Modelling of the groundwater hydrological behaviour of the Langueyú creek basin by using N-way multivariate methods

Rosario Soledad Barranquero,[1]* Rafael Pardo,[2] Marcelo Varni,[3] Alejandro Ruiz de Galarreta[1] and Marisol Vega[2]

[1] *Centro de Investigaciones y Estudios Ambientales, Facultad de Ciencias Humanas, Universidad Nacional del Centro de la Provincia de Buenos Aires, Gral. Pinto 399, CP. 7000, Tandil, Buenos Aires, Argentina*
[2] *Departamento de Química Analítica, Facultad de Ciencias, Universidad de Valladolid, Campus Miguel Delibes, E-47011, Valladolid, España*
[3] *Instituto de Hidrología de Llanuras 'Dr. Eduardo J. Usunoff', Universidad Nacional del Centro de la Provincia de Buenos Aires, República de Italia 780, B7300, Azul, Buenos Aires, Argentina*

Abstract:

The groundwater hydrochemical behaviour of the Langueyú creek basin (Argentina) has been evaluated through a systematic survey, followed by application of hydrological and chemometric multivariate techniques. Ten physicochemical parameters were determined in groundwater samples collected from 26 wells during four sampling campaigns (June 2010; October 2010; February 2011 and June 2011), originating a tridimensional experimental dataset $\underline{\mathbf{X}}$. Univariate statistical and graphical hydrochemical tools (contour maps and Piper diagrams) applied to individual campaigns, allowed to reach some preliminary conclusions. However, a best visualization of the aquifer behaviour was achieved by applying Principal Component Analysis (MA-PCA) and N-way PCA procedures, Parallel Factor Analysis and Tucker3. Results were consistent with two-term models, being Tucker3 [2 2 1] the most adequate, explaining a large amount of the dataset variance (50.7%) with a low complexity. The first Tucker3 [1 1 1] interaction (38.2% of variance) is related with (i) calcium/magnesium *versus* sodium/potassium ion exchange processes; (ii) an increase of ionic concentration and (iii) a decrease of nitrate pollution, all processes along the direction of the groundwater flow. The second [2 2 1] interaction (12.5% of variance), accounts for the predominant role played by conductivity, bicarbonate and magnesium in the dataset. The seasonal variations are closely related to concentration/dilution phenomena originated by the variations of the phreatic levels, although this point will require additional sampling to establish a definitive hydrochemical model. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS   Langueyú creek basin; groundwater; hydrochemistry; N-way PCA; PARAFAC; Tucker3

Received 20 November 2012; Accepted 9 July 2013

INTRODUCTION

A quantitative assessment of the available water resources in the humid Pampean region (Argentina) could conclude that these are abundant, but factors such as quality, availability and accessibility may result in their differential geographical distribution, leading to their scarcity or total absence in some areas, with the negative implications for ecosystems and people affected.

Located in the Argentinean Pampa, the Langueyú creek basin endures hydrological problems because of the presence of the Tandil urban area. The town is located in the Southern extreme of the basin (See Figure 1) and extends for $50\,km^2$ on the Northern slopes of the Tandilian Hill System and on the valley itself. With a population of 111 483 habitants (Vapnarsky and Gorojovsky, 1990; INDEC, 2010), Tandil heavily depends upon groundwater for human consumption and for its economical activities, because there is no nearby available alternative water resource (Barranquero *et al.*, 2006; Barranquero *et al.*, 2008).

The basin has a great agricultural productivity, with a 79.3% of its surface devoted to extensive crop cultivation (wheat, soybean, corn and sunflower). Livestock-related and other agricultural uses (horticulture and fruit trees) cover 14.4 and 6.3% of the land, respectively, mainly in areas in which shallow or emergent bedrock material causes a poor development of the soil. Industrial activities related to metallurgical and food manufacturing cause additional aquifer depletion.

This heavy pressure on the limited groundwater resources makes indispensable the knowledge of the hydrology and hydrochemistry of the Tandil area. Systematic surveys of the aquifer must be carried out,

*Correspondence to: Rosario Soledad Barranquero, Centro de Investigaciones y Estudios Ambientales, Facultad de Ciencias Humanas, Universidad Nacional del Centro de la Provincia de Buenos Aires. Gral. Pinto 399 (CP. 7000). Tandil, Buenos Aires, Argentina.
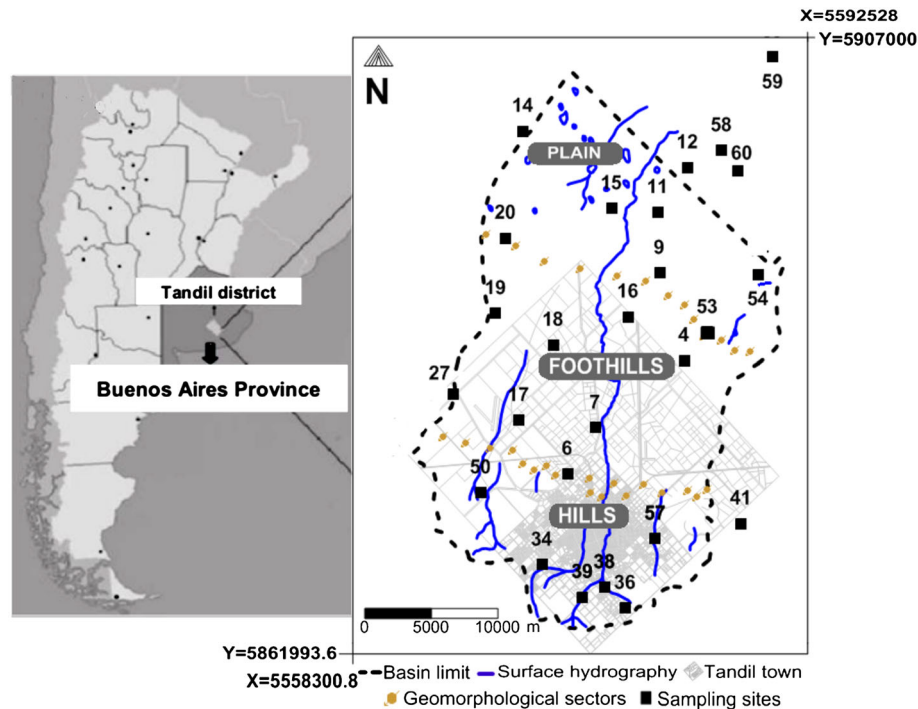E-mail: rosariobarranquero@yahoo.com.ar

Figure 1. Schematic view of Langueyú creek basin, showing the geomorphological sectors and the sampling sites

and from the interpretation of the relationships found amongst hydrochemical parameters, sampling sites and collecting times, the hydrological behaviour of the region can be elucidated.

Besides classical hydrological tools such as hydrochemical indices, contour maps or Piper diagrams (Catalán Lafuente, 1969; Custodio and LLamas, 1983; Appelo and Postma, 2005; Comisión Docente del Curso Internacional de Hidrología Subterránea, 2009), chemometric multivariate methods arise as powerful tools for these studies (Einax, 1995; Massart *et al*., 1998), because they allow to extract conclusions, not evident at a first glance, about the underlying latent variables that explain the variability existent in large datasets. Amongst these multivariate tools, Principal Component Analysis (PCA), designed for bi-dimensional datasets (e.g. *objects × variables*), allows to build simple linear models, providing simultaneously an easy visualization of the relationships amongst objects (groundwater samples) and/ or variables (physicochemical parameters) (Simeonov *et al*., 2003; Singh *et al*., 2004).

The use of multi-way methods such as N-PCA extends the scope of classic PCA to additional dimensions or modes and becomes indispensable when the dataset has a multidimensional structure (e.g. *objects × variables × time*) (Barbieri *et al*., 1999; Leardi *et al*., 2000; Smilde *et al*., 2004; Singh *et al*., 2006a; Peré-Trepat *et al*., 2007; Giussani *et al*., 2008; Pardo *et al*., 2008; Cid *et al*., 2011).

In this work, the hydrological behaviour of the Langueyú creek basin aquifer has been investigated through the interpretation of the spatial and temporal variations of 10 hydrochemical parameters (conductivity, pH, bicarbonate, chloride, sulphate, nitrate, calcium, magnesium, sodium and potassium), determined in groundwater samples collected at 26 wells, during four sampling campaigns in 2010–2011. Classical hydrological graphical methods and chemometric multivariate tools have been applied to visualize and interpret the information contained in the resulting dataset. This study is the first systematic hydrological survey carried out on the Langueyú basin and represents highly valuable information to be used as a baseline for future studies, as well as to assist to the development of more suitable management strategies of these limited groundwater resources.

MATERIALS AND METHODS

Area under study

The Langueyú Creek basin (Figure 1) is located in the Tandilia Hill System (Tandil County, Buenos Aires, Argentina). Its $600 \, km^2$ are divided in three major geomorphological sectors (Ruiz de Galarreta and Banda Noriega, 2005) characterized by the following features:

– Hills (Southern): ridges, isolated hills and well-defined stream valleys.

– Foothills: gentler slopes, some isolated depressions that interrupt the regularity of the relief, and a well-defined drainage network with a distributive design.
– Plains (North sector): gentle slopes, poorly defined drain network with narrow streams, temporary water courses and isolated depressions (Fidalgo *et al.*, 1975).

The basin comprises two hydrogeological units which behave differently with regard to the groundwater admission and circulation processes (Ruiz de Galarreta and Banda Noriega, 2005):

– A crystalline basement composed of tonalitic or granodioritic granitic rocks with different fracturing degrees, originating a cracked environment with secondary porosity and permeability that confers a poor character to the aquifer. The crystalline material emerges in the Southern hills, while in the Northern end, it deepens and forms the basis of the porous aquifer system.
– A Cenozoic sedimentary cover, which mainly consists of Pampean and post-Pampean sediments formed by loessic materials with high contents of calcium carbonate and other materials of fluvial origin.

The aquifer is unconfined (phreatic) and recharges through rainfall infiltration mainly in the hilly Southern sector. The depth of the phreatic surface ranges from 15 m at the South to 2 m at the Northeast. The map of isopiezic lines presented in Figure 2 shows that groundwater flows from South to Northeast in a consistent way with the
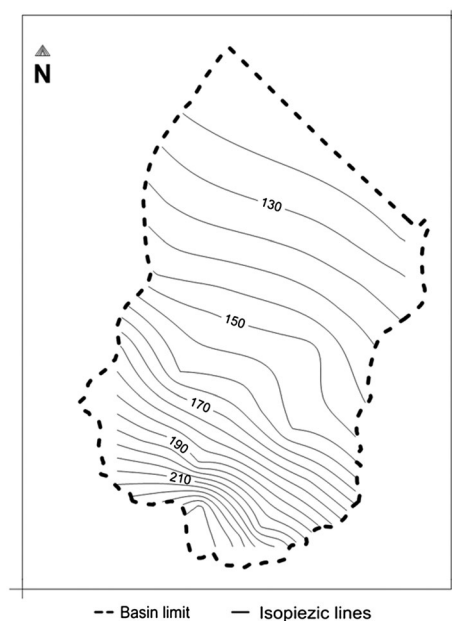
surface morphology, although with lower gradients that vary from 0.015 in the upper basin sector, to 0.0015 in the Northern sector (Ruiz de Galarreta *et al.*, 2007). The regional discharge of the aquifer occurs towards the depression of the Salado River and, locally, towards the perennial water courses and major tributaries of the Langueyú basin. In general, the aquifer behaves as influent with regard to the surface waters, although in some areas the ratio influent/effluent can reverse, because of changes in the hydraulic or geologic-structural conditions and/or of the intensive exploitation of the aquifer.

According to the classification proposed by Thornthwaite and Mather (1957), the area has a subhumid–humid mesothermal climate (Ruiz de Galarreta and Banda Noriega, 2005). The average annual rainfall is 838 mm, with actual and potential evapotranspiration mean values of 694 and 712 mm, respectively. The region presents a small water deficit of 18 mm during the rainy months of January, February and March (warm season) because of evapotranspiration, and an excess of water of 144 mm in the cold season (from May to November).

Sampling and analytical procedures

A preliminary and extensive sampling campaign, carried out between December 2006 and May 2007 (Ruiz de Galarreta *et al.*, 2007), allowed defining a network of 26 wells which were sampled in four successive campaigns: June 2010, October 2010, February 2011 and June 2011. The map of the sampling network is depicted in Figure 1.

Groundwater samples were collected from each well after 4 min of pumping out, kept at 4 °C and carried to the laboratory for analysis. Phreatic levels were also determined in all sampled wells. Ten physicochemical parameters were measured in the water samples: conductivity and pH, measured *in situ*, and bicarbonate, chloride, sulphate, nitrate, calcium, magnesium, sodium and potassium that were determined at the Laboratorio de Análisis Bioquímicos y Minerales (Facultad de Ciencias Veterinarias – UNICEN – Tandil, Argentina) and at the Laboratorio de Técnicas Instrumentales (Universidad de Valladolid, Spain) by using standard analytical methods (APHA, 2005).

Multivariate analysis

The resulting dataset has three dimensions or modes: ten physicochemical parameters or variables (*nvar*), determined in groundwater samples collected at 26 sampling sites (*nsites*), during four sampling campaigns (*ntime*). The resulting data can be arranged into a three-dimensional array $\mathbf{X}$ (*nsites* × *nvar* × *ntime*) or unfolded along the *nvar* dimension, into a bi-dimensional matrix

Figure 2. Hydrographic surface of the Langueyú creek basin

-- Basin limit — Isopiezic lines

$\mathbf{X^{aug}}$ ((*nsites* × *ntime*) × *nvar*) thus having 26 × 4 = 104 rows and ten columns. This j-unfolding procedure, called Matrix Augmentation, allows the application of classical PCA (MA-PCA) to $\mathbf{X^{aug}}$, whereas $\underline{\mathbf{X}}$ can be only studied by true n-way, N-PCA, procedures such as Parallel Factor Analysis (PARAFAC) or Tucker3.

Matrix augmentation PCA. This procedure is based on the decomposition of $\mathbf{X^{aug}}$ according to:

$$x_{ij}^{aug} = \sum_{f=1}^{NF} u_{if}^{aug} v_{jf} + e_{ij}$$

Where u_{if}^{aug} and v_{jf} are the elements of the scores and loading matrices $\mathbf{U^{aug}}$ and $\mathbf{V}$ with ((*nsitess* × *ntime*) × *NF*) and (*nvar* × *NF*) dimensions, respectively, and e_{ij} is the error term, or unexplained variance, of the element x_{ij}^{aug} of the $\mathbf{X^{aug}}$ data array. *NF* is the number of factors, kept as low as possible, but still explaining a significant amount of variance or information. Classic PCA allows to visualize the relationships between rows and columns of bi-dimensional matrixes, so the loadings v_{jf} will provide useful information about the relationships amongst the variables (e.g. physicochemical parameters), but the spatio-temporal information of the other two modes (sampling sites and campaigns) will appear mixed in the u_{if} scores, thus making difficult the interpretation of the model and limiting the usefulness of MA-PCA (Cid *et al.*, 2011).

N-way PCA (N-PCA). N-PCA methods do not require the previous unfolding of the three-dimensional dataset $\underline{\mathbf{X}}$. This allows for a full separation of the information corresponding to each of the three modes: sampling sites (spatial), physicochemical variables (hydrochemical) and sampling campaigns (temporal), thus facilitating the interpretation of the final model. The most used N-PCA algorithms are PARAFAC and Tucker3 that lead to slightly different models.

Tucker3 algorithm decomposes the matrix $\underline{\mathbf{X}}$ according to (Henrion, 1994):

$$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk}$$

Where a_{ip}, b_{jq} and c_{kr} are the elements of the loading matrixes $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ with (*nsites* × *P*), (*nvar* × *Q*) and (*ntime* × *R*) dimensions, respectively, carrying the information contained in the three modes of $\underline{\mathbf{X}}$. The term g_{pqr} denotes the element (p, q, r) of the core array matrix $\underline{\mathbf{G}}$ (*P* × *Q* × *R*), ideally a super-diagonal matrix (Leardi *et al.*, 2000), that accounts for the interactions amongst the three modes, whereas e_{ijk} is an error term as in MA-PCA. The squared g_{pqr} elements reflect the strength and significance of the interactions amongst the three modes, and *P, Q* and

R are kept again as low as possible, while explaining a significant amount of variance (Pravdova *et al.*, 2001).

The decomposition performed by the PARAFAC algorithm is (Bro, 1997):

$$x_{ijk} = \sum_{f=1}^{NF} a_{if} b_{jf} c_{kf} + e_{ijk}$$

The procedure also provides three loading matrices, $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$, with dimensions (*nsites* × *NF*), (*nvar* × *NF*) and (*ntime* × *NF*) whose elements are a_{ip}, b_{jq} and c_{kr}, respectively, while x_{ijk} and e_{ijk} have the same meanings as above.

PARAFAC can be considered as a constrained version of a Tucker3 model having the same number of factors, *NF*, for each mode and with all the superdiagonal elements of the core matrix $\underline{\mathbf{G}}$ equal to 1. Again, *NF* is kept as small as possible.

The interpretation of both models is somewhat different. Whereas PARAFAC can be interpreted in a similar way to PCA, the interpretation of the results of Tucker3 must also take into account the magnitude and sign of the non-zero elements of $\underline{\mathbf{G}}$ (Pardo *et al.*, 2008).

Software

MINITAB 13.0 (Minitab Inc.) and MATLAB 6.0 (The MathWorks Inc.) software were used for statistical calculations. PARAFAC and Tucker3 analysis were carried out with the N-way toolbox for MATLAB (Andersson and Bro, 2000). The contour maps were drawn with SURFER 9 (Golden Software Inc.).

RESULTS AND DISCUSSION

Descriptive statistics and hydrological preliminary study

Table I summarizes the descriptive statistics of the ten physicochemical parameters analysed in the 26x4 = 104 groundwater samples collected at the studied area. Some parameters, such as sulphate, nitrate, calcium, sodium and potassium, have variation ranges greater than one order of magnitude. Figure 3 shows the boxplots, disaggregated by sampling campaign, for all physicochemical parameters. Small temporal variations for all the parameters can be observed, as well as some very skewed distributions, e.g. the corresponding to potassium in the sampling campaign No.3 (February 2011).

The samples can be classified as 'fresh' and 'moderately hard' according, respectively, to their conductivity and 'hardness' levels (Custodio and Llamas, 1983) with 'hardness' levels below 400 mg.l^{-1} CaCO$_3$, the maximum allowed value for drinking waters stated by the Argentinian Food Code CAA (Código Alimentario Argentino) (n.d.). The concentrations of the natural anions always follow the

Table I. Statistical summary of the global dataset (N = 104)

Parameter[a]	Mean	Median	Standard deviation	Minimum	Maximum
Conductivity	820	800	154	480	1332
pH	7,59	7,58	0,26	6,89	8,40
Bicarbonate	485,1	481,9	93,2	268,4	732,0
Chloride	36,7	34,7	16,7	14,0	99,5
Sulphate	17,3	12,7	15,9	2,4	106,7
Nitrate	27,1	23,6	19,1	3,2	111,4
Calcium	40,2	35,8	20,7	10,3	115,7
Magnesium	21,8	20,5	8,9	8,2	45,8
Sodium	117,7	118,3	60,0	17,4	253,4
Potassium	13,5	9,8	13,8	1,3	81,0

[a] All parameters in $mg.l^{-1}$ except conductivity, in $\mu S.cm^{-1}$, and pH dimensionless.
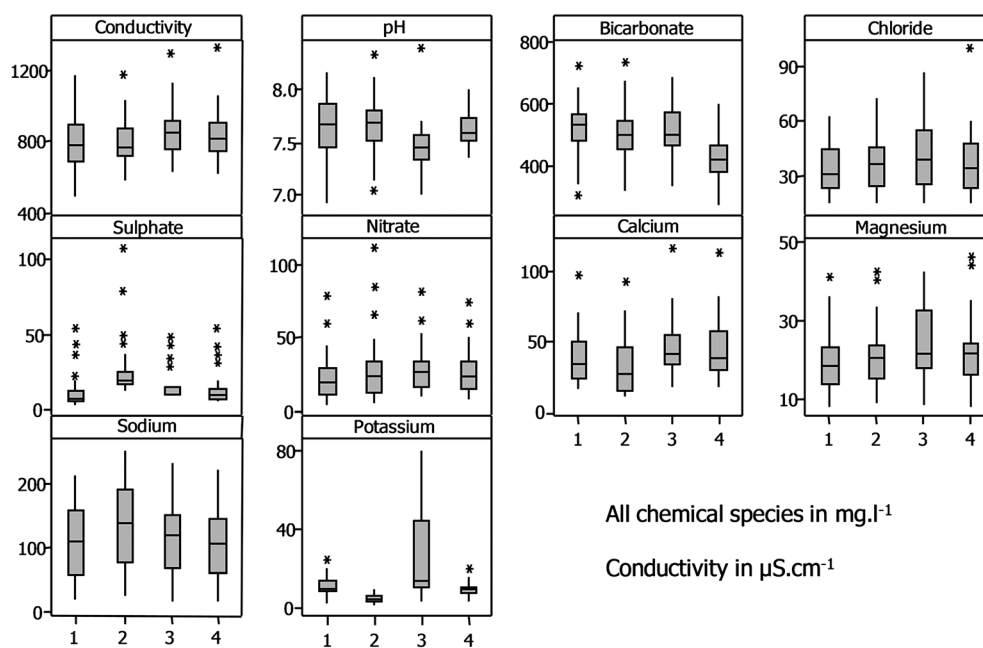

Figure 3. Boxplots of the physicochemical parameters as a function of the sampling campaign

sequence bicarbonate > chloride > sulphate; in the case of the study area, all these anions increase in the flow direction, because the North of the basin is not a regional discharge site and therefore the contact time with sediments is short. Bicarbonate is the most abundant anion, occurring at concentrations ranging from $298.9\,mg\,l^{-1}$ in well 39, located in the recharge area, to $713.7\,mg\,l^{-1}$ in well 14, located in the Northwest corner of the basin, these values corresponding to the June 2010 campaign. On the contrary, sulfate is the anion less abundant in all sampling sites. Most samples present nitrate contents below $50\,mg.l^{-1}$, the maximum allowed CAA level for this campaign.

The hydrological behaviour of the aquifer can be roughly described by considering separately the results corresponding to each single sampling campaign, although this approach leaves out the seasonal variations. Table II shows a statistical summary of the analytical results obtained in the June 2010 campaign. Normal probability distributions (according to the Kolgomorov–Smirnoff test) were found for most variables, except for sulphate and nitrate, whose distributions contain a large number of outlying values (see the 1-labelled boxplots depicted in Figure 3). Despite its usefulness, Table II gives insufficient information to allow extracting definitive conclusion about the aquifer behaviour, which can be only elucidated after studying the spatial distribution pattern of the physicochemical parameters as well as the relationships existent amongst them.

Table II. Statistical summary of the June 2010 dataset (N = 26)

Parameter[a]	Mean	Median	Standard deviation	Minimum	Maximum	p-value*
Conductivity	793	771	163	480	71177	>0.150
pH	7,63	7,66	0,31	6,89	8,15	>0.150
Bicarbonate	511,9	530,7	97,4	298,9	713,7	>0.150
Chloride	33,2	30,8	13,9	14,5	62,0	>0.150
Sulphate	12,2	7,4	13,0	2,4	53,2	<0.010
Nitrate	23,1	19,6	17,0	3,2	79,5	0.050
Calcium	38,4	33,9	18,5	16,4	97,2	>0.150
Magnesium	19,1	18,5	8,2	8,2	41,0	0.100
Sodium	106,6	110,4	54,1	21,4	212,7	>0.150
Potassium	11,3	10,5	4,9	2,9	24,6	>0.150

[a] All parameters in $mg.l^{-1}$ except conductivity, in $\mu S.cm^{-1}$, and pH dimensionless.
*Kolmogorov–Smirnoff normality test

Spatial variability can be studied by means of contour plots such as the displayed in Figure 4, corresponding to the June 2010 campaign and plotted by ordinary kriging, a geostatistical technique that interpolates the value of a given parameter at an unobserved location from observations at known nearby locations. By combining Figures 2 and 4, some patterns arise: the values of the variables conductivity, bicarbonate, chloride and sodium (and also pH, sulphate and potassium, not shown in Figure 4) increase along the flow direction, whereas nitrate and calcium (and magnesium, not shown) follow an opposite tendency.

The rNa^+/rK^+ index has a mean value of 16.9 and generally increases along the South–North axis, i.e. the groundwater direction flow. Although the Tandil area is characterized by basement rocks containing alkaline
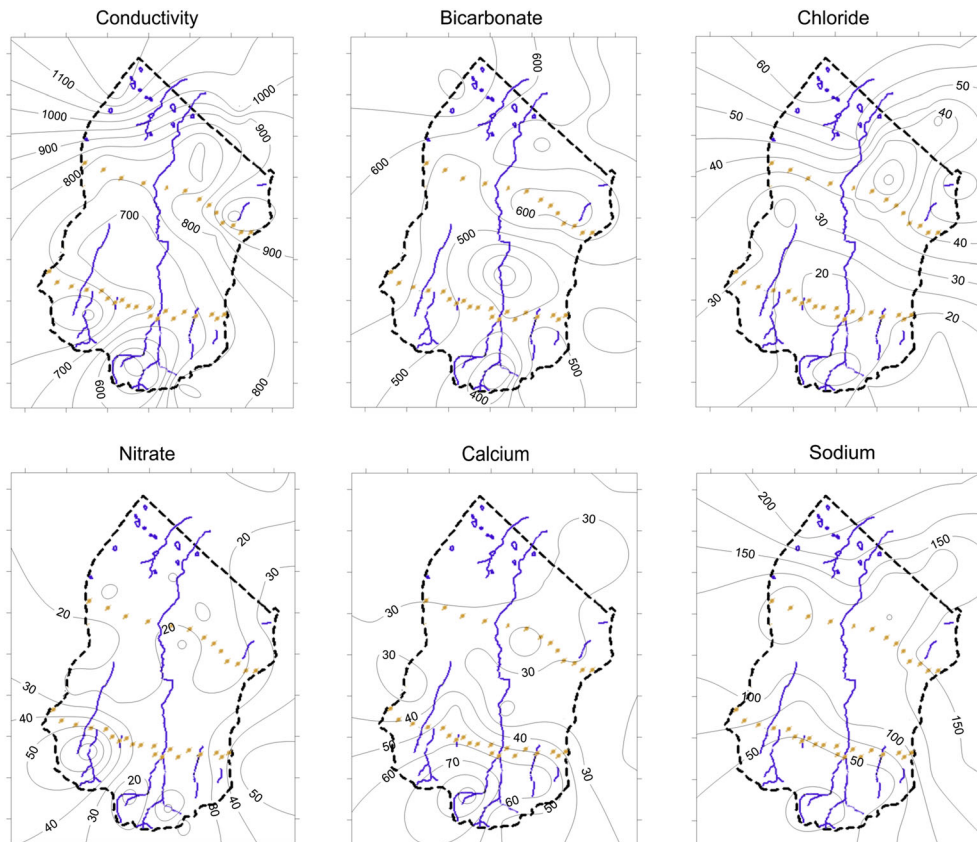

Figure 4. Contour maps of some selected physicochemical parameters for the June 2010 dataset

materials, mainly potassium (Teruggi *et al*., 1958), the investigated region lies on a porous layer of Pampean and post-Pampean sediments. In such cases, and especially when the clay contents are high, potassium becomes retained, originating rNa^+/rK^+ greater than 1.09, the value normally associated to igneous rocks (Catalán Lafuente, 1969).

The compositional characteristics of the aquifer can be observed in the Piper diagrams shown in Figure 5, corresponding to the June 2010 campaign, in which samples originating from the three sectors of the aquifer (hills, foothills and plains) have been plotted in different colours. According to these diagrams, based on the relationships existing amongst the major anions and cations, the groundwater samples can be classified as bicarbonated, either calcium/magnesium (most wells located in the hills) or sodium (rest of samples, located in the foot hills and in the plain). These compositional differences confirm (i) the opposing spatial tendencies

shown by calcium/magnesium *versus* sodium/potassium (Figure 4), and (ii) that the predominant anion is bicarbonate in all the collected samples. The gradual change in groundwater composition, from calcium-magnesium bicarbonated in the Southern sector towards sodium–potassium bicarbonated in the Central and Northern sectors, is also in good agreement with the hydrolithology of the studied area, characterized by a granitic basement that emerges in the Southern area and deepens towards the North, and a Cenozoic sedimentary cover, rich in calcium carbonate (See section 'Area under study').

Similar results were found for the rest of sampling campaigns.

MA-PCA

The above picture is incomplete because the effects caused by seasonal variations can only be ascertained by considering simultaneously the results from successive
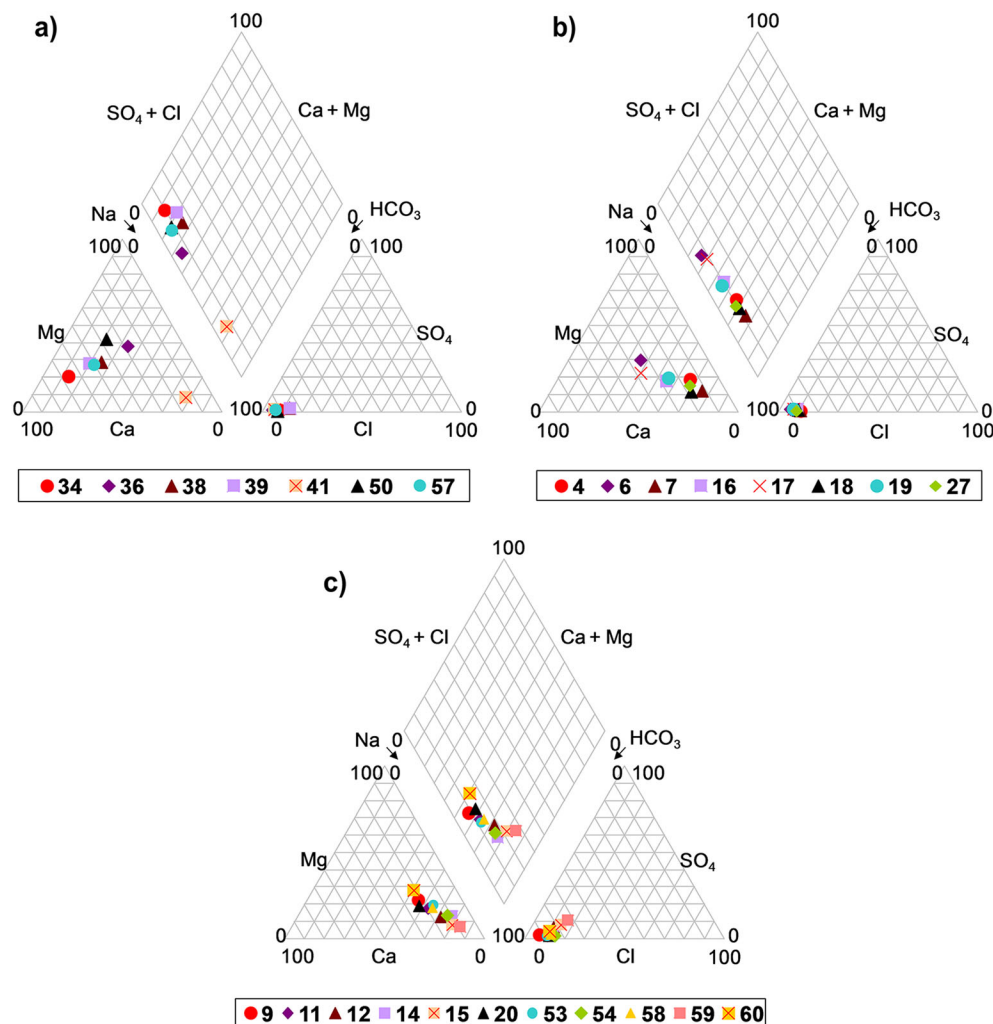

Figure 5. Piper diagrams corresponding to the June dataset a) Hills, b) Foothills and c) Plain

sampling campaigns, i.e. the complete $\underline{\mathbf{X}}$ dataset of dimensions ($26 \times 10 \times 4$).

First, the $\underline{\mathbf{X}}$ matrix was j-unfolded to give the $\mathbf{X}^{\mathrm{aug}}$, with dimensions ($104 \times 10$). This augmented bi-dimensional matrix was analysed by classical PCA using the eingenvalue decomposition of the correlation matrix, which compensates the differences in magnitude and measurement scales existent amongst the physicochemical parameters. In order to avoid the undesirable mixing of logarithmic and linear variables, pH was previously linearized into 'acidity'.

MA-PCA extracted three significant components (PC) with eigenvalues higher than unity. PC1, accounting for 40.5% of the variance, is composed of conductivity, bicarbonate, chloride, sulphate and sodium; PC2 spanning a 16.3% of variance is related with nitrate, calcium, magnesium and potassium and PC3 explains 11.4% of the variance and is composed of acidity and minor contributions of other variables.

Figure 6a shows the loadings plot of the two first PCs. Two main groups of variables are distinguished in the abstract space of the components, separated along the PC1 axis. The first group consists of conductivity, bicarbonate, chloride, sulphate, sodium, and potassium with positive PC1 loadings, whereas the second one is formed by acidity, nitrate, calcium and magnesium having negative PC1 values. PC3 plots confirmed these aggregations.

Figure 6b, which shows the scores plot for the two first PCs, reveals no clear pattern for the groundwater samples, beyond from a significant positive correlation (p 0.000) between the PC1 scores and the South–North axis of the studied area, resulting in an increase of PC1 along the flow direction of the basin. The large number of groundwater samples (104) and the inherent limitation of MA-PCA un-folding, that mixes the spatial (sampling sites) and temporal (sampling campaigns) information in the u_{ij}^{aug} scores (Pardo *et al.*, 2008; Cid *et al.*, 2011), does not allow to extract further conclusions. The refolding and averaging of the scores (Tauler *et al.*, 2000; Pardo *et al.*, 2008), an intermediate method used to separate information of mixed modes after MA-PCA, provided unsatisfactory results in this case.

N-PCA

Unlike MA-PCA, N-PCA models (Tucker3 and PARAFAC) take into account the three-dimensional structure of $\underline{\mathbf{X}}$, allowing to separate the information corresponding to each mode. The dataset was previously j-scaled, with all variables having zero mean and unit variance, thus overcoming, as in MA-PCA, the problems due to the different data magnitude and scales of the physicochemical parameters.

PARAFAC. All possible PARAFAC models were tested, and the *NF* value was selected by means of the core consistence parameter (Bro, 1997). The optimal complexity was found for a two-factor model (core consistency = 100%) explaining 50.9% (38.3% and 12.6%, respectively, for factors 1 and 2) of the total variance. Models with higher dimensionality had lower core consistence values and were consequently rejected (Bro, 1997).

Figure 7 shows a summary of the two-factor PARAFAC model, i.e. the loadings carrying the information of sampling points (**A**), variables (**B**) and time (**C**). The interpretation of PARAFAC models is based, in a similar way to classical PCA, on the existence (or absence) of correlations between the loadings of the three modes, found by comparing simultaneously, within each factor, the magnitudes and signs of the loadings of each PARAFAC mode.

Factor 1 presents positive and almost constant C1 loadings (Figure 7C1), so this factor will be large for two classes of sampling points (Figure 7A1). The first is formed by wells with positive A1 loadings (12, 14, 15, 20, 53, 54, 58 and 59) that are positively correlated with variables (Figure 7B1) having positive B1 loadings
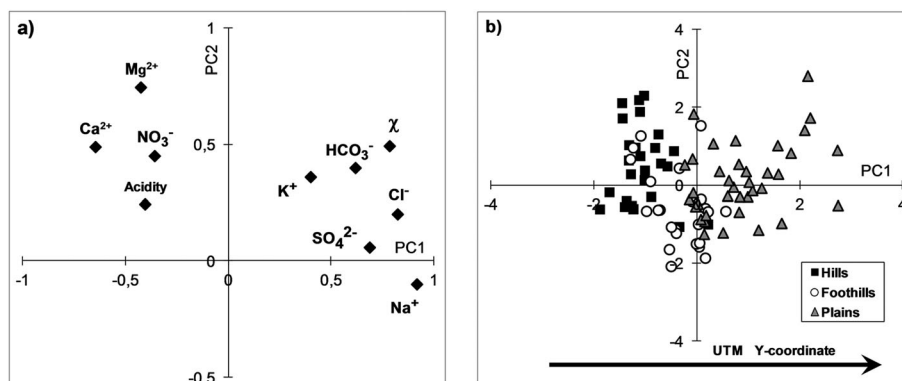

Figure 6. Graphical summary of MA-PCA applied to the augmented dataset $\mathbf{X}^{\mathrm{aug}}$. a) Loadings plot. b) Scores plot
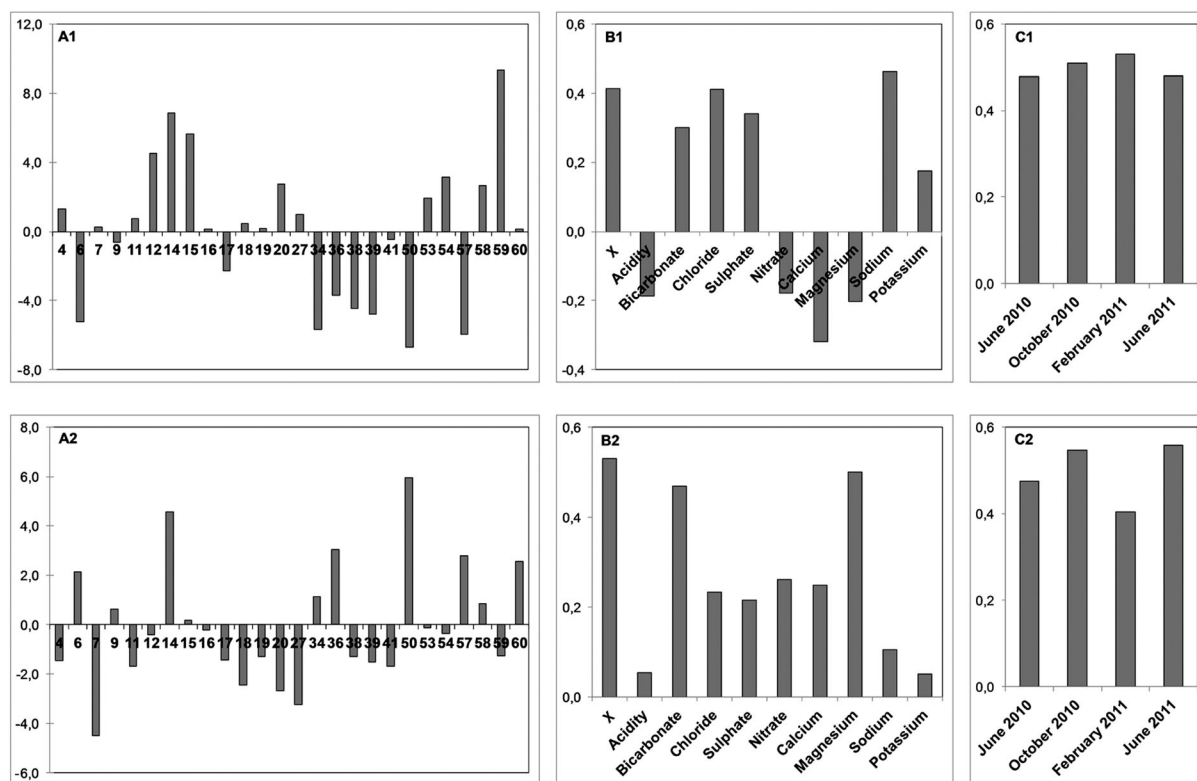
Figure 7. Loadings of the two-factor PARAFAC model in the three A, B and C modes

(conductivity, bicarbonate, chloride, sulphate, sodium and potassium). The second group (Figure 7A1) brings together sampling points with negative A1 loadings (wells 6, 17, 34, 36, 38, 39, 50 and 57) positively correlated to variables (Figure 7B1) also having negative B1 loadings (acidity, nitrate, calcium and magnesium). It is worth to add that B1 PARAFAC loadings have a pattern very similar to the PC1 loadings found in MA-PCA, thus confirming the strength of the relationships amongst the physicochemical parameters included in this factor. On the other hand, Factor 1 is not useful to interpret the hydrochemistry sampling points 4, 7, 9, 11, 16, 18, 19, 27, 41 and 60 as they have negligible or null A1 loadings.

In the case of Factor 2, C2 and B2 loadings resulted positive for all campaign and variables (Figure 7C2 and 7B2, respectively), so this factor will be significant only for points (Figure 7A2) having large positive A2 loadings (6, 14, 36, 50, 57 and 60). These points are related with all physicochemical parameters, mainly with conductivity, bicarbonate and magnesium. The simultaneous examination of the third mode loadings, C1 and C2, (Figure 7C1 and 7C2, respectively) shows their similarity, as well as their common apparent lack of relation with seasonality, thus suggesting a common behaviour of both PARAFAC factors with time.

Tucker3. All possible Tucker3 models having different number of factors in each mode were evaluated. Model [2 2 1] was chosen because it explains 50.7% of variance, i. e. a similar percentage than PARAFAC, with an only dimension in the C-mode, and has a superdiagonal **G** core matrix, thus fulfilling the mathematical conditions for Tucker3 models.

Figure 8 shows the loadings **A**, **B** and **C** of this [2 2 1] model, whose interpretation must also take in account the magnitude and sign of the elements of the resulting core matrix **G**:

$$-19.844 \quad 0.00$$
$$0.00 \quad 11.331$$

The interaction (as the factors are known in Tucker3 modeling) [1 1 1] accounts for a 38.2% of the variance. As all C1 loadings are positive (Figure 8C1) and g_{111} is negative, this interaction relates points (Figure 8A1) with negative A1 loadings (wells 12, 14, 15, 20, 53, 54, 58 and 59) and variables (Figure 8B1) having positive B1 loadings (conductivity, bicarbonate, chloride, sulphate, sodium and potassium), as well as wells (Figure 8A1) with positive A1 loadings (6, 17, 34, 36, 38, 39, 50 and 57) and variables (Figure 8B1) with negative B1 loadings (acidity, nitrate, calcium and magnesium). Therefore, this
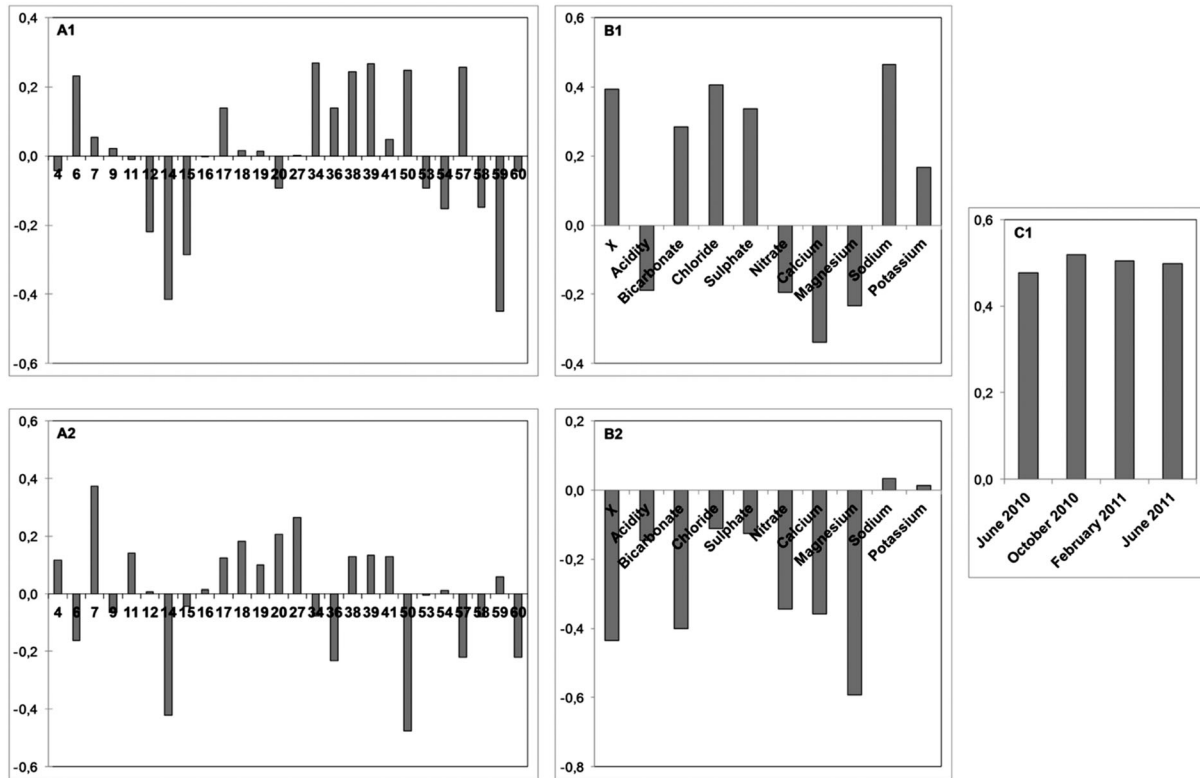
Figure 8. Loadings of the Tucker3 [2 2 1] model in the three A, B and C modes

[1 1 1] interaction is identical to PARAFAC Factor 1 and also presents the same B1 loading pattern found for the MA-PCA PC1 loadings (See section 'PARAFAC'). As before, there is a group of samples presenting small or negligible contribution for this interaction: 4, 7, 9, 11, 16, 18, 19, 27, 41 and 60.

The interaction [2 2 1] spans a 12.5% of variance and has positive values for both, C1 loadings (Figure 8C1) and g_{112}. Since all B2 loadings are negative (Figure 8B2) with the exception of the almost null contributions for sodium and potassium, this interaction will be only significant for points (Figure 8A2) with negative A2 loadings (6, 14, 36, 50, 57 and 60). This second Tucker3 interaction is again very similar to its equivalent PARAFAC Factor 2: A2 loadings are almost identical and B2 loadings are strongly biased towards conductivity, bicarbonate and magnesium in both models, with only slight differences for the remaining variables.

Loadings of the single temporal C-mode show again a very low dependence with season.

Both N-PCA algorithms led to very similar models consisting of two terms (factors or interactions) explaining around 50% of the dataset variance, a percentage similar to the explained by these modelling N-way tools when applied to environmental data (Leardi *et al.*, 2000; Singh *et al.*, 2006b; Cid *et al.*, 2011). However, Tucker3 model is simpler, needing only one

dimension in the temporal C-mode *versus* the two required by PARAFAC.

Hydrological interpretation of the multivariate models.

A coherent picture emerges from the preceding discussions. The hydrochemical parameters are grouped in two main conglomerates that (i) appear systematically in all multivariate models and (ii) behave in opposite ways. The first group is composed of conductivity, bicarbonate, chloride, sulphate, sodium and potassium, and the second one corresponds to acidity, nitrate, calcium and magnesium. These aggregations can be easily visualized in the loadings plot of the two first PCs obtained after applying MA-PCA to $\mathbf{X^{aug}}$ (Figure 6a), or in the almost identical structure of the B1 loadings found when N-PCA is applied to the original $\underline{\mathbf{X}}$ dataset (Figure 7B1 and 8B1, respectively).

The B1 loadings are associated, both in PARAFAC and Tucker3, in the same way with their corresponding A1 loadings (Figures 7A1 and 8A1), revealing now the existence of two groups of sampling points, also having opposing behaviours. The first group comprises wells 6, 17, 34, 36, 38, 39, 50 and 57, all located in the hills and foothills Southern sectors, and the second one is formed by wells 12, 14, 15, 20, 53, 54, 58 and 59 situated at the Northern plains sector, near the zone of discharge of the basin (Figure 1). The remaining wells (4, 7, 9, 11, 16, 18,

19, 27, 41 and 60) are in intermediate spatial positions and are not ruled by this first Factor or interaction.

The significant and positive correlations (at a p = 0.000 level) existent between the A1 loadings, from PARAFAC or Tucker3 models, and the UTM Northing coordinate of the sampling points, confirm the patterns depicted in the contour maps and/or in the Piper diagrams. Figure 9 shows the spatial variation of the A1 loadings in the studied region and also visualizes the strong association of this first factor/interaction with the flow direction of the aquifer. These A1 variations are simultaneously correlated with the increase of conductivity, bicarbonate, chloride, sulphate, sodium and potassium along the flow direction, and the decrease of acidity (i.e. an increase of pH), nitrate, calcium and magnesium, along this same South–North axis.

The opposing behaviour of calcium/magnesium *versus* sodium/potassium can be attributed to the existence of ion exchange processes in which calcium and magnesium, with concentrations above the average in the Southern zone, are replaced by sodium and potassium, thus resulting in a progressive decrease of calcium and magnesium and an increase of sodium and potassium from South to North. Additionally, as the Northern end of the study area was established based on the administrative boundary between Tandil and Ayacucho County, so the groundwater samples collected at the Northern sector are not yet fully chlorinated and still contain high levels of bicarbonate. This anion is the main cause, in combination with sulphate, chloride, sodium and potassium, of the subsequent increase of conductivity and salinity towards the Northern area, as well as of the decrease of acidity (increase of pH) in the same direction.

The behaviour of nitrate, a pollutant of anthropic origin, is similar to the followed by magnesium and calcium. The highest nitrate concentrations (in occasions near or above the legal limit of $50 \, \text{mg.l}^{-1}$) are found in samples collected at the Southern sector, near pollution point sources associated with the town of Tandil, and its contents decrease along the South–North axis in the flow direction.

The C1 loadings carry the information about the temporal or seasonal variability. The first striking characteristic is their almost constant value (Figures 7C1 and 8C1) and, therefore, their apparent lack of dependence from time. However and since this factor is related with hydrological processes connected with the aquifer flow direction, the relationship between C1 loadings and the mean phreatic level variations in each sampling campaign was studied. Negative correlations were found, r = −0.834 (p = 0.116) and r = −0.750 (p = 0.250) for PARAFAC Factor 1 and Tucker3 [1 1 1] interaction, respectively. This negative, although not statistically significant, correlation seems to suggest that C1 term could be associated to processes of saline concentration/dilution caused by variations in the phreatic level, thus confirming its hydrological nature. The lack of statistical significance of the correlations can be due to the small number of campaigns included in the dataset (and the subsequent small number of associated degrees of freedom) although this assumption can only be verified by carrying out further sampling campaigns.

To summarize, the PARAFAC Factor 1 and the Tucker3 [1 1 1] interaction contain information about hydrological processes occurring along the South–North axis in the direction of the flow. The first is natural and
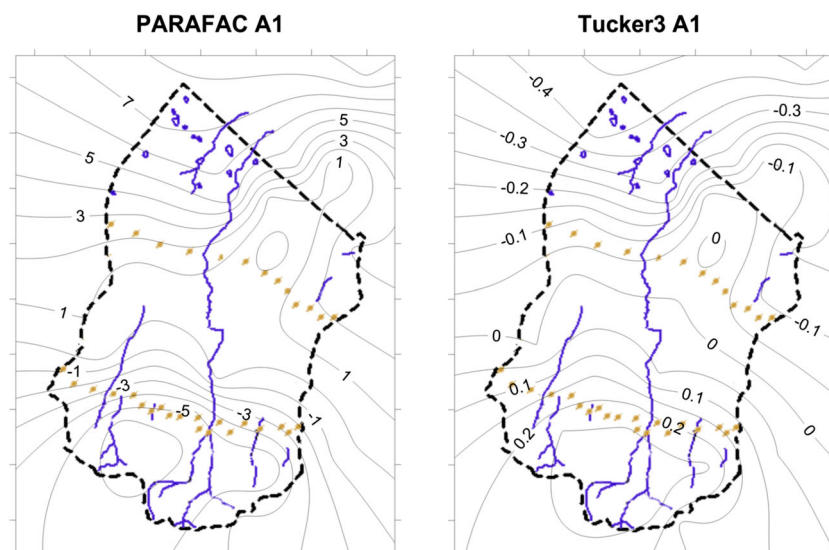

Figure 9. Contour maps of the A-loadings for a) PARAFAC and b) Tucker3 two-factor models

consists of ion exchange between calcium-magnesium *versus* sodium-potassium associated with a natural increase in salinity, whereas the second one is the decreasing of the man-made nitrate pollution. All of them are related with seasonal variations of the phreatic level, which cause the corresponding concentration/dilution phenomena.

On the other hand, the similarity between PARAFAC Factor 2 and Tucker3 [2 2 1] interaction, is less clear than in the previous case. The algebraic differences derived from the different number of dimensions required by both models in the temporal C-mode result in slight differences between their B2 loadings (Figures 7B2 and 8B2, respectively) which, however, share a similar pattern: strong contributions of conductivity, bicarbonate and magnesium and in lesser extension, of calcium and nitrate. In addition, the A-mode loadings of both models are almost identical (Figures 7A2 and 8A2, respectively) and show the existence of a group of sampling points composed of wells 6, 14, 36, 50, 57 and 60, which behave in a similar way. A closer examination of points 6, 14, 36, 50 and 57 shows a behaviour similar to that found for the first factor/interaction. Wells 6, 36, 50 and 57 are located in the Southern sector, where groundwaters are predominantly calcium–magnesium bicarbonated, whereas the point 14 lies in the Northern area, characterized by high conductivity and bicarbonate. The presence in this group of point 60, located in the North end sector, is likely due to its high levels of magnesium.

Since the models require different number of dimensions in the C-mode, larger differences in the interpretation of the temporal behaviour of this second factor/interaction can be now expected. The Tucker3 [2 2 1] interaction uses the same C-loadings than the [1 1 1], so the assertions made above about concentration/dilution phenomena and the variations of the phreatic levels are also of application in this case. On the other side, PARAFAC Factor 2 uses a new set of C2 loadings, somewhat different from those used by Tucker3 (Figures 7C2 and 8C1) and with a weaker correlation with the phreatic level variations (r = −0.103; p = 0.897). This fact makes difficult its interpretation and limits the usefulness of PARAFAC modeling in our study.

The stronger connection with seasonal water table variations leads to a more easy and logical interpretation of the temporal behaviour of this second Tucker3 [2 2 1] interaction, which, although cannot be associated with a definite hydrological process, emphasizes the leading role played by the seasonal and geographical variations of magnesium, bicarbonate and conductivity and in a lesser extension by that of chloride, sulphate and calcium in the studied region, and also makes the simplest Tucker3 [2 2 1] model as the most suitable to describe the experimental dataset.

CONCLUSIONS

The multivariate models applied in this study reveal two underlying factors/interactions in the dataset, leading to linear models with two terms, which explain respectively around 38% and 13% of the information present in the experimental dataset. Tucker3 [2 2 1] model has proven to be the most simple and efficient to explain the hydrological behaviour of the studied aquifer.

The first term summarizes (i) ion exchange processes amongst calcium/magnesium and sodium/potassium, which originate a gradual variation of the chemistry of groundwater samples following the direction of the groundwater flow, from calcium/magnesium bicarbonated in the Southern sector, to sodium/potassium bicarbonated in the North, in concordance with the hydrolithology of the region; (ii) an increase, along the same South–Northeast axis, of the concentrations of all major anions and of the conductivity and (iii) the singular behaviour of the pollutant nitrate that follows an opposite tendency. All these assertions are consistent with the information found by examining contour maps and Piper diagrams of individual sampling campaigns, as well as with the geology of the area.

The second term is not related with a definite hydrochemical process, but accounts for the role played by bicarbonate, magnesium, conductivity and in a lesser extension, nitrate and calcium.

The seasonal variations of the two terms could be only ascertained after using true N-PCA methods. Whereas PARAFAC modeling originated two different temporal behaviours difficult to interpret, Tucker3 led to a [2 2 1] model which is thus preferred to describe the global dataset.

This single seasonal variation is closely and negatively correlated (r = −0.750; p = 0.116) with the variations of the phreatic level in the sampled wells, so it accounts for saline concentration/dilution processes in the studied area. The lack of significance is due to the small number of campaigns carried out to date.

In conclusion, chemometric multivariate techniques are useful tools for the interpretation of hydrochemical datasets. Nevertheless, additional experimental work, through new sampling campaigns, will increase the amount of available information and will allow checking the validity of the seasonal patterns.

REFERENCES

Andersson CA, Bro R. 2000. The N-way toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems* **52**: 1–4. DOI: 10.1016/S0169-7439(00)00071-X. Retrieved from http://www.models. life.ku.dk/nwaytoolbox (access date: 2012/11/6).

APHA. 2005. *Standard Methods for the Examination of Water & Wastewater, (*21st Edition*)*, Eaton AD, Clesceri LS, Rice EW, Greenberg AE, Franson MAH (eds). APHA: Washington, D.C.

Appelo, C, Postma D 2005. *Geochemistry, Groundwater and Pollution*. A A Balkema Publishers: Rotterdam.

Barbieri P, Andersson CA, Massart DL, Predonzani S, Adami G, Reisenhofer E. 1999 Modeling bio-geochemical interactions in the surface waters of the Gulf of Trieste by three-way principal component analysis (PCA). *Analytica Chimica Acta* **398**, 227–235.

Barranquero R, Ruiz de Galarreta A, Banda Noriega R. 2006. Evaluación de nitratos en los pozos de explotación en la ciudad de Tandil, Buenos Aires, Argentina. In *VIII Congreso de ALHSUD*. Asociación de Hidrología Subterránea para el Desarrollo: Asunción, Paraguay; Actas de resúmenes, 29.

Barranquero R, Miguel E, Ruiz de Galarreta A y, Varni M. 2008. Influencia de la explotación local del recurso hídrico subterráneo sobre la hidrodinámica regional en Tandil, Buenos Aires, Argentina. In *IX Congreso Latinoamericano de Hidrología Subterránea*. ALHSUD: Quito, Ecuador; Volumen CD T-37.

Bro R. 1997. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems* **38**: 149–201. DOI: 10.1016/S0169-7439 (97)00032-4.

CAA, n.d. Código Alimentario Argentino, 1969. Ley 18.284/1969 and Decreto 2126/1971.

Catalán Lafuente J. 1969. *Química del agua*. Ed. Blume: Barcelona; 355.

Cid FD, Antón RI, Pardo R, Vega M, Caviedes-Vidal E. 2011. Modelling spatial and temporal variations in the water quality of an artificial water reservoir in the semiarid Midwest of Argentina. *Analytica Chimica Acta* **705**: 243–252. DOI: 10.1016/j.aca.2011.06.013.

Comisión Docente del Curso Internacional de Hidrología Subterránea. 2009. *Hidrogeología. Conceptos básicos de hidrología subterránea*, Publicado por FCIHS (Fundación Centro Internacional de Hidrología Subterránea): Barcelona; 768.

Custodio E y, LLamas M. 1983. *Hidrología Subterránea*, Tomos I y II (ed). Omega: Barcelona; 1157.

Einax JW. 1995. *Chemometrics in Environmental Chemistry*, Springer-Verlag: Berlín Heidelberg; 359.

Fidalgo F, De Francesco F y, Pascual R. 1975. Geología superficial de la Llanura Bonaerense. In *Relatorio Geología Provincia de Buenos Aires. VI Congreso Geológico Argentino*, Imprenta CONI S.A.C.I.F.I: Bahía Blanca, Buenos Aires; 104–106.

Giussani B, Monticelli D, Gambillara R, Pozzi A, Dossi C. 2008. Three-way principal component analysis of chemical data from Lake Como watershed. *Microchemical Journal* **88**: 160–166. DOI: 10.1016/j.microc.2007.11.006

Henrion R. 1994. N-way principal component analysis theory, algorithms and applications. *Chemometrics and Intelligent Laboratory Systems* **25**: 1–23. DOI: 10.1016/0169-7439(93)E0086-J.

INDEC. 2010. http://www.censo2010.indec.gov.ar/preliminares/ cuadro_resto.asp . Date of access: 21/12/2011.

Leardi R, Armanino C, Lanteri S, Alberotanza L. 2000. Three-mode principal component analysis of monitoring data from Venice lagoon. *Journal of Chemometrics* **14**: 187–195.

Massart DL, Vandeginste BGM, Buydens LMC, de Jong S, Lewi PJ, Smeyers-Verbeke J. 1998. *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier: Amsterdam; 876.

Pardo R, Vega, DL, Cazurro C, Carretero C. 2008. Modelling of chemical fractionation patterns of metals in soils by two-way and three-way principal component analysis. *Analytical Chimica Acta* **606**: 26–36. DOI: 10.1016/j.aca.2007.11.004.

Peré-Trepat E, Ginebreda A, Tauler R. 2007. Comparison of different multiway methods for the analysis of geographical metal distributions in fish, sediments and river waters in Catalonia. *Chemometrics and Intelligent Laboratory Systems* **88**: 69–83. DOI: 10.1016/j.chemolab.2006.09.009.

Pravdova V, Walczak B, Massart DL, Robberecht H, Van Cauwenbergh R, Hendrix P, Deelstra H. 2001. Three-way Principal Component Analysis for the Visualization of Trace Elemental Patterns in Vegetables after Different Cooking Procedures. *Journal of food composition and analysis* **14**: 207–225.

Ruiz de Galarreta A, Banda Noriega R. 2005. Geohidrología y evaluación de nitratos del Partido de Tandil, Buenos Aires, Argentina. In *Actas del IV Congreso Argentino de Hidrogeología y II Seminario Hispano-Latinoamericano sobre temas actuales de la Hidrología Subterránea*, Blarasin M, Cabrera A, Matteoda (eds). UNCR: Río Cuarto, Córdoba; 99–108.

Ruiz de Galarreta A, Varni M, Banda Noriega R, Barranquero R. 2007. Caracterización geohidrológica preliminar en la cuenca del arroyo Langueyú, Partido de Tandil, Buenos Aires. In *Actas V Congreso Argentino de Hidrogeología*, Díaz EL, Tomás JR, Santi M, D'Elía M y, Dalla Costa O (Compiladores). Asociación Internacional de Hidrogeólogos: Paraná, Entre Ríos; 119–128.

Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofoniou M, K Th. 2003. Assessment of the surface water quality in Northern Greece. *Water Research* **37**: 4119–4124. DOI: http://dx.doi. org/10.1016/S0043-1354(03)00398-1.

Singh KP, Malik A, Mohan D, Sinha S. 2004. Multivariate Statistical Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality of Gomti River (India) - A Case Study. *Water Research* **38**: 3980–3992. DOI: 10.1016/j.watres.2004.06.011.

Singh KP, Malik A, Singh VK, Basant N, Sinha S. 2006a. Multi-way modeling of hydro-chemical data of an alluvial river system-A case study. *Analytica Chimica Acta* **571**: 248–259. DOI: 10.1016/j. aca.2006.04.080.

Singh KP, Malik A, Singh VK, Sinha S 2006b, Multi-way data analysis of soils irrigated with wastewater–A case study, *Chemometrics and Intelligent Laboratory Systems* **83**: 1–12.

Smilde A, Bro R, Geladi P. 2004. *Multi-way Analysis with Applications in the Chemical Sciences*, Wiley, John Wiley & Sons, Ltd.: England.

Tauler R, Barceló D, Thurman EM. 2000. Multivariate correlation between concentrations of selected herbicides and derivatives in outflows from selected US midwestern reservoirs. *Environmental Science & Technology* **34**: 3307–3314.

Teruggi ME, Mauriño VE, Limousin TA, Schauer O. 1958. Geología de las Sierras de Tandil. *Revista de la Asociación Geológica Argentina* Vol. XIII (**3**): 185–204.

Thornthwaite CW, Mather JR. 1957. *Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance*, Laboratory of Climatology, Publication núm. 10: Centertown, N. J.; 185–311.

Vapnarsky CAy, Gorojovsky N. 1990. *El crecimiento urbano en la Argentina*, Grupo Editor Latinoamericano: Buenos Aires; 159.