute the three ways, or modes, of the array. Different methods, namely 'unfolding' [11,12], Tucker models [11,13–15], and PARFAC/CANDCOMP [16–18], have been applied to multi-way arrays in order to generalize the classical PCA [19].

We choose to apply the Tucker3 model. According to Kiers and Bro [11,18], the Tucker3 model, extensively treated in [14], can improve data-reduction per-

formances of the unfolding method since it models the data set with fewer degrees of freedom, keeping, at the same time, the three-way structure of information. Moreover, Smilde indicates [1] how Tucker3 models — having both the same number of components in each mode and a super-diagonal structure of the core array $\underline{\mathbf{G}}$ (see below) — can provide the analyst with hints for building simpler PARAFAC models.

In the following, we will briefly recall the main elements of the theory about the Tucker3 model, which are useful for our discussion. Such a model can be formulated as a factorization of the data array $\underline{\mathbf{X}} = \{x_{ijk}\}$ of dimensions $(n \times p \times q)$, where x_{ijk} is, in our case, the value of the measurement referring to the ith physicochemical parameter, the jth site and the kth sampling month:

$$x_{ijk} = \sum_{u=1}^{r} \sum_{v=1}^{s} \sum_{w=1}^{t} a_{iu} \cdot b_{jv} \cdot c_{kw} \cdot g_{uvw} + e_{ijk} \quad (i = 1 \ldots n,\ j = 1 \ldots p,\ k = 1 \ldots q) \qquad (1)$$

The values r, s and t are the number of components selected to describe the first, the second and the third mode, respectively, of the data array. The elements a_{iu}, b_{jv} and c_{kw} belong, respectively, to the component matrix $\mathbf{A}$ $(n \times r)$, describing the physicochemical parameters, $\mathbf{B}$ $(p \times s)$, describing the sampling sites, and $\mathbf{C}$ $(q \times t)$, describing the sampling months. Each of these matrices can be interpreted as a loading matrix in the classical two-way PCA. The elements g_{uvw} weighs the products between component u of the first mode, component v of the second mode, and component w of the third one, and are stored in $\underline{\mathbf{G}}$, an array of dimensions $(r \times s \times t)$, called 'the core of the model'. Generally, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are constrained to be orthogonal and the columns of the matrices are scaled to have unit length. In this way, the magnitude of the squared element of the core (g^2_{uvw}) indicates as to what is the importance of the interaction between the components u, v and w, in the model of $\underline{\mathbf{X}}$. The term e_{ijk} is the residual, or error term.

The Tucker3 model can be computed by iterative algorithms [20,21], generally referred to as alternating least squares (ALS) algorithms, and the least square solution allows the partitioning of the sum of squares of $\underline{\mathbf{X}}$ as

$$\mathrm{SS}(\underline{\mathbf{X}}) = \mathrm{SS}(\text{model}) + \mathrm{SS}(\text{residuals}) \qquad (2)$$

The ratio SS(model)/SS($\underline{\mathbf{X}}$) can be used to evaluate the descriptive performance of the model. Moreover, the product of the total number of components in each different mode $(r \times s \times t)$ can be used as an indication of the number of possible interactions, and hence of the complexity of the model. Plotting the ratio SS(model)/SS($\underline{\mathbf{X}}$) versus the product $(r \times s \times t)$ for possible models can help in evaluating as to which models realize good compromises between the explained fraction of the sum of square of the data and undesired complexity.

For the specific case study, the data array was pre-processed by a procedure called 'j-normalization' [4,14]; physicochemical parameters were standardized within the combined modes of sampling stations and months, allowing us to focus on the spatial and temporal variations [14,15]. The $\underline{\mathbf{X}}$ array was unfolded to have dimensions $(r \times st)$, and the $(r \times st)$ unfolding was centered and scaled to unit variance row-wise. In this way, differences in the physicochemical parameters that are irrelevant for the present purposes, being due to the choices of unit of measurement and due to the range of the data for each single physicochemical variable, are removed. The algorithms that have been applied are implemented on the *N*-Way Toolbox [22] for the MATLAB® [23] computing environment.

2.3. Missing value treatment

Within the data array, there were some missing values; 143 missing values were detected in a total of 2112 measurements (data array: 11 (parameters) × 8 (sites) × 24 (months)), i.e. 6.8% of the required entries were missing.

The resulting composite procedure for handling missing data emerges from a trial and error process in which the results from the imputation methods were checked against available experimental information as follows.

In the first approach, the algorithm that builds the Tucker model and that can impute values for missing data by means of an expectation maximization (EM) scheme has been applied in a straightforward manner to the data array with all 143 missing values. The selection of the number of components to be considered in each mode has been supported by the examination of the plot of the ratio SS(model)/SS($\underline{\mathbf{X}}$) ver-

sus the product ($r \times s \times t$) for models having one to five components in each of the modes, driving to the same choice as that discussed in detail in the results. Missing values that were clustered in blocks of samplings of all sites in a certain month for a parameter were detectable as hard-to-interpret discontinuities in the trends of temporal components of the **C**matrix. For some of the physicochemical parameters presenting such missing blocks — namely dissolved oxygen, chlorophyll-a and phaeopigments — in situ measurements were available from multiparametric Idronaut CTD 401 probe equipped with Sea-Tech fluorometer and PNF-300 fluorometer; hard-to-interpret discontinuities correspond to imputed data having mismatching trends with in situ measurements. Moreover, values imputed for undetermined data known to be below the limit of detection (LOD) of the considered analytical methods were, in several cases, unrealistically high.

In the second approach, we have tried to obtain realistic estimates for the missing values by application to the unfolded array of iterative schemes built in two-way PCA algorithms implemented in the Missing Toolbox [24] and in PLS Toolbox [25] for MatLab®. Estimates obtained using one to five principal components drove to the same kind of problems of mismatching trends and unrealistic values for data known to be below the LOD.

Consequently, in order to overcome these problems, we decided to address separately the problems brought by different classes of missing values. These classes are (i) data below the LOD, (ii) data missing as blocks for samples collected at all sampling sites in certain months (due mostly to operator or instrumental faults), but for which estimates from in situ measurements were available, and (iii) all remaining missing values.

The three classes of missing data were treated as follows:

1. To fill in the data recorded as below the LOD, (44 cases), numbers lying between the LOD and a value two orders of magnitude smaller than the LOD [26] were randomly generated. The alternative choice of filling the data with fixed values corresponding to the LOD of the parameter yields only a difference on the second figure of merit of the amount of variation explained by the model.
2. When two estimates were available for a parameter, one coming from more precise laboratory measurement and the other from in situ probe recording, the missing values were estimated by orthogonal distance regression [27,28], using the probe measurements as predictors. This regression method was chosen since it is suitable for dealing with predictors and target variables, both being subject to experimental errors. This allowed us to impute values for chlorophyll-a and phaeopigments in August–September 1995 and dissolved oxygen in January and February 1997, when laboratory data were not available for any sampling site. In fact, a lack of blocks of values structured in this way does not allow the proper application of the EM method since the assumption of 'missing at random' is violated [29].
3. The EM estimates for the remaining missing values (53 cases) were computed by the algorithm that builds the Tucker models [23].

3. Results and discussion

In general, one prefers parsimonious models, i.e. models with few components, and models which describe a relatively high fraction of variation in the model. A compromise is needed as more variation is explained when the model contains more components. Several possible models, having different numbers of components in each mode, were considered. We have evaluated all possible models with r, s, $t = 1$–5, considering both fitting performances (Eq. (2)) and complexity. In Fig. 2, the ratio SS(model)/SS($\underline{\mathbf{X}}$) versus the

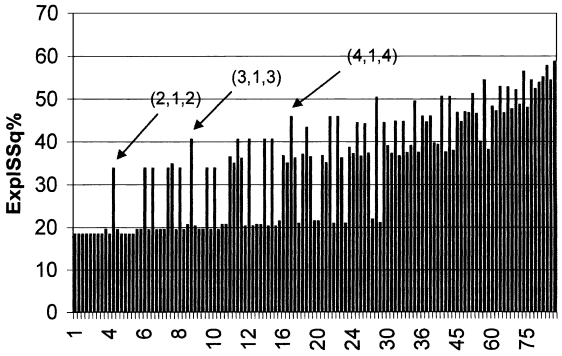

Fig. 2. Diagram of the modeled sum of squares (%) as function of the product of the number of components in different modes for the considered Tucker models. SSqs of models with (2,1,2), (3,1,3) and (4,1,4) components are indicated by an arrow.

product ($r \times s \times t$) is plotted for all these models: one observes that a significant increase in fitting is gained only by losing much of the parsimony of description. In this case, the models explain 18 to 59% of the total sum of squares of the pre-processed data.

The variation accounted for by each single element of the core was examined for the models fitting the data the best (sum of squares >40%). For all these models, the elements explaining the largest fractions of the total sum of squares are due to the first three components of the first mode (physicochemical variables) and the third one (months), while only one component is important for the second mode (sampling sites). We, therefore, decided to base our interpretation on the model which has three components in the physicochemical parameters mode, one component in the sampling site mode, and three components in the time mode. This model, indicated for short as Model 3-1-3, explains 41% of the variation: Fig. 2 shows that this is one of the best compromises, considering both fit and complexity. Models with higher number of components were discarded because they were less interpretable. We further verified the dimensionality of the model by applying a generally applicable approach that utilizes the rotational abilities of the Tucker3 model. We estimated a model with four components in each mode. In such a core, the variation is dispersed over the many (64) elements. In order to see as to what extent the core could be simplified, we optimized the variance-of-squares measure [30] by orthogonal rotations to preserve fit. In the initial core, 89% of the sum of squares of the core entries were dispersed over four components in the first mode, two in the second mode and three in the third mode. Upon rotation, three factors in the first mode, one factor in the second mode and three factors in the third mode covered 94% of the core variation. Thus, by optimizing the variance-of-squares of the core to reveal the latent number of significant combinations of factors, the Tucker model could be simplified and it could still provide a satisfactory fit to data.

A systematic and safe approach for finding the optimal model dimensionality is required as the optimal model dimensionality for exploratory purposes lies between a high fit to data on one hand and a low number of factors representing the systematic variation in the data on the other hand. If too many factors are included, there is the risk of over-interpretation causing

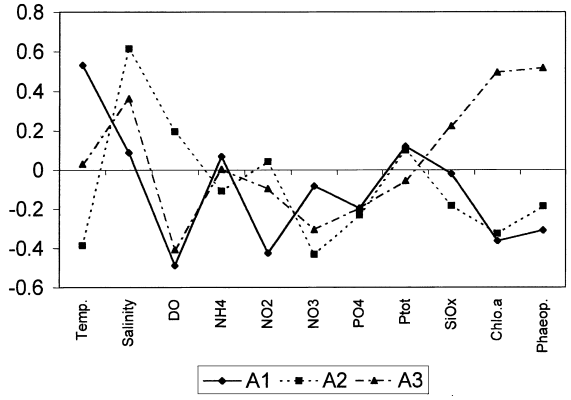

Fig. 3. Plot of loading values for the three components (A1, A2 and A3) describing the physicochemical parameters.

extrapolated conclusions that may be correct only for the data set at hand but not valid for future, and essentially similar data sets. Furthermore, if the data cannot be projected onto a low-dimensional space, the model becomes too difficult to overview and interpret and the analyst may be better off keeping with the original raw data. Thus, only the combinations of factors that have high weights in the core array should be used for interpretation. Accordingly, factor rotation is applied to simplify the core. When the core has been brought to its simplest form, we may interpret the distribution over factors. As described, the approach based on rotations verified the dimensionality found from evaluating fit over model dimensionality.

First, we will discuss the components of each mode separately, and then the relevance of the interactions between the different modes will be addressed.

The loadings of each physicochemical variable on the three components of matrix **A** are displayed in Fig. 3. The first component A1 has high positive loadings of temperature, and high negative loadings for dissolved oxygen, chlorophyll-a, phaeopigments and nitrite. This latent variable explains some of the effects of temperature on the composition of the waters. The contrast in sign between the temperature of water and the amount of dissolved gases can be explained by the fact that a gas dissolves less at high temperature. The contrast between temperature and chlorophyll is interpreted in terms of the activities of phytoplankton in seawaters, which is particularly relevant for the algal bloom of early spring [31] when waters are cold and rich in oxygen. The loading of nitrite (intermediate

oxidation state for nitrogen) is linked to the loading of dissolved oxygen.

Salinity is positively loaded on the component A2, while temperature, nitrate and chlorophyll-a have high negative loadings. Since the Gulf receives an input of freshwaters from the Isonzo river, this latent variable can be interpreted as being related to the river input of relatively warmer water low in salinity, rich in nitrate, from leachate of agricultural soils. Nitrates are nutrients feeding the phytoplankton, which in turn produces chlorophyll pigments.

The negative loadings of oxygen and nutrients (NO_3 and PO_4) on A3, and the positive ones of salinity, chlorophyll-a and phaeopigments suggest that the third latent variable could be explained in terms of the metabolic and the catabolic activity of planktonic species. In fact, phytoplankton consumes NO_3 and PO_4, producing chlorophylls; the decay of dead organisms consumes dissolved oxygen, and produces phaeopigments. All this happens when waters have a relatively high salinity.

To describe the spatial information, the loadings of each sampling site on the single component B1 extracted for the second mode are reported in Fig. 4.

It can be observed that all loadings have a positive sign, which means that this component simply ranks all stations; such a component is sometimes referred to as a 'size component' [28]. We note that sites that are spatially close have similar loadings so that the stations could be roughly grouped in the following three classes: class I (loadings greater than 0.35) contains sites 7, 8 and 4, in class II (loadings between 0.35 and 0.30) there are sites 1, 2, 3 and 6, while site 5 belongs to class III (loading lower than 0.30). This classification is reported on the map of Fig. 1: a spatial gradient of the influence of the river estuary on the seawaters is clearly shown.

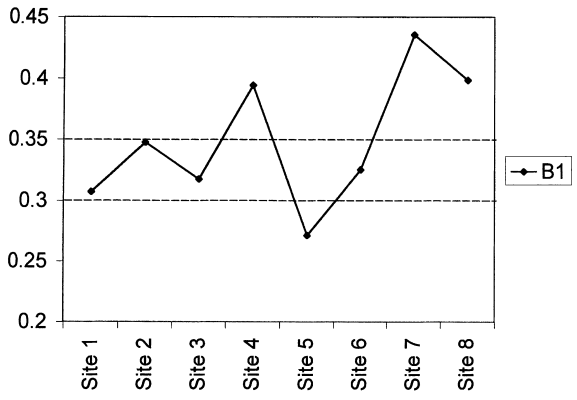

Fig. 4. Plot of loading values for the component (B1) describing the spatial pattern.

Table 1
Core array (**G**) of the 3-1-3 model in the unfolded form

	C1 B1	C2 B1	C3 B1
A1	19.2605	0.0000	0.0000
A2	0.0000	18.1401	0.0000
A3	0.0000	0.0000	11.8597

The loadings of the components explaining the variation of data in time (third mode) are shown in Fig. 5. The first temporal component C1 shows seasonal oscillations; there are positive values for summer and autumn, while winter and spring months have negative values. Loadings on C2 are high and positive in winter, high and negative in late spring (June 1995, May 1996). Loadings on C3 are positive for May and November 1995 and December 1996, and negative in June 1995 and April 1996.

The core array **G** of the model, which can be visualized as a parallelepiped having, as dimensions, 3 (physicochemical components) × 1 (spatial component) × 3 (temporal components), is displayed in the unfolded form in Table 1. The element g_{uvw} reflects the extent of the interaction between A_u, B_v and C_w ($u = 1,2,3$, $v = 1$, $w = 1,2,3$). It is noteworthy that, in the present case, there are only three relevant elements in the core, the ones for which $u = w$. Recalling Eq. (1), we note that the sign of g_{uvw} depends on the algebraic multiplication of the signs of A_u, B_v and C_w. Since B_v is always positive, it will be sufficient to focus on the signs of A_u and C_w.

The element of the core g_{111} (19.2605) explains the interaction among the first factors of each of the modes. The seasonal variation of the physicochemical pattern of the waters (A1), related to the variation of temperature, affects all locations (B1), according to the spatial gradient displayed in Fig. 1 (waters near the river estuary are more oxygenated), and oscillates in time (C1) between the maxima in summer–autumn and the minima in winter–spring. During summer and autumn (positive C1), high values of temperature have, as a consequence, low values of dissolved oxygen in the waters (A1); in winter and spring, the effect of

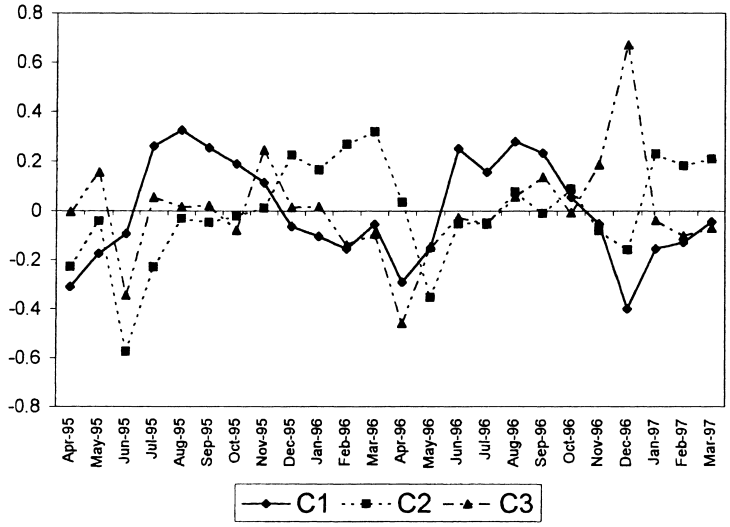

Fig. 5. Plot of loading values for the three components (C1, C2 and C3) describing the temporal pattern.

temperature is opposite. The positive value of the element g_{212} (18.1401) can be explained by the fact that in winter (positive values of C2), the waters have a relatively high salinity, while temperature, nutrients and phytoplankton activity are low (A2). The positive g_{212} also implies that, in late spring (negative values of C2), the waters of the Gulf become less saline, richer in nitrates and warmer (change of sign for A2) as a consequence of the input of freshwaters from the Isonzo river bringing nitrate fertilizers drained from the agricultural soils; waters flowing through the soils become warmer. The high chlorophyll concentration in the waters demonstrates that the physicochemical pattern of the waters is favorable for phytoplankton activity. Again, the importance of this phenomenon depend on the spatial pattern, described by B1. Since all interactions involve B1, there being only one B component, this is true for all interactions and will no longer be mentioned.

The last relevant interaction among components in the Model 3-1-3 is accounted for by the element of the core g_{313} (11.8597), which shows that the metabolic and catabolic activities of phytoplankton (A3), consuming nutrients and dissolved oxygen and producing pigments, were particularly relevant in May and November 1995 and in December 1996 (positive C3); in June 1995 and April 1996 (negative C3), this decrease in nutrients and dissolved oxygen is contrasted and masked by the river water input and seasonal increase in oxygen (inversion of sign for A3). The three-dimensional scatter plot of the components C1, C2 and C3 of Fig. 6 shows June 1995, April 1996, and December 1996 as anomalous sampling conditions within the monitoring period. These three outliers can be interpreted with the help of meteorological data. In June 1995, the freshwater input from the Isonzo river was exceptionally abundant (high negative C2), and this is reflected mainly by salinity and nitrate parameters, scoring, during this month, the lowest and the highest values, respectively, of the whole survey, this being associated with low values of C3, related to phytoplankton catabolism (consumption of dissolved oxygen). In April 1996, low temperatures caused the C1 value to be low, that is, caused the value for dissolved oxygen to be high, and the C3 value to be low. December 1996 was the warmest winter month in the last 35 years [32]: this special meteoclimatic situation provoked an anomalous algal bloom, with the related metabolic and catabolic activities. For these reasons, December 1996 has loadings similar to spring months in C1, and to summer–autumn ones in C3.

4. Conclusions

This study represents a further step in our project aimed at finding methods describing the dynamics and fate of chemicals in a highly anthropized area, the

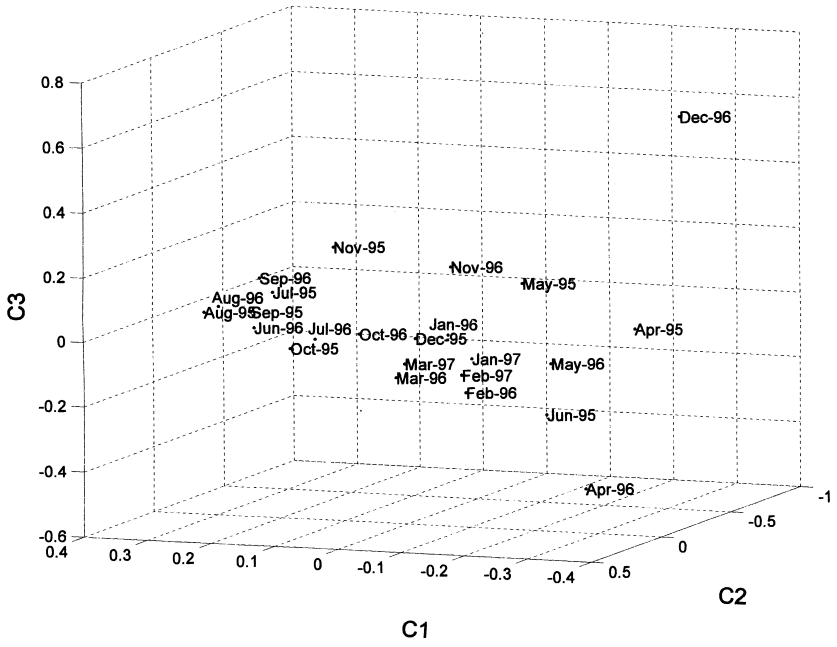

Fig. 6. Three-dimensional score plot of components C1, C2 and C3, pointing out anomalous sampling conditions.

Gulf of Trieste [33,34]. From a methodological point of view, it demonstrates how a Tucker3 model succeeds in capturing a significant part of the systematic variations present in this data set, collected during 2 years in open-field conditions. Such a result is also obtained if the model accounts for a relatively small portion of the total sum of squares of the data since Tucker3 tends to model data in a more parsimonious way with respect to the classic two-way PCA.

After having tried to apply some algorithms to fill in all missing data, we decided to apply different approaches to handle missing values generated in different stages of the experimental collection. We remark that algorithms filling holes in the data sets have to be supervised carefully in order to obtain consistent results. The choice of the appropriate number of components for each mode is a critical point in building a model for the data set since it represents our comprehension of the structure of the information present in the considered data. A simple plot of degree of fit versus complexity of the examined models helped us in this selection, pointing at the Model 3-1-3 as the best interpretable model; and the optimization of the variance of squares of the core confirmed the choice of the number of combination of factors.

The issue of validation is a cardinal point in the current work. In addition to validating the model dimensionality, the solutions have been validated in terms of their geophysical context. The patterns and systematic variations that have been resolved substantiate findings from other research projects. Mechanisms of, e.g., light, temperature and flow, are found to consistently explain the covariation in the observed data. We have outlined a robust and feasible approach to the analysis of a very complex data structure that, quickly and in a cognitive easy way (graphical outputs), presents the significant variations of data and allows for interpretation in the context of the Gulf of Trieste. The path from variable scaling and pre-treatment, handling of missing values, estimation of model dimensionality and presentation of the solutions is generally applicable to multi-way analysis; also for monitoring studies concerning different kind of ecosystems [35] and in areas other than environmental studies.

It is worth noting that, in the considered model, the factors defining the physicochemical composition of these surface waters are all related to the chlorophyll concentration, and consequently, to both the photosynthetic activity of phytoplankton and the primary

production. This gives us a highly detailed image of the processes involved in chlorophyll production in the Gulf, and at the same time, it realizes an effective synthesis of information from the data collection.

Since Tucker models allow exploration when and where these factors are important, such models seem to be suited for further examination of the complex pattern ruling the basic stage of the trophic chain in seawaters. The model also allowed highlighting of the environmentally anomalous situations by means of the examination of the three-dimensional plot of the temporal components.

We have to note that, in this study, only surface waters were considered, without taking into account the vertical dynamics in the water column. These surely condition the quality of these waters, for instance, at the overturn of waters, generally occurring in spring and autumn, due to the disappearance of thermal stratification [36]. Tucker models for four-way arrays are currently being studied, allowing us to consider different layers, and hopefully, to integrate vertical dynamics in our models.

Acknowledgements

The authors are grateful to the Italian National Research Council (C.N.R.) for allowing the usage of the data, Dr. Massimo Celio for in situ measurements, and the Flemish Scientific Research Fund (FWO) for the financial support to this study. C.A. Andersson thanks PMP for funding.

References

[1] A. Smilde, Chemometr. Intell. Lab. Syst. 15 (1992) 143.
[2] Y. Zeng, P. Hopke, Chemometr. Intell. Lab. Syst. 7 (1990) 237.
[3] Y. Zeng, P. Hopke, Atmospheric Environ. 26A (1992) 1701.
[4] R. Henrion, G. Henrion, G. Onuoha, Chemometr. Intell. Lab. Syst. 16 (1992) 87.
[5] R. Leardi, C. Armanino, L. Alberotanza, Book of Abstracts CAC '96, Tarragona, 1996, p. 76.
[6] J.H. Carpenter, Limnol. Oceanogr. 10 (1965) 141.
[7] K. Grasshoff. M. Ehrhardt, K. Krenling, Methods of Seawater Analysis, VCH, Weinheim, 1983.
[8] T.R. Parsons, Y. Maita, C.M. Lalli, Manual of Chemical and Biological Seawater Analysis, Pergamon Press, Oxford, 1985.
[9] G. Magazzù, Metodi per lo studio del plancton e della produzione primaria, G.M., Roma, 1978.
[10] C.N.R., M.U.R.S.T., PRISMA. http://adria.irpem.an.cnr.it/
[11] H.A.L. Kiers, Psychometrika 56(3) (1991) 449.
[12] S. Wold, P. Geladi, K. Esbensen, J. Ohman, J. Chemometr. 1 (1987) 41.
[13] L. Tucker, Psychometrika 31 (1966) 279.
[14] P.M. Kroonenberg, Three Mode Principal Component Analysis, DSWO Press, Leiden, 1983.
[15] R. Henrion, Chemometr. Intell. Lab. Syst. 25 (1994) 1.
[16] R.A. Harshman, UCLA Working Papers in Phonetics 16 (1970) 1.
[17] J.D. Carrol, C.C. Chang, Psychometrika 35(3) (1970) 283.
[18] R. Bro, Chemometr. Intell. Lab. Syst. 38 (1997) 149.
[19] I.T. Joliffe, Principal Component Analysis, Springer, New York, 1986.
[20] C.A. Andersson, R. Bro, Chemometr. Intell. Lab. Syst. 46 (1998) 93.
[21] P. Paatero, C.A. Andersson, Chemometr. Intell. Lab. Syst. 47 (1999) 17.
[22] R.Bro, C.Andersson, *N*-way Toolbox. http://newto.foodsci.kvl.dk/Matlab/Nwaytoolbox/index.htm
[23] The MathWorks, Matlab 5.0, Natick, MA, 1997.
[24] C.A. Andersson, The Missing Toolbox. http://newton.mli.kvl.dk/missing/
[25] B. Wise, PLS Toolbox, Eigenvector Research, Manson, USA.
[26] R. Aruga, Anal. Chim. Acta 354 (1997) 255.
[27] P.T. Boggs, C.H. Spiegelman, J.R. Donaldson, R.B. Schnabl, J. Econometrics 38 (1988) 169.
[28] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1998.
[29] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, Wiley, New York, 1987.
[30] R. Henrion, C.A. Andersson, Chemometr. Intell. Lab. Syst. 47 (1999) 189.
[31] J.P. Riley, G. Skirrow, Chemical Oceanography, vol. 1, Academic Press, London, 1965, p. 575.
[32] F. Stravisi, Chair of Metereology, University of Trieste, personal communication, 1997.
[33] E. Reisenhofer, G. Adami, A. Favretto, Fresenius' J. Anal. Chem. 354 (1996) 729.
[34] P. Barbieri, G. Adami, A. Favretto, E. Reisenhofer, Fresenius' J. Anal. Chem. 361 (1998) 349.
[35] P. Barbieri, G. Adami, E. Reisenhofer, Annali di Chim. 89 (1999) 1.
[36] P. Barbieri, G. Adami, S. Predonzani, E. Reisenhofer, D.L. Massart, J. Environ. Monit. 1 (1999) 69.