A methodological approach for spatiotemporally analyzing water-polluting effluents in agricultural landscapes using partial triadic analysis (PTA)

J. J. Jiménez[1,*], N. Darwiche-Criado[1], R. Sorando[2], F. A. Comín[2], J. M. Sánchez-Pérez[3]

[1] *Pyrenean Institute of Ecology-National Spanish Research Council, IPE-CSIC, Av. Nuestra Señora de la Victoria s/n, 22700, Jaca (Huesca), Spain*
[2] *Pyrenean Institute of Ecology-National Spanish Research Council, IPE-CSIC, Av. Montañana 1005, 50080 Zaragoza, Spain*
[3] *Université de Toulouse, INPT, UPS, ECOLAB (Laboratoire Ecologie Fonctionnelle et Environnement), Ecole Nationale Supérieure Agronomique de Toulouse (ENSAT), Avenue de l'Agrobiopole BP 32607 Auzeville Tolosane, 31326 Castanet Tolosan Cx, France*

* Corresponding author

Instituto Pirenaico de Ecología (IPE-CSIC)

Av. Nuestra Señora de la Victoria s/n,

22700 Jaca (Huesca)

Spain

Telephone: +34 976 369393 (Ext. 881142)

Fax: +34 974 363222

Email: jjimenez@ipe.csic.es

ABSTRACT

Multivariate techniques for two-dimensional data matrices are normally used in water-quality studies. However, if the temporal dimension is included in the analysis, other statistical techniques are recommended. In this study, partial triadic analysis (PTA) was used to investigate the spatial and temporal variability in water-quality variables sampled in a northeastern Spain river basin. The results highlight the spatiality of the physical and chemical properties of water at different sites along a river over one year. PTA allowed us to clearly identify the presence of a stable spatial structure that was common to all sampling dates across the entire catchment.

Variables such as electrical conductivity (EC) and Na^+ and Cl^- ions were associated with agricultural sources, whereas total dissolved nitrogen (TDN), NH_4^+-N concentrations and NO_2^--N concentrations were linked to polluted urban sites; differences were observed between irrigated and non-irrigated periods. The concentration of NO_3^--N was associated with both agricultural and urban land uses. Variables associated with urban and agricultural pollution sources were highly influenced by the seasonality of different activities conducted in the study area. In analyzing the impact of land use and fertilization management on water runoff and effluents, powerful statistical tools that can properly identify the causes of pollution in watersheds are important. PTA can efficiently summarize site-specific water chemistry patterns in an applied setting for land- and water-monitoring schemes at the landscape level. The method is recommended for land-use decision-making processes to reduce harmful environmental effects and promote sustainable watershed management.

KEYWORDS: Water quality, Agricultural intensification, Statistical methods, Partial triadic analysis, Spatial analysis.

INTRODUCTION

Water quality is considerably dependent on anthropogenic activities and changes in land use and management practices, with agricultural land use as a primary determinant (Niemi *et al.*, 1990; Lenat and Crawford, 1994; Tong and Chen, 2002; Brainwood *et al.*, 2004). Water quality is also influenced by watershed runoff discharge (Caccia and Boyer, 2005; Zhang *et al.*, 2007) due to the excessive use of mineral fertilizers and manure that are not efficiently or timely applied. As a result, agricultural practices in watersheds are the primary sources of nutrients worldwide (Baker, 1992), particularly of nitrogen in European aquatic environments (Grizzetti *et al.*, 2005). Although point- and non-point-source pollution contribute to the degradation of water quality, the latter is much more difficult to attribute to a given source (Seeboonruang, 2012).

Causality is difficult to demonstrate because it requires unraveling the interactions of a wide range of human-influenced variables, such as land use, fertilizer application, soil type and hydrological pathways that link the land to a particular stream (Casey and Clarke, 1979; Dermine and Lamberts, 1987). Similarly, the temporal patterns of non-point-source pollutants can identify the transport characteristics associated with hydrological processes (Kang and Lin, 2007). Consequently, to improve agricultural management at the watershed scale with less harmful impacts on ecosystems, powerful statistical tools should be used to properly identify the causes of water pollution so that appropriate land use and fertilization measures can be suggested to end-users.

Multivariate analyses available for water-quality assessments at the watershed scale are generally two-dimensional (Singh *et al.*, 2004; Gourdol *et al.*, 2013). However, when sampling is repeated at the same sites on different dates, the resulting data matrix is three-dimensional, i.e., the sampling sites, variables and dates form a cube. These spatial and temporal dimensions are inherent in ecology when assessing ecological processes, although the classical statistical techniques that are used generally continue to be two-dimensional. Unfortunately, an incomplete picture is obtained of the multivariate space-time variation within the above-mentioned datacube (Thioulouse and Chessel, 1987). Commonly, conventional multivariate techniques have been applied to groundwater geochemistry studies (Güler *et al.*, 2002; Cloutier *et al.*, 2008). However, partial triadic analysis (PTA) has the potential to improve the interpretation of both spatial and temporal changes in geochemistry, with broad applications for assessing point-source and diffuse groundwater contamination (Gudmundsson *et al.*, 2011; Gourdol *et al.*, 2013). This spatio-temporal perspective is expected to improve decision-making in watershed management and to achieve more sustainable use of territories. Thus, the main objective of this work was to prove the usefulness of PTA in efficiently monitoring the common spatial and temporal structures of water-quality variability at the catchment scale in a Mediterranean area with increasing high-input agricultural practices.

METHODS

Study area

The Flumen River is located in Huesca Province in the north-central region of the large Ebro River Basin (NE Spain) (Fig. 1A). The river, which is 120-km long, and the Isuela River tributary drain a watershed area of 1,430 km^2. The Flumen River originates in "Sierra de Guara", a calcareous pre-Pyrenean mountain chain (1,250 m.a.s.l.) with forest and pasture cover; following the mountainous region of the basin, the river flows through flat, agricultural plains. In this final route, the river crosses quaternary glacis and alluvial fans that overlay a tertiary structure composed of conglomerates, sandstones and clays. Saline mudstones and gypsum deposits observed in the lower region of the basin influence the water quality of the lowest reaches of the river (Martín-Queller *et al.*, 2010). The Isuela River, which runs parallel to the Flumen River for one-third of its length, is the only perennial tributary; it joins the Flumen River in the flat area of the basin. Other tributaries include seasonal streams that permanently discharge water during the agricultural irrigation period (April-October).

The flow of the Flumen River is controlled by three reservoirs located in the upper third of the river, i.e., “Santa Maria de Belsué”, “Cienfuens” and “Montearagón”, with water storage capacities of 13, 1 and 51 hm^3, respectively. The Isuela River is regulated by the “Arguis” reservoir (2.7 hm^3), which is also located in the upper region of the river. Furthermore, in the lower half of the basin (Northern Monegros County), a complex network of irrigation canals distributes the water transported by a large irrigation canal, i.e., the Monegros canal, which is created by the confluence of two other large canals that transport water from the Cinca and Gállego Rivers (outside the Flumen watershed). The flow of the Flumen River is partly determined by irrigation water outside the watershed, being minimum in “Barbués” (center of the lower basin) and prominent in “Albalatillo” (Fig. 1A) (Sorando *et al*., unpubl.)

The climate over the entire study area is Mediterranean, with irregular seasonal and inter-annual rainfall (Comín and Williams, 1994). The average annual temperature and rainfall in the basin over the last 70 years were 10.5 – 13.9ºC and 987 – 402 mm, respectively, in the north and south regions of the basin (AEMET data). Annual precipitation in the basin during the study period was slightly lower for the south region of the basin, i.e. 372.6 mm (Fig. 2).

The upper region of the watershed is dominated by oak woods and shrublands (7.8% of the total watershed area), whereas the middle region is an urban-dominated area that includes the town of Huesca (52,354 inhabitants) and substantial animal farming and agricultural activity. The Isuela River passes through the town, which discharges effluents from the wastewater treatment plant into the river. Wastewater treatment plants are also found in the villages of Grañén and Lalueza (4,428 and 1,149 inhabitants, respectively) in the central region of the basin. In the lower area, irrigation is used for rice (Oryza sativa, 7.1% of the total watershed area), maize (Zea mays, 7.2%) and alfalfa (Medicago sativa, 13.2%), which are the most common crops. Cereals such as Triticum spp. (7.7% of the total watershed area) and Hordeum vulgare (32.5%) are also cultivated using only rainfall in the drylands along the margins of the lower region of the basin. Small and large farms and pig farming dominate all livestock husbandry in the region, particularly in the northern portion of Monegros County. Intensive pig production systems have considerably increased in the lower reaches of the basin (Martín-Queller *et al.*, 2010).

Sampling strategy

The sampling stations are distributed throughout a 120-km river network, which includes tributaries and the main stem of a river that travels through forested, urban, and agricultural areas (Table 1).

Water samples were collected directly by hand in weak acid pre-cleaned polyethylene bottles that were previously rinsed three times with distilled water. The bottles were filled with running water from the river, as far as possible from the edge of the river shoreline and avoiding big particles entering the bottles, and then transported to the lab in a cooler at 4ºC and later filtered with a fiberglass filter (Whatman GF/F 0.7 µm). We followed a two-step procedure based on previous studies which showed a clear differentiation of the water characteristics between two periods of the year mostly related with huge inflows of water exceeding agricultural irrigation (April-September) and non-irrigation period (October-March) (Martin-Queller et al. 2010; Darwiche et al. 2015):

Sampling group A: Water-quality measurements were collected at 15 sampling sites, i.e., nine along the Flumen River (F1-F9) and six in the Isuela River (I1-I6), on three dates in the non-irrigation season (November 2009, January 2010 and February 2010). These sampling sites were first selected based on the watershed origin (Flumen and Isuela Rivers) according to the major land use present in a given area, the presence of governmental gauging stations (three in the Flumen River and one in the Isuela River) and where point pollution was clearly identified, i.e., discharge from the wastewater treatment plant of Huesca. All data were used for PTA-1.

Sampling group B: After analyzing water samples from group A, 11 out of 15 sites were selected in the middle and lower regions of the basin according to their pollution potential. Samples were collected in 2010 in the following months: April, June, July, August, September and October. All data were used for PTA-2.

The rationale for performing two separate PTAs for sampling periods A and B was to first obtain a global baseline perspective of the Flumen Basin during the non-irrigation season for all sampling stations. Second, we focused on the increase in the water pollution during intensive irrigation in the lower region of the watershed (Table 1).

The water characteristics recorded *in situ* using portable calibrated electronic apparatus (YSI®ProPlus Multiparameter) were temperature (T), pH, electrical conductivity (EC) and dissolved oxygen (DO). Suspended solids (SS) were determined by means of gravimetric method, filtration through 0.45µm and drying out at 105ºC, difference of filter weight before and after filtering. Phosphate was determined spectrometically by colorimetry using the ascorbic acid method, total dissolved phosphorus (TDP), and total dissolved phosphorus (TDP) and total phosphorous (TP) as phosphate after acid digestion of, respectively, filtered and non-filtered water aliquots. Dissolved ammonium (NH_4^+-N), nitrite (NO_2^--N), nitrate (NO_3^--N), chloride (Cl^-), sulfate (SO_4^{2-}-S), sodium (Na^+), potassium (K^+), calcium (Ca^{2+}), magnesium (Mg^{2+}) concentrations were determined by ion chromatography (Metrohm 861 Advanced Compact IC). Dissolved organic carbon (DOC) and total dissolved nitrogen (TDN) were determined by catalytic combustion using a Multi-N/C 3100 analyzer (Analytik Jena®, Jena, Germany), fluoride (F^-), bromide (Br^-). The total phosphorous (TP) and the alkalinity (Alk) were determined using an unfiltered water sample. Alkalinity was determined by pH potentiometric automatic titration with H2SO4 (Metrohm®, Herisau, Switzerland). All the water variables were determined using standard methods described in APHA (1998) and Moreno-Mateos *et al.* (2008).

An initial graph that depicts the values of these variables shows dissimilarity among the sampling stations (Fig. 1B).

Statistical analysis: partial triadic analysis (PTA)

An assessment of watershed hydrological patterns primarily relies on multivariate statistical approaches, such as principal component analysis (PCA) (Valder *et al.* 2012) or other classical multivariate techniques (e.g., linear regression analysis, cluster analysis, or discriminant analysis). Also, repeated measures analysis is common in statistical analysis of time series data at multiple sites (multiple objects) with multiple explanatory data (Vonesh and Chinchilli, 1997). In combination with these, simultaneous analyses of both the spatial and temporal variability in longitudinal data can be achieved with PTA. When repeated measurements are performed within a spatial structure, PTA allows for depicting of temporal variability of the multivariable spatial structure and/or the spatial structure of the temporal trajectories (Rossi *et al*. 2014). This multivariate analysis was initially developed by Escoufier (1973) and was later integrated into the statistical method of ACT-STATIS ("Analyse Conjointe de Tableaux – Structuration des Tableaux à Trois Indices de la Statistique") by L'Hermier des Plantes (1976). PTA can be viewed as a particular simplified case of Tucker three-mode factor analysis (Tucker 1966). The first example was given by Jaffrenou (1978) and was later developed by Thioulouse and Chessel (1987), Thioulouse *et al.* (2004), Kroonenberg (1987; 1989), Dolédec (1988), Lavit (1988), Centofanti *et al.* (1989), Kiers (1991), Rossi (2003), Jiménez *et al.* (2006), Decaëns *et al*. (2009), Mendes *et al.* (2010), Marques *et al*. (2011), and Rossi *et al*. (2014).

In PTA, "partial" indicates simplified and "triadic" refers to the three-mode analysis (Kroonenberg 1989). Indeed, PTA is a PCA that is performed on data matrices with a triple-array or three-dimensional structure (Fig. 3), while the latter uses two-dimensional data matrices. This triple-array table is viewed as a sequence of two-way tables (Thioulouse *et al.* 2004). The objective of PTA is to define the common structure of several tables that share rows (sites or sampling points) and columns (variables), i.e., a datacube table that is expressed with different dates, in which only the main pattern described by the first axis is retained for interpretation (Rossi 2003). All tables, i.e., $\mathbf{X}_1$, $\mathbf{X}_2$, …, $\mathbf{X}_t$, contain observations of p variables measured at s sites at each of t times, e.g., tables for analyzing the spatial distribution of soil organisms and the temporal stability of the observed spatial pattern (Rossi, 2003; Jiménez *et al.*, 2006; Decaëns *et al.* 2009), spatio-temporally analyzing dynamic phytoplankton communities (Rolland *et al.* 2009; Bertrand and Mummy, 2010), assessing low and high flows (Gudmundsson *et al.*, 2011), and spatio-temporally analyzing hydrogeochemical parameters (Gourdol *et al.*, 2013).

PTA-1: This PTA refers to water-quality data from sample group A, i.e., from all stations during the low agricultural-activity period. Three tables for different sampling dates were used: November 2009, January 2010 and February 2010. Missing values of T, DO and SS due to multi-parameter probe malfunctions in November 2009 and January 2010 were treated with the "mice" package which allows filling empty cells in the data matrix by multivariate imputation by chained equations (Buuren and Groothuis-Oudshoorn, 2011). The statistical software "R" was used (R Core Team 2013).

PTA-2: Water-quality data that correspond to sample group B, which were collected during the intense farming activity period, were used to test the hypothesis that spatial and temporal trends of the analyzed variables from PTA-1 are similar. Six temporal data matrices were used: April, June, July, August, September and October 2010.

Each set of samples collected from sample groups A and B represents the same sites that were sampled at various times to analyze the same variables. All data matrices collected at all sampling dates were merged to perform the PTA. In other words, the data do not represent one sampling collection on one date but three and six dates for samples of groups A and B, respectively. The PTA reveals the spatial and temporal differences among the variables and sites.

Three steps characterize the PTA: interstructure, compromise and intrastructure.

<u>Interstructure</u>

The first step of the PTA determines the common information or structure present in the different matrices (Fig. 3A) using PCA. The objective is to provide a global description of the sampling points as a function of the typology of the sampling dates and to extract information common to all variables (Fig. 3B). This process provides a statistical analysis that is an alternative to the PCA application to several temporal matrices, and it has the advantage of grouping all sampling dates into the same PCA. Therefore, the temporal dynamics are not excluded with this statistical tool, in contrast to other methods in which only regular tendencies are analyzed. In other words, PCA analyses could be performed, one for each period (sampling dates t_1, t_2, …, t_n), but the factorial axes are different in each PCA for each temporal data matrix.

In addition, a duality diagram or statistical triplet (X, D_p, D_n) can be used to define a multivariate data analysis from a geometrical point of view, where X is the n × p table to analyze, D_p is the diagonal p × p symmetrical matrix (positive and definite) of column weights, and D_n is an n × n matrix of weights on the "observations".

The interstructure analysis is based on the concepts of vectorial variance Var_V, vectorial covariance Cov_V, and vectorial correlation coefficient R_V (Escoufier, 1973). The vector variance of table $\mathbf{X}_j$ is given by:

$$= tr()/p =) = p$$

where Var is the vectorial variance of table $\mathbf{X}_j$.

The vector covariance between tables $\mathbf{X}_j$ and $\mathbf{X}_k$ (duality diagrams) is the sum of the correlations between identical pairs of variables:

$$= tr()/p =)$$

where $Cov_v(X_j, X_k)$ is the vectorial covariance of tables $\mathbf{X}_j$ and $\mathbf{X}_k$.

A cosine matrix was first created to analyze the similar structures of the tables. Matrices $\mathbf{X}(k)$ and $\mathbf{X}(k')$ were normalized such that the sum of squares of their elements equals 1 and the inner product between both matrices equals the cosine between the matrices, i.e., the R_V coefficient, which is a measure of similarity between squared symmetric matrices (Escoufier 1973). Thus, the R_V coefficient between two tables ranges between 0 and 1, and it is equivalent to the r-correlation coefficient between two variables. This coefficient matrix allows for a comparison of the sampling dates and representation of the proximity between tables depending on the analyzed variables (Robert and Escoufier, 1976). Consequently, the R_V coefficient is the mean of the correlations between identical pairs of variables:

$$=$$

$$=$$

The calculation of the vectorial correlation coefficient matrix (R_V) between sampling dates allows for a comparison of the sampling dates and representation of the proximity between dates that depend on the analyzed variables. The function of this step is to assign a weight to each sampling date sub-matrix (ακ coefficients).

<u>Compromise</u>

The second step is the compromise analysis, which involves PCA of a new data table (compromise table) that results from reorganizing the variable-sample scores (Fig. 3C). This analysis involves construction of a mean matrix of maximum inertia (referred to as the compromise matrix). The compromise analysis allows for a multivariate synthesis of the information expressed through axis I (Compromise 1) of the date ordination analysis and provides an idea of the structures that are common to all tables and a simultaneous representation of individuals and variables (Thioulouse and Chessel, 1987; Bertrand and Mummy 2010). In other words, this analysis provides a description of the sampling sites as a function of the typology of variables and identification of the variables responsible for similar patterns at different dates (Fig. 3D). Therefore, our approach focuses on analyzing the spatial patterns and temporal variability/stability of the variables linked to water quality in agricultural effluents.

The compromise table provides the best summary properties of the initial temporal matrices. The R_V coefficient indicates the extent to which the compromise expresses the information contained in each sub-matrix (the cos² between a sub-matrix and the compromise table). Next, the matrix that represents the vectorial correlations between the different sampling dates and sub-matrices (R_V coefficient) provides an indication of the strength of the links among the different sub-matrices from the various sampling dates (Rolland *et al.*, 2009). This step describes the sampling sites as a function of the typology of variables and identifies the variables responsible for similar patterns on different dates (Jiménez *et al.*, 2006; Decaëns *et al*. 2009).

Intrastructure

The last step is known as the intrastructure or the reproducibility of the compromise. In this representation, the row and column loadings of all tables are graphically displayed on the first two principal components of the compromise matrix as additional elements (Thioulouse and Chessel, 1987). This step summarizes the variability in the series of tables around a common structure defined by the compromise, highlighting which elements best fit (or do not fit) the structure of the compromise. The rows and columns of all tables of the series are projected onto the factorial plane of the PCA of the compromise as additional elements (Thioulouse *et al.* 2004). The intrastructure analysis shows the departure of the spatial structure observed at each date from the spatial structure common to all sampling dates.

All of the computations involved in the PTA were directly processed with the Ade-4 package (Thioulouse *et al.* 1997) in R statistical software (R Core Team 2013).

Correlogram analysis

The resulting compromise table from the PTA can be used with spatially explicit statistical functions, such as the correlogram, the function that represents the spatial pattern of a given variable and its significance (Sokal and Oden 1978; Legendre and Fortin 1989; Overmars *et al.* 2003). The degree of spatial autocorrelation for each PTA was assessed using Moran's I (Moran 1948) spatial autocorrelation statistics (Cliff and Ord, 1981), which were computed using the positive and negative sample scores of the first axis of the compromise table (Decaëns and Rossi 2001). This procedure aims to reveal the degree of spatial autocorrelation in the common structure described by the PCA of the compromise table.

Moran's I index is given by:

$$I =$$

for $h \neq i$, where y_h and y_i denote the values of the observed variable at sites h and I, d is the distance class, and w is the weight.

These indices are plotted in a graph called a correlogram, which is used to quantify the spatial dependency of the variables per distance class or lag. Only the pairs of sites (h, i) within the stated distance class (d) are taken into account when calculating any given coefficient (Legendre and Legendre, 1988). Moran's I usually takes on values in the interval (-1, +1), although values < -1 or > 1 might be obtained. High values of I, either negative or positive, indicate strong autocorrelation (Legendre and Legendre, 1988).

The overall statistical significance is tested with the Bonferroni-corrected probability procedure (Oden 1984). The corrected P* is α' = α/k, where k is the number of distance classes, and $\alpha<0.05$ is the global significance level (Oden, 1984). The computation of spatial autocorrelation indices was performed with the ncf package in the R statistical software (R Core Team, 2013).

RESULTS

A statistical summary of all of the data and variables is provided in Table 2. The variability in our data was within that observed using historical data collected monthly at the gauging stations during 2007-2014 (Fig. 4).

Spatio-temporal pattern at the watershed scale

A common pattern was detected across the different sampling dates during the study period. Two axes accounted for 73.6% and 14.2% of the total data variability in the interstructural analysis of PTA-1. The representation of the eigenvectors in Euclidean space revealed that all sampling dates displayed positive scores on axis I, indicating the presence of a structure common to all dates (Fig. 5A). This so-called "inter-table size effect" indicates that no inversion of the temporal structure of the analyzed variables occurred.

This distribution was consistent with the R_V coefficient matrix (Table 3a, Fig. 5B), which exhibited a strong correlation (R_V = 0.631) between November and January, whereas the weakest correlation (R_V = 0.577) occurred between January and February. The matrix corresponding to November 2009 most greatly contributed to the temporal dynamics of the variables, as given by the highest value of $\cos^2$.

In the compromise analysis (Fig. 5C), the main spatio-temporal patterns of the variables were highlighted by extracting only the first two axes, which explained 80.7% of the total inertia of the PCA performed on the compromise matrix. The first axis (56.7% of the total variance) separated the variable SS (negative side of axis I) from the remainder of the variables on the positive side, whereas the second axis (24.0% of the total variance) was characterized primarily by pH and NH_4^+-N, as opposed to NO_2-N. Two main groups were distinguished with respect to the second axis, and those variables were related to salt and ion concentrations (EC, Ca^{2+}, Mg^{2+}, Na^+, SO_4^{2-}-S and F^-); NO_3^--N occupied the positive side of axis II, whereas variables related to nutrients (TP, TDP, PO_4^{3-}-P, TDN, NO_2^--N, NH_4^+-N, and DOC) and K^+ were grouped on the negative side of axis II (Fig. 5C). The distribution of the 15 sampling stations (Fig. 5D) in the factorial plane formed by the first two axes showed a clear separation between the stations located in non-urban areas and areas with non-intensive agricultural use (F1, F2, F3, F4, F5, I1 and I2), as opposed to the remainder of the sites along the negative side of axis I. Sampling stations I3 and I4, which correspond to urban contributions, slightly participate in the data structure; they were related to the variable "dissolved solids". Along the second axis, stations F8 and F9 were opposed to I5 and I6; the latter indicate sampling stations that received inputs from the Isuela River after passing through the city of Huesca and the wastewater treatment plant. These stations were related to the group of variables that consist of phosphorus compounds, NO_2^--N, NH_4^+-N and K^+ (Fig. 5C). Sampling stations F8 and F9 were located in areas with intensive agricultural use and were related to saline compounds and NO_3^--N.

In the last step, the intrastructure helps reveal which initial table fits the model expressed in the previous step. The original tables were projected as complementary tables onto axis I of the PCA of the compromise table. The structure was strongest in

November 2009, when the physical and chemical properties of the water indicated higher pollution compared with January and February 2010 (Fig. 6), when agricultural activity was lower. Sites located upstream (half of the negative side of the first axis) of the watershed are less polluted, whereas the stations located downstream (F8, F9, I5 and I6) that receive sewage and fertilization inputs have greater pollution.

Influence of agricultural use on the common spatial structure of water quality

In PTA-2, the analysis of the interstructure of the sampling dates that correspond to the sampling stations of the agricultural zone revealed similar measured-variable dynamics (Fig. 7). We retained the first two axes of the interstructure analysis, which explained 69.3% and 9.7% of the total variance. The representation of the eigenvectors in Euclidean space revealed that all sampling dates displayed positive scores on axis I of PTA-2, indicating the presence of a structure common to all dates (Fig. 5A). Again, the “intertable size effect” appeared, i.e., no inversion of the structure occurred from one date to another (Fig. 7A).

The R_V coefficient ($\cos^2$) is an indicator of the extent to which the compromise expresses the information contained in every table. According to the information in Fig. 7B, the sub-matrices for September, July, June and April (R_V = 0.737) substantially contributed to the definition of the compromise, whereas those of August and October had lower weights (R_V = 0.307) in the development of the compromise (Table 3b).

In the compromise analysis (Fig. 7C, D), two axes were retained that explained 70.9% of the total inertia. The first axis of the compromise PCA (Fig. 7C) explained 44.5% of the total inertia and revealed a clear organization of the variables related to human activities, i.e., electrical conductivity and nitrogen compounds (NO_3^--N and TDN) and salinity (Mg^{2+}, SO_4^{2-}-S and Alk) on the positive half of the first axis in opposition to SS (Fig. 7C). The second axis (Fig. 7C) accounted for 26.4% of the total inertia, and it was interpreted as a higher availability of nutrients, such as NH_4^+-N, PO_4^{3-}-P, and TP on the positive side of axis II, pH on the negative side of the axis and, to a lesser extent, the T, NO_2^--N, DOC concentrations and DO.

The map of the factorial coordinates of the 10 sampling stations (Fig. 7D) also revealed a clear spatial variation. Sampling stations located in the region of the basin with highly intensive agricultural use (IC1, F8, F9, and IC2) were linked to ions related to salinity (Na^+, Cl^-, EC, Br^-, and F^-), whereas sampling stations IW (the discharge from the wastewater treatment plant) and I6 (in the Isuela River downstream of the wastewater treatment plant) were associated with TDN and NO_3^--N and NH_4^+-N, PO_4^--P, TP and K^+, respectively.

Finally, the intrastructure analysis was achieved by projecting the six temporal matrices onto the factorial plane of the compromise (Fig. 8) to display their similarity with the common spatio-temporal pattern extracted in the PTA; in other words, the temporal stability of the process was analyzed only if the variables were projected in the same places in the factorial plane. Overall, the structure was strongest in July, August, September and October 2010, when the physical and chemical properties of the water indicated higher pollution compared with April and June 2010 (Fig. 8). The projection intrastructure analysis showed how the different variables contributed to the definition of the factorial axes for each sampling date. In all cases, stations F4 and F5 (negative half of the first axis) were displayed in opposition to sites F8, F9, IC1, IC2 and IW, which received larger inputs of fertilization and sewage.

Correlogram analysis

A significant positive autocorrelation was observed at a distance lag of 10-11 km, and negative autocorrelation was observed at greater distances (17 km) for axis I of the compromise in PTA-1 (Fig. 9a). However, the correlograms computed with the sample site scores on the compromise PCA were not globally significant at the Bonferroni-corrected probability level of $p^* = 0.002$ (25 distance classes) and $p^*= 0.0029$ (17 distance classes) for PTA-1 (Fig. 9a) and PTA-2 (Fig. 9b), respectively.

DISCUSSION

Spatio-temporal structure and stability of water-quality data

The results highlight the spatial structure of the physical and chemical water properties of different sites in the Flumen River Basin over one year. PTA allowed us to clearly identify the presence of a significant spatial structure across the entire catchment that was present on all sampling dates. Both PTAs provided a good summary of the spatio-temporal structure of water variables for two time periods and all sampling sites. The interstructure analysis of PTA-2 basically followed the same pattern as PTA-1 (Fig. 5), even though the latter refers to winter months only. Moreover, differences were observed in the distribution of the variables that were dependent on the sampling dates; these differences could be partially related to the agronomic calendar of existing land use in the territory.

- <u>PTA-1</u>

Our analysis showed that land use had a strong influence on the spatio-temporal pattern of water quality as given by the observed typology of sampling stations in the factorial plane. Differences were observed in the distribution of the variables among the three sampling dates (Fig. 5). Indeed, the upper watershed sites (I1, I2, F1, F2, F3, F4 and F5), which corresponded to forest and low-intensity agricultural areas, were all similar and relatively diluted (Fig. 5D), i.e., they were not affected by temporal variation, in contrast to the zones in the lowest region of the basin. Sites downstream of the wastewater treatment plant outfall (I5 and I6) dramatically deviated from the upper watershed group, presumably due to input of wastewater effluent upstream of I5, which contributes DOC and nutrients. Determining whether this observation represents the impacts of urban land use or the impact of a point source of poor-quality water is beyond of the scope of this study; however, the I3 and I4 urban stations were distinctly different from I5 and I6. The characteristics of sampling stations F6 and F7 were between those of the upper watershed sites and I5-I6, indicating that they are composed of a mixture of these two sources, i.e., F6 and F7 are located just downstream of the confluence of I6 and F5 (Fig. 5D).

Finally, stations F8 and F9 (agricultural sampling sites) located in the lower region of the basin (North Monegros County) showed an increase in major ions (NO_3^--N and saline compounds) relative to the other regions of the watershed in November 2009 and January 2010 (Fig. 5). A dilution of salt concentration occurs with intensive irrigation during the spring and summer, i.e., EC decreases from 1,550 µS cm^{-1} in the non-irrigation period to 450 µS cm^{-1} in the irrigation season (Martín-Queller *et al.*, 2010). Monteagudo *et al.* (2012) reported that irrigation practices are more influential in NO_3^--N export than in non-irrigated agriculture. A large group of salt compounds was correlated with non-irrigation dates in PTA-1, whereas Na^+ and Cl^- ions were linked to summer months (Figs. 4 and 5). In the factorial plane of PTA-1, the variables were clearly organized into two groups: nitrogenous compounds (aligned with sampling stations that receive urban inputs) and saline variables (more closely correlated with sampling stations located in agricultural zones, i.e., F8 and F9). This effect also could be related to the cumulative effect of saline compounds across the lower basin area; however, stations F6, F7, I5 and I6 were not highly correlated with salt compounds.

This result is possibly associated with the temporal fertilization of cereals (*Triticum* spp. and *Hordeum vulgare)* and alfalfa (*Medicago sativa*). A correlation was also observed in the R_V coefficient matrix (R_V=0.631, Table 3b). These results, which are essentially a site-specific interpretation of longitudinal changes in water quality, were efficiently summarized in the PTA.

- <u>PTA-2</u>

In PTA-2, the interstructure analysis and the representation of variables in the factorial plane of the compromise revealed a correlation between the physico-chemical water parameters and the agricultural practices performed on each sampling date (Figs. 4 and 5). A stable temporal pattern of the measured variables at particular sampling stations of the basin is shown in PTA-2. The described structure represents a relatively important proportion of the total variability in the initial matrices (44.5% of 69.3% explained by axis I of the interstructure). The sub-matrices that correspond to April, June, July and September 2010 best represent the temporal dynamics of the measured parameters as given by their high values of cos², with the highest R_V coefficient observed between April and June 2010 (Table 3b). This result appears to explain why TDN and NO_3^--N, EC and alkalinity were aligned with the sampling dates (Figs. 4 and 5) during the irrigation season, when the principal agricultural practices are conducted. The base fertilizer for rice is applied in April, whereas fertilization of maize crops occurs in June. These results allow us to infer that the trends adopted by the analyzed variables followed a spatio-temporal pattern marked by the seasonal agronomic calendar of adjacent land use.

The sampling stations with the highest row score on axis I in PTA-2 are expected to have the highest EC and major ion concentrations; these locations corresponded to IC1, IC2, F8 and F9, i.e., stations at the lowest region of the basin that are most impacted by agricultural use. In contrast, sampling stations with the lowest scores on axis I exhibited the lowest EC and ion concentrations and corresponded to F4 and F5. Consequently, this axis is interpreted as intensive agricultural use; in other words, the results reflect the cumulative influence of human activities on water chemistry in the lowland areas of the basin. The EC increment represents the progressive water enrichment by major ions and NO_3^--N (Fig. 7). With respect to axis II, NH_4^+, TP and PO_4^{2-} were positive contributors, whereas pH was negatively correlated. In summer, the concentrations of NH_4^+-N, TP and PO_4^{3-}-P on axis II of PTA-2 appear to be important to the water chemistry and could be linked to the increasing number of tourists, residential buildings and recreation areas during this period (Perona *et al.*, 1999).

This effect was apparent at stations I6 and F6, which received contributions from the wastewater treatment plant of Huesca, the largest city in the study area. The findings are also supported by the different placements of NO_3^--N in each multifactorial analysis, which could be explained by the contribution of the Huesca wastewater plant to the IW station (PTA-2). High concentrations of NO_3^--N in the streams of urbanized areas were also reported by Osborne and Wiley (1988) and Sliva and Williams (2001).

The pH and suspended solids (SS) did not follow any spatiotemporal pattern with significant effects in the Flumen River Basin, which might indicate a water buffering effect. The value for SS was separated from the remainder of the variables in the PTA-2 factorial plane. The concentration of SS exhibited a high inter-annual variability that is likely more closely linked to the number of punctual flood events during the year than to seasonal variations because nearly all of the SS are transported during these events (Rovira and Batalla, 2006; Oeurng *et al.*, 2011).

Finally, the differences observed in both PTAs could also reflect dissimilar sources of urban pollution, i.e., in the factorial plane, these variables were not aligned with the IW station. This observation corresponds to a point source of pollution, whereas NH_4^+-N, TP and PO_4^{3-}-P may arise from different urban runoff regimes and non-point sources (Sliva and Williams, 2001). The primary sources of reactive PO_4^{3-}-P and NH_4^+-N are urban inputs (Brainwood *et al.* 2004; Mendiguchía *et al.*, 2007; Neal *et al.*, 2000), which are also the sources for NO_2^--N (Martín-Queller *et al.*, 2010). This close relationship with urban pollution sources was clearly shown in PTA-1 (Fig. 5C), in which the variables were linked to I5 and I6 (urban stations).

Spatial autocorrelation of the common structure described in the PTA

The sampling sites distributed throughout the river basin are highly spatially clustered. The largest forested areas are headwater sites (I1, I2, and F1 in Fig. 1). Two rainfed sites are immediately downstream of the Montearagon Reservoir. The urban sites are all located within a few kilometers in the upper reaches of the river (I3, I4, IW, and I5). The irrigated sites are distributed in the lower half of the watershed (I6-I9). The likely spatial relationship between sites complicates their interpretation as a function of land use. With respect to the impact of land use on water quality and the inherent spatio-temporal autocorrelation of sites, the correlogram showed positive and negative spatial autocorrelations at distances of 11 km and 17 km, respectively. Because the correlogram was not statistically significant, our results indicated that different processes (natural and anthropogenic) affect the water quality across the river network.

Although no correlation analysis was performed between changes in water quality and land use, our results indicate that the PTA efficiently summarized changes in the water quality that might be influenced by the range of land uses along the stream network.

CONCLUSIONS

The proposed methodology allowed us to identify a common multivariate spatial structure and to assess the temporal stability of the spatial structure of the measured water variables. The PTA constitutes a robust technique for land and water management monitoring programs that evaluate the ecological state of ecosystems and agro-ecosystems. This method is a potentially beneficial tool for decision-making by watershed managers and for use by engineers and scientists in evaluating water-quality impacts and addressing natural and anthropogenic influences in watershed management.

Our results showed that in urban and agricultural areas, the observed trend of the analyzed variables followed spatio-temporal patterns that are possibly marked by the seasonality of land use. The stability of the spatio-temporal structure of water-quality data appeared to be linked to agricultural use (seasonal land management) and human activities. The cumulative effect of pollutants derived from agricultural and urban uses was observed in stations located in the lowest regions of the basin. Nevertheless, land use cannot be treated as the sole driving force of the variability in the analyzed parameters, and other hydrogeological or biochemical processes might also be important.

PTA can be used as a convenient space-time framework tool for assessing the dynamics of the relationships among variables, sites and time (Doledec and Chessel, 1987; Centofanti *et al.*, 1989). In our study, with regards to water quality at the catchment scale, we assessed:

1) The temporal variability in a common spatial pattern;

2) The common spatial structure of a pattern via removal of the temporal effect;

3) The temporal stability of the common structure, i.e., over months or the calendar of agricultural practices, and

4) The spatial autocorrelation (and significance) of the global spatio-temporal structure with the correlogram using the row scores of the first axis extracted in the compromise.

In conclusion, knowledge of basin characteristics and spatio-temporal trends of pollutant transfer is essential when applying effective corrective measures that minimize the effects of water pollution. PTA can be used for efficiently summarizing site-specific water-quality patterns in an applied setting for land- and water-monitoring schemes at the landscape level, in isolation or in combination with other available tools to assess spatial and temporal variations of water quality at landscape scales.

ACKNOWLEDGMENTS

This work is a component of the AGUAFLASH project funded by the Program of Territorial Cooperation "Interreg IVB-SUDOE" (SOE1/P2/F146) and EU FEDER. The authors thank "Comarca de Los Monegros" for their continuous support and cooperation in the development of this project and C. Pedrocchi, J. Cervantes, S. G. Eismann, M. García, S. Pérez and A. Barcos for their comments and assistance during the field campaign and laboratory work. The assistance provided by Jean-Pierre Rossi in the correlogram computation is also acknowledged. Key formal cooperation during this research was provided by Mr. A. Calvo at CHE. The authors also thank AEMET (Spanish Meteorological Agency) for the meteorological data. All co-authors declare that they have no competing issues or conflicts of interest.

REFERENCES

APHA 1998. Métodos normalizados para el análisis de aguas potables y residuales, American Public Health Association, Madrid, (in Spanish).

Baker, L.A. 1992. Introduction to nonpoint source pollution in the United States and prospects for wetland use. Ecol. Eng. 1:1-26.

Bertrand, F., and M. Maumy. 2010. Using partial triadic analysis for depicting the temporal evolution of spatial structures: assessing phytoplankton structure and succession in a water reservoir, CS-BIGS 4(1):23-43.

Brainwood, M.A., S. Burgin, and B. Maheshwari. 2004. Temporal variations in water quality of farm dams: impacts of land use and water sources. Agr. Water Manage. 70(2):151-175.

Caccia, V.G., and J.N. Boyer. 2005. Spatial patterning of water quality in Biscayne Bay, Florida as a function of land use and water management. Mar. Pollut. Bull. 50(11):1416-1429.

Casey, H., and R.T. Clarke. 1979. Statistical analysis of nitrate concentrations from the River Frome (Dorset) for the period 1965-76. Freshwater Biol. 9(2):91-97.

Centofanti, M., D. Chessel, and S. Dolédec S. 1989. Stabilité d’une structure spatiale et compromis d’une analyse statistique multi-tableaux: application à la physico-chimie d’un lac réservoir. Rev. Sci. Eau 2:71-93.

Cliff, A.D., and J.K. Ord. 1981. Spatial processes: models and applications. Pion Ltd., London.

Cloutier, V., R. Lefebvre, R. Therrien, and M. Savard. 2008. Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. J. Hydrol. 353:294-313.

Comín, F.A., and W.D. Williams. 1994. Parched continents: Our common future? A paradigm of planetary problems. In: R. Margalef, editor, Limnology now: a paradigm of planetary problems. Elsevier Science B.V., New York. p. 473-527.

Darwiche-Criado, N., F.A. Comín, R. Sorando, and J.M. Sánchez-Pérez. 2015. Seasonal variability of NO_3^- mobilization during flood events in a Mediterranean catchment: The influence of intensive agricultural irrigation. Agr. Ecosys. Environ. 200:208-218.

Decaëns, T., and J.-P. Rossi. 2001. Spatio–temporal structure of earthworm community and soil heterogeneity in a tropical pasture. Ecography 24:671-682.

Decaëns, T., J.J. Jiménez, and J.-P. Rossi. 2009. A null-model analysis of the spatio-temporal distribution of earthworm species assemblages in Colombian grasslands. J. Trop. Ecol. 25:415-427.

Dermine, B., and L. Lamberts. 1987. Nitrate nitrogen in the Belgian course of the Meuse River – Fate of the concentrations and origin of the inputs. J. Hydrol. 93(1-2):91-99.

Dolédec, S. 1988. Les analyses multi-tableaux en écologie factorielle. II. Stratification longitudinale de l'Ardëche à partir de descripteurs physico-chimiques. Acta Oecol. 9:119-135.

Dolédec, S., and D. Chessel. 1994. Co-Inertia analysis: an alternative method for studying species-environment relationships. Freshwater Biol. 31:277-294.

Escoufier, Y. 1973. Le traitement des variables vectorielles. Biometrics 29:750-760.

Gourdol, L., C. Hissler, L. Hoffmann and L. Pfister. 2013. On the potential for the Partial Triadic Analysis to grasp the spatio-temporal variability of groundwater hydrochemistry, Appl. Geochem. 39:93-107.

Grizzetti, B., F. Bouraoui, and G. De Marsily. 2005. Modelling nitrogen pressure in river basins: A comparison between a statistical approach and the physically-based SWAT model. Phys. Chem. Earth 30(8-10):508-517.

Gudmundsson, L., L.M. Tallaksen, and K. Stahl. 2011. Spatial cross-correlation patterns of European low, mean and high flows. Hydrol. Process. 25:1034-1045.

Güler, C., G.D. Thyne, J.E. McCray, and A.K. Turner. 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeol. J. 10:455-474.

Hunsaker, C.T., and D.A. Levine. 1995. Hierarchical approaches to the study of water quality in rivers. BioScience 45(3):193-203.

Jaffrenou, P.A. 1978. Sur l'analyse des familles finies des variables vectorielles. Bases algébriques et application à la description statistique. Thèse de 3^{eme} cycle, Université des Sciences et Technique du Languedoc, Montpellier-II. (in French).

Jiménez, J.J., T. Decaëns, and J.-P. Rossi. 2006. Stability of the spatio-temporal distribution and niche overlap in neotropical earthworm assemblages. Acta Oecol. 30: 299-311.

Kang, S,, and H. Lin. 2007. Wavelet analysis of hydrological and water quality signals in an agricultural watershed. J. Hydrol. 338(1-2):1-14.

Kiers, H.A.L. 1991. Hierarchical relationships among three-way methods. Psychometrika 56 (3):449-470.

Kroonenberg, P.M. 1989. The analysis of multiple tables in factorial ecology. III. Three mode principal component analysis: “analyse triadique complète”. Acta Oecol. 10(3):245-256.

L’Hermier des Plantes. 1976 Structuration des tableaux à trois indices de la statistique. Théorie et applications d'une méthode d'analyse conjointe. Thèse de $3^{\text{ème}}$ cycle, USTL, Montpellier,

Lavit, C. 1988. Analyse conjointe de tableaux quantitatifs. Masson, Paris. 240 p.

Lenat D., and J.K. Crawford. 1994. Effects of land use on water quality and aquatic biota of three North Carolina Piedmont streams. Hydrobiologia 294:185-199.

Legendre, P., and M.J. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80:107-138.

Legendre, P., and L. Legendre. 1998. Numerical ecology. 2^{nd} Ed., Elsevier Science, Amsterdam. 853 p.

Marques, S.C., M.A. Pardelo, S. Mendes, and U.M. Azeiteiroa. 2011. Using multitable techniques for assessing the temporal variability of species–environment relationship in a copepod community from a temperate estuarine ecosystem. J. Exp. Mar. Biol. Ecol. 405:59-67.

Martín-Queller, E., D. Moreno-Mateos, C. Pedrocchi, J. Cervantes, and G. Martínez. 2010. Impacts of intensive agricultural irrigation and livestock farming on a semi-arid Mediterranean catchment. Env. Monit. Assess 167:423-435.

Mendes, S., J.F. Gómez, M.J. Pereira, U.M. Azeiteiro, and M.P. Galindo-Villardón. 2010. The efficiency of the partial triadic analysis methods: an ecological application, Biometr. Letters 47:83-106.

Mendiguchía, C, C. Moreno, and M. García-Vargas. 2007. Evaluation of natural and anthropogenic influences on the Guadalquivir River (Spain) by dissolved heavy metals and nutrients. Chemosphere 69(10):1509-1517.

Monteagudo, L., J.L. Moreno, and F. Picazo. 2012. River eutrophication: Irrigated vs. non-irrigated agriculture through different spatial scales. Water Resour. 46(8):2759-2771.

Moran, P.A.P. 1948. The interpretation of statistical maps. J. Royal Stat. Soc. Ser. B-Method. 10:243-251.

Moreno, D., U. Mander, F.A. Comín, and C. Pedrocchi. 2008.Relationships between landscape pattern, wetland characteristics, and water quality in agricultural catchments. J. Environ. Quality 37:2170-2180.

Neal, C., H.P. Jarvie, S.M. Howarth, P.G. Whitehead, R.J. Williams, M. Neal, M. Harrow, and H. Wickham. 2000. The water quality of the River Kennet: initial observations on a lowland chalk stream impacted by sewage inputs and phosphorus remediation. Sci. Tot. Environ. 251/252:477-495.

Niemi, G., P. DeVore, N. Detenbeck, D. Taylor, A. Lima, and J. Pastor. 1990. Overview of case studies on recovery of aquatic systems from disturbance, Environ. Manage. 14(5):571-587.

Oeurng, C., S. Sauvage, A. Coynel, E. Maneux, H. Etcheber, and J.M. Sánchez-Pérez. 2011. Fluvial transport of suspended sediment and organic carbon during flood events in a large agricultural catchment in southwest France. Hydrol. Process. 25(15):2365-2378.

Osborne, L.L., and M.J. Wiley. 1988. Empirical relationships between land use/cover and stream water quality in an agricultural watershed. J. Environ. Manage. 26(1):9-27.

Overmars, K.P., G.H.J. de Koning, and A. Veldkamp. 2003. Spatial autocorrelation in multi-scale land use models. Ecol. Model. 164: 257-270.

Perona, E., I. Bonilla, and P. Mateo. 1999. Spatial and temporal changes in water quality in a Spanish river. Sci. Tot. Environ. 241(1-3):75-90.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org/.

Robert, P., and Y. Escoufier. 1976. A unifying tool for linear multivariate statistical methods: the RV-coefficient. Appl. Stat. 25(3):257-265.

Rolland, A., F. Bertrand, M. Maumy, and S. Jacquet. 2009. Assessing phytoplankton structure and spatio-temporal dynamics in a freshwater ecosystem using a powerful multiway statistical analysis. Water Resour. 43:3155-3168.

Rossi, J.-P. 2003. The spatiotemporal pattern of a tropical earthworm species assemblage and its relationship with soil structure. Pedobiologia 47(5-6):497-503.

Rossi, J.-P., M. Nardin, M. Godefroid, M. Ruiz-Diaz, A.-S. Sergent, A. Martinez-Meier, L. Pâques, and P. Rozenberg. 2014. Dissecting the space-time structure of tree-ring datasets using the Partial Triadic Analysis. PLoS ONE 9(9): e108332.

Rovira, A., and R.J. Batalla. 2006. Temporal distribution of suspended sediment transport in a Mediterranean basin: The Lower Tordera (NE Spain). Geomorphology 79(1-2):58-71.

Seeboonruang, U. 2012. A statistical assessment of the impact of land uses on surface water quality indexes. J. Environ. Manage. 101:134-142.

Sliva, L., and D. Dudley-Williams. 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Resour. 35(14):3462-3472.

Singh, K.P., A. Malik, D. Mohan and S. Sinha. 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study. Water Resour. 38, 3980–3992.

Sokal, R.R., and N.L. Oden. 1978. Spatial autocorrelation in biology. 1. Methodology. Biol. J. Linn. Soc. 10:199-228.

Thioulouse, J., and D. Chessel. 1987. Les analyses multitableaux en écologie factorielle I.-De la typologie d'état à la typologie de fonctionnement par l'analyse triadique. Acta Oecol. 8(4):463-480.

Thioulouse, J., D. Chessel, S. Dolédec, and J.M. Olivier. 1997. ADE-4: a multivariate analysis and graphical display software. Stat. Comp. 7(1):75-83.

Thioulouse, J., M. Simier, and D. Chessel. 2004. Simultaneous analysis of a sequence of paired ecological tables. Ecology 85(1):272-283.

Tong, S.T.Y., and W. Chen. 2002. Modeling the relationship between land use and surface water quality. J. Environ. Manage. 66(6):377-393.

Tucker, L. 1966. Some mathematical notes on three-mode factor analysis, Psychometrika 31:279-311.

Valder, J.F., A.J. Long, A.D. Davis, and S.J. Kenner. 2012. Multivariate statistical approach to estimate mixing proportions for unknown end members. J. Hydrol. 460-461:65-76.

van Buuren, S., and K. Groothuis-Oudshoorn. 2011. MICE: Multivariate Imputation by Chained Equations in R. J. Stat. Softw. 45(3):1-67.

Vonesh, E.F., and V.M. Chinchilli. 1997. Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker Inc., New York.

Zhang, M.K., Wang, L.P., and Z.L. He. 2007. Spatial and temporal variation of nitrogen exported by runoff from sandy agricultural soils. J. Environ. Sci. 19:1086-1092.

Figure captions

Fig. 1. A) Pictorial representation of sampling stations in the Flumen catchment in the Ebro Basin and B) concentration of select variables in November 2009 (raw data). NO3 = nitrate; SO4 = sulfate; NH4 = ammonium; TDN = total dissolved nitrogen; DOC = dissolved organic carbon; EC = electrical conductivity; and TP = total phosphorous. The symbol sizes are relative to the maximum value of the variable, e.g., $2x[NO_{3\ conc}]/[NO_{3\ max.conc}]$). The sampling stations along the Flumen River are F1-F9, whereas I1-I6 represent sampling stations along the Isuela River; IC corresponds to irrigation channels, and IW corresponds to the sampling point located close to the sewage treatment plant for the city of Huesca along the Isuela River.

Fig. 2. Average precipitation during the study period in the whole basin.

Fig. 3. Graphical layout of the general sampling scheme and different steps performed in partial triadic analysis. From the datacube and initial matrices of the different sampling dates (X_1, X_k…., X_t), the interstructure matrix Y was constructed and analyzed using a simple PCA. The data were first centered (removal of mean) and standardized, i.e., divided by the standard deviation (Centofanti *et al.*, 1989). The extraction of the compromise table Z and the compromise tables were derived from the coordinates of the variables at the sites on the principal components of the PCA of Y (construction of the first compromise table from the first principal component of the simple PCA of Y is depicted). The compromise tables were subsequently analyzed using a simple PCA, and the reproducibility of the compromise analysis constitutes the

intrastructure analysis (Adapted from Centofanti *et al.* 1989; Rossi, 2003 and Gourdol *et al.*, 2013).

Fig. 4. Box-plot of the historical records for the physical and chemical properties of water at four permanent governmental gauging stations across the catchment for comparison. The total number of observations, the median (straight line), and the 5% and 95% percentiles for the period 2007-2014 are shown (data from CHE).

Fig. 5. (A, top left) Temporal interstructure derived from each sampling station table. Ordination of sampling dates on the factorial plan defined by the first two axes of the PCA on the interstructure matrix in PTA-1, and eigenvalues associated with each axis; (B, bottom left) projections of the variables in the first plane (axes I–II) of the compromise table and histogram of eigenvalues that identify the prominence of the first two axes that define the average spatio-temporal structure; (C, top right) projections of the sampling stations in the first plane (axes I–II) of the compromise table; and (D, bottom right) weight of each table (α_κ) in the construction of the compromise and the quality index of the compromise structure ($\cos^2$) for each original sampling date matrix.

Fig. 6. (left) Reproducibility of the compromise for each of the sampling stations on axes I and II for PTA-1 for the sites and (right) analyzed variables. The row and column loadings of all tables are projected onto the first two principal components of the compromise matrix as additional elements (Thioulouse and Chessel, 1987).

Fig. 7. (A, top left) Temporal interstructure derived from each sampling station table. Ordination of sampling dates on the factorial plan defined by the first two axes of the

PCA on the interstructure matrix in PTA-2, and eigenvalues associated with each axis; (B, bottom left) projections of the variables in the first plane (axes I–II) of the compromise and histogram of eigenvalues identifying the prominence of the first two axes that define the average spatio-temporal structure; (C, top right) projections of the sampling stations in the first plane (axes I–II) of the compromise table; and (D, bottom right) weight of each table (α_κ) in the construction of the compromise and quality index of the compromise structure (cos^2) for each original sampling date matrix.

Fig. 8. Reproducibility of the compromise for each of the six sampling dates on axes I and II for PTA-2, summarizing the variability in the series of tables surrounding the common structure defined by the compromise.

Fig. 9. Correlograms using the scores of the first axis extracted in the compromise table in (a) PTA-1 and (b) PTA-2 with 999 permutations. The black dots indicate significant autocorrelation at $p<0.05$ for a given distance class. The correlogram was not statistically significant at the Bonferroni-corrected p-level of $p^*=0.0020$ (25 distance classes) and $p^*=0.0029$ (17 distance classes) for PTA-1 and PTA-2 under the null hypothesis.

Table captions

Table 1. Percentage of area occupied by the primary land uses associated with the sampling stations in the Flumen catchment.

Table 2. Descriptive summary statistics of all analyzed water-related variables.

Table 3. Matrix of vectorial correlation coefficients (R_V) between tables for PTA-1 (a) and PTA-2 (b).

Table 1.

Sampling Station	Canola-Polish	Maize	Winter Wheat	Italian Ryegrass	Alfalfa	Winter Barley	Sunflower	Rice	Grain Sorghum	Pasture	Soybean	Mixed Forest	Water	Residential
I1	--	--	--	26.1	--	--	--	--	--	20.6	--	53.3[*]	--	--
F1	--	--	--	--	--	--	--	--	--	44.5	--	55.5[§]	--	--
I2		0.4		9.2				7.0		21.3		58.3[¶]	3.8	
I3	0.4	0.1	11.0	6.2	0.6	25.0	2.4	3.3	0.0	21.8	0.1	26.7	2.3	0.2
I5	0.4	0.1	10.3	5.8	0.7	27.5	2.3	3.1	0.1	21.2	0.2	24.9	2.1	1.2
I6	0.3	0.3	9.7	5.3	2.2	33.0	1.9	3.0	0.1	19.1	0.3	19.1	1.6	4.1
F5	0.4	1.0	9.6	6.0	2.7	25.3	1.5	3.0	0.1	19.3	0.6	25.9	2.9	1.5
F6	1.2	1.4	9.4	5.0	4.3	30.0	3.9	2.9	0.1	15.6	1.3	21.3	2.3	1.2
F7	1.0	2.3	8.9	5.2	7.1	30.4	3.4	5.2	0.2	14.2	1.4	17.7	2.0	1.0
F9	0.5	6.5	7.3	5.8	11.8	29.2	2.4	6.7	0.3	14.8	1.1	11.8	1.2	0.7

[*] Pinus sylvestris, Buxus sempervirens, Ilex aquifolium;
[§] Quercus ilex, Pinus sylvestris, Buxus sempervirens;
[¶] Quercus faginea, Q. ilex.

Table 2.

Variables	Mean	Standard deviation	Skewness	Kurtosis
NO_3-N [1]	3.86	4.77	5.30	40.36
PO_4^{3-}-P [1]	0.09	0.14	2.49	6.33
TDP [1]	1.02	6.75	7.37	53.34
TP [1]	0.18	0.27	4.76	32.47
Cl^- [1]	63.39	47.06	1.12	1.54
SO_4^{2-}-S [1]	61.10	40.62	0.86	0.45
Na^+ [1]	55.12	41.37	1.35	2.50
K^+ [1]	5.48	4.48	1.53	1.77
Ca^{2+} [1]	110.92	34.98	0.27	-0.40
Mg^{2+} [1]	29.26	13.34	0.91	0.82
DOC [1]	4.97	2.22	2.71	13.91
TDN [1]	4.55	4.29	1.34	2.01
ALK [1]	253.62	40.81	0.36	-0.18
SS [1]	62.92	99.04	3.90	16.35
Br^-	0.26	0.13	1.17	1.25
F^- [1]	0.10	0.17	6.53	56.05
T	14.56	4.93	-0.41	-1.10
pH	8.04	0.28	-0.76	1.44
EC [2]	889.95	316.57	0.33	-0.01
DO [1]	10.00	2.26	0.84	0.88
NH_4^+-N [1]	1.41	3.74	3.94	17.33
NO_2^--N [1]	0.24	0.96	8.55	81.13

* TDP = total dissolved phosphorous; TP = total phosphorous; DOC = dissolved organic carbon; TDN = total dissolved nitrogen; ALK = alkalinity; SS = suspended solids; T = temperature; EC = electrical conductivity; DO = dissolved oxygen.

[1] $mg \cdot l^{-1}$

[2] $\mu S \cdot cm^{-1}$

Table 3a.

	Nov-09	Jan-10	Feb-10
Nov-09	1		
Jan-10	0.631	1	
Feb-10	0.604	0.577	1

Table 3b.

	Apr-10	Jun-10	Jul-10	Aug-10	Sep-10	Oct-10
Apr-10	1					
Jun-10	0.737	1				
Jul-10	0.575	0.688	1			
Aug-10	0.467	0.493	0.553	1		
Sep-10	0.656	0.732	0.708	0.486	1	
Oct-10	0.456	0.452	0.331	0.307	0.432	1

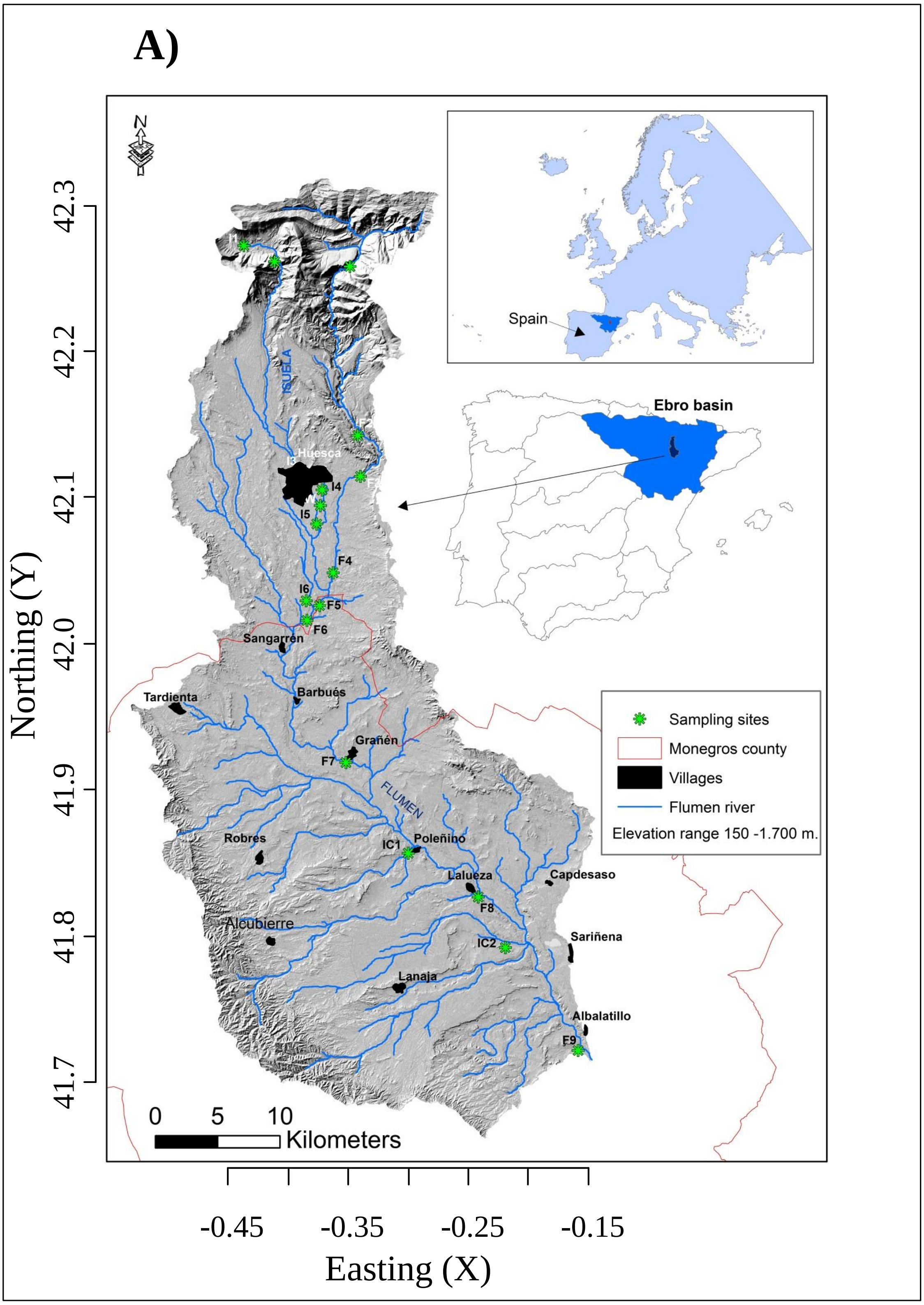

A)
Spain
Ebro basin
Huesca
ISUELA
I4
I5
F4
I6
F5
F6
Sangarrén
Barbués
Tardienta
Grañén
F7
FLUMEN
Robres
IC1
Poleñino
Lalueza
Capdesaso
F8
Alcubierre
IC2
Sariñena
Lanaja
Albalatillo
F9
Sampling sites
Monegros county
Villages
Flumen river
Elevation range 150 -1.700 m.
0
5
10
Kilometers
42.3
42.2
42.1
42.0
41.9
41.8
41.7
Northing (Y)
-0.45
-0.35
-0.25
-0.15
Easting (X)

Figure 1A

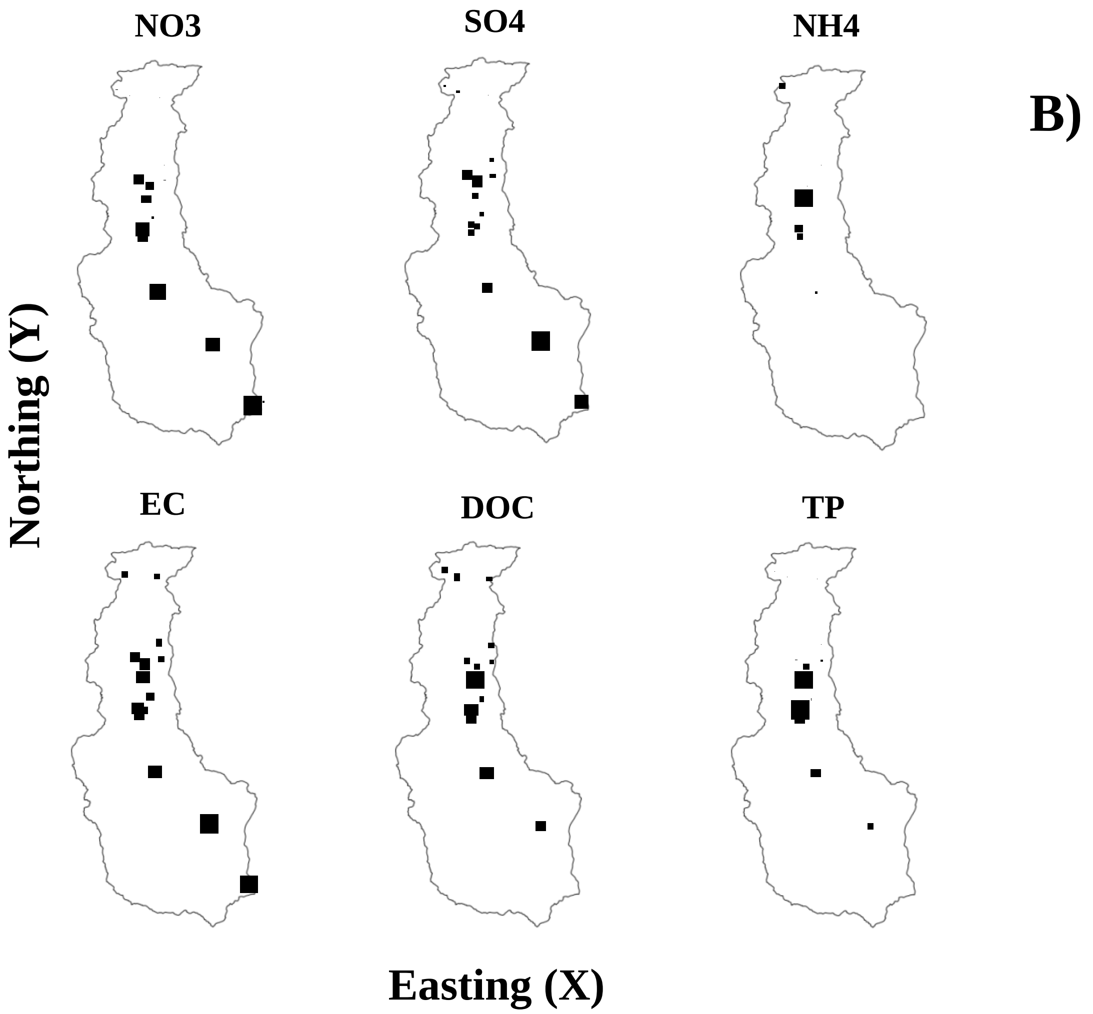


Figure 1B

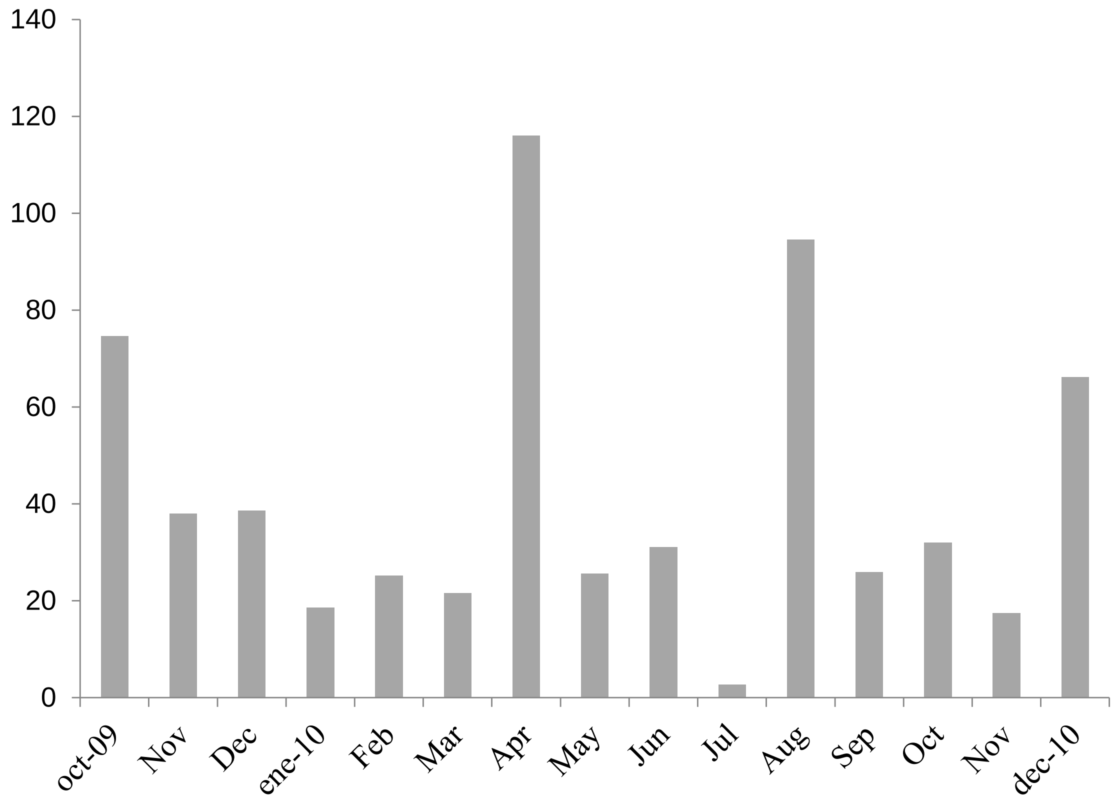

140
120
100
80
60
40
20
0
oct-09
Nov
Dec
ene-10
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
dec-10

Figure 2

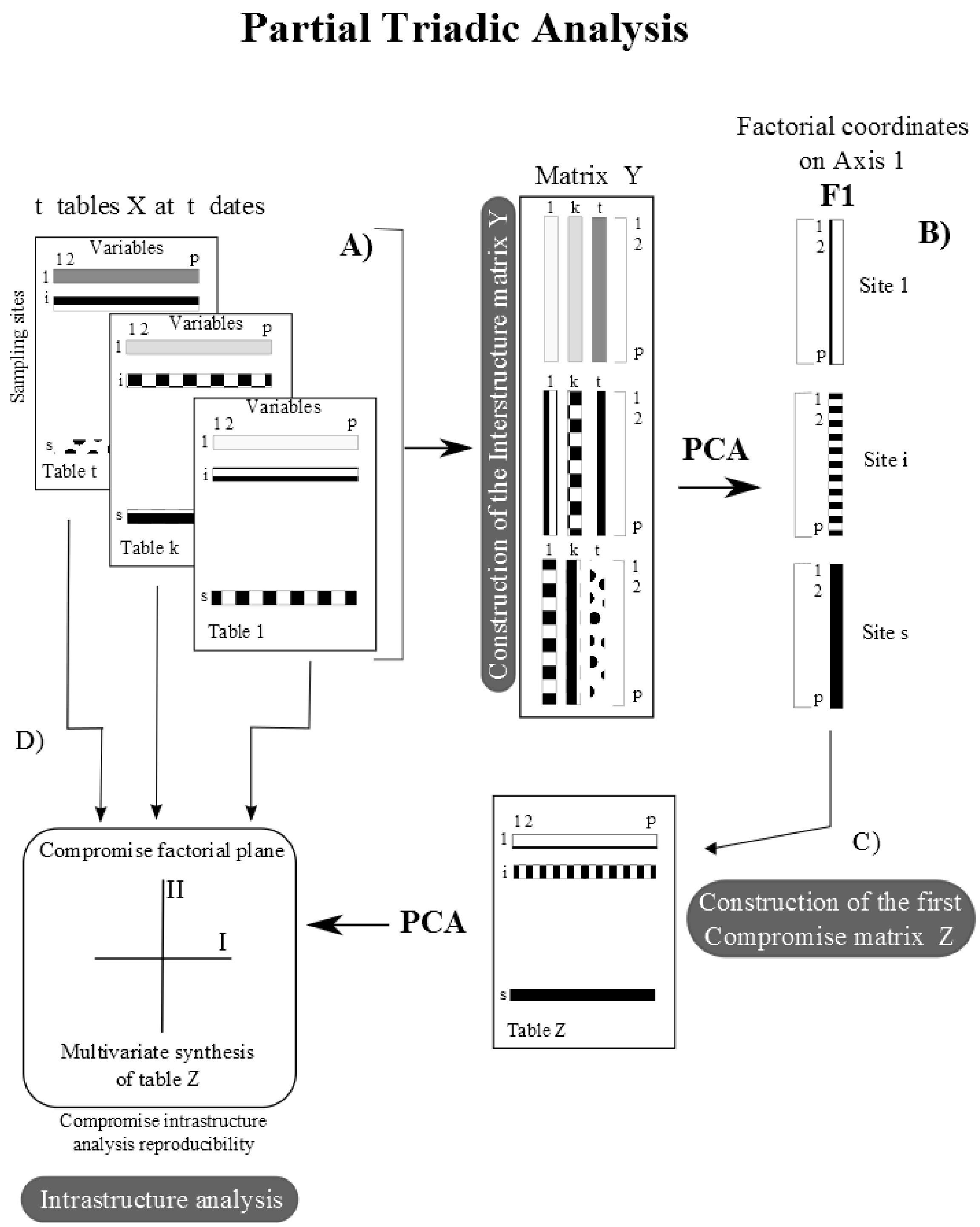
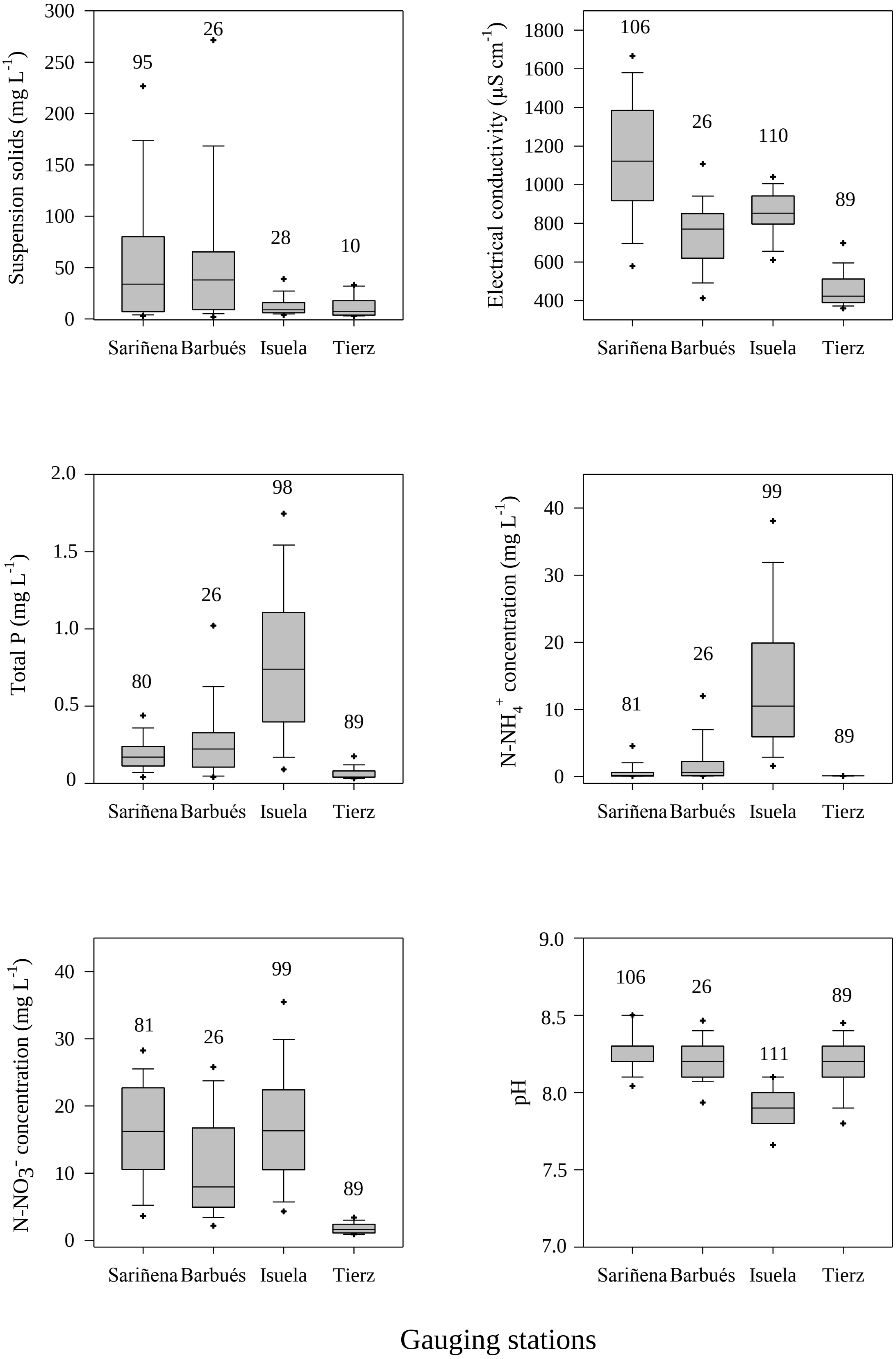

Partial Triadic Analysis
t tables X at t dates
Variables
Sampling sites
Table t
Table k
Table 1
A)
Construction of the Interstructure matrix Y
Matrix Y
Factorial coordinates on Axis 1
F1
B)
Site 1
Site i
Site s
PCA
C)
Construction of the first Compromise matrix Z
Table Z
PCA
D)
Compromise factorial plane
II
I
Multivariate synthesis of table Z
Compromise intrastructure analysis reproducibility
Intrastructure analysis

Figure 3.



896

897 Figure 4.

898

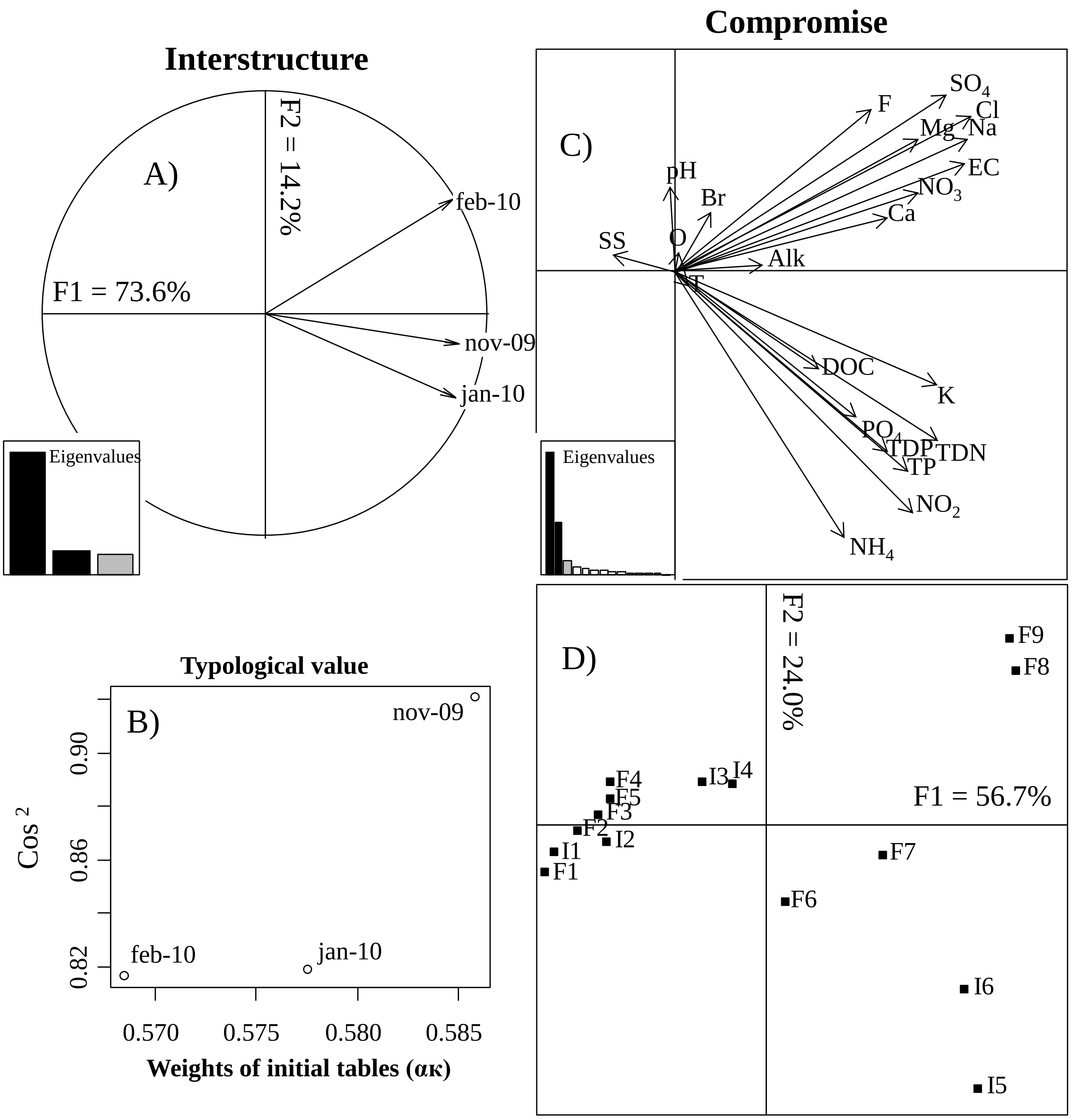

Interstructure
Compromise
A)
F2 = 14.2%
F1 = 73.6%
feb-10
nov-09
jan-10
Eigenvalues
C)
SO4
F
Cl
Mg
Na
EC
NO3
Ca
pH
Br
O
SS
Alk
T
DOC
K
PO4
TDP
TDN
TP
NO2
NH4
Eigenvalues
Typological value
B)
nov-09
feb-10
jan-10
Cos 2
0.90
0.86
0.82
0.570
0.575
0.580
0.585
Weights of initial tables (αk)
D)
F2 = 24.0%
F1 = 56.7%
F9
F8
I3
I4
F4
F5
F3
F2
I2
I1
F1
F7
F6
I6
I5

Figure5.

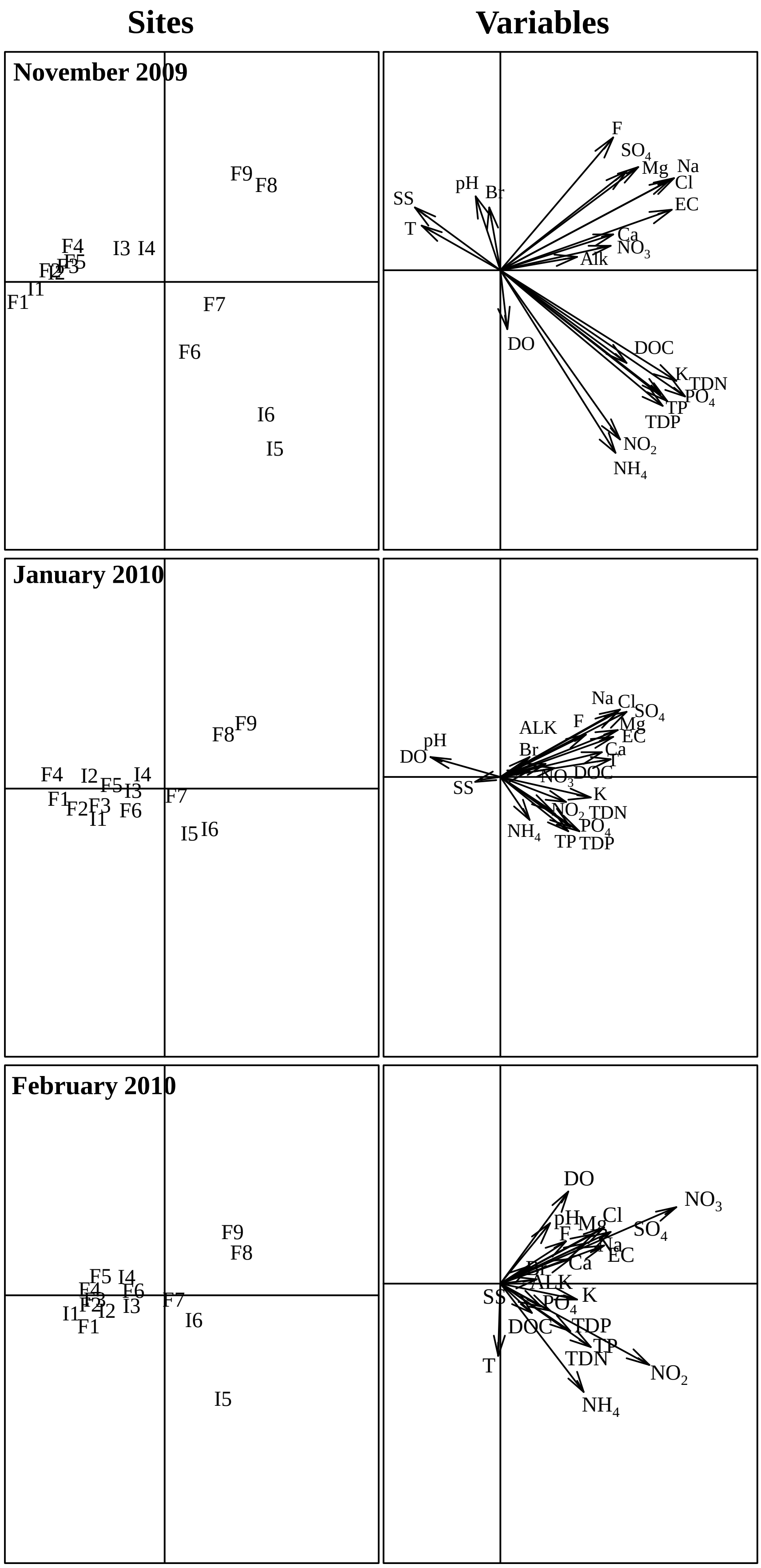

Sites
Variables
November 2009
F9 F8
F4 I3 I4
F2 I2 F3 F5
I1
F1
F7
F6
I6
I5
F
SO4
Mg
Na
Cl
EC
pH
Br
SS
T
Ca
NO3
Alk
DO
DOC
K
TDN
PO4
TP
TDP
NO2
NH4
January 2010
F8 F9
F4 I2 F5 I4
I3
F1 F2 F3 F6 F7
I1
I5 I6
Na Cl SO4
ALK F Mg
EC
pH
DO
Br
Ca
SS
NO3 DOC
K
NO2 TDN
NH4 PO4
TP TDP
February 2010
F9
F8
F5 I4
F4 F6
F3
F2 I2 I3 F7
I1
F1
I6
I5
DO
NO3
pH Mg Cl
F
SO4
Na
EC
Ca
Br
ALK
K
SS
PO4
DOC
TDP
TP
TDN
NO2
T
NH4

Figure6.

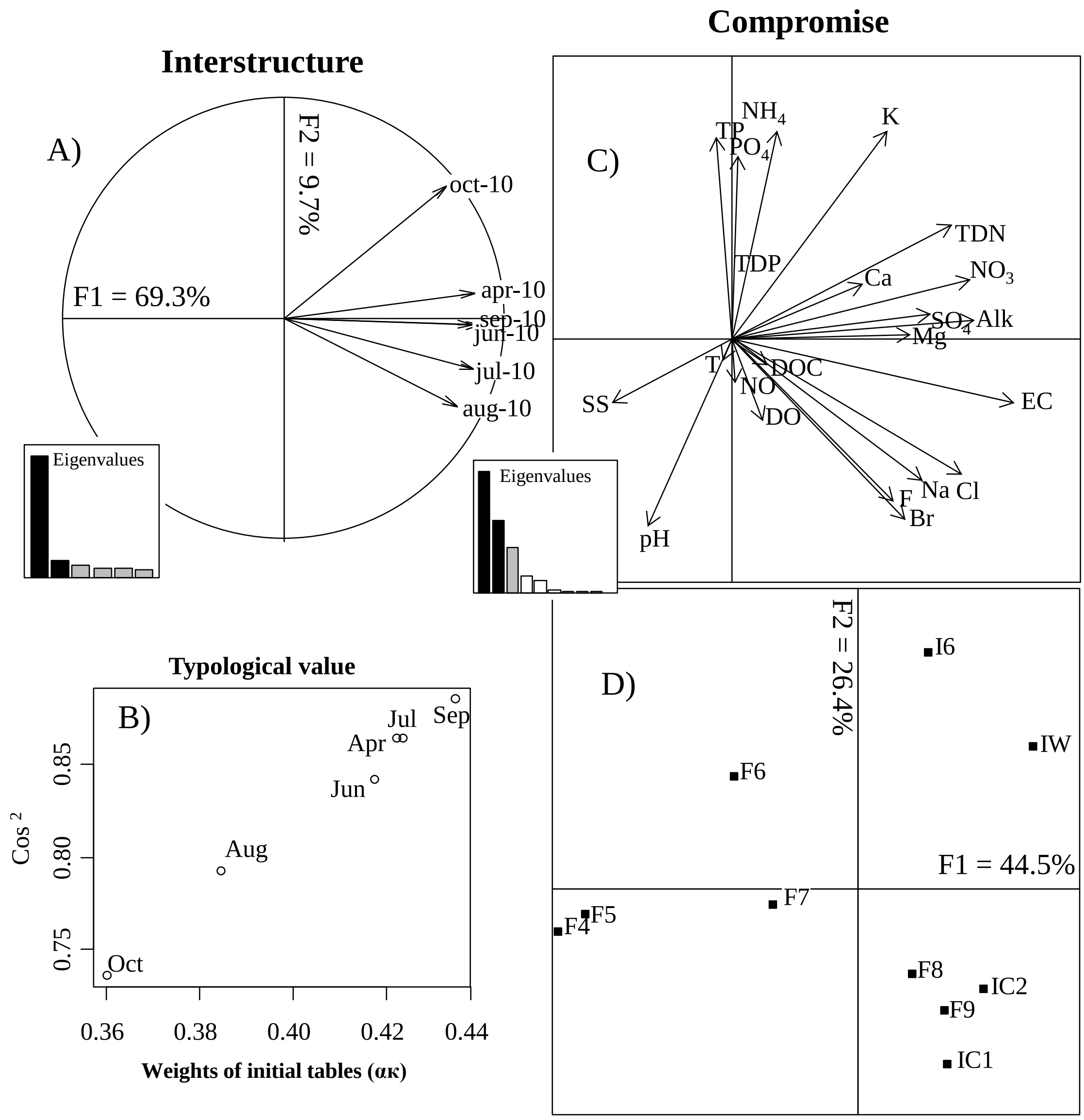

Interstructure
A)
F2 = 9.7%
F1 = 69.3%
oct-10
apr-10
sep-10
jun-10
jul-10
aug-10
Eigenvalues
Compromise
C)
NH4
TP
PO4
K
TDP
TDN
NO3
Ca
SO4
Alk
Mg
T
DOC
NO
SS
EC
DO
Na Cl
F
Br
pH
Eigenvalues
Typological value
B)
Jul
Sep
Apr
Jun
Aug
Oct
Cos 2
0.75
0.80
0.85
0.36
0.38
0.40
0.42
0.44
Weights of initial tables (αк)
D)
F2 = 26.4%
I6
IW
F6
F1 = 44.5%
F7
F5
F4
F8
IC2
F9
IC1

Figure7.

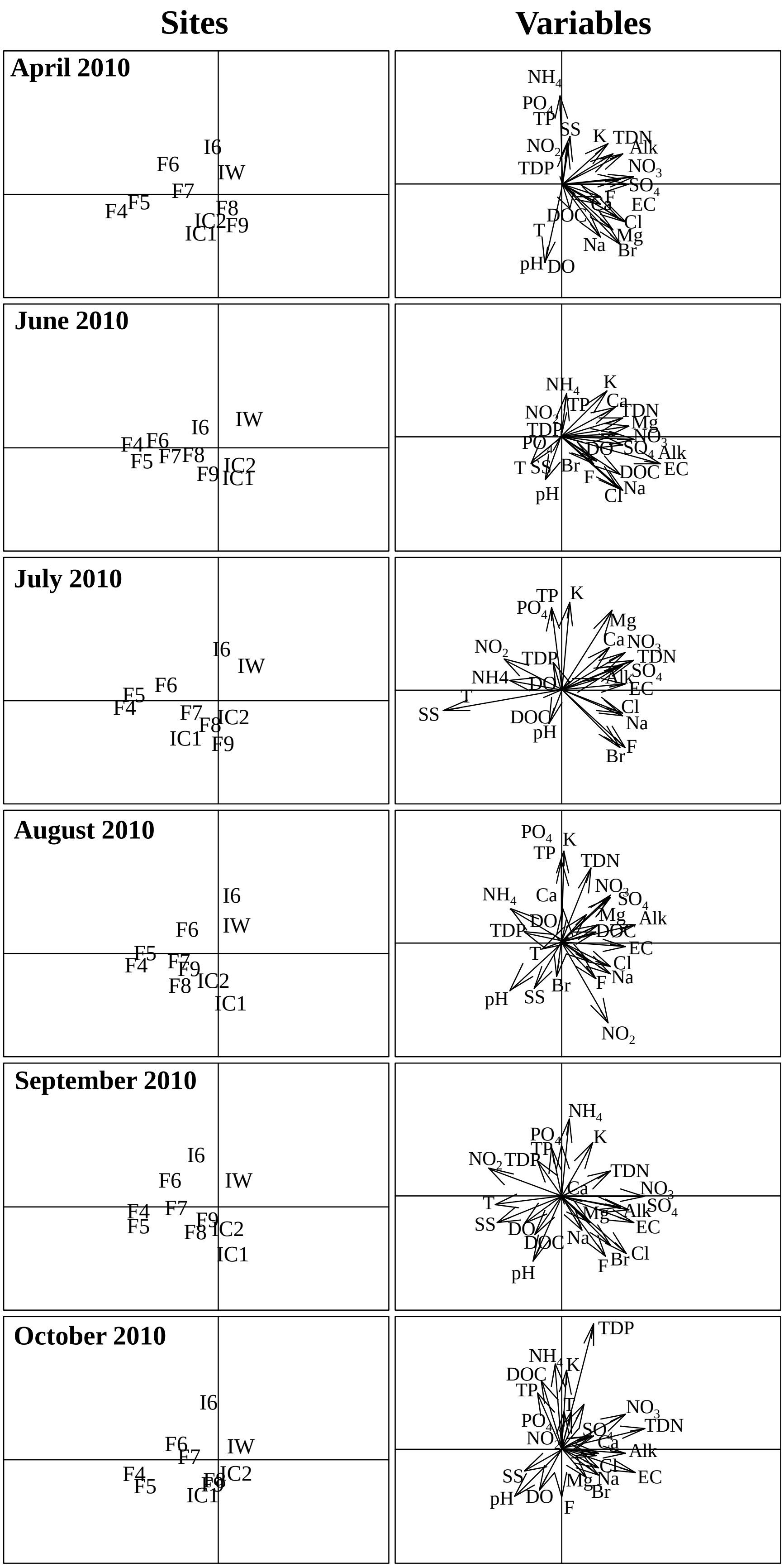

Sites
Variables
April 2010
June 2010
July 2010
August 2010
September 2010
October 2010
I6
IW
F4
F5
F6
F7
F8
F9
IC1
IC2

Figure8.

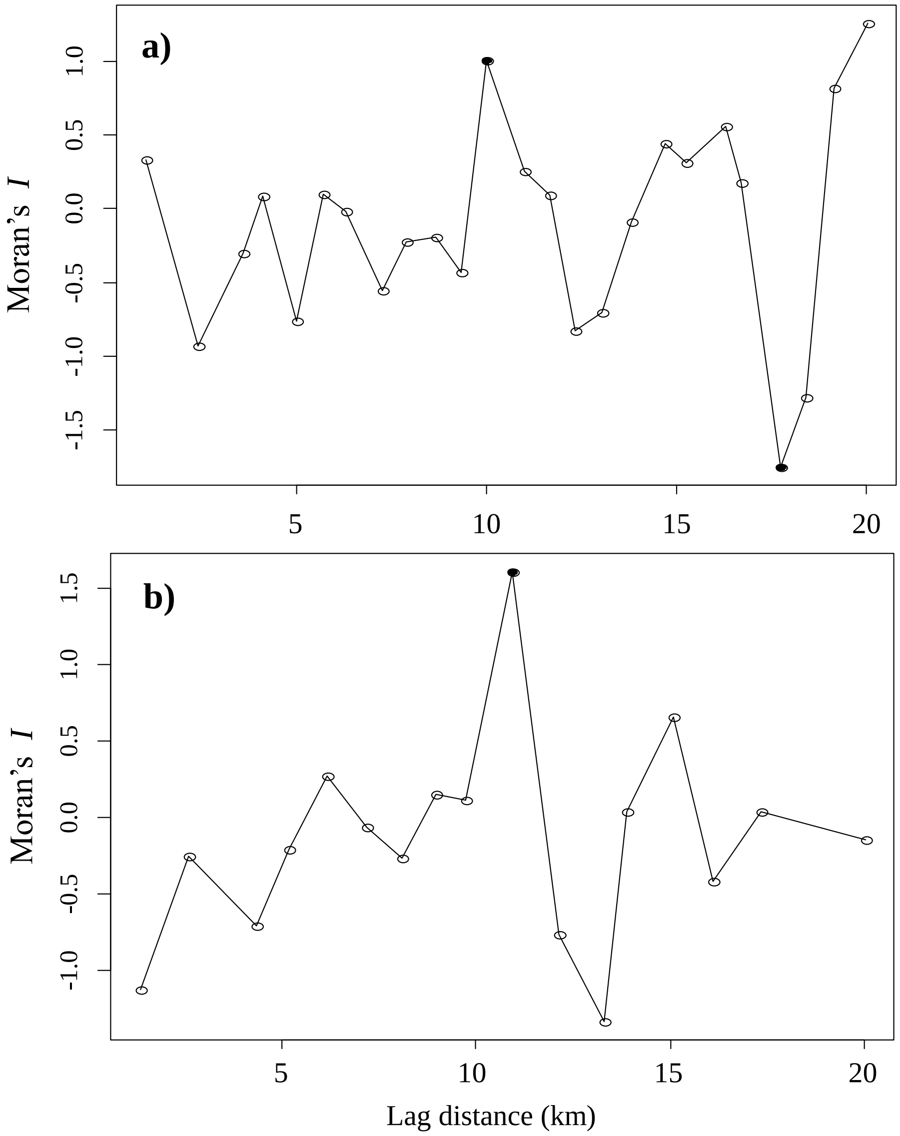


Figure9.