ELSEVIER

Chemometrics and Intelligent Laboratory Systems 62 (2002) 89–100

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

A three-way principal factor analysis for assessing the time variability of freshwaters related to a municipal water supply

P. Barbieri [a,*], G. Adami [a], S. Piselli [b], F. Gemiti [b], E. Reisenhofer [a]

[a] *Department of Chemical Sciences, University of Trieste, Via Giorgieri 1, 34127 Trieste, Italy*
[b] *AC.E.GA.S S.p.A., Via Maestr: del Lavoro 8, 34123 Trieste, Italy*

Received 27 July 2001; accepted 15 January 2002

Abstract

Chemical analyses (total hardness, HARD; dissolved oxygen, DO; chlorides; sulfates; nitrates; nitrites; ammonia; orthophosphates; and UV-absorbing organic constituents, UV-ORG), physical data (turbidity, TURB; temperature, TEMP; conductivity, COND), and biological monitors (total and faecal coliforms, FAEC; faecal streptococci, STREPTO) constitute the 15 parameters, monitored with monthly frequency in the space of 4 years on freshwaters sampled at seven sites in a karstic area of northeastern Italy. The data set was used for a three-way principal factor analysis aimed at exploring the pattern of information about the environmental quality of the monitored freshwaters, since four wells are feeding the municipal water supply of the Province of Trieste, and the other water courses can influence them. The selected three-way (3,3,2) model uses three components for describing the analytical parameters, three for temporal variations and two for spatial variations. The method optimising the 'variance of squares' of the core elements has permitted a simple and meaningful interpretation of the Tucker-3 solution. The procedure succeeded in decomposing the overall temporal variation in three parts, thus highlighting nonperiodic critical events, a periodic seasonal component and a constant term. The seasonality has been confirmed by the examination of the autocorrelation function of the second temporal component. An environmental interpretation and an estimate of the relative relevance of phenomena conditioning the considered water body, detected by the multiway analysis, have been proposed.

Keywords: Multiway PF analysis; Tucker-3 models; Freshwater quality; Water monitoring

1. Introduction

For years, we have studied the quality of freshwater of the karstic area surrounding the town of Trieste (northeastern Italy) [1–4]. Due to the calcitic–dolomitic nature of this area, many of the water courses flowing here are subterranean: i.e., the case of the Timavo River that has a surficial course of about 50 km in Slovenia, then sinks into a limestone fissure near the Italian border, and assumes a variety of routes, mainly hypogeous, before emerging near the coast and flowing into the Adriatic Sea. This condition is favourable for preserving the quality of these waters, but, at the same time, hinders not only a detailed knowledge of the hydrology of the subterranean water courses, but also the sampling and monitoring operations necessary for an adequate understanding of the behaviour and properties of this complex hydrological system. A further factor of

* Corresponding author. Tel.: +390-406-763-954; fax: +390-406-763-903.
E-mail address: barbierp@univ.trieste.it (P. Barbieri).

0169-7439/02/$ - see front matter
PII: S0169-7439(02)00007-2

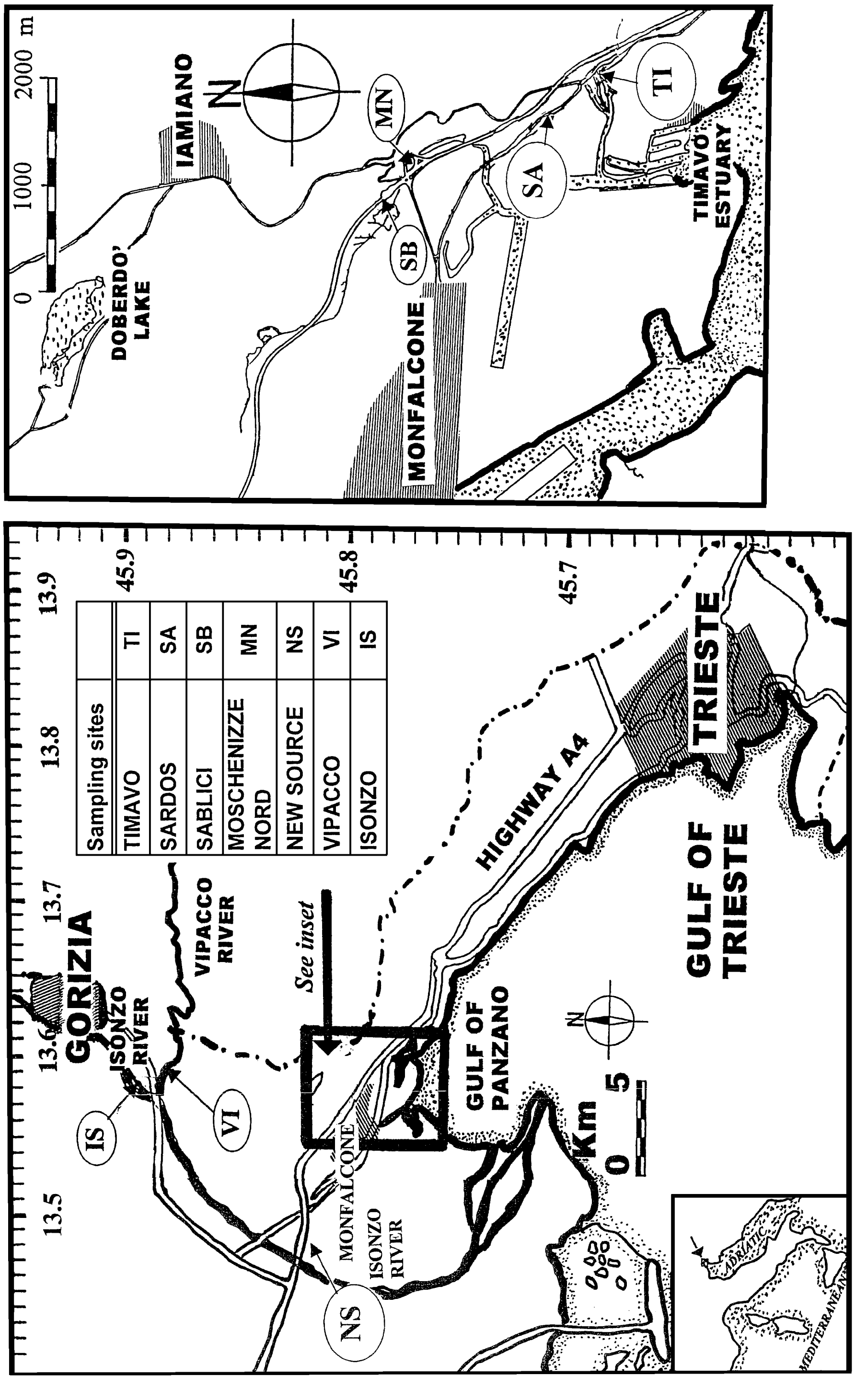

Fig. 1. Sampling sites.

complexity is due to the permeability of the karstic soils that induces mutual overflowing among the contiguous watersheds of this area. In a previous work [2], we have verified these overflowing phenomena, which obviously depend on the different seasonal rain conditions and produce occasional intrusions of waters from the northern Isonzo and Vipacco rivers into the southern karstic wells related to Timavo River.

Besides the speculative interest for such a com-plex hydrological system, these waters are relevant because they feed the municipal supply of the Province of Trieste. In particular, the waters of the three wells indicated in Fig. 1 as SB, MN, and SA are collected since the 19th century for drinking use. For satisfying the always increasing need of water, new wells were drilled in recent years on the left side of Isonzo River: the addition of this new source of freshwater (displayed as NS in Fig. 1) assures, since the late 1980s of the past century, a delivery capacity that largely exceeds the water request (estimated about 150,000 mc/day) of the Province of Trieste, also in the summer dry period when water use reaches its maximum.

In this work, we consider not only the freshwaters sampled at the abovementioned sites, but also for the sake of comparison, the waters taken at three monitoring stations relative to the three main rivers of this area: Isonzo, Vipacco, and Timavo rivers, respectively (IS, VI, and TI in Fig. 1).

We report here a study based on chemical analyses (total hardness, dissolved oxygen, chlorides, sulfates, nitrates, nitrites, ammonia, orthophosphates, and UV-absorbing organic constituents), on physical data (turbidity, temperature, conductivity), and on biological monitors (total and faecal coliforms, faecal streptococci) relative to these waters. All data have been determined on samples collected with monthly frequency within 4 years (from January 1995 to December 1998) at the seven stations abovementioned.

While our previous works were based on monitoring campaigns focused on seasonal critical events covering at most the space of 2 years, the present work faces time series of monthly samplings being collected during 4 years. The longer temporal range being covered should evidence in a clearer way both regularities and anomalous events in the variations of the measured parameters.

It has already been shown [4–10] that the three-way principal component analysis (PCA) succeeds in rationalising efficaciously the information underlying data arrays of analytical parameters collected at various sampling times and at various sites. The (i) chemical–physical–biological parameters, (ii) the sampling months, and (iii) the sampling sites constitute the three ways that allow us to identify the data under study. The two types of multiway methods that have been mainly applied in environmental pattern recognition are Tucker-3 and PARAFAC [11,12]. Recently, the multilinear engine ME has been proposed [13].

In environmental field, Tucker-3 models, considering the same number of components (often two) for each way with diagonalisation of the core, have been mostly used [7–10,14]. It has also been shown [4] how Tucker-3 models are selected, having the same number of significant components in each way identifying the data and, in case they have a simple information structure, detected after diagonalisation of cubic cores of the model, it can be appropriate to use simpler PARAFAC models.

In the present paper, we aim at exploring the amount of information provided by multiway PCA for environmental quality monitoring with relatively long time series, considering also the possibility of *different* number of components in each way. In the case that different numbers of components are retained for the three ways, the “variance of squares” method [15] of rotation could be used for improving the interpretability of the model. This kind of rotation seems very effective [6]; however, it has not yet been applied widely for modelling environmental data.

The number of components to be retained as significant for each way can be selected by means of graphical methods helping the visualisation of the compromise between fitting performances and complexity for different models, as we proposed in previous works [4,6].

2. Experimental

2.1. Sampling

Sampling operations were made at seven sites. Sardos (SA), Sablici (SB), Moschenizze Nord (MN) are the three historical karstic wells that feed the municipality of Trieste together with the more recent northern source (NS). The other three sites are devoted to monitor the main rivers flowing in this area, i.e., Isonzo (IS), Vipacco (VI), and Timavo (TI) rivers (see Fig. 1). Water samples were taken with monthly frequency from January 1995 to December 1998, and analysed within 48 h in the Laboratory for Analysis and Control of ACEGAS of Trieste. The analytical determinations followed the official procedures of the Italian Law [16] and the standard methods of the American Public Heath Association [17].

Turbidity (TURB, measure unit: Jackson turbidity unit, JTU), temperature (TEMP, °C), and conductivity (COND, μS/cm, corrected to 25 °C) were determined in situ, while all the other parameters were measured in the laboratory. They were: chlorides (Cl, mg/l), sulfates (SO_4, mg/l), total hardness (HARD, °F), dissolved oxygen (DO, mg/l), nitrates (NO_3, mg/l), nitrites (NO_2, mg/l), ammonia (NH_3, mg/l), orthophosphates (PO_4, mg/l), UV-absorbing organic constituents determined by spectrophotometry at 253.7 nm wavelength using cells of 5 cm path length (UV-ORG, A), total coliforms (COLI, most probable number MPN/100 ml), faecal coliforms (FAEC, MPN/100 ml), and faecal streptococci (STREPTO, MPN/100 ml).

2.2. Statistical analysis

In this work, a three-way principal component analysis using the Tucker-3 model [5,11,18] will be applied to a data set, which represents 15 parameters reflecting the water quality, collected for 48 months from January 1995 to December 1998 at seven sites. The dimensions of the three ways (i.e., the quality parameters, the sampling months, and the sampling sites that identify each datum) are 15, 48, and 7, respectively, for a total of 5040 data. The missing data are 43 (0.85% of the total).

The Tucker-3 model constitutes a factorisation of the $X=\{x_{ijk}\}$ data array of $(n\times p\times q)$ dimensions. In our case, x_{ijk} is the value of the chemical, physical, or biological parameter i (going from 1 to $n=15$), on j month (from 1 to $p=48$), at k site (from 1 to $q=7$), accordingly the following equation:

$$x_{ijk}=\sum_{u=1}^{r}\sum_{v=1}^{s}\sum_{w=1}^{t}a_{iu}b_{jv}c_{kw}g_{uvw}+e_{ijk}.$$

In this equation, the r, s and t indexes represent the number of components chosen for describing the first, the second, or the third way of the data array, respectively, while a_{iu}, b_{jv}, and c_{kw} are the elements of the three component matrices **A**, **B**, and **C**. The $A(n\times r)$ matrix describes the measured parameters, while $B(p\times s)$ the sampling months, and $C(q\times t)$ the sampling sites. Each of these matrices can be interpreted in the same manner as a loading matrix of the classical two-way PCA, since they are all column-wise orthogonal. The equation term g_{uvw} represents an element of G, an array with $(r\times s\times t)$ dimensions, called the 'core' of the model. g_{uvw} weighs the products of the u component of the first way by the v component of the second way, and the w component of the third way. The component matrices **A**, **B**, and **C** are constrained to be orthogonal, and the matrix columns are scaled to unity length. In this way, the squared value of the core element, i.e., (g^2_{uvw}), shows the entity of the interactions among the u, v, w components of the $X=\{x_{ijk}\}$ data array. The last element, e_{ijk}, constitutes the residual, i.e., the part of data not represented by the model. The Tucker-3 model is computed by an iterative procedure based on the 'alternating least square' (ALS) algorithm [12,18], and the solution permits to partition the sum of squares of the X elements as:

$$\mathrm{SS}(X)=\mathrm{SS}(\mathrm{model})+\mathrm{SS}(\mathrm{residual}).$$

The SS(model)/SS(X) ratio can be used for evaluating the strength of the model in representing its objects. In the following, we will call this ratio 'explained variation' of the model. The data array was pretreated, by centring and scaling each chemical, physical, and biological parameter [4,6,18], so as to remove differences due to their different units of measure.

The **A**, **B**, **C** matrices and the G core array can be rotated, as well as in classical factor analysis. A recently proposed rotation method, the "variance of squares" [15], optimises the variance of the squared core elements, by distributing the total variance among a little number of elements that permits to obtain models more easy to interpret. This rotation method can be used with advantage on models with different component numbers in the different ways. All calculations were performed by MatLab 5.0

computing environment [19], using the N-way toolbox of Andersson and Bro [20,21], which estimates the missing data by an expectation/maximisation-type algorithm.

3. Results and discussion

We will rationalise here the spatial–temporal variations of freshwaters related to municipal water supply of Trieste on the basis of 15 chemical–physical–biological parameters, intended as indicators of water quality. These parameters were determined in samples collected at seven monitoring stations with monthly frequency, for a total of 48 sampling events. The basic statistics of these experimental data are reported in Table 1.

We have taken into account many Tucker-3 models, with different component numbers in the different ways. As a rule, we prefer to handle models with few components, but, at the same time, it is advisable to use models providing maximum explained variation, i.e., SS(model)/SS(X) ratio. Since more variation is explained when more components are included in the model, it is necessary to reach a compromise. Thus, we have computed the SS(model)/SS(X) ratios for all possible models with r, s, and t varying from 1 to 5. The ($r \times s \times t$) product of the component numbers in each different way can be considered as an indicator for the number of possible interactions and, consequently, for the complexity of the model. The models here considered explicate a percent of the SS ranging from 28% to 69%.

In previous works [4,6], we proposed a diagram for the selection of optimal models, where the SS(model)/SS(X) ratios were plotted vs. the ($r \times s \times t$)products, retained as indicator of complexity. In the diagram, the SS(model)/SS(X) ratios of the models were sorted along the x-axis on the basis of growing ($r \times s \times t$) products so that from left to the right, the values of the indicator of complexity increased monotonically. Relevant gains in the fitting performances at the increase of complexity could be appreciated. Even clearer, for highlighting interesting models and for the selection of the number of factors to be considered in each of the three ways, is the scatter plot of SS(model)/SS(X) ratios vs. ($r \times s \times t$) products, reported in Fig. 2 for our data set.

The models providing the highest SS(model)/SS(X) ratios for each ($r \times s \times t$) product can be evidenced in this plot. They present the more interesting fit complexity trade-off. The model (3,3,2)—indicated in Fig. 2 by an arrow—uses three components for describing the analytical parameters, three components for the temporal variations, and two components for spatial variations, and it explicates the 50.17% of the data variation. It is the model with lower complexity ($3 \times 3 \times 2 = 18$ possible interactions to be considered) overcoming 50% of the parameter of fit, and it has been chosen as a reasonable compromise to be discussed in more detail.

With the aim of getting an easily interpretable solution, we have rotated the **A**, **B**, **C** matrices of this (3,3,2) model by optimising the 'variance of squares' [15]. With this procedure, the number of core elements with significant figures is minimised by orthogonal transformations. The result is that the model is rearranged so as to obtain a little number of chemical–physical–biological factors, which are well characterised both temporally and spatially. The so obtained factors are plotted in Figs. 3–5.

The three factors of parameters are displayed in Fig. 3. We note that A1 shows negative values for TEMP, COND, Cl, SO_4, HARD, NO_3 (i.e., parameters related to the water salinity), while positive values for NO_2, UV-ORG, COLI, FAEC, STREPTO (i.e., indicators of contamination), and DO that is high in surficial waters. The second factor A2 has positive values for TURB, DO, NO_3, STREPTO (signs of runoff of agricultural soils). The third factor has the highest loadings for TURB and NH_3 (parameter related to sewage contamination).

The three temporal factors are plotted in Fig. 4. The first factor B1 shows few positive peaks, occurring without evident periodicity. On the contrary, the second factor B2 alternates positive values in colder months with negative ones in warmer season. No temporal variations are related by the third factor B3, which has rather constant values.

The two spatial factors are plotted in Fig. 5. The factor C1 has positive values for VI and IS (surficial rivers), while negative ones for the karstic wells (TI and SA), and very low loading for the deepest NS. The second factor C2 displays a very high value for VI.

It could be interesting to verify in more detail the seasonality emerging from the multiway treatment.

Table 1
Basic statistics of the 15 parameters, relative to seven sampling sites: yearly mean values (*m*) and standard deviations (S.D.); n.d. = not determinable (below LOD); n.r. = negative reactions for bacterial tests

Site	Year		TURB	TEMP	COND	CL	SO_4	HARD	DO	NO_3	NO_2	NH_3	PO_4	UV-ORG	COLI	FAEC	STREPTO
TI	1995	*m*	6.1	11.7	386	6.5	12.5	21.8	9.1	6.6	< 0.01	< 0.03	0.04	0.11	363	60	30
		S.D.	13.0	1.1	26	2.0	4.3	1.4	0.9	0.8	n.d.	n.d.	0.02	0.04	391	42	43
	1996	*m*	5.6	11.9	383	6.8	12.0	21.5	8.8	6.5	< 0.01	< 0.03	< 0.03	0.11	558	117	54
		S.D.	7.1	1.0	33	2.1	4.7	2.0	0.8	1.2	n.d.	n.d.	n.d.	0.06	596	113	63
	1997	*m*	1.8	11.9	375	7.2	10.7	21.2	8.0	6.3	< 0.01	< 0.03	0.04	0.09	150	54	28
		S.D.	1.8	0.8	17	1.8	1.0	1.7	1.0	1.1	n.d.	n.d.	0.02	0.04	124	70	35
	1998	*m*	4.1	12.4	387	7.2	11.9	21.9	9.0	6.7	< 0.01	< 0.03	< 0.03	0.09	449	64	32
		S.D.	5.2	1.0	22	1.8	4.8	1.3	1.5	0.5	n.d.	n.d.	n.d.	0.05	683	101	44
SA	1995	*m*	2.6	12.5	385	7.2	10.5	21.9	8.8	7.5	< 0.01	< 0.03	< 0.03	0.05	299	46	25
		S.D.	3.5	0.4	31	1.3	1.8	1.8	0.6	1.5	n.d.	n.d.	n.d.	0.02	346	48	41
	1996	*m*	2.4	12.4	396	7.4	10.4	22.2	9.0	6.8	< 0.01	< 0.03	< 0.03	0.06	282	58	44
		S.D.	1.8	0.5	35	1.5	0.8	2.1	0.5	1.2	n.d.	n.d.	n.d.	0.02	182	53	50
	1997	*m*	1.0	12.6	368	6.8	10.7	20.8	8.1	6.7	< 0.01	< 0.03	< 0.03	0.06	105	29	26
		S.D.	0.5	1.2	43	0.9	1.0	2.3	0.7	1.0	n.d.	n.d.	n.d.	0.02	84	27	25
	1998	*m*	1.9	12.7	378	7.5	10.0	21.5	8.2	7.0	< 0.01	< 0.03	< 0.03	0.06	375	61	24
		S.D.	1.7	1.2	41	1.8	0.8	2.4	3.0	0.6	n.d.	n.d.	n.d.	0.02	565	84	35
SB	1995	*m*	0.9	12.1	341	4.9	10.3	19.1	8.7	6.6	< 0.01	< .03	< 0.03	0.06	301	43	25
		S.D.	0.4	0.8	27	0.8	0.9	1.6	0.6	0.8	n.d.	n.d.	n.d.	0.02	298	35	53
	1996	*m*	0.9	12.0	341	4.9	10.3	19.1	8.7	6.6	< 0.01	< 0.03	< 0.03	0.06	301	43	25
		S.D.	0.4	0.8	27	0.8	0.9	1.6	0.6	0.8	n.d.	n.d.	n.d.	0.01	298	35	53
	1997	*m*	0.6	11.8	323	5.0	9.5	18.4	8.0	6.1	< 0.01	< 0.03	< 0.03	0.06	110	26	17
		S.D.	0.3	1.0	31	0.9	0.8	1.7	1.0	1.0	n.d.	n.d.	n.d.	0.01	50	18	17
	1998	*m*	0.9	12.2	333	4.6	8.8	19.1	8.2	6.5	< 0.01	< 0.03	< 0.03	0.07	375	68	24
		S.D.	0.5	0.9	39	0.5	0.8	2.2	0.9	0.6	n.d.	n.d.	n.d.	0.03	655	96	33
MN	1995	*m*	0.9	12.1	340	4.6	10.5	19.1	8.5	6.8	< 0.01	< 0.03	< 0.03	0.06	277	38	23
		S.D.	0.5	0.7	27	1.0	1.0	1.6	0.8	0.7	n.d.	n.d.	n.d.	0.02	279	36	55
	1996	*m*	0.8	11.8	349	4.5	9.6	19.9	8.5	6.6	< 0.01	< 0.03	< 0.03	0.06	218	43	34
		S.D.	0.2	1.0	34	0.7	0.9	2.2	0.6	0.7	n.d.	n.d.	n.d.	0.01	131	49	50

	1997	*m*	0.5	11.9	323	4.9	9.3	18.6	8.0	6.2	< 0.01	< 0.03	< 0.03	0.06	94	18	15
		S.D.	0.3	1.0	31	0.9	0.7	1.6	1.0	1.0	n.d.	n.d.	n.d.	0.01	48	13	13
	1998	*m*	1.1	12.2	335	4.4	8.8	19.1	7.3	6.6	< 0.01	< 0.03	< 0.03	0.06	312	60	22
		S.D.	0.6	0.7	38	0.7	0.8	2.3	2.6	0.5	n.d.	n.d.	n.d.	0.03	516	91	31
DN	1995	*m*	0.7	12.0	374	19.1	12.1	18.8	8.0	7.9	< 0.01	< 0.03	< 0.03	< 0.01	n.r.	n.r.	n.r.
		S.D.	0.5	0.4	14	1.5	0.9	0.4	0.2	0.8	n.d.	n.d.	n.d.	n.d.	n.r.	n.r.	n.r.
	1996	*m*	0.2	12.1	389	21.7	12.3	19.4	8.1	8.4	< 0.01	< 0.03	< 0.03	< 0.01	n.r.	n.r.	n.r.
		S.D.	0.1	0.5	11	1.8	0.6	0.5	0.4	0.3	n.d.	n.d.	n.d.	n.d.	n.r.	n.r.	n.r.
	1997	*m*	0.2	12.1	387	23.2	13.0	19.7	7.7	8.7	< 0.01	< 0.03	< 0.03	< 0.01	n.r.	n.r.	n.r.
		S.D.	0.2	0.6	12	1.1	0.4	0.2	0.6	1.2	n.d.	n.d.	n.d.	n.d.	n.r.	n.r.	n.r.
	1998	*m*	0.2	12.4	380	18.4	12.3	19.4	6.5	9.1	< 0.01	< 0.03	< 0.03	< 0.01	n.r.	n.r.	n.r.
		S.D.	0.1	0.8	19	2.2	0.6	0.5	3.2	0.9	n.d.	n.d.	n.d.	n.d.	n.r.	n.r.	n.r.
VI	1995	*m*	3.8	11.2	332	3.9	11.3	18.4	10.5	7.6	0.08	0.04	0.08	0.13	30692	3318	313
		S.D.	4.9	6.4	18	1.2	1.9	1.1	1.4	2.1	0.09	0.03	0.03	0.05	35371	3172	318
	1996	*m*	4.4	12.0	328	3.6	10.9	18.2	10.4	6.5	0.05	0.05	0.06	0.12	27350	3379	418
		S.D.	7.3	5.9	36	0.9	2.1	1.9	1.4	1.4	0.05	0.04	0.03	0.03	35539	4602	706
	1997	*m*	1.4	12.5	316	3.8	11.0	17.9	10.0	5.6	0.07	0.04	0.18	0.15	35808	5400	323
		S.D.	0.7	5.8	22	0.5	1.2	1.2	1.8	1.6	0.06	0.03	0.27	0.04	84040	14075	480
	1998	*m*	26.3	12.7	318	3.6	10.3	17.8	10.7	6.9	0.06	0.04	0.08	0.17	28283	2183	190
		S.D.	55.3	6.3	20	1.0	1.8	1.4	1.8	3.2	0.04	0.01	0.05	0.06	41353	2440	226
IS	1995	*m*	4.3	10.1	264	2.8	8.0	14.9	10.9	3.4	0.04	< 0.03	0.05	0.11	47250	7925	514
		S.D.	6.2	4.9	20	1.5	1.3	0.8	2.3	0.5	0.04	n.d.	0.02	0.03	35428	6977	492
	1996	*m*	4.3	10.0	264	2.8	8.0	14.9	10.9	3.4	0.04	< 0.03	0.05	0.08	47250	7925	514
		S.D.	6.2	4.9	20	1.5	1.3	0.8	2.3	0.5	0.04	n.d.	0.02	0.02	35428	6977	492
	1997	*m*	1.5	11.3	264	2.8	7.0	15.0	10.8	3.3	0.03	0.04	0.06	0.11	34583	9267	518
		S.D.	0.8	4.6	30	1.3	1.9	1.1	1.4	0.4	0.04	0.02	0.03	0.05	41649	18554	319
	1998	*m*	7.0	11.0	247	1.7	6.0	14.6	11.4	3.6	0.02	< 0.03	0.04	0.10	26500	3367	429
		S.D.	8.4	4.5	20	0.6	1.3	0.9	1.3	0.6	0.02	n.d.	0.02	0.03	20774	3589	550

Measure units are reported in Experimental section.

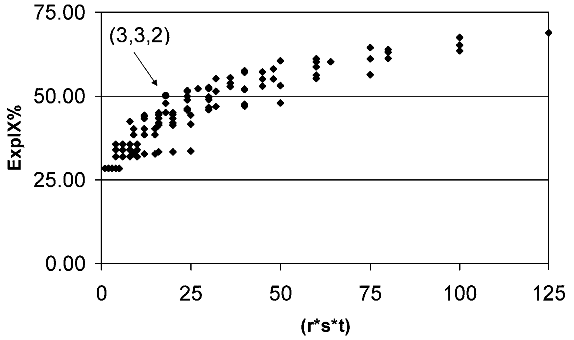

Fig. 2. Scatter plot of modelled sum of squares (%) as function of the product of the number of components in different modes for the considered Tucker models. SS of the model (3,3,2) components is indicated by an arrow.

For this purpose, we have performed an autocorrelation study on the values of factor B2. The autocorrelation coefficients [22] for different lags are plotted in Fig. 6 together with the confidence limits. Autocorrelation coefficients have significant positive values for lags corresponding to a delay of 12 ± 1 months, while significantly negative ones for 6 ± 1 months: all confirm the annual periodicity for our data set.

The environmental meaning of all these factors will be discussed commenting the parameter–temporal–spatial interactions related to the core elements of the (3,3,2) model, which are reported in Table 2. The percentage importance of each core-array element is also displayed in the same table. We see here that the maximum importance (55.67%) for interpreting the variance that the model is able to explicate is associated to (i) the (A1, B3, C1) core elements. They follow — as of importance — (ii) the (A3, B1, C2) core elements (15.52%), (iii) the (A2, B2, C2) elements (9.52%), (iv) the (A3, B3, C2) elements (6.85%), and (v) the (A2, B2, C1) elements (5.85%). The remaining core elements show rapidly decreasing importance: more than half of them have importance smaller than 1 ‰.

With regard to the first (55.67% as importance) core array (A1, B3, C1), A1 (first component characterising the analytical parameters) has positive loadings for TURB, DO, NO_2, NH_3, PO_4, UV-ORG, COLI, FAEC, and STREPTO for parameters all referable (with the exception of DO) to bad water quality, while it has negative loadings for TEMP, COND, Cl, SO_4, HARD, NO_3 (see Fig. 3). B3 (third component characterising the temporal variations) has loading values practically constant along all the 48 considered months (see Fig. 4), indicating scarce variability in time for this core element. Finally, C1 (first component describing the site variations) has positive loadings for Isonzo and Vipacco rivers, while negative ones for the typically karstic waters, for Timavo river and all the wells, with a minimum value for the deepest one NS (see Fig. 5). Therefore, the main systematic variation of the data set, emerging from this most important core element, appears correlated to the contrast between the surficial rivers (Isonzo and Vipacco) on one side and all the other 'Timavo-like' freshwaters on the other side. The surficial rivers have lower salinity, but are more turbid, more oxygenated, and more exposed to contamination by nitrites, ammonia, phosphates, organic constituents, coliforms, and faecal streptococci with respect to the hypogeous Timavo River and the karstic wells.

The (A3, B1, C2) core elements—second in importance (15.52%)—has A3 'parameter' component with positive loadings for TURB, TEMP, COND, Cl, SO_4, HARD, NO_3, NO_2, NH_3, PO_4, UV-ORG, COLI, FAEC, STREPTO, while negative values for DO. The B1 'temporal' component has negative loadings along the cold months and occasional positive ones in December 1995, September 1996, July 1997, and April 1998 (see Fig. 4). The C2 'spatial' component displays a very high value for Vipacco River only. This data set variation, characterised by the (A3, B1, C2) components, can be referred to turbidity phenomena in Vipacco River (which has torrent-like behaviour) occasionally producing waters of poor quality. The high A3 loadings are related to high concentra-

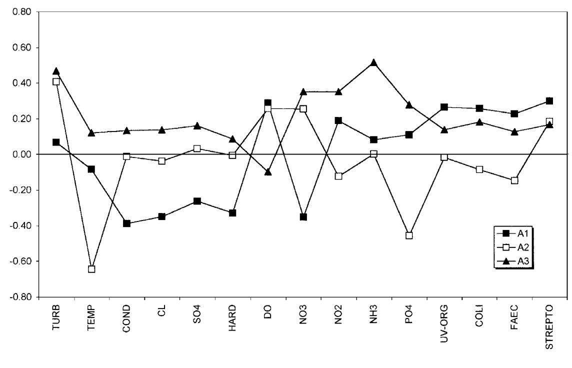

Fig. 3. Three factors of chemical–physical–biological parameters.

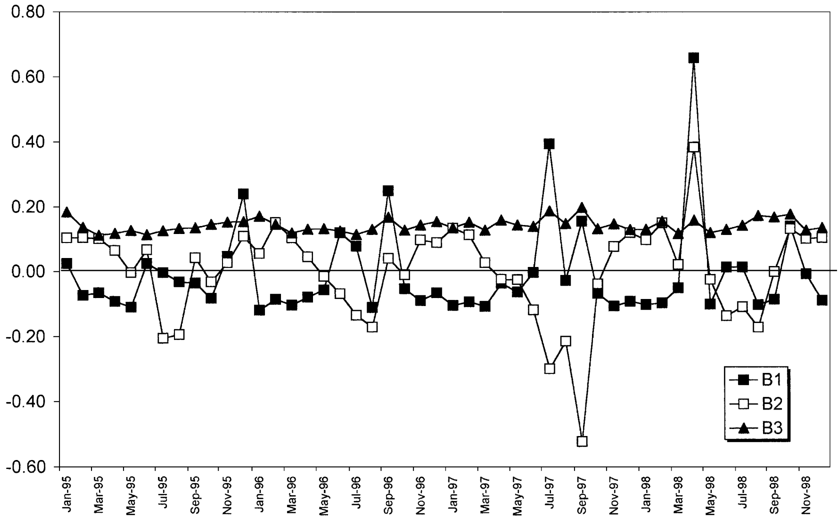

Fig. 4. Three factors of temporal parameters.

tions of solutes as nitrates, nitrites, ammonia, phosphates, chlorides, and sulfates, together with coliforms and streptococci, partly due to leaching of soils and partly to sewage runoff.

The third and fourth core elements are, as the second one, referable to the peculiarity of Vipacco rivers, since they share the same spatial component C2. In particular, the (A2, B2, C2) core elements (9.56%) has A2 'parameter' component with high TURB, DO, NO_3, and low TEMP, PO_4. The B2 'temporal' component shows positive values for cold months, while negative ones in warm months. Exceptional values correspond to April 1998 (spring flood) and September 1997 (end summer dry period): soil leaching produces turbidity and high concentration of nitrate fertilisers. The fourth (A3, B3, C2) core element has positive values for all parameters (particularly high for ammonia and turbidity) and a negative value only for dissolved oxygen. B3 has positive, fairly constant values suggesting that Vipacco River constitute a steady source of contamination, not only during critical (B1) or seasonal (B2) episodes.

The fifth core element (A2, B2, C1) share the same 'parameter' and 'temporal' components with the third one. These were interpreted as representing an agricultural pollution typical of cold, rainy months. The 'spatial' component C1 stresses, this time, the contrast

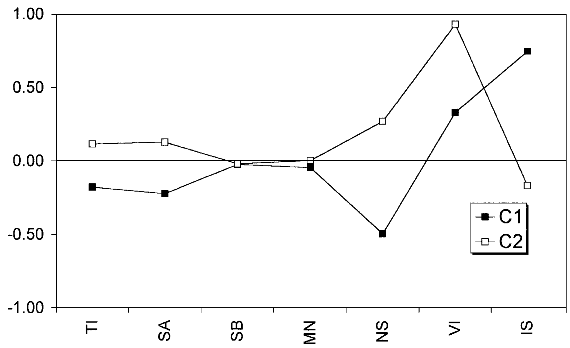

Fig. 5. Two factors of spatial parameters.

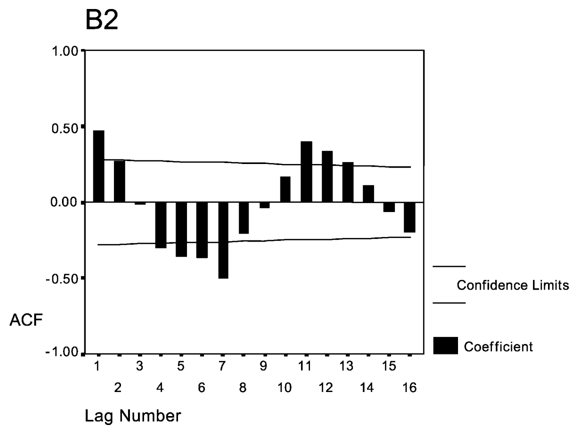

Fig. 6. Autocorrelation coefficients for the temporal component B2 and their confidence limits.

Table 2
Core-array elements of the (3,3,2) model obtained by 'variance of squares' rotation method

	C1	C1	C1	C2	C2	C2
	B1	B2	B3	B1	B2	B3
A1	1.13	−0.11	**37.25**	7.56	−2.43	1.44
	(0.05)	(0.00)	(**55.67**)	(2.29)	(0.24)	(0.08)
A2	−0.38	**12.08**	−0.18	−0.62	**15.40**	−1.62
	(0.01)	(**5.85**)	(0.00)	(0.02)	(**9.56**)	(0.11)
A3	8.98	−3.30	1.04	**19.70**	1.11	**13.10**
	(3.23)	(0.44)	(0.04)	(**15.52**)	(0.05)	(**6.85**)

The percentage importance of each core-array element is displayed in brackets. Bold characters refer to the five element cores discussed in the text.

between surficial rivers (Vipacco, but also Isonzo) and the high quality water reservoirs of NS.

At this point, some comments can be useful regarding the benefits of the multiway approach with respect to more conventional two-way factorial techniques, where data are organised in a table whose columns represent measured parameters, while rows represent samples that are collected at different times and sites. Several two-way principal component analysis and principal factor analysis models have also been built for the considered data set. For comparison purposes, a model can be considered, extracting three Varimax rotated [22] principal factors, which succeeds in finding reasonably interpretable latent variables, explaining 58% of the total variance on the whole. These new variables represent, in decreasing order of importance, variation of water composition due to (a) bacterial and chemical contaminations (high loadings for COLI, FAEC, STREPTO, NH_3, NO_2, PO_4), (b) change of ionic contents (high loadings for COND, Cl, SO_4, HARD, NO_3), and (c) influence of meteorological events on the water parameters (high loadings for TEMP (with negative sign), DO, TURB). The graphic inspection of factor scores allows the discussion of both site specific and temporal variations of the water composition. Such scores indeed mix the temporal and spatial information on samples that should be isolated a posteriori by separate visualisations. Scatter plots and sequence plot of the scores of three factors for the seven sampling sites (not reported here, for the sake of brevity) should be examined in order to gain a detailed insight in the information content of the data set. Most of the results obtained by considering the output of the three-way PFA have been discovered as well as with two-way PCA; however, the number of scatter plots and sequence plots to be considered in the two-way case is huge: three bivariate scatter plots (PF1 vs. PF2, PF1 vs. PF3, PF2 vs. PF3) and three sequence plots (PF1 vs. time, PF2 vs. time, PF3 vs. time) for each of seven sites sum to 42 plots. On the contrary, the three figures numbered as 3, 4, and 5 in this paper, plus the core matrix reported in Table 2, are sufficient to discuss the proposed three-way model. Moreover, some of the features clearly obtained by our Tucker-3 model—for instance, the separation of nonperiodic, periodic, and constant temporal factors and the exceptionality of the VIPACCO site—result only implicitly in the two-way models.

On the whole, the most important benefit of multiway PCA on data collected with a multiway design is the clear identification of factors characterising each of the way designed for the data collection. This is not possible in applying conventional two-way PCA to the same data since two or more ways result to be mixed.

4. Conclusions

- This study has shown a factor analysis procedure suitable for an exploratory analysis of a set of data describing variations of 15 freshwater quality parameters, sampled monthly for 4 years, at seven sampling sites.
- The visual inspection of the scatter-plot reporting explained variation of data vs. the product of the number of components in the three modes retained as indicator of complexity for 125 models has supported the choice of three factors for characterising the quality parameters, three for their temporal variations, and further two for their spatial pattern.
- The method optimising the "variance of squares" of the core elements has allowed a meaningful and simple interpretation of the Tucker-3 solution for the (3,3,2) model.
- In particular, the procedure succeeded in decomposing nicely the overall temporal variations in three parts, thus highlighting nonperiodic critical events, a periodic seasonal component, and a constant term.
- The seasonality has been confirmed by the study of the autocorrelation function of the dimensional component B2 of the Tucker-3 model, showing sig-

nificant positive coefficients for lags of 12±1 months and negative ones for lags of 6±1 months.

• A quantitative estimate of the environmental relevance of phenomena conditioning the considered water body, given by the squared elements of the rotated core and their environmental interpretations, has been proposed here.

• The main detected variations of the data set are correlated to a contrast between river waters, exposed to seasonal variations and to exceptional meteoric phenomena (as rainfall inducing turbidity and leaching of agricultural soils) on one side, and profound waters of the karstic wells on the other side.

• The Vipacco River is responsible for the highest variations in parameters since during episodic events (its floods), waters of poor quality appear.

• Besides the environmental considerations strictly connected to the present case study, in general terms, we remark the suitability of the proposed procedure to handle data coming from environmental monitoring programmes that are typically collected on the basis of a multiway sampling plan. The ability to detect periodic patterns of multivariate time series within a pool of sampling points that can be hindered by both anomalous events and site specific features constitute a relevant advantage of the procedure.

• Again, in general terms, while analysing environmental monitoring data that report about multiple measurements in several locations and sampling times, the output of the Tucker-3 principal factor analysis is relatively more compact than the one from classical two-way factorial techniques as PCA and PFA for detecting spatial and temporal patterns from the same data. Moreover, factors extracted by multiway models are explicitly linked to the modes used to identify the data and this fact can help the interpretation of numerical/graphical results.

• The statistical multivariate control of environmental dynamic systems, as the water system related to the municipal water supply here described, is one of the main aims of monitoring programmes. Multiway factorial techniques can play a role in this field, besides other system identification techniques [23]; however, some refinement of the presented modelling procedure are needed. Once the major sources of nonperiodical variation and eventual outliers in the multiway data structure had been identified, these should be removed, then reconstructing the variance–covariance structure filtered from anomalous events. Besides factorisation, the problem of outlier detection for such multiway data arises here. There is room for theoretical work on the subject. The filtered variance–covariance structure could be used for simulating the behaviour of the system in absence of critical events—possibly identifiable as pollution—and then for a further characterisation of regional typical values for the monitored parameters. Probabilities of the occurrence of natural nonperiodic events—as the alteration of quality standards due only to meteorological events—could be also estimated, thus stepping forward towards a true environmental control grounded on statistical bases. The understanding and rationalisation of modifications of environmental quality parameters is a process intrinsically dependent on the comprehension and visualisation of the multiway structure of the gathered information. All of these make the multiway chemometrical techniques very promising tools to be integrated in environmental decision support systems.

References

[1] E. Reisenhofer, G. Adami, P. Barbieri, Toxicological and Environmental Chemistry 54 (1996) 233–241.

[2] E. Reisenhofer, G. Adami, P. Barbieri, Water Research 32 (1998) 1193–1203.

[3] P. Barbieri, G. Adami, E. Reisenhofer, Annali di Chimica 88 (1998) 381–391.

[4] P. Barbieri, G. Adami, E. Reisenhofer, Annali di Chimica 89 (1999) 639–648.

[5] R. Henrion, Chemometrics and Intelligent Laboratory Systems 25 (1994) 1.

[6] P. Barbieri, C.A. Andersson, D.L. Massart, S. Predonzani, G. Adami, E. Reisenhofer, Analytica Chimica Acta 398 (1999) 227–235.

[7] V. Simeonov, P. Barbieri, B. Waltzak, D.L. Massart, Toxicological and Environmental Chemistry 79 (2001) 55–72.

[8] V. Librando, G. Drava, M. Forina, Annali di Chimica 88 (1998) 867–878.

[9] R. Leardi, C. Armanino, S. Lanteri, L. Alberotanza, Journal of Chemometrics 14 (2000) 187–195.

[10] R. Henrion, G. Henrion, G.C. Onuoha, Chemometrics and Intelligent Laboratory Systems 16 (1992) 87–94.

[11] A. Smilde, Chemometrics and Intelligent Laboratory Systems 15 (1992) 143.

[12] C.A. Andersson, R. Bro, Chemometrics and Intelligent Laboratory Systems 42 (1998) 93–103.

[13] Y.-L. Xie, P.K. Hopke, P. Paatero, L.A. Barrie, S.-M. Li, Atmospheric Environment 33 (1999) 2229–2239.

[14] G. Henrion, D. Nass, G. Michael, R. Henrion, Fresenius' Journal of Analytical Chemistry 352 (1995) 431–436.
[15] R. Henrion, C.A. Andersson, Chemometrics and Intelligent Laboratory Systems 47 (1999) 189–204.
[16] P.R. 236/88: Decree of President of Italian Republic, May 24, 1988, no. 236.
[17] A.D. Eaton, L.S. Clesceri, American Public Health Association, Water environment federation, American Water Works Association, "Standard methods for the examination of water and wastewater", Washington, DC (1995).
[18] P.M. Kroonenberg, Three Mode Principal Component Analysis, DSWO Press, Leiden, 1983.
[19] The MathWorks, Matlab 5.0, Natick, MA (1997).
[20] C.A. Andersson, R. Bro, N-way Toolbox; http://models.kvl.dk/srccode.html (1999).
[21] C.A. Andersson, R. Bro, Chemometrics and Intelligent Laboratory Systems 52 (2000) 1–4.
[22] 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.
[23] L. Ljung, System Identification: Theory for the User, Prentice-Hall, Upper Saddle River, NJ, 1999.