

# APPLICATION OF CORRESPONDENCE ANALYSIS IN THE ASSESSMENT OF GROUNDWATER CHEMISTRY

Running Head: *Application of Correspondence Analysis in the Assessment of Groundwater Chemistry*

Fernando António Leal Pacheco<sup>(a)</sup>

---

<sup>(a)</sup>Department of Geology, Trás-os-Montes e Alto Douro University, 5000 Vila Real, Portugal

Phone (59) 320186

Fax: (59) 320480

E-mail: fpacheco@utad.pt

**Paper number: 96-100**

## **ABSTRACT**

Correspondence Analysis was used to determine the sources and processes that may explain the variation observed in data sets of groundwater analyses. The following method was adopted: (1) based on the observation of “sympathies” and “antipathies” between loadings signs, correspondence factors were represented by parameters with some pertinent geochemical meaning; (2) the relation between factors and parameters was then checked by Multiple Linear Regression, where factors acted as independent variables and parameters as dependent variables. Sample scores of the selected parameters gave us the ability to separate polluted from non-polluted waters, to identify areas where pollution is dominated by agriculture and areas where contamination is controlled by domestic effluents, and to describe weathering and agricultural activities in the sampled area. The practical example presented here anticipates the impact of agriculture and urban pollution on the chemistry of 160 water samples collected in a granitoid area at central Portugal (Fundão).

*Key words:* Correspondence Analysis, Multiple Linear Regression, hydrogeochemistry, weathering, pollution.

## INTRODUCTION

The composition of groundwaters is affected by a several processes that include wet and dry deposition of atmospheric salts, evapotranspiration, and water-soil and water-rock interactions. In inhabited areas, especially where there is intensive agriculture or industry, the water composition may be affected by fertilizers and manure, by solid waste leachates, and by domestic and industrial effluents.

The contribution of each source or process to the groundwater chemistry may be quantitatively assessed by mass balance models, as early recognized by Garrels and Mackenzie (1967) in the study of a pristine area (Sierra Nevada, USA) and lately by Pacheco and Van der Weijden (1996) in the study of an area with sizeable anthropogenic input (Fundão, Portugal).

Correspondence Analysis and other eigenvector techniques may be used as methods for assessing the controls on groundwater composition, especially in the study of large hydrochemical data sets. Frequently, the approach is strictly statistical and although the analysis can select several factors more or less objectively the interpretation of these factors in terms of actual controlling sources and processes is highly subjective.

Pacheco and Portugal Ferreira (1996) studied the groundwaters from a pristine area in Portugal and established simple linear regression models between correspondence factors and several relevant hydrochemical parameters (e.g., the molar bicarbonate to silica ratio, which is characteristic for particular water-mineral interactions, as pointed out by Garrels, 1967). Based on the meaning of these parameters, Pacheco and Portugal Ferreira (1996) concluded that the variation in the concentrations of major cations, major anions and dissolved silica could be approximately attributed to silicate weathering of metamorphic rocks-forming minerals and soil

fertilization (a minor contribution), and classified their data set as a system of double influence. Alencão, Sousa Oliveira, and Pacheco (1996) applied a similar methodology to a small set of stream waters draining granites and schists and found that the direct discharge of domestic effluents into streams and rivers could be demarcated by Correspondence Analysis as a third independent influence. Therefore, they classified their stream water data set as a system of triple influence. The present paper (1) corroborates the results obtained by the above cited authors in the case of a large groundwater data set (160 samples), (2) provides a comprehensive review of the rationale for the selection of hydrochemical parameters, (3) argues about the option for Correspondence Analysis in detriment of more standard eigenvector techniques such as Principal Components Analysis or Factor Analysis, (4) and discusses the choice of variables to include in the model.

## **METHODOLOGY**

### *Eigenvector Techniques*

In general, variables in hydrochemical data sets are intercorrelated, but the correlations are hard to see and the data are difficult to interpret as they stand. For instance, the number of variables with which the investigator has to grapple is frequently large and the common scatter diagrams between pairs of variables fail to show a complete or at least comprehensive picture of the whole data variation. Eigenvector techniques such as Correspondence Analysis (CA), Principal Components Analysis (PCA) or Factor Analysis (FA) transform these original variables onto a new set of variables (generally called vectors; or correspondence factors, principal components

or factors, respectively for CA, PCA and FA) so that a major portion of data variation is concentrated on just a few of them (commonly two or three). This reduced set of vectors may then be represented on a single diagram and the groupings formed interpreted in terms of some physical or chemical process. Basically, vectors are a new set of orthogonal axes geometrically characterized by the eigenvectors and eigenvalues: (1) the elements of the eigenvectors are the cosines of the angles between the original and transformed axes; (2) the eigenvalues measure the spread of the data in the direction of the vectors. For example, the first eigenvector is the direction of maximum spread of the data in terms of a  $p$ -dimensional space (where  $p$  is the number of original variables) and this spread is measured by the first eigenvalue; the second vector has maximum spread at right angles to the first vector, and so on. Eigenvalues are used to measure the proportion of data used by each vector. The first vector will contain the most information and succeeding vectors will contain progressively less information. The methods used to decide about the number of vectors that should be retained and submitted to interpretation are frequently called stopping rules (Jackson, 1991). One of the most popular is the SCREE test (Cattell, 1966). This is a graphical technique which plots the  $p$  eigenvalues in ordinate and their ranks in abscissa. In a typical SCREE plot, the  $k$  highest eigenvalues will nearly fit to a straight line of high slope and the lowest  $p-k$  eigenvalues to a straight line of low slope. The break in slope between these two lines is sometimes called an elbow and is used as stopping rule. According to it, only the  $k$  highest eigenvectors should be considered for interpretation.

The mechanics of CA is briefly outlined below. The details of this technique are beyond the scope of this paper and can be found elsewhere, for example in the original work of Benzecri (1973, 1977), or in later studies such as the analytical discussion of Heiser and Meulman (1983), the geometric interpretation of Greenacre and Hastie (1987), or the tutorial paper of Van der

Heijden, Falguerolles and Deleeuw (1989). The initial raw data for CA consist of a matrix  $X_{n \times p}$ , where  $n$  is the total number of observations and  $p$  the total number of variables. Traditionally,  $X$  is a matrix of qualitative data or quantitative data treated as qualitative (categorical data), but no practical objections exist to the use of “raw” quantitative data (Davis, 1986). In the present paper, the observations are represented by water samples and the variables by the analytical concentrations of major inorganic compounds in those samples. The method of CA comprises four consecutive steps:

- 1)  $X$  is scaled in a manner that treats rows and columns identically. This is done by dividing the elements of  $X$  ( $x_{ij}$ ) by the product of the square roots of the row and column totals:

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}} \times \sqrt{\sum_{j=1}^p x_{ij}}} \quad (1)$$

Scaled values are saved in the matrix  $Z_{n \times p}$ .

- 2) A similarity matrix,  $R_{p \times p}$ , is obtained by multiplying  $Z$  by its transpose ( $Z^T$ ):

$$R = Z^T \times Z \quad (2)$$

- 3) The  $p$  eigenvalues and eigenvectors are calculated from  $R$  (some eigenvector extracting techniques are described in Press and others, 1989). By definition, eigenvectors have unit length and must be scaled in order to be converted onto correspondence factors. This is done by multiplying each eigenvector by its corresponding singular value, which is the square root of the corresponding eigenvalue. Factor elements are called loadings and measure the extent to which the original variables are associated with the factors. Besides, if two loadings have the same sign the variables are positively correlated and if they have opposite signs the variables are negatively correlated. Factor loadings are saved in the matrix  $W_{p \times p}$ .

- 4) The samples coordinates (the so-called correspondence factor scores) are finally computed and saved in the matrix  $F_{n \times p}$ :

$$F = Z \times W \quad (3)$$

PCA and FA are two major competitors of CA. The original works date back to the turn of the century (Pearson, 1901; Spearman, 1904; respectively for PCA and FA) and later derivations of FA (especially the method of Principal Factor Analysis - PFA - of Thompson, 1934) are widely used in hydrogeochemistry (Reeder, Hitchon and Levinson, 1972; Klován, 1975; Jöreskog, Klován and Reymont, 1976; Likens and others, 1977; Miller and Drever, 1977a; Drever, 1988). The mathematics of PCA is in most similar to that of CA. Basically, the methods differ in the matrix that is used to calculate the eigenvalues and eigenvectors (R). In PCA, this is a variance-covariance matrix when the p variables are all measured in the same units or a correlation matrix otherwise. FA also uses variance-covariance or correlation matrixes to calculate the eigenvalues and associated eigenvectors, but is somehow a different technique. For a better understanding of the differences between PCA and FA, a few concepts must be introduced:

**System variability and system structure:** System variability measures data variation and is equal to the sum of the variances of the original variables. System structure measures the interdependencies between original variables and is equal to the sum of the covariances. The diagonal elements of a variance-covariance matrix describe system variability and the off-diagonal elements describe system structure.

**Common vectors** are the k vectors obtained by the stopping rule.

**Common variability, common structure and communality:** Common variability and common structure are the proportions of the system variability and system structure

explained by the common vectors. Communality is the contribution of each original variable to the common variability. These quantities are calculated by multiplying a matrix of loadings,  $W_{pk}$ , by its transpose:

$$R^* = W_X W^T \quad (4)$$

The diagonal elements of  $R^*$  are the communalities, the sum of these elements is the common variability and the sum of the off-diagonal elements is the common structure.

**Residual variability** and **residual structure** are the part of the system variability and system structure unexplained by the common vectors. These are calculated by:

$$E = R - R^* \quad (5)$$

where  $E$  is the matrix of residuals.

Equation 5 may be turned around and written as:

$$R = R^* + E \quad (6)$$

This equation says that the variance-covariance matrix is made up of the part explained by the common vectors (first term of the right-hand side) plus the proportion unexplained by them (second term). The main difference between PCA and FA stands on the fact that PCA continues extracting principal components until the residual variability has been reduced to some prescribed amount determined by the stopping rule, while the residual structure is left unconstrained. On the other hand, FA tries to reduce the residual structure as much as possible. The minimization of residual covariances cannot be done by any direct method and for that reason FA has to be carried out iteratively. There are a number of methods to do this, generally called estimation procedures (Jackson, 1991). PFA is one of these methods and operates as follows: (1) The diagonal elements of  $R$  are replaced with estimates of the communalities. A



number of initial estimates for the communalities have been proposed over the years, but the square multiple correlation coefficient of each original variable with all the others is the most widely used at this point in time.; (2) The eigenvalues and associated eigenvectors are obtained from the resultant matrix on the basis of the stopping rule; (3) From these, new estimates of the communalities are obtained (Equation 4) and placed at the diagonals of R. The process is repeated until convergence is attained between “old” and “new” communalities; (4) At this point, the common variability and structure are calculated as well as the residuals (Equations 4 and 5). If full convergence has been attained, matrix E is diagonal (all off-diagonals are equal to zero), meaning that the system structure has been completely explained. It is important to mention that once the diagonals of E are minimized (PCA) the off-diagonals are going to be reduced, and vice versa (FA). Each method does best what is supposed to do, but the results obtained by the two methods will generally be very similar. For that reason, only CA and PCA will be further considered.

### *The Option for CA*

CA and PCA belong to that category of methods in which utility is judged by performance and not by theoretical considerations (Davis, 1986). In this study, three performance criteria were tested in order to select the most appropriate technique to use in the assessment of groundwater chemistry: (1) The interpretability of vector one. Despite the fact that interpretability of vectors may not be considered an objective matter (whatever method is used), this seems a reasonable criterion as in most cases vector one explains about half of the system variability; (2)

Reproducibility of results in identical situations; (3) Robustness regarding the units in which data are expressed.

Four hydrochemical data sets were selected for the performance analysis: Sousa Oliveira (1995), Pacheco and Portugal Ferreira (1996), Alençoo, Sousa Oliveira and Pacheco (1996), Pacheco and Van der Weijden (1996); this last data set (the Fundação data set) is given in the Appendix. In all cases the bedrocks were silicate (granites and/or metamorphic) rocks and land use affected the groundwater chemistry to a certain extent. The original variables were the major inorganic compounds  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ ,  $\text{HCO}_3^-$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$  and dissolved silica ( $\text{SiO}_2$ ). The results obtained when data are expressed in mg/l are shown in Figures 1a,b (changing the measurement scale from mg/l to  $\mu\text{mol/l}$  had little effect on the results). As expected, the results obtained with both methods are identical for all data sets. The exception is the loading of  $\text{SiO}_2$  when PCA is applied to the data set of Alençoo, Sousa Oliveira and Pacheco (1996).

CA plots the loadings of  $\text{HCO}_3^-$  and  $\text{SiO}_2$  at the positive side and the loadings of  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$  and  $\text{NO}_3^-$  at the negative side of the first factor axis (Figure 1a). This means that the concentrations of bicarbonate and silica are positively correlated on factor one ( $F_1$ ) and the same is true for the concentrations of chloride, sulphate and nitrate.  $\text{Na}^+$  has always very low factor loadings and the distribution of the other cations depends on the data set:  $\text{K}^+$  is at the negative side of the  $F_1$  axis for Sousa Oliveira (1995) and at the positive side for the other data sets, and the distributions of  $\text{Mg}^{2+}$  and  $\text{Ca}^{2+}$  are reversed. The results of CA are consistent with the generally held assumptions about the sources of dissolved components in contaminated groundwaters from silicate rocks: (1) All dissolved  $\text{HCO}_3^-$  and  $\text{SiO}_2$  are assumed to be products of chemical weathering of silicate minerals (Garrels and Mackenzie, 1967; Barnes and O'Neil, 1971; Miller and Drever, 1977b; Pacheco, 1995; Pacheco and Van der Weijden, 1996; among

others). Alkalinity of groundwater may vary due to sizeable input of limestone dust, application of calcium carbonate on agricultural land, dissolution of trace amounts of calcite in joint planes, secondary precipitation of carbonate, selective uptake of nitrate or cations in exchange with bicarbonate or protons (respectively), but in the vast majority of cases these processes are assumed to produce little effect on the level of alkalinity acquired by water-rock interaction under soil  $P_{CO_2}$ . The  $SiO_2$  concentrations may change by secondary precipitation of silica but that can be neglected for dilute shallow groundwaters; (2) The  $Cl^-$ ,  $SO_4^{2-}$  and  $NO_3^-$  concentrations are mostly attributed to atmospheric and pollution sources, especially fertilizers and domestic effluents (Zobrist and Stumm, 1981; Matthess, 1982; Appelo and Postma, 1993; Frapporti, 1994; Pacheco, 1995; Pacheco and Van der Weijden, 1996; and references therein). Contributions made by dissolution of trace amounts of Cl- and S-minerals (e.g., halite, gypsum, pyrite), frequently present in metamorphic rocks, are swamped by the anthropogenic contributions and may be ignored; (3) The cation concentrations are related to both natural and anthropogenic/atmospheric sources as the negative electric charge associated with  $HCO_3^-$ ,  $Cl^-$ ,  $SO_4^{2-}$  and  $NO_3^-$  has to be compensated by an equivalent positive charge derived from  $Na^+$ ,  $K^+$ ,  $Mg^{2+}$  and  $Ca^{2+}$ . In brief, when results are obtained by CA,  $F_1$  may be easily interpreted as a “source factor” which can be used separate between samples with pollution-dominated chemistries and samples with weathering-dominated chemistries, respectively the ones with negative and positive  $F_1$  scores.

An equivalent interpretation for the first principal component ( $pc_1$ ) is hardly attained. Figure 1b shows that 8 (of 9) original variables have positive moderate to high  $pc_1$  loadings, meaning that they are all positively correlated. Besides, for all data sets but one (Pacheco and Portugal Ferreira, 1996), the  $pc_1$  scores were found to be highly correlated with the samples electroconductivities (Ec). The relation between  $pc_1$  and Ec was already noticed by Van der

Weijden and others (1983). Chemical weathering, use of fertilizers on farmland, pollution by sewage systems, apart or in combination, can produce waters with different Ec values. Different degrees of evapotranspiration will also give rise to differences in these values. But there is no easy way to estimate the contribution of each process to the Ec just looking to the samples  $pc_1$  scores. For that reason, the interpretation of  $pc_1$  will depend on prior knowledge about the geology, land use and human occupation of the study area. For instance, Pacheco and Van der Weijden (1996) identified 13 groups of water samples using the RST algorithm, a clustering algorithm based on the mathematical concept of equivalent relations between objects. Using a mass balance algorithm (the SiB algorithm), seven of these groups were associated with petrological units (bedrock groups), two with basic dikes (structure groups), and one with pollution (pollution group); the remaining three groups could not be associated with any source or process. Figure 2 plots the Ec values of Fundão water samples against the corresponding  $pc_1$  scores. Different symbols were used to represent the bedrock, structure and pollution groups, respectively circles, bullets and diamonds. The distribution of these symbols suggests that chemical weathering is responsible for the low (bedrock groups) to moderate (structure groups) electroconductivities and pollution for the high electroconductivities. When the relation between  $pc_1$  and Ec is checked for the Alenção, Sousa Oliveira and Pacheco (1996) data set one finds that granite waters are responsible for the lower conductivities and schist waters for higher conductivities. For the other two data sets no simple explanation could be found for the variation in Ec. The exigency of a detailed background knowledge about the study area for the interpretation of  $pc_1$  reduces the applicability of PCA as an “exploratory” technique for the assessment of groundwater chemistry, and for that reason CA was used in this paper.

### *Choice of Variables for the CA model*

As previously mentioned, the concentrations of major dissolved components are constrained by the charge balance condition, which is represented by the following equality (concentrations in the equivalent scale):

$$[\text{Na}^+] + [\text{K}^+] + [\text{Mg}^{2+}] + [\text{Ca}^{2+}] = [\text{HCO}_3^-] + [\text{Cl}^-] + [\text{SO}_4^{2-}] + [\text{NO}_3^-] \quad (7)$$

This “closed sum” condition demands that the sources and processes responsible for the variation in the anion concentrations match those which are responsible for the variation in the cation concentrations (and vice versa). In fact, because the negative electric charge in water is not free to vary without the corresponding variation in the positive charge, one could define the cations as the “positive face” of the inorganic groundwater chemistry and the anions as its “negative” counterpart. This means that, in principle, there should be no further gain when CA is applied to a data set using both the cations and anions, with respect to those cases where only the cations or the anions are used. In other words, the conclusions taken from the interpretation factors should be the same no matter the model in use. To test this hypothesis, three separate analyses were carried out using the Fundão data set: (1) All the 9 major inorganic compounds were included in the analysis (global model); (2) Just the four major anions and silica were used (anions model); (3) Just the four major cations and silica were used (cations model). The SCREE plots are shown in Figure 3a. The five circles representing the anions model almost coincide with the first five black bullets of the global model. Further correspondence between the anions and the global models exist regarding the loadings in the first four factors (the common factors according to the SCREE test); in Figure 3b, the trends in the anions model (white ribbons) follow the anions trends in the global model (dashed ribbons). A different set of factors is obtained with the cations

model (diamonds in Figure 3a). A single factor accounts for about 80% of the data variation and three “satellite” factors explain the remaining 20%. It was also found that the first factor is highly correlated with Ec (Pearson coefficient of 0.8 for 99.999% of probability). The cleavage between the cations model and the anions/global models is very striking. It seems like the variation in the anion concentrations swamps the variation in the cation concentrations so that when both the cations and the anions are taken together (global model) only the variation in the anions is seen.

The results obtained with the anions and the global models support the working hypothesis. They are identical and therefore the interpretation of the four common factors leads to the same conclusions. The main advantage of the anions model with respect to the global model is that the first explains 100% of the system variability and structure while the second leaves unexplained the variability and structure associated with the last five correspondence factors (cf. Figure 3a). The results of the cations model do not contradict the initial hypothesis. In this case, a major portion of the system variability is connected to a physical parameter (Ec) which, as previously seen (Figure 2), is related to a number of sources and processes, but its variation is not readily interpretable.

For the reasons presented above, the definition of hydrochemical parameters discussed below will stand on the results obtained with the anions model.

#### *Rationale for the Selection of Hydrochemical Parameters*

The application of CA to a set of n groundwater samples from crystalline rocks that were analytically measured for  $\text{HCO}_3^-$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{NO}_3^-$  and  $\text{SiO}_2$  may result in four common factors

(Figure 3a). The relation between factor i ( $F_i$ ) and the original variables is described by Equation 3 which can be re-written as:

$$F_i = w_{i,HCO_3} \left[ HCO_3^- \right] + w_{i,Cl} \left[ Cl^- \right] + w_{i,SO_4} \left[ SO_4^{2-} \right] + w_{i,NO_3} \left[ NO_3^- \right] + w_{i,SiO_2} \left[ SiO_2 \right] \quad (8)$$

As mentioned, the  $w_i$ -coefficients (factor loadings) measure the extent to which each variable is associated with  $F_i$ . It was also told that if two loadings have the same sign the variables are positively correlated and if they have opposite signs the variables are negatively correlated. From the observation of these “sympathies” and “antipathies” between factor loadings, Equation 8 may be transformed onto other equations which have some real geochemical meaning.

A common observation for the first factor is that bicarbonate and silica are sympathetic, and antipathetic in regard to chloride, sulphate and nitrate (cf. Figure 1a). In other words, weathering is in opposition to pollution on factor one. The hydrochemical parameter selected to represent this factor was the w%-Pollution defined by:

$$w\% - \text{Pollution} = \frac{\text{Pollution}}{\text{Weathering} + \text{Pollution}} \times 100 \quad (9)$$

where,

$$\begin{aligned} \text{Pollution} &= w_{1,Cl} \left[ Cl^- \right] + w_{1,SO_4} \left[ SO_4^{2-} \right] + w_{1,NO_3} \left[ NO_3^- \right] \\ \text{Weathering} &= w_{1,HCO_3} \left[ HCO_3^- \right] + w_{1,SiO_2} \left[ SiO_2 \right] \end{aligned}$$

Samples with w%-Pollution scores less than 50% will have weathering-dominated water chemistries and samples with w%-Pollution scores greater than 50% will have pollution-dominated water chemistries. The extent to which each component contributes to the w%-Pollution is determined by the  $w_1$  values.

In general, silica and bicarbonate have different loading signs on the second factor. This antipathy between natural components suggests that factor two may describe weathering in the

study area. Garrels (1967) showed that the ratio  $[\text{HCO}_3^-]/[\text{SiO}_2]$  (in mole values) is a good diagnostic parameter for particular water-mineral interactions and for that reason this ratio was chosen to represent factor two.

In the third factor, chloride is frequently in opposition to sulphate and nitrate. All three components are present in common fertilizers, but chloride may in a great deal be derived from domestic effluents and atmospheric input. For this reason, factor three was interpreted as the factor that discriminates sources of pollution, and the parameter chosen to perform this separation was the w%-Agriculture defined by:

$$\text{w\% - Agriculture} = \frac{\text{Agriculture}}{\text{Dom. Effluents / Atm. Input} + \text{Agriculture}} \times 100 \quad (10)$$

where,

$$\begin{aligned} \text{Agriculture} &= w_{2,\text{SO}_4} [\text{SO}_4^{2-}] + w_{2,\text{NO}_3} [\text{NO}_3^-] \\ \text{Dom. Effluents / Atm. Input} &= w_{2,\text{Cl}} [\text{Cl}^-] \end{aligned}$$

Samples with w%-Agriculture greater than 50% are assigned to agricultural activities and samples with w%-Agriculture less than 50% are attributed to urban pollution plus atmospheric input.

Finally, in factor four, sulphate and nitrate have opposite loading signs and the sulphate to nitrate ratio ( $[\text{SO}_4^{2-}]/[\text{NO}_3^-]$ ) was proposed to represent this factor. The antipathy between components derived from fertilizers suggests that factor four may describe land use in the study area. For example, the  $[\text{SO}_4^{2-}]/[\text{NO}_3^-]$  ratio may be used to separate areas occupied by dry farming from those used for irrigated crops or vine yards as these crops are normally fertilized by products with different sulphur and nitrogen compositions.

Hydrochemical parameters represent sources or chemical processes whose impact on groundwater chemistry is related with the portion of variation explained by the corresponding



factors. This variation decreases from factor one to factor four, which means that the impact of selected parameters also declines from the w%-Pollution to the ratio  $[\text{SO}_4^{2-}]/[\text{NO}_3^-]$ . Furthermore, the “low variation” parameters represent particular features of the “high variation” parameters. In brief, (a) The w%-Pollution describes a general feature, the separation between polluted and non polluted waters; (b) The ratio  $[\text{HCO}_3^-]/[\text{SiO}_2]$  and the w%-Agriculture furnish more detailed information about polluted and non-polluted waters ( $[\text{HCO}_3^-]/[\text{SiO}_2]$  characterizes weathering and the w%-Agriculture discriminates sources of pollution); (c) Finally, the mole ratio  $[\text{SO}_4^{2-}]/[\text{NO}_3^-]$  characterizes one specific source of pollution, namely agriculture. This hierarchical (or tree) structure of hydrochemical parameters is represented in the dendrogram of Figure 4.

There are several exceptions to the sympathies and antipathies commonly observed between factor loadings:

**Factor absence:** Sometimes, the sympathies and antipathies of a particular factor cannot be identified. For example, Alenção, Sousa Oliveira and Pacheco (1996) were not able to identify the sympathy between bicarbonate and silica that is commonly observed for factor one. Factor absence is related to what one may call "variance deficit", which occurs when parameter scores are all very similar. In the case of Alenção, Sousa Oliveira and Pacheco, the w%-Pollution could not be used to represent factor one because all water samples were heavily polluted.

**Factor exchange:** Sometimes, factor two exchanges with factor three, which means that, in these cases, the bicarbonate to silica ratio is attributed to factor three whereas the w%-Agriculture is assigned to factor two. Factor exchange is inherently associated to parameter variation. In general, (a) If water samples are collected from a large number of rock types that are affected by different weathering rates, the variance in the

bicarbonate to silica ratio may be high, and this ratio may then represent factor two, otherwise, it will represent factor three; (b) Similarly, if pollution is essentially the result of atmospheric input, it might be expected that w%-Agriculture is assigned to factor three (or be absent), but if domestic effluents and agriculture are also important sources of contamination, variance in the w%-Agriculture will increase and, in principle, this parameter will be attributed to factor two.

**Factor summation** occurs when two factors are condensed onto a single factor. Summed factors are represented by summed parameters. For example, factor three may in some cases be the sum of (common) factors three and four. In this case, the parameter w%-Agriculture +  $[\text{SO}_4^{2-}]/[\text{NO}_3^-]$  (or  $[\text{HCO}_3^-]/[\text{SiO}_2] + [\text{SO}_4^{2-}]/[\text{NO}_3^-]$ , if factor exchange occurs) will be chosen to represent factor three. The reason for this summation may be purely statistical: (1) When all factors have a definite interpretation, there is no factor left to represent the variation associated with analytical errors or other sources of variability not related to chemical processes (the residuals); (2) In case this residual variation is high, it may swamp the variance associated with factor four; (3) Consequently, factors three and four will join, and the new factor four will gather the non-process variations.

**Parameter substitution:** In some cases, the antipathy between sulphate and nitrate in factor four (or factor three, when factor summation occurs) may not be clear. This probably means that fertilization is not dominated by the pair sulphate-nitrate, but by the sulphate-chloride or nitrate-chloride pairs. In these cases, the sulphate to chloride ( $[\text{SO}_4^{2-}]/[\text{Cl}^-]$ ) or the nitrate to chloride ( $[\text{NO}_3^-]/[\text{Cl}^-]$ ) ratios will be chosen to represent factor four, in substitution for the sulphate to nitrate ratio.

Changes in the use of fertilizers may also affect the w%-Agriculture parameter. If the fertilizers composition is largely dominated by nitrate, the sulphate loading in factor three may change its sign, becoming equal to the chloride's sign. In these cases, it may be more realistic to compute the w%-Agriculture with sulphate attributed to Dom. Effluents/Atm. Input instead to Agriculture (cf. Equation 10).

### *Significance of Hydrochemical Parameters*

Despite the number of exceptions, the selected parameters are chemically sound because they are reasonably attributed to sources or chemical processes. It is important to mention, however, that this set of parameters is not unique. In general, a parameter is suited for selection (or is significant) if correlated to one (and just one) factor.

Multiple Linear Regression (MLR) may be used to test the significance of hydrochemical parameters (for details about MLR, see Jobson, 1991). The MLR model for the w%-Pollution is represented by:

$$\text{w\% - Pollution} = b_0 + b_1F_1 + b_2F_2 + b_3F_3 + b_4F_4 \quad (11)$$

(where the b values are model coefficients and the F values are factors) and similar equations may be written for the other parameters. After solving Equation 6 in regard to the b-coefficients, each  $b_j$  may be standardized in order to represent the impact that  $F_j$  has on the variance of w%-Pollution. Then, the goodness-of-fit of the model is estimated by the Adjusted Coefficient of Multiple Determination ( $R^2$ ) which measures the variation in w%-Pollution explained by the factors. When all the standardized b values (the so-called B values) are close to zero except  $B_1$ , the w%-Pollution is qualified to represent factor one if  $R^2 \approx 100\%$ . In general, when all the B values except  $B_i$  are close to zero, a parameter  $P_i$  is suited to represented factor  $F_i$ .

### *Factor Representation of Systems of Double and Triple Influence*

The main difference between systems of double and triple influence stands on the number of pollution sources that affect the water chemistry, two sources (atmospheric input plus soil fertilization) in the first case and three sources (plus domestic effluents) in the second case. Because the separation between waters with fertilizer-dominated chemistries and those with effluent and/or atmosphere-dominated chemistries is based on the w%-Agriculture parameter, the distinction between systems of double and triple influence are expected to become visible through the analysis of factor three (Equation 10), or factor two if exchange occurs. In general, the factor associated with the w%-Agriculture should be absent when systems are of double influence (Pacheco and Portugal Ferreira, 1996) and present when systems are of triple influence (Alenção, Sousa Oliveira and Pacheco, 1996).

When systems are of triple influence, it should be possible to classify the water samples according to their dominant influence. A method is suggested that takes into account the scores in the w%-Pollution and w%-Agriculture parameters (Table 1): a) The class weathering includes all samples with w%-Pollution less than 50%; b) The class effluents, all the polluted samples (w%-Pollution > 50%) that have w%-Agriculture less than 50%; c) And the class fertilization, the remaining samples (w%-Pollution > 50% and w%-Agriculture > 50%).

### **APPLICATION**

A set of 160 groundwater samples were collected in the Fundão area (central Portugal, Figure 5a) by Van der Weijden and others (1983). The samples locations are plotted in Figure 5b and the chemical analyses are given in the Appendix. This data set was extensively studied by Pacheco and Van der Weijden (1996) who used a combination of a novel grouping algorithm (the RST-algorithm) and a novel weathering algorithm (the SiB-algorithm) to assess the contributions made by chemical weathering and anthropogenic inputs to groundwater composition in this granitoid area. Using the grouping algorithm, they identified groups of waters with similar chemistries in the data set, and with the weathering algorithm they related the medians of each identified group to water-mineral interactions in the area. Background information on the mineralogy of rocks and soils of the area (Costa and others, 1971; Portugal Ferreira, 1982; Portugal Ferreira, Alves and Macedo, 1985) allowed them to chose the most realistic water-mineral interactions. Despite the high background concentrations derived from pollution, Pacheco and Van der Weijden (1996) obtained realistic results, later corroborated by Pacheco (submitted, 1996) using the classical plot - mole ratio  $[\text{HCO}_3^-]/[\text{SiO}_2]$  vs.  $[\text{HCO}_3^-]$  (mg/l) - of Garrels (1967). In the present paper, the impact of individual sources of pollution is briefly evaluated on the basis of results obtained by CA.

### *The Fundão CA-model*

CA was applied to the 160 groundwater samples collected in the Fundão area. The computer program used to perform the analysis was adapted from Benzecri (1973). The variables included in the CA model were the major anions and dissolved silica (concentrations in mg/l). The results are shown in Table 2.

From the observation of sympathies and antipathies between factor loadings (Table 2), and taking into account the significance of hydrochemical parameters as tested by MLR (Table 3), the factors are represented by:

The **w%-Pollution** (factor one), with

$$\text{Pollution} = 443.0[\text{Cl}^-] + 370.8[\text{SO}_4^{2-}] + 748.9[\text{NO}_3^-]$$

$$\text{Weathering} = 313.4[\text{HCO}_3^-] + 421.0[\text{SiO}_2]$$

The **w%-Agriculture** (factor two), with

$$\text{Agriculture} = 63.4[\text{SO}_4^{2-}] + 591.6[\text{NO}_3^-]$$

$$\text{Dom. Effluents/Atm. Input} = 377.0[\text{Cl}^-]$$

The  $[\text{HCO}_3^-]/[\text{SiO}_2] + [\text{SO}_4^{2-}]/[\text{Cl}^-]$  (factor three)

**Residuals** (factor four)

### *Areal Distribution of Influences*

The distribution of element concentrations in the Fundão area has been visualized by Van der Weijden and others (1983) using SYMAP plots (Dougenik and Sheenan, 1976). In these plots, the concentration ranges are divided into ten equal steps with the darkest shading depicting the highest concentrations. The distribution maps obtained for the pollutants chloride, sulphate and nitrate show a set of small spots of very high concentrations which coincide with the main villages in the area. Pacheco and Van der Weijden (1996) measured the percentage of pollution in the Fundão groundwater samples using the parameter %-Pollution =  $([\text{Cl}^-] + [\text{SO}_4^{2-}] + [\text{NO}_3^-]) / ([\text{HCO}_3^-] + [\text{Cl}^-] + [\text{SO}_4^{2-}] + [\text{NO}_3^-]) \times 100$ . They plotted the %-Pollution scores on a map and separated the areas where water composition is dominated by water-rock interaction (%-Pollution

< 50%) from those where the composition is dominated by pollution (%-Pollution > 50%). The results confirmed that the polluted areas roughly coincide with the local villages and densely farmed areas, but it was also seen that these areas were further extended in the direction of the main tributaries of the Meimoa river (cf. the drainage pattern of the Fundão area in Figure 5b). In the present paper, a detailed version of the map obtained by Pacheco and Van der Weijden (1996) is produced by applying the methodology described in Table 1. In Figure 6, the areas with pollution-dominated water are further separated into areas where pollution is mainly produced by domestic effluents and areas where pollution is caused by agriculture. The results suggest that pollution caused by agriculture is essentially concentrated near the villages whereas pollution caused by domestic effluents basically follows the direction of streams and streamlets in the area.

#### *Distribution of the $[SO_4^{2-}]/[Cl^-]$ Ratios in the Samples with Agriculture-Dominated Chemistries*

The climate in the Fundão area is very good for agriculture. Four main types of agriculture may be distinguished: dry farming (rye, barley, etc), irrigated crops (potato, bean, etc), fruit orchards (apple, cherry), and other crops (vine and olive yards, etc). Fertilizers commonly applied in the first three types of cultures have  $[SO_4^{2-}]/[Cl^-]$  ratios of 0.0, 1.4 and 0.9, respectively (Van der Weijden and others, 1983). Figure 7 shows the distribution of the  $[SO_4^{2-}]/[Cl^-]$  ratios for the samples with agriculture-dominated chemistries (26 samples, as determined by the methodology specified in Table 1). The significance of this distribution should be critically considered given the reduced number of samples that have been used to produce the histogram, but it may be provisionally stated that the impact of agriculture on the groundwater chemistry is determined by the fruit orchards, followed by the irrigated crops, dry farming and other cultures.

## CONCLUSIONS

The chemical parameterization of groundwater data sets, as performed by relating hydrochemical parameters with correspondence factors, is very effective. It is remarkable how the variation in the chemical composition of large groups of water samples can be imputed to a small set of sources and chemical processes represented by straightforward parameters. The application of this methodology to the 160 Fundão groundwater samples produced promising and mostly convincing results, as the areas with weathering-dominated water chemistries were clearly demarcated from those where water is affected by fertilization or domestic effluents, and the relative impact of specific crops could be evaluated.



**APPENDIX:** Fundão data set. Chemical analyses (in  $\mu\text{mol/l}$ ) of 160 water samples from the Fundão area. The sample numbers are given in the first column and their locations are shown in Figure 5b. In the second column (T) the type of sampled water is given: 2-fresh wells; 3-springs. Reproduced from Pacheco and Van der Weijden (1996).

Nr	T	Na <sup>+</sup>	K <sup>+</sup>	Mg <sup>2+</sup>	Ca <sup>2+</sup>	Cl <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	NO <sub>3</sub> <sup>-</sup>	HCO <sub>3</sub> <sup>-</sup>	SiO <sub>2</sub>
28	3	613	39	245	433	447	178	371	780	656
30	3	691	38	242	468	493	229	460	844	639
31	2	613	15	129	315	430	93	387	490	506
32	3	491	31	100	260	287	65	371	390	558
35	3	639	25	183	320	353	171	221	729	614
39	2	406	17	121	128	235	65	216	620	260
41	3	332	19	74	180	192	99	55	261	463
42	3	396	8	58	8	132	63	139	370	421
45	2	1070	55	387	488	680	232	121	1280	571
51	2	817	36	275	650	508	229	189	780	100
59	2	1683	52	683	1142	2152	318	150	2260	674
60	3	433	13	126	217	252	79	63	561	524
61	3	439	11	93	158	235	42	18	580	560
63	2	704	76	300	622	803	367	998	229	399
66	3	530	7	192	325	301	162	366	480	474
67	3	591	9	129	236	247	2	»0	1052	684
71	3	470	16	192	199	186	217	5	639	626
72	3	330	6	19	13	166	28	1	239	478
74	3	424	14	80	146	152	22	32	661	609
75	2	313	7	105	88	141	31	1	480	399
76	2	374	11	150	200	158	46	5	810	503
77	3	792	20	204	405	560	119	258	851	499
78	2	1796	197	400	630	1693	236	875	918	438
79	3	407	11	113	136	184	63	121	410	634

Nr	T	Na <sup>+</sup>	K <sup>+</sup>	Mg <sup>2+</sup>	Ca <sup>2+</sup>	Cl <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	NO <sub>3</sub> <sup>-</sup>	HCO <sub>3</sub> <sup>-</sup>	SiO <sub>2</sub>
84	3	509	7	171	221	313	161	211	451	426
85	3	425	9	88	221	172	47	82	590	606
86	3	570	25	183	200	284	18	60	870	663
87	3	470	2	92	123	247	89	1	451	613
90	3	535	11	150	181	252	63	47	761	506
92	2	452	14	92	145	189	58	»0	580	552
96	2	591	11	75	73	341	29	32	480	652
99	2	752	21	171	248	577	183	37	600	353
202	2	856	43	419	708	1320	654	840	139	573
203	2	677	25	196	272	379	209	185	610	440
204	2	527	11	114	126	258	865	158	352	657
205	2	748	23	296	375	545	202	380	716	485
206	2	731	21	156	207	689	145	37	472	441
207	2	708	13	236	422	631	238	500	244	489
208	3	433	11	75	88	184	86	92	328	474
209	2	567	21	107	213	235	161	240	367	532
210	3	362	11	45	75	186	38	82	389	626
211	2	716	69	232	381	402	192	71	1080	587
212	3	397	27	70	130	86	86	143	357	603
213	2	830	32	337	556	703	216	855	429	405
214	3	528	11	112	229	224	219	203	215	437
215	3	457	345	125	170	123	87	18	690	564
216	3	518	14	213	356	192	113	53	1113	660
217	3	541	11	90	225	247	45	3	787	654
218	2	604	88	302	1306	493	114	181	2995	635
219	3	710	53	238	457	775	69	877	167	465
220	2	1968	155	2163	1133	2524	524	1081	3656	264
221	2	495	11	94	139	212	68	3	477	522
222	3	538	23	172	285	264	97	216	642	634
223	3	530	11	218	308	258	185	435	326	485

Nr	T	Na <sup>+</sup>	K <sup>+</sup>	Mg <sup>2+</sup>	Ca <sup>2+</sup>	Cl <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	NO <sub>3</sub> <sup>-</sup>	HCO <sub>3</sub> <sup>-</sup>	SiO <sub>2</sub>
224	3	359	57	162	386	132	69	77	1156	411
225	3	393	15	71	155	135	113	58	372	472
226	2	387	16	132	66	114	79	85	436	545
227	3	409	27	80	186	126	77	226	367	581
228	3	667	41	201	328	272	171	177	836	750
229	3	880	40	140	230	545	81	205	557	666
230	3	666	11	136	210	206	159	21	664	546
231	3	592	15	127	330	181	224	124	626	508
232	2	565	33	75	213	178	40	435	334	745
233	2	1227	82	603	1165	803	257	2903	690	687
234	3	443	11	89	215	161	139	132	433	670
235	3	473	25	43	147	158	35	65	523	668
236	2	1265	74	503	1192	1377	572	1387	602	586
237	3	427	12	74	154	120	48	248	400	207
238	3	403	11	68	109	143	68	68	438	535
239	3	633	62	219	392	330	137	500	600	558
241	3	256	12	33	41	80	12	61	231	514
242	3	215	11	9	22	66	5	71	136	445
243	2	1097	109	486	763	890	258	435	1529	776
244	3	656	18	105	254	270	168	250	323	476
245	2	536	29	153	265	316	87	131	692	608
246	3	493	11	99	170	129	76	131	564	519
247	3	593	13	125	269	287	186	167	454	532
248	2	656	24	196	347	198	192	500	526	415
249	3	418	12	83	83	109	85	52	408	560
250	3	711	31	163	124	273	139	324	187	579
251	3	316	16	15	24	95	8	35	203	467
252	3	283	11	12	23	106	3	66	128	414
253	2	356	11	24	31	92	33	61	249	619
254	3	345	16	69	45	118	37	29	296	600

Nr	T	Na <sup>+</sup>	K <sup>+</sup>	Mg <sup>2+</sup>	Ca <sup>2+</sup>	Cl <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	NO <sub>3</sub> <sup>-</sup>	HCO <sub>3</sub> <sup>-</sup>	SiO <sub>2</sub>
255	3	383	11	66	47	126	36	66	293	672
256	3	385	11	38	32	132	38	35	236	520
257	3	474	11	72	60	146	50	140	243	613
258	3	184	11	12	14	98	5	66	59	237
259	3	635	29	210	234	267	87	190	723	740
260	3	481	12	102	140	143	31	43	648	760
261	2	1177	93	296	517	947	183	452	675	620
262	2	2312	113	700	1254	3385	375	1516	1048	617
263	2	1846	93	1051	1827	3213	725	1241	1724	740
264	3	549	16	114	252	338	102	250	451	550
265	3	407	11	59	90	155	22	9	567	697
266	2	1123	36	586	1018	1320	892	532	1278	486
267	2	938	46	294	78	574	56	182	1278	739
268	2	740	27	231	377	537	220	282	526	581
269	2	3889	193	663	1794	6886	559	1048	877	567
270	2	1016	55	799	855	1348	338	726	1169	452
271	3	481	18	72	856	218	25	139	449	842
272	3	365	26	85	126	264	6	187	367	573
273	3	359	19	65	144	166	12	4	554	723
274	3	492	11	108	105	310	52	150	470	530
275	3	417	27	89	107	287	13	105	516	662
276	3	547	15	135	223	488	35	176	531	615
277	3	447	49	143	191	367	59	113	606	736
278	3	359	11	54	75	207	18	113	375	760
279	3	690	22	158	212	602	67	118	688	692
280	3	271	43	43	90	155	11	13	434	583
402	3	165	24	25	34	116	7	22	150	211
404	3	266	24	43	105	192	10	32	308	399
406	3	987	33	127	195	634	75	60	853	692
407	2	1152	67	446	676	757	200	106	2081	757

Nr	T	Na <sup>+</sup>	K <sup>+</sup>	Mg <sup>2+</sup>	Ca <sup>2+</sup>	Cl <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	NO <sub>3</sub> <sup>-</sup>	HCO <sub>3</sub> <sup>-</sup>	SiO <sub>2</sub>
408	3	430	34	181	76	273	70	112	551	711
410	3	290	27	78	47	109	12	74	272	530
411	2	177	48	122	75	202	46	30	214	209
415	3	465	6	115	42	248	17	8	470	612
420	3	313	15	53	108	213	23	14	390	340
421	3	317	31	70	97	192	22	23	353	339
423	3	241	18	93	29	89	16	29	262	352
424	3	468	26	78	119	236	39	81	365	445
425	3	375	32	115	183	92	10	11	819	812
427	3	354	38	61	101	165	5	30	433	534
430	3	645	29	89	302	245	162	34	725	464
432	3	368	39	43	57	164	9	25	338	689
433	3	327	11	59	64	108	41	24	280	524
434	3	239	31	48	52	80	8	13	292	524
435	3	293	38	39	67	65	5	5	421	581
438	2	454	47	68	57	155	30	30	430	487
439	3	448	55	68	134	130	13	»0	714	709
440	3	398	36	68	134	129	29	7	636	729
441	3	472	47	129	237	172	33	8	957	875
442	3	680	58	166	283	846	25	75	607	838
443	3	475	45	156	147	384	33	33	549	569
444	3	301	38	95	60	160	13	36	351	442
446	3	488	34	102	171	260	14	33	651	887
447	2	707	72	207	373	542	95	107	974	548
452	3	669	50	198	332	344	176	105	838	774
453	3	607	36	197	253	280	46	17	1124	752
457	3	304	36	62	146	94	1	1	566	568
458	3	330	35	66	148	196	9	22	526	670
463	2	313	37	71	82	245	33	31	231	366
514	2	371	31	96	188	164	66	27	558	497

Nr	T	Na <sup>+</sup>	K <sup>+</sup>	Mg <sup>2+</sup>	Ca <sup>2+</sup>	Cl <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	NO <sub>3</sub> <sup>-</sup>	HCO <sub>3</sub> <sup>-</sup>	SiO <sub>2</sub>
522	2	442	30	241	294	184	47	32	1150	600
523	2	600	39	118	282	307	149	34	650	860
524	2	576	63	145	293	252	114	21	919	679
525	2	400	47	86	142	128	69	48	503	554
530	2	669	37	168	316	462	178	62	643	742
534	2	471	36	95	150	281	81	100	310	604
535	2	548	113	245	404	737	143	105	625	568
536	2	370	41	99	254	240	55	42	529	431
539	2	566	51	100	289	697	113	150	600	375
540	2	319	84	124	186	228	23	63	520	806
573	2	177	35	28	29	199	10	19	96	280
574	3	634	80	223	241	930	95	29	652	515
575	2	596	81	223	224	433	130	30	876	514
583	2	880	166	738	452	2390	134	284	305	415
589	3	302	88	121	146	208	19	24	502	679
591	2	450	134	183	285	419	209	20	572	228

## **ACKNOWLEDGMENTS**

I wish to thank C.H. Van der Weijden of the Utrecht University (Holland) for his geochemical advice. I also thank two anonymous referees for their constructive remarks and suggestions on an earlier version of this paper.

## REFERENCES

- Alencão, A.M.P., Sousa Oliveira, A., and Pacheco, F.A.L.,** 1996, Seasonal variations on systems of triple influence: In First European Conference on Geostatistics for Environmental Applications (geoENV96), Instituto Superior Técnico, Lisbon.
- Appelo, C.A.J., and Postma, D.,** 1993, Geochemistry, groundwater and pollution: Balkema, Rotterdam, 536p.
- Barnes, I., and O'Neil, J.R.,** 1971, The relationship between fluids in some fresh Alpine-type ultramafics and possible modern serpentinization, western United States: Geol. Soc. Am. Bull., v. 80, p. 1947-1960.
- Benzecri, J.P.,** 1973, L'analyse des données, v. 2 - L'analyse des correspondences: Dunod, Paris, 578p.
- Benzecri, J.P.,** 1977, Histoire et préhistoire de l'analyse des données, v. 5 - L'analyse des correspondences: Cahiers de l'analyse des données, v. 2, p. 9-40.
- Cattell, R.B.,** 1966, The scree test for the number of factors: Mult. Behav. Res., v. 1, p. 245-276.
- Costa, C.V., Pereira, L.G., Portugal Ferreira, M.R., and Santos Oliveira, J.M.,** 1971, Distribuição de oligoelementos nas rochas e solos da região do Fundão: Memórias e Notícias (Coimbra University), v. 71, p. 1-37.
- Davis, J.C.,** 1986, Statistics and data analysis in geology: John Wiley & Sons, New York, 646 p.
- Drever, J.I.,** 1988, The geochemistry of natural waters: Prentice Hall, New Jersey, 437p.
- Dougenik, J.A., and Sheenan, D.E.,** 1976, SYMAP user's reference manual: Technical report, Massachusetts University, 170p.



- Frapporti, G.**, 1994, Geochemical and statistical interpretation of the dutch national ground water quality monitoring network: Ph. D thesis, Utrecht University, 119p.
- Garrels, R.M.**, 1967, Genesis of some ground waters from igneous rocks: In P.H. Abelson (ed), *Researches in Geochemistry*, Wiley, New York, v. 2, p. 405-420.
- Garrels, R.M., and MacKenzie, F.T.**, 1967, Origin of the chemical compositions of some springs and lakes: In R.F. Gould (ed), *Equilibrium Concepts in Natural Water Systems*, *Advances in Chemistry Series 67*, Am. Chem. Soc., Washington, p.222-242.
- Greenacre, M.J., and Hastie, T.**, 1987, The geometric interpretation of correspondence analysis: *J. Amer. Stat. Assoc.*, v. 82, p. 437-447.
- Heiser, W.J., and Meulman, J.**, 1983, Analyzing rectangular tables by joint and constrained multidimensional scaling: *J. Econometrics*, v. 22, p. 139-167.
- Jackson, J. E.**, 1991, *A user's guide to principal components*: John Wiley & Sons, New York, 569p.
- Jobson, J.D.**, 1991, *Applied multivariate data analysis*, v. 1 - Regression and experimental design: Springer-Verlag, New-York, 621p.
- Jöreskog, K.G., Klován, J.E., and Reymont, R.A.**, 1976, *Geological factor analysis*: Elsevier, Amsterdam, 178p.
- Klován, J.E.**, 1975, R- and Q-mode factor analysis: In R.B. McCammon (ed), *Concepts in geostatistics*, Springer-Verlag, New York, p. 21-69.
- Likens, G.E., Bormann, F.H., Pierce, R.S., Eaton, J.S., and Johnson, N.M.**, 1977, *Biogeochemistry of a forested ecosystem*: Springer-Verlag, New York, 135p.
- Matthess, G.**, 1982, *The properties of groundwater*: John Wiley & Sons, New York, 405p.

- Miller, W.R., and Drever, J.I.**, 1977a, Water chemistry of a stream following a storm, Absaroka Mountains, Wyoming: Geol. Soc. Am. Bull., v. 88, p. 286-290.
- Miller, W.R., and Drever, J.I.**, 1977b, Chemical weathering and related controls on surface water chemistry in the Absaroka Mountains, Wyoming: Geochim. Cosmochim. Acta, v. 41, p. 1693-1702.
- Pacheco, F.A.L.**, 1995, Interacção água rocha em unidades do Grupo Peritransmontano (Serra da Padrela, Vila Pouca de Aguiar): M Sc thesis, Trás-os-Montes e Alto Douro University, 123p.
- Pacheco, F.A.L.**, submitted, 1996, Clustering groundwater data sets using the concept of equivalence class: Computers & Geosciences, paper number ms#96.112.
- Pacheco, F.A.L., and Portugal Ferreira, M. R.**, 1996,. Sistemas hidroquímicos com dupla influência, sua definição e interpretação no contexto da análise de correspondências: In 3º Congresso da Água/VII Silubesa, Associação Portuguesa de Recursos Hídricos (APRH), LNEC, Lisbon, v. 3, p. 177-185.
- Pacheco, F.A.L., and Van der Weijden, C. H.**, 1996, Contributions of water-rock interactions to the composition of groundwater in areas with sizeable anthropogenic input. A case study of the waters of the Fundão area, central Portugal: Water Resour. Res., v. 32, p. 3553-3570.
- Pearson, K.**, 1901, On lines and planes of closest fit to systems of points in space: Phil. Mag., Series B, v. 2, p. 559-572.
- Portugal Ferreira, M.R.**, 1982, A magmatic arc in the Iberian Segment of the Hercynian Chain: I- the northwest-southeast lineament between Oporto (Portugal) and Zarza la Major (Spain): Memórias e Notícias (Coimbra University), v. 94, p. 31-50.
- Portugal Ferreira, M.R., Ivo Alves, E., and Regêncio Macedo, C.A.**, 1985, A zonalidade interna de um plutonito, estruturas condicionantes e idades de evolução (plutonito do Fundão, Portugal central): Memórias e Notícias (Coimbra University), v. 99, p. 167-186.

- Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T.**, 1989, Numerical Recipes in Pascal: Cambridge University Press, Cambridge, 759 p.
- Reeder, S.W., Hitchon, B., and Levinson, A.A.**, 1972, Hydrogeochemistry of the surface waters of the Mackenzie river drainage basin, Canada. Factors controlling inorganic composition: *Geochim. Cosmochim. Acta*, v. 36, p. 825-865.
- Sousa Oliveira, A.**, 1995, A hidrogeologia da região de Pedras Salgadas: M Sc thesis, Trás-os-Montes e Alto Douro University, 230p.
- Spearman, C.**, 1904, General intelligence, objectively determined and measured: *Amer. J. Psychol.*, v. 15, p. 201-293.
- Thompson, G.H.**, 1934, Hotelling's method modified to give Spearman's  $\rho$ : *J. Educ. Psych.*, v. 25, p. 366-374.
- Van der Heijden, P.G.M., de Falguerolles, A., and Deleeuw, J.**, 1989, A combined approach to contingency table analysis using correspondence analysis and log-linear analysis: *Appl. Stat.*, v. 38, p. 249-292.
- Van der Weijden, C.H., Oosterom, M.G., Bril, J., Walen, C.G., Vriend, S.P., and Zuurdeeg, B.W.**, 1983, Geochemical controls of transport and deposition of uranium from solution. Case study: Fundão, Portugal: Technical Report, Utrecht University, Institute of Earth Sciences, Department of Geochemistry, EC contract 007.79.3 EXU NL.
- Zobrist, J., and Stumm, W.**, 1981, Chemical dynamics of the Rhine catchment area in Switzerland. Estimation of the "pristine" Rhine river input to the ocean: Proceedings of the review and workshop on river inputs to ocean systems, Food and agriculture organization, Rome.

## **TABLE LEGENDS:**

**Table 1:** Weathering, effluents and fertilization classes of influence as established by the w%-Pollution and w%-Agriculture scores (Equations 9 and 10).

**Table 2:** Results obtained with CA (eigenvalues and factor loadings were multiplied by 1000).

**Table 3:** MLR model relating hydrochemical parameters with factors.

## FIGURE CAPTIONS

**Figure 1a:** Distribution of the CA  $F_1$  loadings in four different data sets.  $n$  = Number of groundwater samples; % $F_1$  = Percentage of data variation explained by  $F_1$

**Figure 1b:** Distribution of the  $pc_1$  loadings in four different data sets.  $N$  = Number of groundwater samples; % $pc_1$  = Percentage of data variation explained by  $pc_1$ .

**Figure 2:** Relation between the  $pc_1$  and the Ec scores of the Fundão groundwater samples.

**Figure 3a:** SCREE plots obtained for the Fundão data set.

**Figure 3b:**  $F_1$ - $F_4$  loadings trends of the global and anions models.

**Figure 4:** Dendrogram showing the tree structure of hydrochemical parameters. Adapted from Alenção, Sousa Oliveira and Pacheco (1996).

**Figure 5a:** Geographical map of central Portugal indicating the position of the Fundão area. Reproduced from Van der Weijden and others (1983).

**Figure 5b:** Blow-up of the Fundão area with the major villages (dashed polygons) and connecting roads (double lines), the rivers Zêzere and Meimoa and their tributary streams and streamlets. Distribution of 160 springs and wells water samples (labeled circles). Modified after Van der Weijden and others (1983).

**Figure 6:** Areas with weathering-, urban pollution- and agriculture-dominated groundwater chemistries as established by the method outlined in Table 1.

**Figure 7:** Distribution of the ratios  $[SO_4^{2-}]/[Cl^-]$  in the samples with agriculture-dominated chemistries (samples with w%-pollution > 50% and w%-Agriculture > 50%).

**Table 1**

Class	w%-Pollution	w%-Agriculture
Weathering	<50	
Effluents	>50	<50
Fertilization	>50	>50

**Table 2**

F	Variance distribution			Correspondence factor loadings (x1000)				
	$\lambda$ (x1000)	%F	Cum-%F	HCO <sub>3</sub> <sup>-</sup>	Cl <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	NO <sub>3</sub> <sup>-</sup>	SiO <sub>2</sub>
1	190.5	55.3	55.3	313.4	-443.0	-370.8	-748.9	421.0
2	74.8	21.7	77.0	92.0	377.0	-63.4	-591.6	-160.3
3	44.9	13.0	90.0	-202.6	178.3	-346.5	35.7	210.3
4	34.6	10.0	100.0	164.6	14.6	-394.1	235.5	-114.7

Symbols: F - factor;  $\lambda$  - eigenvalue; %F- percentage of data variation explained by F; Cum-%F - cumulative %F.

**Table 3**

Hydrochemical parameter	B <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	R <sup>2</sup> (%)
w%-Pollution	51.4	1.4	0.6	99.1
w%-Agriculture	22.1	68.5	12.0	77.9
$[\text{HCO}_3^-]/[\text{SiO}_2]+[\text{SO}_4^{2-}]/[\text{Cl}^-]$	0.2	0.2	3.2	81.4

Symbols: B<sub>i</sub> - standardized regression coefficient for factor i; R<sup>2</sup> - adjusted coefficient of multiple determination; w%-Pollution and w%-Agriculture are calculated by Equations 9 and 10, respectively.



**Figure 1a**

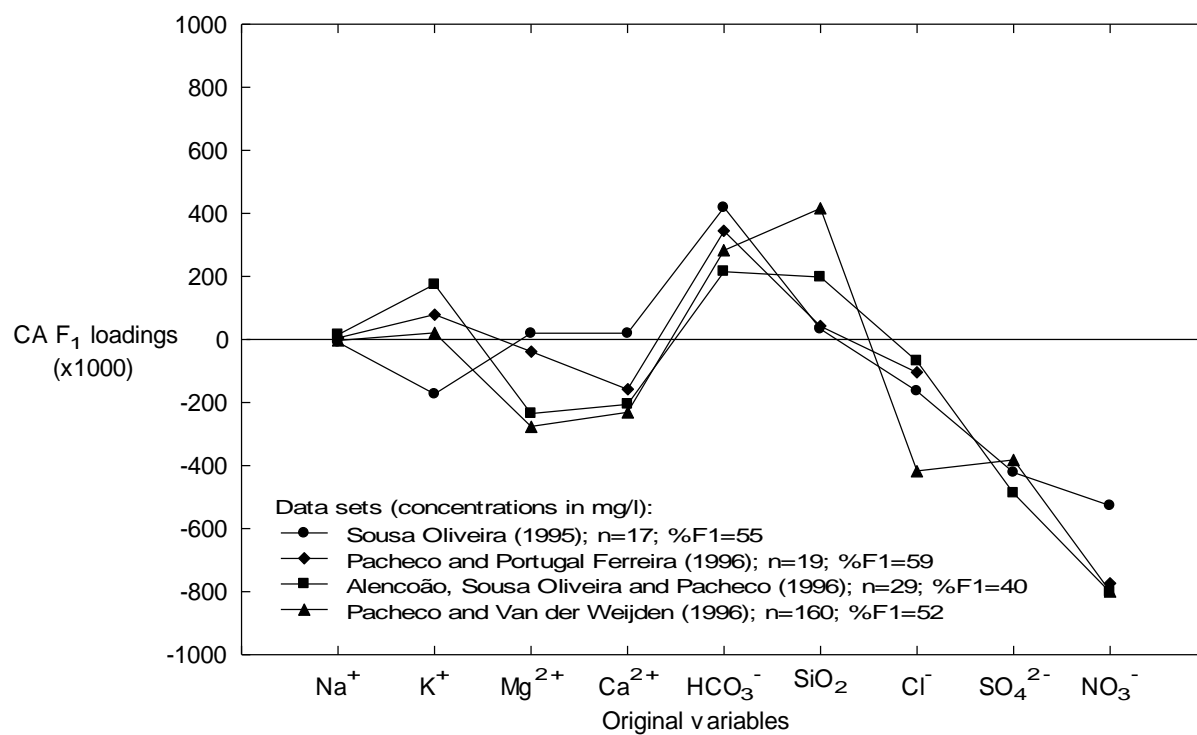
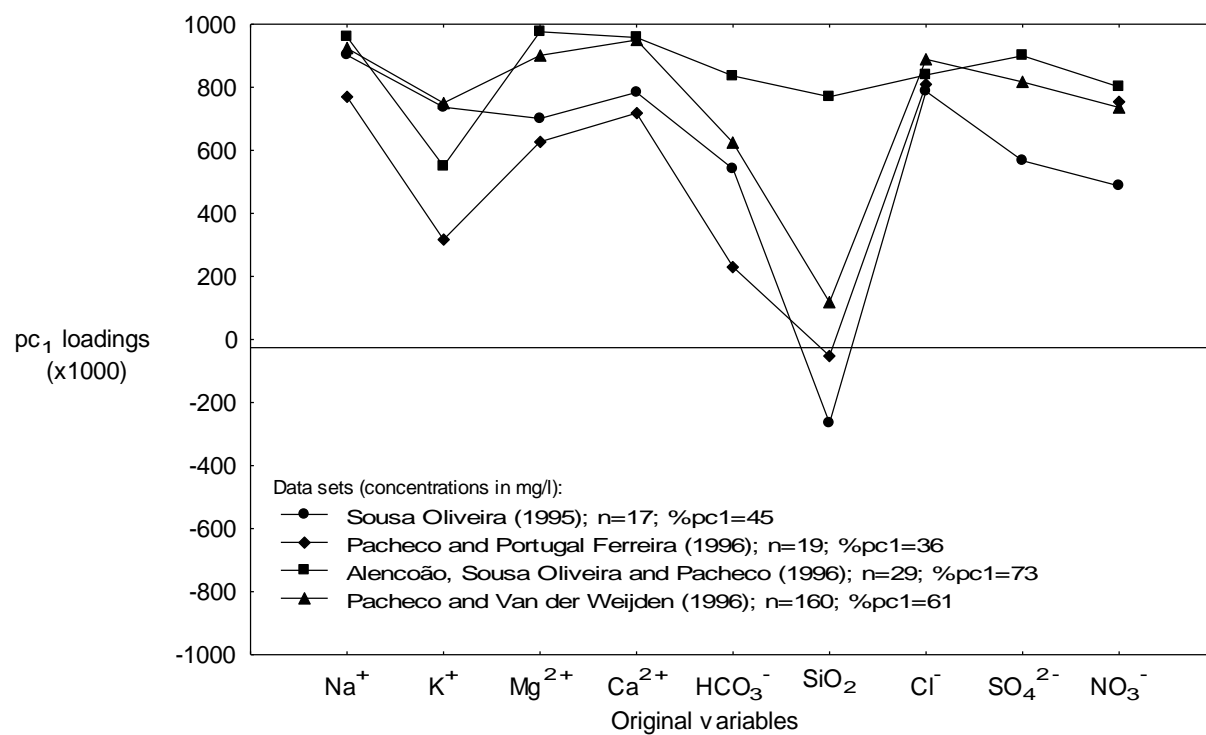
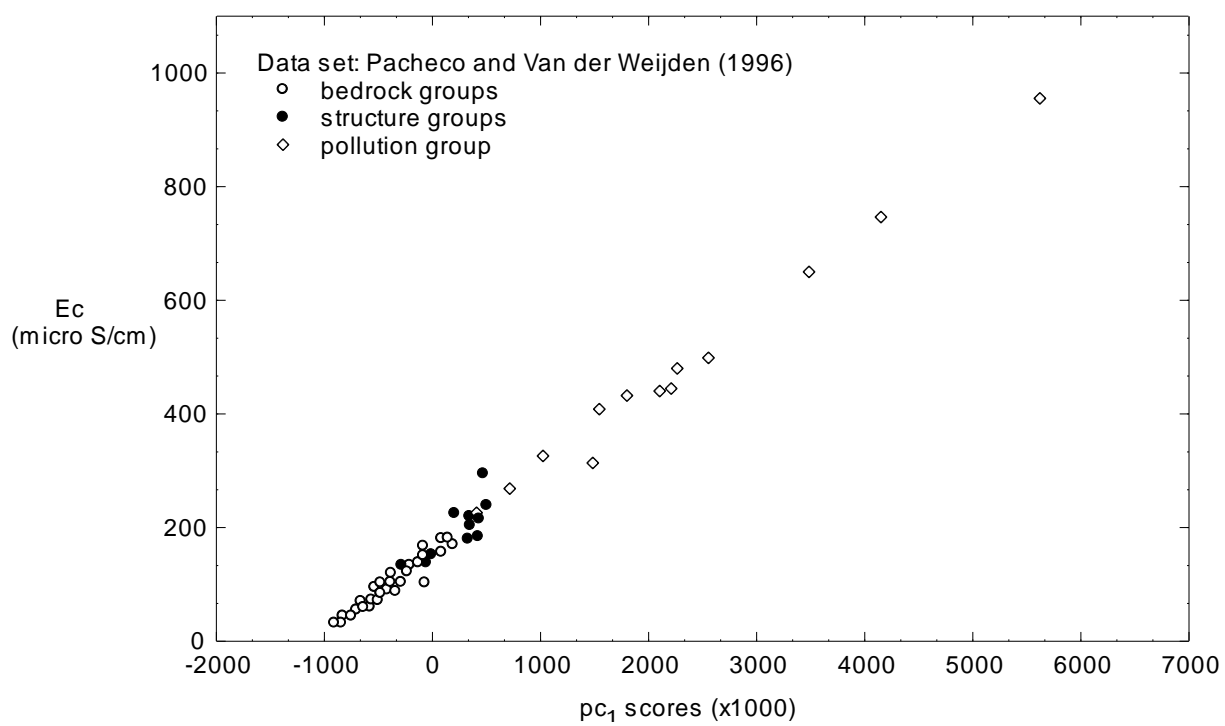


Figure 1b



**Figure 2**



**Figure 3a**

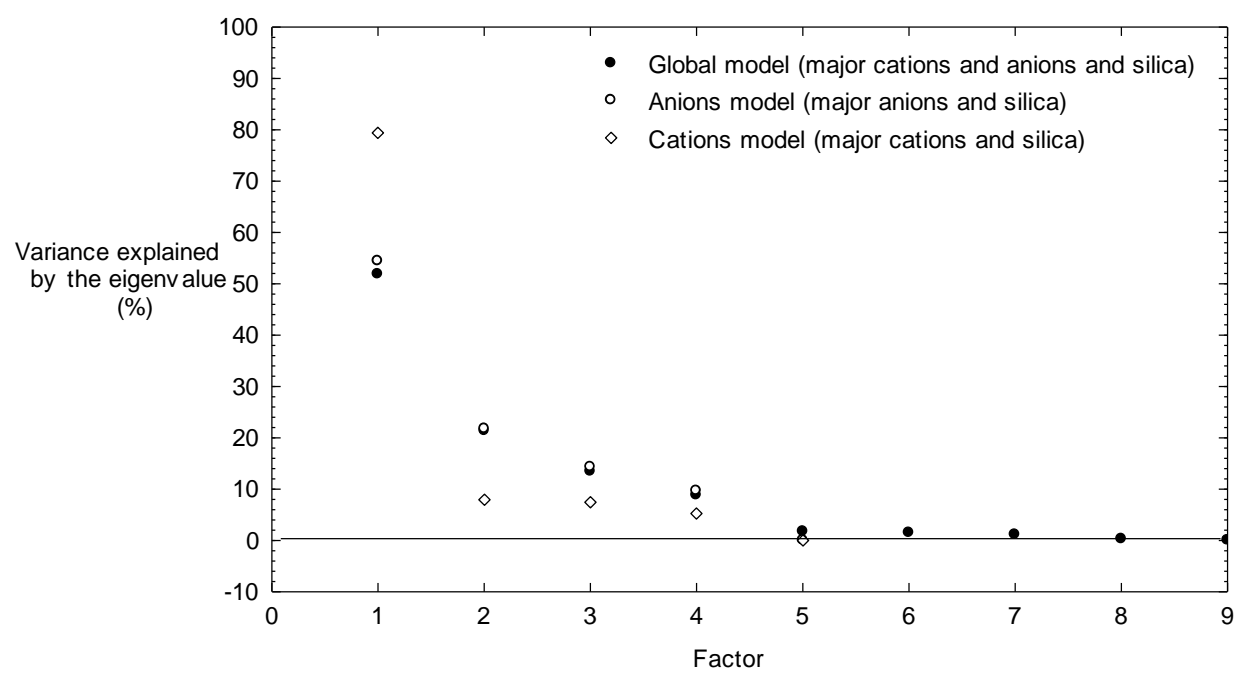
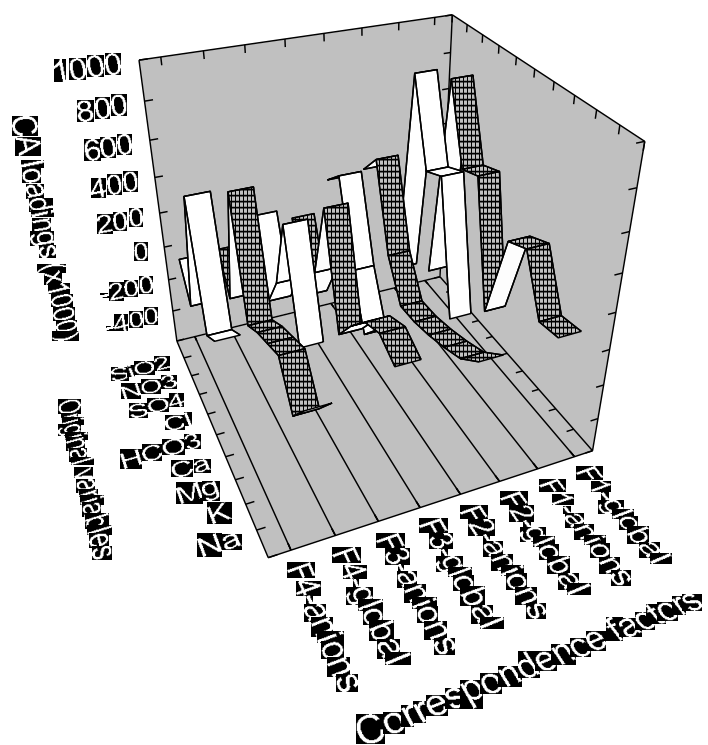
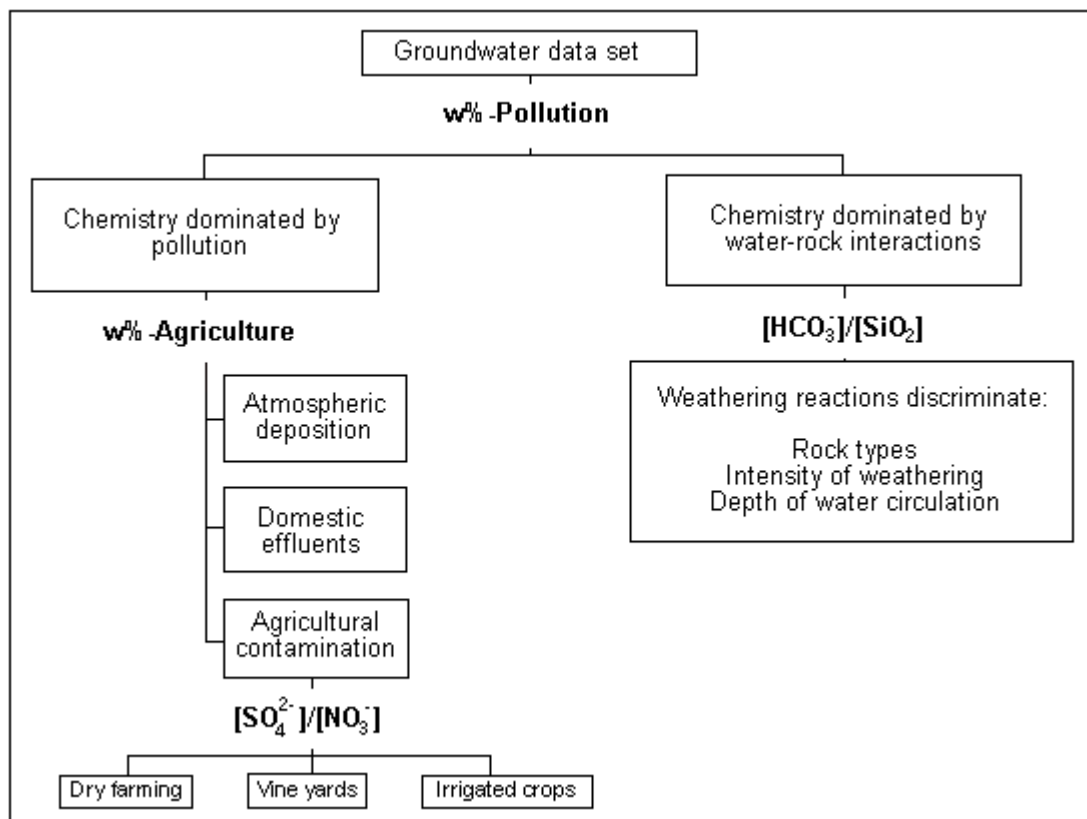


Figure 3b



**Figure 4**



**Figure 5a**

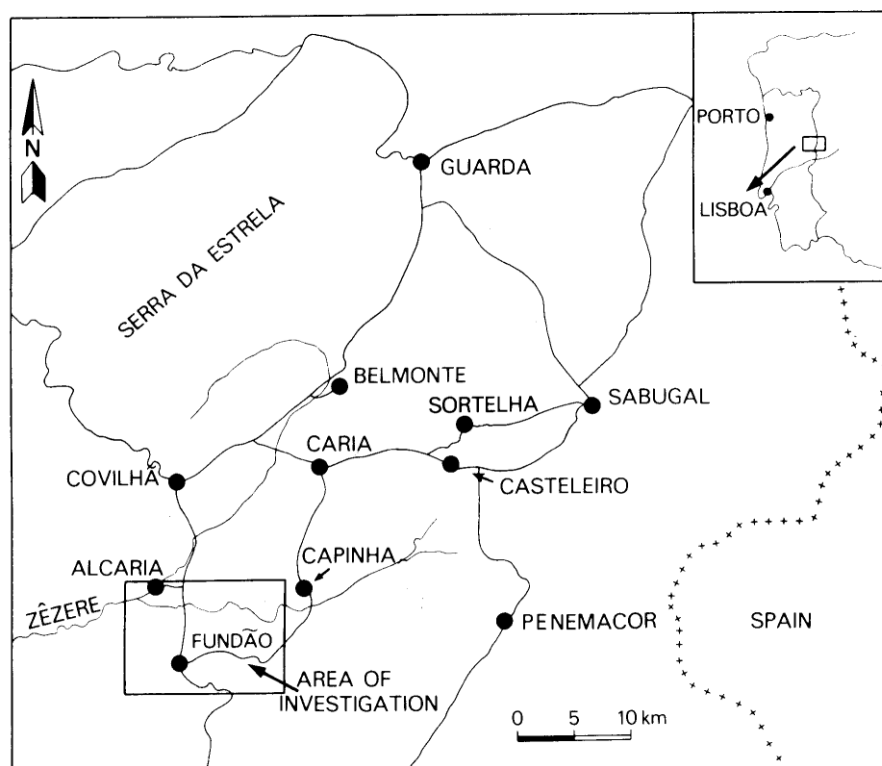
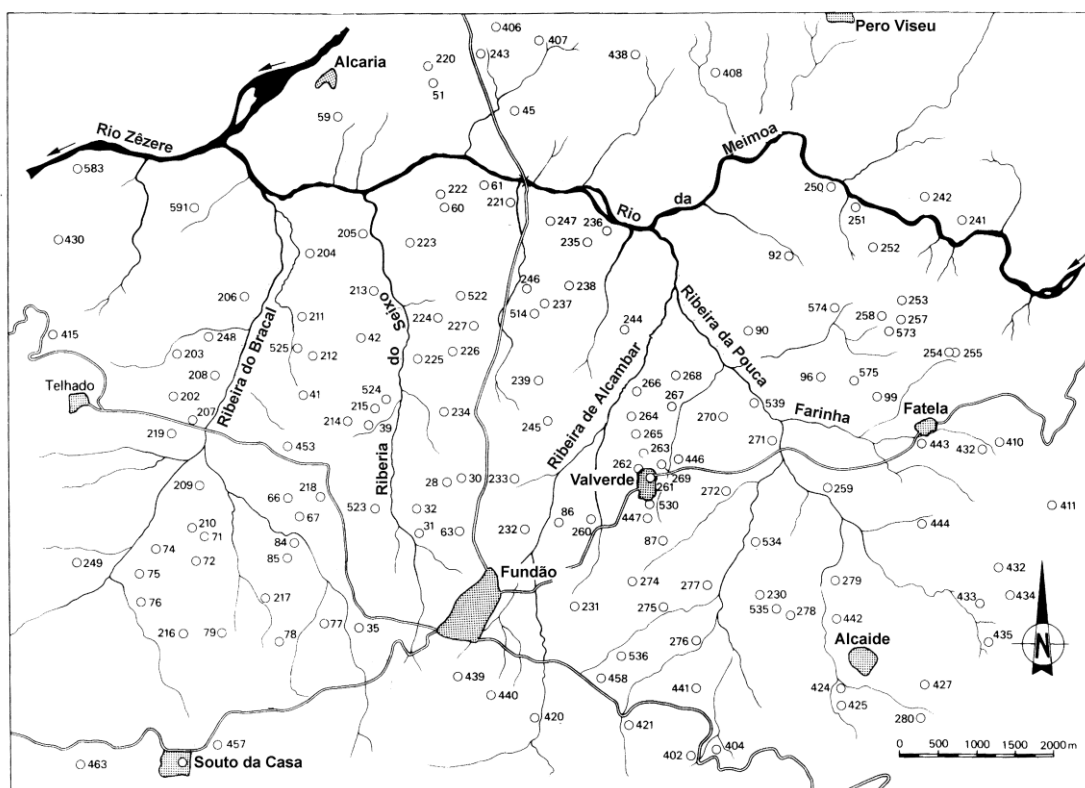
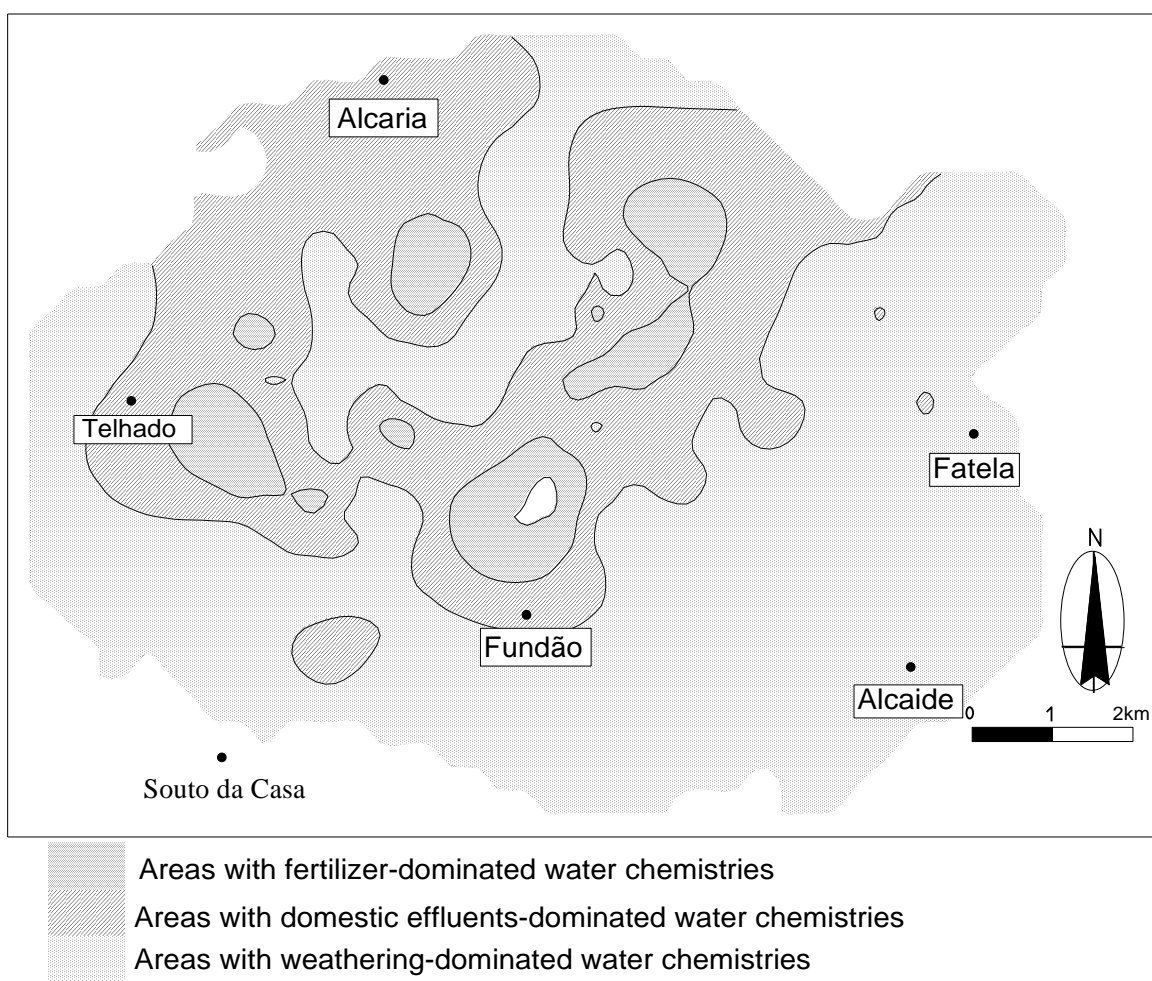


Figure 5b





**Figure 6**



**Figure 7**

