

Elements for a methodology to interpret hydrochemical data.

1 Summary

The EWFD(European Water Framework Directive) mandates reaching a good status for water bodies. So, monitoring of our water sources in order to administer, protect and plan water use in associated areas is mandatory. In order to accomplish the EWFD objectives the Agència Catalana de l'Aigua has a large ground water quality network monitoring points all over the territory. This network consists of a series of wells that measures a large number of parameters which lead to a huge amount of data to be interpreted. The size of the database poses a tremendous challenge for interpreting, and thus, there is a need to explore the possibility of improving data representation and initial data analysis.

The aim of this thesis is to identify a series of techniques, tools or methods to ease a systematic interpretation of generic groundwater chemical data sets, assuming that the user might not be an expert in managing such kind of data. So, the following tasks have been performed: target area definition and study; techniques and tools decision; data acquisition, conversion and filtering and, finally, technique application.

As a result of these process, an appropriate data format has shown to be be mandatory in order to correctly export the data set to the needed codes. When applied, the most easily understood techniques have usually been graphic methods. In some cases their interpretability has been conditioned by data quality or code lacks so both fields are intended to be improved in future work lines. These includes taking advantage of Aquachem's modelling techniques and ArcGIS plotting advantages. On the other hand, the most interesting multivariate technique has been factor analysis, which has no unique interpretations and might have to be more studied in. So, on the whole, to validate the conclusions of this work it might be interesting to apply this same methodology on different data sets.

Contents

| | | |
|-------|--|----|
| 1 | Summary | 2 |
| 2 | Aknowledgements | 7 |
| 3 | Introduction | 8 |
| 4 | Objectives | 10 |
| 5 | Target area | 12 |
| 5.1 | Geology | 13 |
| 5.2 | Hydrogeologic units | 17 |
| 5.2.1 | Fluviodeltaic Area of the Ter river | 19 |
| 5.2.2 | Lower Ter paleogene | 27 |
| 6 | Methods | 30 |
| 6.1 | Data treatment and filtering | 30 |
| 6.2 | Evaluated software | 33 |
| 6.3 | Classical study techniques for chemical analysis | 38 |
| 6.3.1 | Column or Collins diagrams | 38 |
| 6.3.2 | Triangular or Piper diagrams | 40 |
| 6.3.3 | Circular and Pie diagrams | 44 |
| 6.3.4 | Poligonal, radial and modified Stiff diagrams | 47 |
| 6.3.5 | Vertical line and Schoeller-Berkaloff diagrams | 49 |
| 6.3.6 | Scatter plots | 51 |
| 6.3.7 | Frequency diagrams | 55 |
| 6.3.8 | Hydrograms | 56 |
| 6.4 | Multivariate statistics techniques. | 62 |
| 6.4.1 | Factor analysis | 65 |
| 6.4.2 | Cluster analysis | 73 |
| 6.5 | Modelling techniques | 77 |
| 6.5.1 | Saturation indexes | 77 |
| 6.5.2 | Sample mixing | 78 |
| 7 | Discusion of methods and conclusions. | 79 |
| 8 | Future work lines | 82 |

List of Figures

| | | |
|----|--|----|
| 1 | Target area with sample locations. | 12 |
| 2 | Geological map with stratigraphic units where samples are located. | 14 |
| 3 | Underground water areas with sample locations. | 18 |
| 4 | Deep and shallow Baix Ter aquifers map, provided by Geoservei, S.L. | 23 |
| 5 | Isopaques map, provided by Geoservei, S.L. | 24 |
| 6 | Baix ter aquifers lithostratigraphy , provided by Geoservei, S.L. | 25 |
| 7 | Baix Ter aquifers piezometry in year 1990 , provided by Geoservei, S.L. | 26 |
| 8 | Partial view of the original data set. | 30 |
| 9 | Partial view of the rewritten data set. | 31 |
| 10 | Rewritten and filtered data set. Although columns like Al, Sb or As seem void, most of their values are in lower rows of the matrix and they do not show up in this image. | 32 |
| 11 | Data base, Stiff diagram projected on map, piper diagram, histogram, radial diagram and depth profile done with Aquachem. | 34 |
| 12 | SPSS working environment. | 35 |
| 13 | ArcGIS interface. | 36 |
| 14 | R interface. | 37 |
| 15 | Column diagram made with ArcGIS | 39 |
| 16 | Piper diagram done with Aquachem | 41 |
| 17 | Piper diagram for year 2004. | 42 |
| 18 | Piper diagram with R | 43 |
| 19 | Pie chart of year 1995-2004 done with Aquachem | 44 |
| 20 | Pie chart of year 2004 done with ArcGIS | 45 |
| 21 | Radial diagram chart. | 47 |
| 22 | Stiff chart. | 48 |
| 23 | Schoeller diagram years 1996-2004. | 49 |
| 24 | Schoeller diagram year 2004 | 50 |
| 25 | Na-Cl scatter plot with Aquachem. | 51 |
| 26 | Mg(mg/l)-Cl(μ g/l) Scatter plot with spss | 52 |
| 27 | Scatter plot matrix done with SPSS | 53 |
| 28 | Scatter plot log-matrix done with SPSS | 54 |
| 29 | Al (μ g/l) histogram made with Aquachem | 55 |
| 30 | Al (μ g/l) histogram made with SPSS in μ g/l. | 55 |
| 31 | Ca(mg/l) vs. time | 56 |

| | | |
|----|--|----|
| 32 | Multiparameter station 17195-1 vs. Time. Low peak in year 99 analysis due to “wet” period. | 57 |
| 33 | Accumulated daily rain events for station 17195-1 prior to the 1998 sampling(red line). Dry period, 25 days without rain. | 57 |
| 34 | Accumulated daily rain events for station 17195-1 prior to the 1999 sampling(red line) in mm. Recent rain event, 47,5mm 7 days before. | 58 |
| 35 | Accumulated daily rain events for station 17195-1 prior to the 2000 sampling(red line). Dry period, only 12,5mm rain 10days before. | 58 |
| 36 | Multiparameter station 17195-3 vs. Time. High concentration in year 2000 analysis due to “dry” period. | 59 |
| 37 | Accumulated daily rain events for station 17195-1 prior to the 1999 sampling(red line) in mm. Wet period, recent rain event (47,5mm) 7 days before. | 59 |
| 38 | Rainfall events for station 17195-3 prior to the 2000 sampling(red line) in mm. Dry period, 19days without rain. | 60 |
| 39 | Rainfall events for station 17195-3 prior to the 2001 sampling(red line) in mm. Wet period, 3 high rain events, last 12 days before the analysis date. | 60 |
| 40 | Multiparameter station 17195-3 vs. Time. Changes in K , Na and Mg are more evident and in HCO_3 and SO_4 smoothened. | 61 |
| 41 | n-points cloud in p-dimensional space. | 62 |
| 42 | Reduction of a n point cloud to a 1D subspace. | 63 |
| 43 | Reduction of a n point cloud to a 2D subspace. | 64 |
| 44 | Projection of n points cloud onto a random line vs. projection of n points cloud onto line that maximizes the variance of the projected points. | 65 |
| 45 | Table of communalities. Its closeness to 1 gives an idea of how good a variable is explained by the obtained solution. | 67 |
| 46 | % of variance explained by each factor. In this case, only the 4 first factors are considered for the solution. | 67 |
| 47 | Components of the factors. The parameters with highest values are highlighted. So, a marine first factor can be deducted. | 68 |
| 48 | Factor analysis made with R. First factor clearly continues as marine but others change. | 69 |
| 49 | Factor loadings given by ArcGIS. | 70 |
| 50 | Representation of first factor with ArcGIS. | 70 |
| 51 | Representation of second factor with ArcGIS. | 71 |
| 52 | Representation of third factor with ArcGIS. | 72 |

| | | |
|----|--|----|
| 53 | Cluster analysis dendrogram. | 73 |
| 54 | Cluster analysis results, upper half. | 75 |
| 55 | Cluster analysis resulting dendrogram, lower half. | 76 |
| 56 | Saturation index siderite | 77 |
| 57 | Linear pH mixing | 78 |

2 Acknowledgements

First of all, I would like to thank Carles Ayora, Jesús Carrera, and Xavier Sánchez for jumping in head first to help guide me, and for all those brain storming sessions we had together.

I would like to thank, Agència Catalana de l'Aigua and Geoservei Projectes i Gestió Ambiental, S.L for their support of this work. I would specially like to thank Jordi Montaner for all his help. It was absolutely invaluable. I would also like to thank Josep Ma. Niñerola and Josep Fraile for their interest in this study.

Thanks to all my workmates who have encouraged me specially during the writing of this document and who have helped me revise my text. Special thanks to Anais, Belen, Beniamino, Carlos, Hans, Jordi, Jose Luís, Montse, Quim, Rodrigo and Úrsula, for their company in our office during some of my writing moments. And thanks to Josep and Ernest for being understandable when I needed to work on this. Thanks also to Manolo for showing his solidarity with this work and to Cesca for always being so helpful and encouraging when I needed office material. And, specially, thanks to all the rest of workmates I didn't mention, for understanding that I forgot to include them.

More thanks go to my parents and brothers, Jacqueline, Luc, Krl, Diego and Jan for those pleasant Sunday lunches we have had together. Thanks also go out to Flapy for getting even fatter and looking more like a black pudding, which provided me a source of inspiration.

Above all, many thanks to all my friends who have made me have many good times. I would specially want to thank, as a whole, Zaero for all these coffee breaks we've made together; Samba for being so funny when she gets exasperated with me; Albert for his maccaroni; Albert for that delicious burger he offered me; el Pendó for our casual meetings; Pendejo and Dolly for those Pic-zas sessions; Wapi for always having been there; Manu for his bread rolls; Lluís for his poetry; Binita for her Grinberg skills; Supertere, Ama10 and Teirraks for being on the other side; Papi for those family dinners we've had together; Teo for his online company while I hope we will soon see each other; Will for those Duvels we have to drink since a year ago; Sunyer for that cinema we have to go since two years ago; Gabri for not coming to ask about Carpa when I needed to work; Thohntha for being such a nice flatmate; future-Nacho for giving me Nachete a year ago; Tocinillo for being so sweet, Cosita for being so lovely, and Roxo for not being annoying 100% of the time, only 99%; my six little kitties for being my second family; Piarra for his Piumcataratapums; Micky for remembering me when I had to give in this work and, finally all the rest of friends I forgot to thank and that will understand that not naming them does not mean I don't thank them.

3 Introduction

The EWFD (European Water Framework Directive) mandates reaching a good status for water bodies. So, monitoring of our water sources in order to administer, protect and plan water use in associated areas is mandatory.

The Catalan Water Agency, also known as ACA (Agència Catalana de l'Aigua), has 53 ground water bodies identified in Catalunya which are important for supplying water for urban, industrial and agricultural use. They represent approximately 35% of used hydric resources, around 200 hm³. In order to accomplish the EWFD objectives there is a large ground water quality network monitoring points all over the territory. This network consists of a series of wells which are controlled with a pre-established periodicity and with the same diagnostic objective.

Basic network General parameters are analyzed in all the sub-networks of the basic network: conductivity, pH, hardness, TOC, major anions and cations (Ca, Mg, Na, K, HCO₃, SO₄, Cl), nitrogen compounds (NO₃, NH₄ and NO₂) and the metals Fe and Mn. Moreover, volatile organochlorinates, industrial solvents, dioxines and other metals are analyzed if considered convenient.

Salinity network Controls the state of seawater intrusion in Catalan coastal aquifers. It is divided into different salinity sub-networks defined by the coastal hydrogeological units.

Network of nitrate-vulnerable areas Its origin is Royal Decree 261/1996, which transposes Directive 91/676/CEE for the protection of waters against pollution caused by agricultural nitrates. Six vulnerable zones have been established that make up this network. The parameters controlled are: major anions and cations, nitrogen compounds (NO₃, NO₂, NH₄) and metals present in animal feeding (Fe, Mn, Co, Zn, Se and Cu).

Pesticide network This specific network was created with the objective of underscoring the incidence of pesticides in areas with significant agricultural activity. Levels are controlled for 25 organochlorinate compounds and 19 organophosphates in the insecticide group and 7 triazines in the herbicide group. Every year two controls are organized in pesticides sub-networks, the first in the months of April, May, June and July and the second in September and October.

This large number of parameters together with a large number of observation sites lead to a huge amount of data to be interpreted. All these data need to be analyzed with the objectives of understanding the functioning of the water bodies, so as to determine evolution trends and, eventually, define management actions. Water agencies find themselves faced to a herculean task due to the complexity of such job which often has to be performed by people not familiar with this area.

There are many different ways to focus groundwater interpretation and lots of techniques, tools and applications that can be used in order to reach these objectives. Classical groundwater analysis techniques as well as statistical methods might be interesting instruments for this purpose.

4 Objectives

The aim of this thesis is to identify a series of techniques, tools or methods to ease a systematic interpretation of generic groundwater chemical data sets, assuming that the user might not be used to manage such kind of data. As this methodology is supposed to be applied on huge amount data sets such as those managed by the ACA, our work has been focussed to deal with data coming from its database. In order to reach the aim of this thesis a sequence of different tasks has been defined:

- Target area definition and preamble study.

As a first step, the study area has to be defined and studied in order to know in advance which kind of characteristics can be found in the analyzed data set and which not. So, geological and, more specifically, hydrogeological information has to be collected from different sources. Moreover, an on site visit, guided by Jordi Montaner (Geoservei S.L.), has been done. Gathering all this information will give us a first approach to this hydrological area.

- Choice of techniques, tools and applications to evaluate.

When this first study is done, the techniques, tools and methods to use have to be chosen. This includes the graphics to be drawn, the statistical methods to apply as well as the software which might be helpful to develop these tasks. So, considering which programs are available and their characteristics, a series of analytical steps to perform will be designed.

- Data acquisition, filtering and treatment.

In order to perform the steps decided beforehand, an appropriate data set is indispensable. After acquiring groundwater analysis from the ACA web site, the format will be analyzed and adapted to the software with which the data will be managed. So, knowing the requirements of the applications that will be used and the original data set characteristics, a code to automatically convert data to the convenient format will be designed. A subsequent data filtering in order to exclude non valid values will also be necessary. This will be done just after the data acquisition and processing as well as during the application of the analysis techniques. Incongruent results will help us to determine invalid data during the analyzing process.

- Application of chosen techniques, tools and applications.

After the target area study, the choice of methods and the data acquisition and formatting, the methodological part of this thesis will start. The idea consists in performing systematically all the considered techniques on our data set using the different available codes. As a result, we will have a considerable amount of graphics and statistical results to compare.

- Comparison of results.

The obtained results will be used to make a comparison between the different applied techniques to decide which are more useful and easier to interpret. On the other hand, which software has been more helpful and also easier to use will be determined as well as the specific lacks of each one.

- Conclusion drawing.

The final idea is to determine which techniques combined with which codes might be the most “user friendly” methodology. So, all obtained results will be considered together with the difficulty or easiness to obtain them as well as how much time consuming they were. Also, the conclusions drawn with each technique and how useful it was to interpret will be taken into account in order to draw the final conclusions of this work.

5 Target area

The target area of our study is the Fluviodeltaic Area of the Ter river, an aquifer system located in the region of Baix Empordà, north of the Les Gavarres massif foothills. It's geographic location is shown in figure 1.

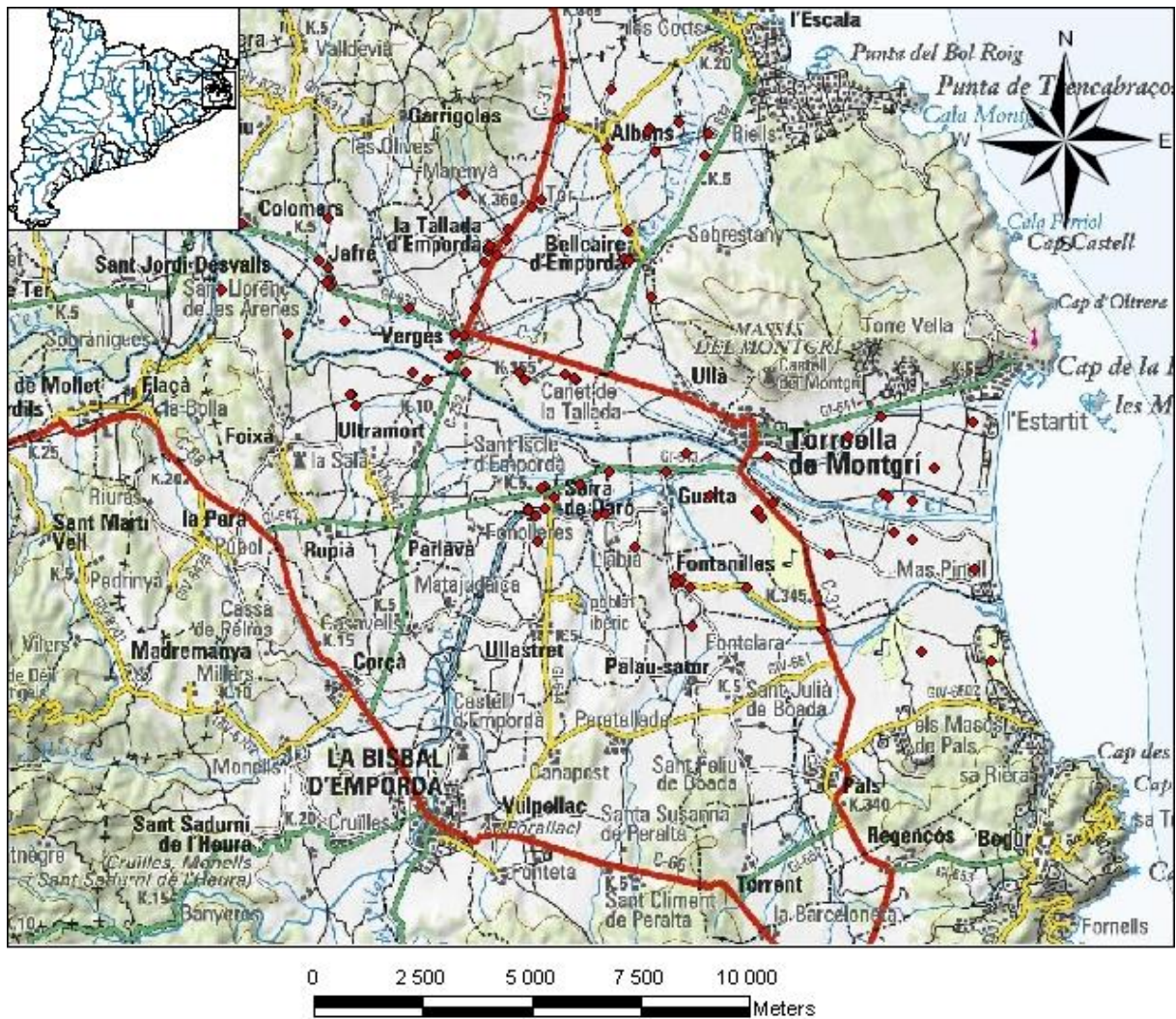


Fig. 1: Target area with sample locations.

5.1 Geology

A geologic introduction to this area has been extracted from the IGME¹ chart corresponding to Torroella de Montgrí [5] and geologic maps available at the ICC² web page [9] and is explained below.

Geologically there are three units with different stratigraphic and structural characteristics:

1. Les Gavarres massif.
2. Montgrí massif.
3. Empordà depression.

The main characteristic of the region is the existence of an autochthonous wide Paleogene series that lies discordantly over Paleozoic deposits consisting of a set of shales, affected by contact metamorphism due to granitoid post-tectonic intrusions that surface outside the study area. The Baix Empordà depression is filled by a series of deposits in continental and sometimes marine (Neogene) facies affected by late distensive fracturation stages. There exists a series of volcanic materials associated to those features which affect directly or indirectly to the tertiary series and that appear mainly close to the paleozoic deposits.

Another interesting characteristic to emphasize is the presence of the allochthonous Montgrí cretacic massif on the upper continental paleogene series on the northeast. The fluvial net, relief and climatic conditions during the Quaternary period have contributed to developing several fluvial, marine and eolic deposits which are very representative in the horizontal and vertical spread. As seen on figure 1, data samples are very spread, so they are located on different geological formations that are shown in figure 2.

The stratigraphic units where the wells are located (figure 2) have been studied with more detail.

¹ Instituto Geológico y Minero de España

² Institut Cartogràfic de Catalunya

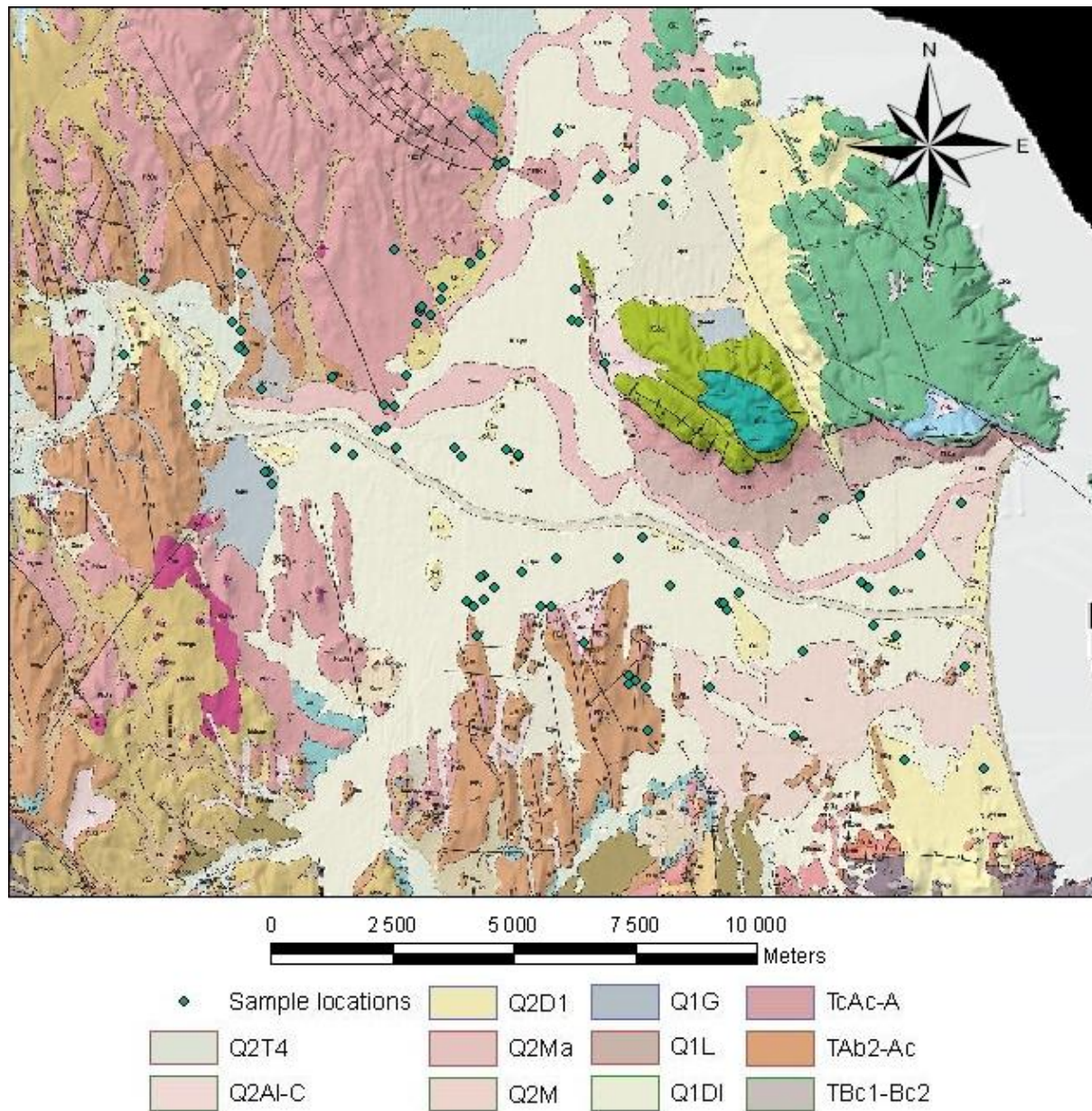


Fig. 2: Geological map with stratigraphic units where samples are located.

Stratigraphic units

Q_2T_4 **Fluvià river's 4th terrace.** It is the most modern terrace, approximately 2-3m above the river level. The actual river channel flows excavated in it. It consists of an heterometric and heterogeneous set of gravels and loss pebbles included in sands, clays and silts. It is located southwest of Colomers changing in that area to deltaic plain deposits corresponding to the lowest river trench.

$Q_2Al - C$ **Aluvial-coluvial.** It is very usual to find deposits with this combined genesis along streams. Its distinction is very difficult so both are grouped together. This facies consists of a set of fluvial characteristics mixed with lateral contributions. These are gravels, sands and limes in a clayey matrix that contains laterally many angular or subangular pebbles.

Q_2D_1 **Fixed Dunes.** These eolic formations have a remarkable development in this zone appearing from La Escala to de Montgrí massif. With fewer extension it also appears between the Ter and Daró rivers and upstream of the Ter river, between Jafre and Colomé. It is also present sporadically in the Albons corridor, in small depressions coincident with abandoned channels. The principal components of these eolic sands are quartz, metamorphic rock fragments and, with fewer proportions, biotite.

Q_2Ma **Abandoned meanders.** These old channels appear at the Ter river when it acquires low slopes, at Jafre and Verges. There are two principal abandoned channel courses. One follows approximately the current channel while the other one goes through the Albons corridor, splits and follows sinuous courses with higher curvatures. These old channels are easily explainable due to the low slope of the deltaic plain, the short depth of the river channel and the closeness to the coastline with few base level variations. This makes the river follow a neither definitive nor definite course, flowing through an almost flat extension. These abandoned meanders are easily identified by aerial photography and the presence almost constantly of medium sized sandy formations of almost 8m thickness through the Albons corridor confirms the existence of these channels. Lithologically it consists of sands filled with fine elements(clay and silt) mainly due to anthropic action.

Q_2M **Marshes.** They are basically in two spots: at the SE angle, close to the Daró river and N, developed at the Roses bay. They correspond to salty muds with sedimentation of a sandy fraction and a limey fraction. This last one fixes the abundant vegetation of bushes, reedbeds, brushes, etc. These places are flooded with salty waters where limes and grey muds are predominant with high abundance of organic matter.

Q_1G Accumulation glacis. Close to Foixà, between Juià and Celrà and at the northern part of Montgrí there exists a series of deposits with high superficial development and gentle slope. Lithologically it consists of an ochre and red clayey set filling subrounded and rounded discontinuous pebble levels with, sometimes, thin chalky levels interpreted as paleosoil. This happens at the Foixà glacis and at the Juià and Celrà too, where the current fluvial net has cut through these deposits.

Q_1L Screens. At the southern slope of the Montgrí massif, from Ullà until almost Estartit, there exists a series of clayey and ochre limestone deposits with irregularly distributed angular pebbles. There are also present some old and ephemeral eroded glacis deposits. This unit is considered as piedmont deposits coming from the meteorization of the southern border of the Montgrí massif.

Q_1Dl Deltaic plain deposits. They are located in the Albons corridor and the lower course and mouth of the Ter river, forming most of the alluvial plain, between the Fluvià and Ter rivers. It consists of fine grain deposits, muds, silts and sands with a considerable soil developed on it. These are the filling materials of the depression that spreads out from the Roses bay to the Pals beach, filled with deposits of the most important fluvial channels.

T_{21-3}^{Ac-A} Upper Eocene-Oligocene red claystone with conglomerate levels and grey sandstone. This unit's lithology is very variable because of the continental regime and orogenic movements on it. On the southern half its lithology consists of a reddish set due to its content of clay in which there are sandstones and conglomerate levels with scarce lateral continuity. In the northern half, close to the Montgrí, its formed of chalky cemented conglomerates. This series obeys to a continental fluvial environment, with frequent conglomerate channels and flooding zones, indicating presence of emerged massifs as well as a regional climax tectonics moment.

T_{22-21}^{Ab2-Ac} Clays and blue-grey limestones alternately with sandstones. Locally conglomerate intercalations at the top. This unit consists of a very characteristic grey-blue marine series formed by clays, organogenic limestones and some sandstones. It presents ripples, burrows and large scale cross bedding structures. This set is a typical continental shelf marine environment where it is usual to find foreshore bars and shoreface deposits. The existence of thick conglomerate levels responds to the existence of torrential braided delta fans. These are streams that flow into the sea whose front is reworked by the swell, acquiring sometimes some deltaic front aspect.

$T_{12-12}^{Bc^1Bc^2}$ **Sands, conglomerates, clays and marls with chalky levels.** These deposits present a high heterogeneity. The border facies are more detritic while the middle depression facies have finer granulometry and frequent carbonatic levels. The marly levels correspond to more basal trenches. As a whole, it is considered as a continental sedimentary environment.

5.2 Hydrogeologic units

As a consequence of the latest environmental law reforms, every public organism must have their environmental information available for anyone who wants to use it. So, the ACA has on its web page [6] extensive information about these two water bodies that has been used to describe them next. Further information has also been provided by Geoservei, S.L. [8] and is also provided in the following section.

The Ter and Fluvià rivers, generate vast alluvial plains on their lowest course. These plains define the morphological characteristics of most of the area. We can find unconfined and confined aquifers. Water quality is threatened by irrigation return flow and river recharge. Moreover, there is risk of saline intrusion in the coast. Figure 3 shows which 2 of the 53 ground water bodies identified in Catalunya are involved in this area: the Fluviodeltaic Area of the Ter river and the Lower Ter Paleogenes.

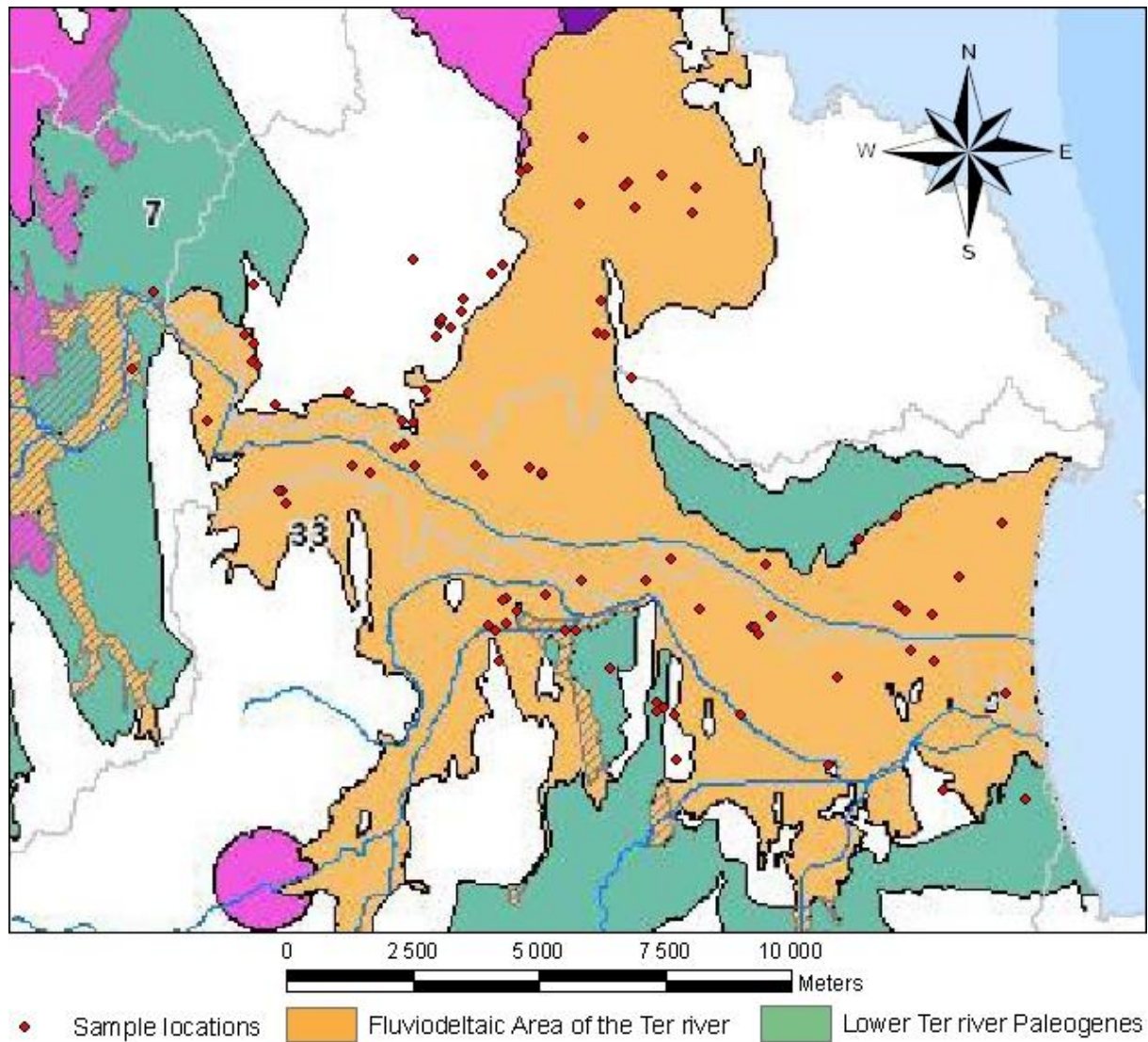


Fig. 3: Underground water areas with sample locations.

5.2.1 Fluviodeltaic Area of the Ter river

Characteristics summary:

- Hydrogeological area code: 402 .
- Total surface: $165Km^2$.
- Predominant lithology: Alluvial.
- Main Hydraulic Characteristics: Unconfined and confined aquifers with predominance of unconfined.
- Presence of multiple-layer aquifers.
- Main flow: Porous.
- General flow direction: W-E(Ter) and S-N(Daró) with local and temporal variations.
- Permeability: 100-1000 m/d
- Transmissivity: $2500-11000m^2/d$
- Storage coefficient: 15-20
- Variable influent-effluent interaction with the Alt Empordà and Baix Empordà marshes and the Boada and Bellcaire ponds.
- Hydrochemical facies: Sodic and calcic chlorates.
- Other: Littoral zone with marine intrusion risk.

As seen in figure 3, this water body is the main location of our samples. So, the characteristics of most them will be similar to it. It is contained in the fluvial terraces T_1 and T_2 , alluvial units associated to the Celrà basin, and in sediments corresponding to the T_1 and T_2 fillings of the Ter and Daró river plains. The dominant lithology consists of gravels and sands with presence of limestones and clays. This set of sediments can be grouped according to its lithology and internal geometry in three sedimentary aquifers:

- The surface aquifer of the Celrà basin, which has physical continuity with the Surface aquifer of the Baix Ter and Daró alluvials.
- Surface aquifer of the Baix Ter and Daró alluvials. It consists of an unconfined surface aquifer with 15-20m of thickness formed by gravels at the Celrà basin and more sandy sediments at the distal areas.
- Deep Aquifer of the Baix Ter river alluvial plain, a deep confined aquifer with 20m of thickness, consisting of sands and gravels and connected to the surface aquifer in proximal areas. It spreads out from the middle area of the plain, at Torroella de Montgrí, towards the sea, disappearing before the coast line.

At the Celrà-Verges-Torroella de Montgrí and Daró river areas the set of materials forms one only unconfined aquifer. In the Ter river domain, at the plain, and from Torroella de Montgrí towards the sea, there are two distinct aquifers, separated until the area where the deep aquifer disappears by an intermediate unit consisting of 15-25m thick limestone-clay lithology with sand intercalations with a multi-layer aquitard behavior.

The limits of these water mass correspond to litostratigraphic geologic contacts or, occasionally, hydrodynamic limits. The northern limit is defined by the connection area between the Ter and Fluvià plains with a low hydraulic gradient and an undefined flow area that separates the Fluvià and Ter river domains. Westwards and at the SW its contour is defined by the cartographic limit of the quaternary filling of the alluvial plain. The South limit coincides with the contour of the quaternary filling of the Daró plain.

The natural recharge of this area is produced by direct rain infiltration, infiltration from the Ter and Daró rivers, recharge from the Pla de la Mata travertins and Terri alluvials aquifer and from the Upper barthonian detritic aquifer. The recharge is done mainly through the area that comes up to surface and through the Ter and Daró rivers at the influent stretches. The discharge is mainly at the surface aquifer close to the distal stretch of the Ter river and from the surface aquifer towards the Mediterranean sea. Due to the geological variability, the predominant flow is porous. There is regional flow from W to E, following the Ter course, and from S to N, following

the Daró course, with temporary periods and non delimited stretches with influent or effluent regimes. The shallow aquifer hydraulic gradient is about 10^{-3} order.

The piezometric story has some remmarcable facts affecting the deep aquifer at the Gualta and L'Escala zones. In 1969 there some starting detections of an incipient piezometric depression due to pumping at Gualta. During the 1982 to 1985 years the Gualta funnel is very marked and at the supplying well zone of L'Escala really outstanding. At the years 1988-1989 the funnels become more pronounced. Nowadays both funnels keep there. On the whole there exists continuous flow from the Pla de Mata travertins and Terri alluvials aquifer towards the unconfined shallow aquifer of the Celrà basin and the Ter plain. There is also continuous discharge flow from the shallow aquifer towards the sea.

The non saturated zone lithology is mainly limestone-sandy and its thickness ranges from 5 to 14,7m at the Celrà basin and 0 to 5m at the Baix Ter zone, depending on the proximity to the coast line. The multi-layer aquitard stretch that separates the unconfined shallow aquifer from the semiconfined deep aquifer at the middle part of the plain can reach 15-20m of thickness and its lithology is limestone-clayey with intercalations of sandy levels . Close to the coastline, the deep aquifer disappears completely in favour of these fine granulometry sediments.

The Ter river stretch through the Celrà basin and the Daró river are predominantly influent superficial waters. Most of the Ter river during its way through the alluvial plain is mainly effluent. At the intermediate part of the Ter river there is variable effluent-influent flow.

Historical chemical status: (1982-2002)

| Parameter | Value | Units |
|------------------|--------|-------|
| HCO ₃ | 348,00 | mg/l |
| Ca | 118,00 | mg/l |
| Cond-lab | 751,00 | uS/cm |
| Cl | 55,20 | mg/l |
| Mg | 7,29 | mg/l |
| pH-lab | 7,25 | u.pH |
| K | 3,71 | mg/l |
| Na | 72,50 | mg/l |
| SO ₄ | 72,00 | mg/l |
| Fe | 180,00 | μg/l |
| Mn | 40,00 | μg/l |

The most important affections to this water mass are due to coastal zone water extractions, which produce saline water intrusion, and to intensive agriculture. Use of purines, fertilizers and plaguicides affects the quality of the underground waters. It affects mainly the shallow aquifer which presents the highest mean nitrate levels, over 50mg/l, and less to the deep aquifer, with a mean nitrate level of 24,6mg/l. This agricultural affection also reflects on some detectable plaguicide concentrations in the deep and shallow aquifers.

Current chemical status:³

| Parameter | BTAPDA | CBSA | BTDAPSA |
|-------------------------|----------|--------|---------|
| pH(u.pH) | 7,52 | 7,57 | 7,46 |
| CE(uS/cm) | 1798,0 | 854,7 | 2181,29 |
| HCO ₃ (mg/l) | 427,68 | 362,03 | 388,8 |
| Cl(mg/l) | 364,2 | 65,9 | 639,3 |
| SO ₄ (mg/l) | 167,67 | 89,33 | 149,29 |
| Ca(mg/l) | 174,33 | 149,07 | 148,24 |
| Mg(mg/l) | 49,28 | 14,33 | 59,81 |
| Na(mg/l) | 186,48 | 52,34 | 60,61 |
| K(mg/l) | 11,88 | 4,55 | 19,04 |
| NO ₃ (mg/l) | 24,6 | 43,4 | 33,2 |
| NH ₄ (mg/l) | 0,83 | 0,09 | 0,8 |
| Fe(μ g/l) | 14951,41 | - | 7237,0 |
| Mn(μ g/l) | 764,44 | 35,5 | 603,0 |

Average chloride concentrations are higher than 250mg/l and sulphate and nitrate levels are high on specific samples, up to 525mg/l and 107mg/l respectively.

As this is the main location of our samples, further information has been provided by Geoservei, S.L. The following figures, from 4 to 7 help to get a general vision of these water bodies:

³ BTAPDA=Baix Ter alluvial plain deep aquifer; CBSA=Celrà basin shallow aquifer;BTDAPSA=Baix Ter&Daró alluvial plains shallow aquifer

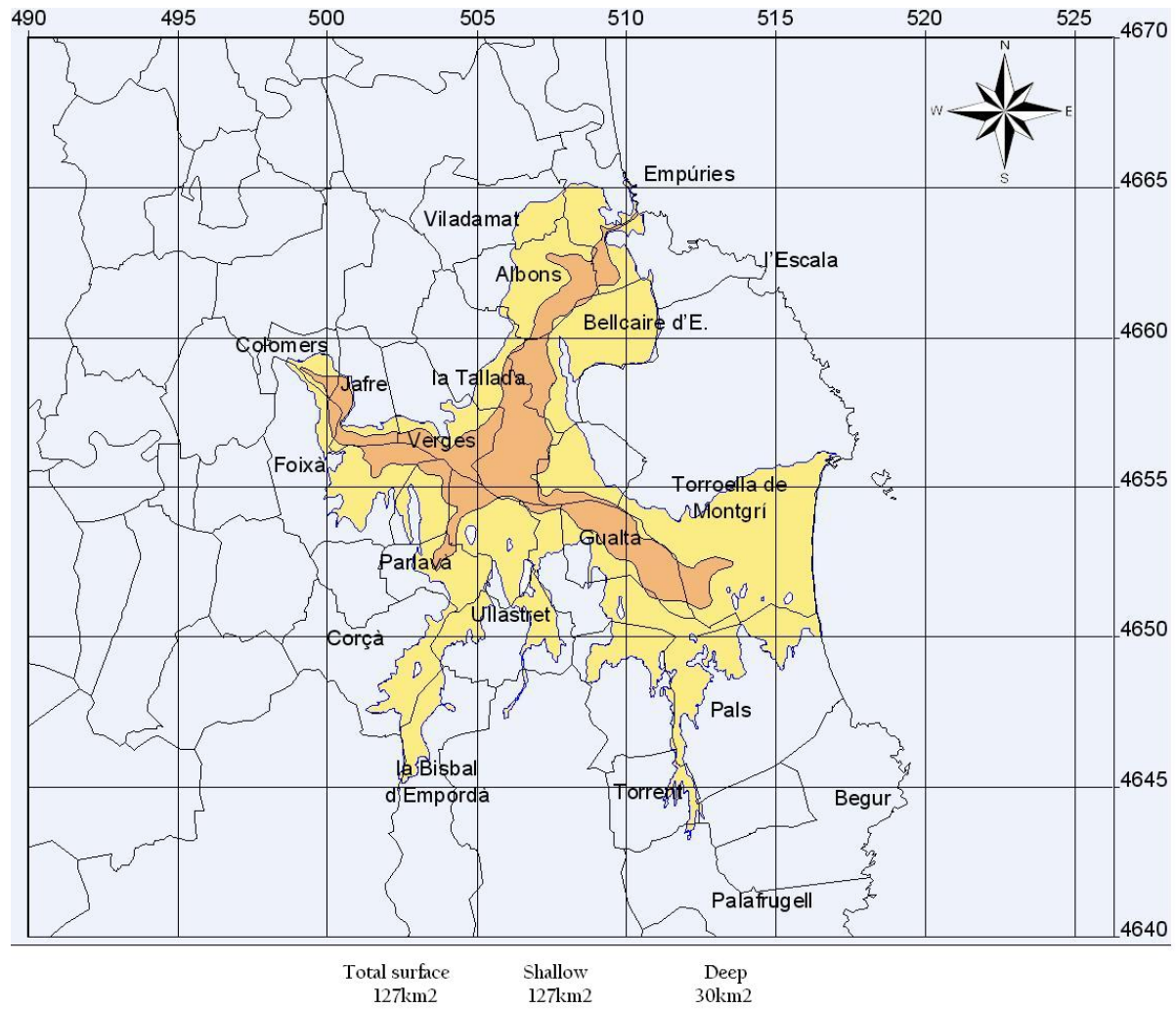


Fig. 4: Deep and shallow Baix Ter aquifers map, provided by Geoservei, S.L.

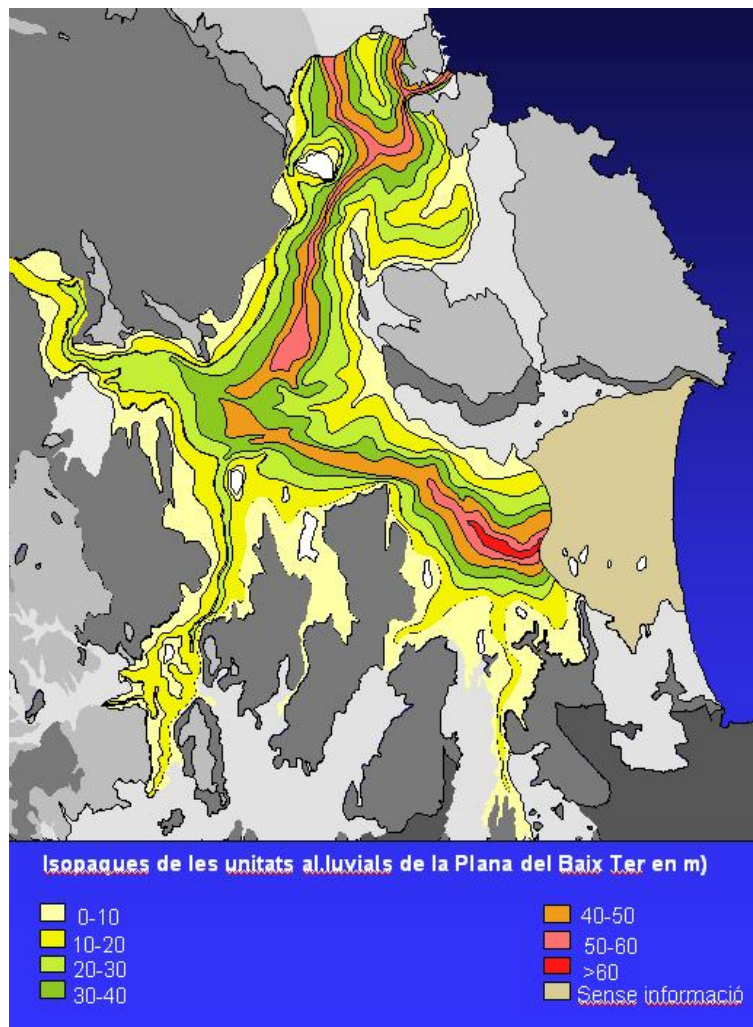


Fig. 5: Isopaques map, provided by Geoservei, S.L.

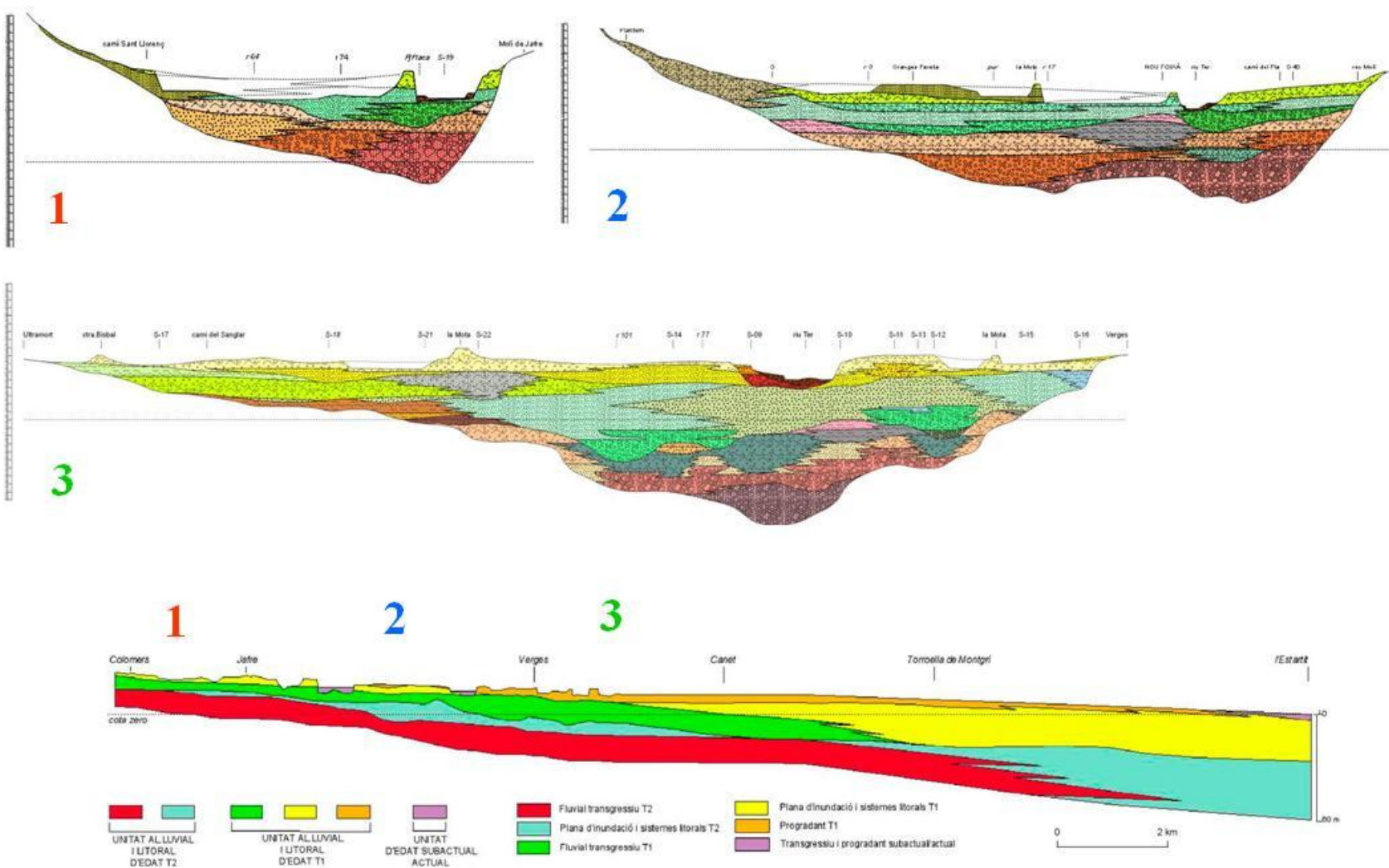


Fig. 6: Baix ter aquifers lithostratigraphy , provided by Geoservei, S.L.

5.2.2 Lower Ter paleogene

Characteristics summary:

- Hydrogeological area code: 201.
- Total surface: 116Km^2 from which only 49Km^2 come to surface.
- Predominant lithology: Non-alluvial detritic.
- Main Hydraulic Characteristics: Associated unconfined and confined aquifers with predominance of the confined ones.
- Presence of multiple-layer aquifers.
- Main flow: Fissured.
- General flow direction: associated with the fractures, to the center of the plain and eastwards.
- Permeability: No data.
- Transmissivity: $1-8000\text{m}^2/d$
- Storage coefficient: No data.
- No interaction with superficial waters.
- Hydrochemical facies: Calcic bicarbonates.

This water mass is defined by a set of sedimentary aquifers located in the domain of the Empordà trench depression. The geologic units implied in this unit have diverse lithology.

- Upper Bartonian deposits constituted by fine granulometry arcose stoneware, marls, conglomerates and red clays interpreted as open marine platform, deltaic front and deltaic plain, respectively.
- Lower Bartonian detritic units formed by coarse to fine sandstones and ochre microconglomerates. This unit's base presents an up to 13m thick level of bioclastic reddish limestones with ferruginous nodules, constituted as a condensation level. It consists of a superficial marine environment with strong mareal streams.
- Torrent limestones, corresponding to the carbonated units of the Girona formation that comes to surface in small areas or appears at low depth.

The limits of this water mass correspond to geological limits. This is, lithostratigraphic limits of the aquifer units that constitute this water mass and the structural limits that separate or isolate hydraulically these units. In this sense, the lithostratigraphic limits are determined by the own extension of the lithostratigraphic units that constitute the different aquifers of the geomorphologic field located mainly between the W and S margin of the Baix Empordà depression.

Natural recharge is due to direct rainfall infiltration in the 3 aquifers and, probably, by recharge of the bartonian suprajacent aquifers to the Torrent limestones. There is recharge in all the extents of the aquifers in touch with surface. There might be discharge at the contacts with the Baix Ter aquifers. Due to geological variation horizontally and vertically, main flow is fissured. At the lower Bartonian ochre sandstones aquifer there can be deduced from piezometric data that underground flow has a mainly northward direction, from the northern reliefs of the Pregavarres to the middle of the lower Ter river stretch plain. At the Torrent paleogene limestone aquifer it can be interpreted from historic data that during the summer periods water level drops from 29,85 to 59,00m.

The thickness and lateral variation of the non-saturated area for this water mass varies from meters to tens of meters. There are local areas where water outflows. This water unit has no interaction with superficial waters. The only affection to be considered for this water mass might be due to coastal zone water extractions, which could produce saline water intrusion.

Historical chemical status: (1982-2002)

| Parameter | Value | Units |
|------------------|--------|-------|
| HCO ₃ | 414,05 | mg/l |
| Ca | 115,00 | mg/l |
| Cond-lab | 836,00 | uS/cm |
| Cl | 50,10 | mg/l |
| Mg | 31,63 | mg/l |
| pH-lab | 7,56 | u.pH |
| K | 9,34 | mg/l |
| Na | 26,41 | mg/l |
| SO ₄ | 27,47 | mg/l |

Recent chemical status: ⁴

| Parameter | UBDA | LBOS | TPLA |
|--------------------------------------|--------|--------|--------|
| pH(u.pH) | 7,8 | 7,6 | 8,2 |
| CE(uS/cm) | 836,5 | 1079,8 | 788,0 |
| HCO ₃ (mg/l) | 414,05 | 363,89 | 447,32 |
| Cl(mg/l) | 50,1 | 88,0 | 60,4 |
| SO ₄ ²⁻ (mg/l) | 87,5 | 94,44 | 68,0 |
| Ca ²⁺ (mg/l) | 115,3 | 144,49 | 134,0 |
| Mg ²⁺ (mg/l) | 45,9 | 31,63 | 21,1 |
| Na(mg/l) | 26,41 | 57,67 | 59,2 |
| K(mg/l) | 9,34 | 20,67 | 2,79 |
| NO ₃ (mg/l) | 24,0 | 12,1 | 35,5 |
| NH ₄ (mg/l) | < 0,08 | 20,67 | 0,2 |
| Fe(μ g/l) | - | 445,00 | 104 |
| Mn(μ g/l) | - | 60,78 | 46 |

The most important impact on chemical state is the high nitrate levels with average levels higher than 50mg/l and seldom values above 400mg/l. Sulfates have occasional values higher than 250mg/l.

⁴ UBDA=Upper bartonian detritics aquifer;LBOS=Lower bartonian ochre sandstones; TPLA=Torrent paleogene limestones aquifer.

6 Methods

This chapter contains a brief description of the data acquisition and treatment as well as codes used for representation. Actual interpretation is made by representing chemical data in a way that graphs “speak by themselves”. That is, graphical representation should be informative about the study goals (types of waters, undergone processes, end-members, etc.) The types of graphs used in this thesis will be depicted in following sections.

6.1 Data treatment and filtering

Our original data set consists of a set of hydrochemical analysis of the Ter river fluviodeltaic area during the 1995-2005 period. These data have been downloaded from the web site of the ACA. The original exported data format is an excel file where each row corresponds to one parameter of one analysis on a specific date. It consists of 8534 entries corresponding to 630 different analysis⁵. The overall appearance of this file can be seen in figure 8:

| CODI POU | FONDARIA | UTM X | UTM Y | PARAMETRE | DATA | VALOR | UNITATS |
|----------|----------|--------|---------|----------------|------------|-------|---------|
| 17004-2 | 40 | 507350 | 4662500 | CE - lab | 25/09/1995 | 2730 | uS/cm |
| 17004-2 | 40 | 507350 | 4662500 | Clorurs | 25/09/1995 | 644.1 | mg/l |
| 17004-2 | 40 | 507350 | 4662500 | Nitrats | 25/09/1995 | 6 | mg/l |
| 17004-2 | 40 | 507350 | 4662500 | pH - lab | 25/09/1995 | 7.6 | u. pH |
| 17068-1 | 5.9 | 498420 | 4657920 | Amoni | 25/05/1995 | 0 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Bicarbonats | 25/05/1995 | 430.7 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Calci | 25/05/1995 | 170.5 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | CE - lab | 25/05/1995 | 974 | uS/cm |
| 17068-1 | 5.9 | 498420 | 4657920 | Clorurs | 25/05/1995 | 35.1 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Duresa total | 25/05/1995 | 510.2 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Magnesi | 25/05/1995 | 20.5 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Nitrats | 25/05/1995 | 63.1 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Oxidabilitat | 25/05/1995 | 0 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | pH - lab | 25/05/1995 | 6.98 | u. pH |
| 17068-1 | 5.9 | 498420 | 4657920 | Potassi | 25/05/1995 | 1 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Sodi | 25/05/1995 | 20.1 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Sulfats | 25/05/1995 | 82.6 | mg/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Ferro total | 25/05/1995 | 0 | ug/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Manganès tota | 25/05/1995 | 0 | ug/l |
| 17068-1 | 5.9 | 498420 | 4657920 | Conductivitat | 25/05/1995 | 1005 | uS/cm |
| 17068-1 | 5.9 | 498420 | 4657920 | Nivell piezomé | 25/05/1995 | 4.81 | m |
| 17068-1 | 5.9 | 498420 | 4657920 | pH - camp | 25/05/1995 | 7.18 | u.pH |
| 17068-1 | 5.9 | 498420 | 4657920 | Potencial Red | 25/05/1995 | 316.7 | mV |
| 17068-1 | 5.9 | 498420 | 4657920 | Temperatura | 25/05/1995 | 15.5 | °C |

Fig. 8: Partial view of the original data set.

⁵ We consider one analysis as the set of parameters analyzed in one same location in a specific date.

Most programs need a data input format where each row corresponds to one analysis and each column to one parameter. So, the original file has been modified and rewritten as needed. In order to this in an automatized way, an own FORTRAN program specific to read the old format and convert it to the desired one has been programmed(tesinabo.exe). As a result we have a text file that has been easily imported to excel. This final data matrix contains 630 rows(analysis) and 101 columns(parameters) as can be seen in figure: 9:

| Analysis | stationi | X | Y | 4,4-D | 4,4'-D | a-hexa | Aldrin | Alumir | Ametr | Amoni | Antim | Arsèn |
|------------|----------|--------|---------|-------|--------|--------|--------|--------|-------|-------|-------|-------|
| 25/05/1995 | 17068-1 | 498420 | 4657920 | | | | | | | 0 | | |
| 25/05/1995 | 17068-2 | 499900 | 4656900 | | | | | | | 0 | | |
| 26/09/1995 | 17070-1 | 507320 | 4653731 | | | | | | | | | |
| 26/09/1995 | 17081-2 | 509100 | 4654150 | | | | | | | | | |
| 26/09/1995 | 17081-5 | 510830 | 4652690 | | | | | | | | | |
| 26/09/1995 | 17199-1 | 510961 | 4654068 | | | | | | | | | |
| 26/09/1995 | 17199-14 | 515721 | 4651507 | | | | | | | | | |
| 26/09/1995 | 17199-18 | 512810 | 4654540 | | | | | | | | | |
| 26/09/1995 | 17199-31 | 515662 | 4654884 | | | | | | | | | |
| 26/09/1995 | 17199-7 | 513862 | 4652342 | | | | | | | | | |
| 26/09/1995 | 17199-9 | 513560 | 4655000 | | | | | | | | | |
| 27/09/1995 | 17081-6 | 510684 | 4652802 | | | | | | | | | |
| 27/09/1995 | 17081-7 | 510745 | 4652830 | | | | | | | | | |
| 27/09/1995 | 17124-1 | 514500 | 4649600 | | | | | | | | | |
| 27/09/1995 | 17124-6 | 516100 | 4649400 | | | | | | | | | |
| 27/09/1995 | 17124-7 | 512226 | 4650095 | | | | | | | | | |
| 27/09/1995 | 17199-20 | 512379 | 4651813 | | | | | | | | | |
| 28/05/1996 | 17004-1 | 508905 | 4661748 | 0 | 0 | 0 | 0 | | | 0 | | |
| 28/05/1996 | 17004-7 | 506246 | 4661877 | | | | | | | 0 | | |
| 28/05/1996 | 17004-8 | 508389 | 4661102 | 0 | 0 | 0 | 0 | | | 0.25 | | |
| 28/05/1996 | 17085-1 | 500847 | 4658107 | | | | | | | 0 | | |
| 28/05/1996 | 17211-1 | 503800 | 4656431 | 0 | 0 | 0 | 0 | | | 0 | | |
| 29/05/1996 | 17070-1 | 507320 | 4653731 | | | | | | | 4.29 | | |
| 29/05/1996 | 17124-7 | 512226 | 4650095 | | | | | | | 0.87 | | |

Fig. 9: Partial view of the rewritten data set.

Many of these columns correspond to rare parameters that have a scarce number of analysis. For example, plaguicides and other contaminants that are not usually analyzed have too few results, so they have been removed. However, the data matrix was not yet appropriate. On the other hand, there were also lots of rows corresponding to analysis with too few analyzed parameters. So, a second filtering criteria has been done to remove the rows corresponding to analysis with less than 7 values. As a result we have a data matrix like the one seen in figure 10, containing 130 analysis and 36 variables.

| Analysis_date | stationid | Al | NH4 | Sb | As | Balanc | Ba | HCO3 | Cd | Ca |
|---------------|-----------|----|-----|------|----|--------|----|-------|----|-------|
| 25/05/1995 | 17068-1 | | | 0 | | | | 430.7 | | 170.5 |
| 25/05/1995 | 17068-2 | | | 0 | | | | 309.4 | | 105.7 |
| 28/05/1996 | 17004-1 | | | 0 | | 0.067 | | 259 | | 96.7 |
| 28/05/1996 | 17004-7 | | | 0 | | 0.085 | | 339 | | 186.6 |
| 28/05/1996 | 17004-8 | | | 0.25 | | 0.197 | | 294 | | 102.1 |
| 28/05/1996 | 17085-1 | | | 0 | | 0.415 | | 433 | | 171.3 |
| 28/05/1996 | 17211-1 | | | 0 | | 0.369 | | 416 | | 171 |
| 29/05/1996 | 17070-1 | | | 4.29 | | 0.537 | | 481 | | 233.2 |
| 29/05/1996 | 17124-7 | | | 0.87 | | 0.833 | | 362 | | 88.3 |
| 29/05/1996 | 17191-1 | | | 0 | | 0.657 | | 363 | | 135.2 |
| 29/05/1996 | 17195-2 | | | 0 | | 0.391 | | 427 | | 140.7 |
| 29/05/1996 | 17199-1 | | | 0 | | 0.789 | | 454 | | 185.7 |
| 29/05/1996 | 17199-14 | | | 1.19 | | 1.845 | | 506 | | 224.8 |
| 29/05/1996 | 17199-20 | | | 0 | | 0.55 | | 546 | | 150.9 |
| 29/05/1996 | 17199-3 | | | 0 | | 0.073 | | 376 | | 138.1 |
| 29/05/1996 | 17199-31 | | | 3.52 | | 0.186 | | 791 | | 238.9 |
| 29/05/1996 | 17199-7 | | | 1.48 | | 0.278 | | 632 | | 127.8 |
| 30/05/1996 | 17081-6 | | | 0 | | 1.015 | | 410 | | 136.7 |
| 30/05/1996 | 17195-1 | | | 0 | | 0.303 | | 413 | | 176.2 |
| 30/05/1996 | 17195-3 | | | 0 | | 1.176 | | 325 | | 117.5 |
| 05/06/1996 | 17068-3 | | | 0 | | 0.666 | | 351 | | 156.4 |
| 13/01/1998 | 17195-1 | | | 0.1 | | 2.284 | | 404 | | 169.4 |
| 13/01/1998 | 17211-1 | | | 0.1 | | 0.631 | | 416 | | 174.8 |
| 15/01/1998 | 17070-1 | | | 1.59 | | 1.918 | | 278 | | 104.3 |

Fig. 10: Rewritten and filtered data set. Although columns like Al, Sb or As seem void, most of their values are in lower rows of the matrix and they do not show up in this image.

This has been the starting data set to apply the different interpretation techniques. As a result of the filterings, all analysis from years 1997 and 2005 have been discarded and only two from year 1995 have been kept. So, we can consider our data set valid during the years 1996 and 1998-2004. As explained in chapter 2, new filterings have been done during the interpretation process. Incongruent results were usually been caused by lack of accuracy of many analysis when their values ranged below detection limits.

6.2 Evaluated software

Several software programs have been used to analyze our data set. Each one has been useful not for some specific tasks and for other ones not. One of the aims of this work is to find the bugs, pros and cons of the different tools used to develop the data analysis. Therefore, this section contains a brief description of the software that has been tested just in case the reader does not know in advance what it can be used for. The information explained next can be consulted with more detail on their respective companies web sites.

The following softwares have been tested:

- Aquachem.

AquaChem defines itself in its web site [10] as a groundwater software package featuring a customizable database of physical and chemical parameters and a selection of analysis, calculation, modeling, and graphing tools. This analysis tools cover a wide range of functions and calculations used for analyzing, interpreting and comparing water quality data. This includes simple unit transformations, charge balances, statistics and sample mixing to more complex functions such as correlation matrices and geothermometer calculations. These analytical capabilities are complemented by a selection of commonly used plotting techniques to represent the chemical characteristics of aqueous geochemical and water quality data.

Data Management

Built around a customizable MS Access database that can be configured to include more than 1000 alphanumeric parameters per sample. Parameters are divided into the following parameter groups: station description parameters, sample description parameters, measured parameters, and modeled parameters. Each parameter group contains pre-defined parameters which can be used to create a customized data structure for different sampling and reporting requirements. Data may be imported and exported in various formats including MS Access, MS Excel and .TXT files.

Data Analysis

AquaChem uses the common measured values (cations and anions) for each sample to calculate additional geochemical values such as water type, sum of Anions, sum of Cations, Ion Balance, TDS, hardness, Alkalinity, Common Ion Ratios, Sodium Adsorption Ratio, Magnesium Hazard, Oxygen saturation and Scaling Indexes such as Langlieier Index, Ryznar Stability Index, Puckorius Scaling Index and Larson-Skoid Index. Data Analysis features include: Compare Samples, Correlation Matrix, Water Quality samples, Mix Samples, Data Reliability check, Rock Source Deduction. Statistical Calculations AquaChem also includes a list of statistical features, optimized to work with water quality data sets. Statistical features included with this software

are: Trend analysis (Mann Kendall, Sen's Test), Outlier Tests (Extreme value, Discordance, Rosner, Walsh test), Test For Normality (Studentized range test, Geary's, Shapiro, D'Agostino test. Water Quality Modeling

AquaChem provides a link to one of the popular public domain geochemical modeling programs, PHREEQC available from the USGS. PHREEQC generated values (saturation indices, activities) can be read back to the database and used for reporting, plotting or statistics. One of AquaChem's most interesting features is its ability to manage, create, and graphically display a wide range of water quality data. This includes a selection of more than 23 industry-standard plots including: Piper, Schoeller, Scatter, Box & Wisker, time series, histogram, Stiff, Radial and Pie chart. Some of these are shown in figure 11.

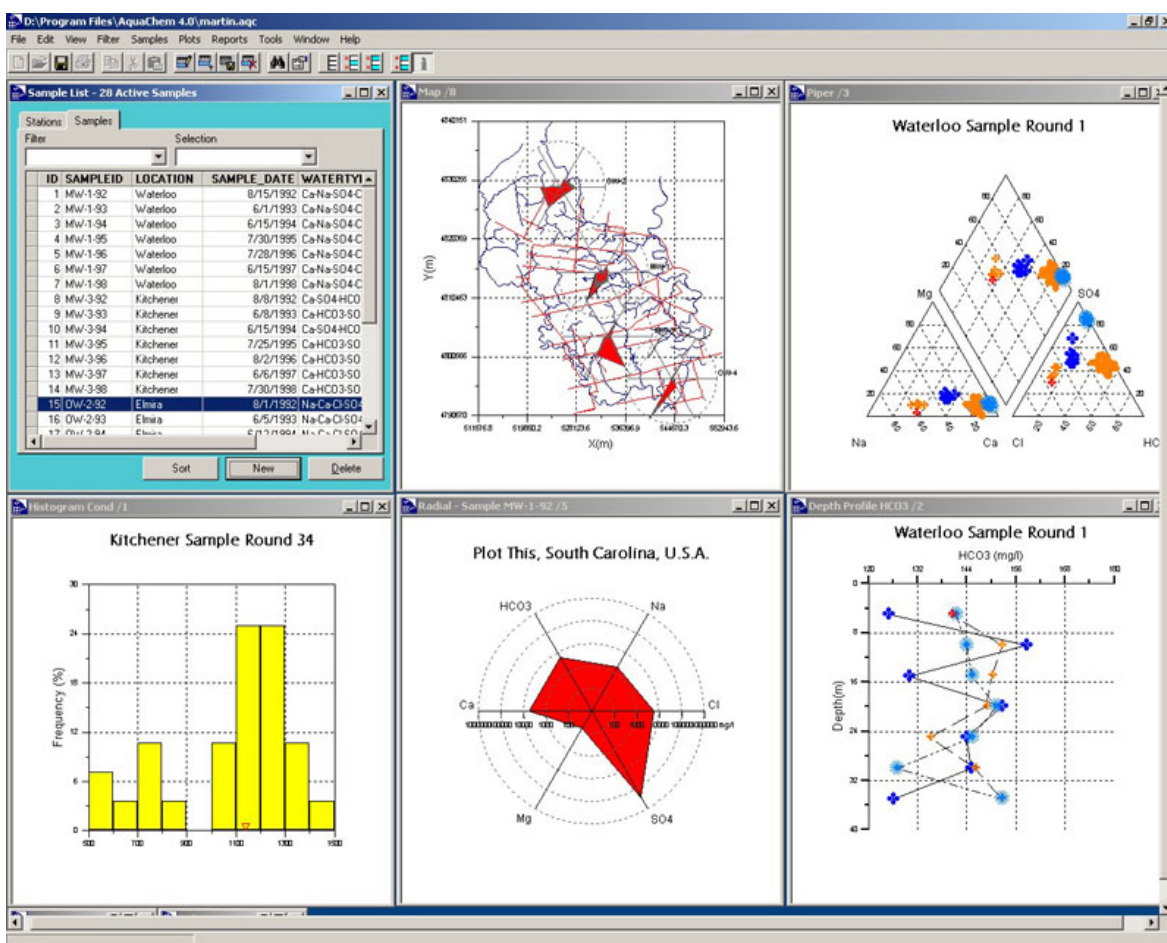


Fig. 11: Data base, Stiff diagram projected on map, piper diagram, histogram, radial diagram and depth profile done with AquaChem.

- SPSS.

SPSS (originally, Statistical Package for the Social Sciences) is, as described in one of most popular divulgative web pages, Wikipedia[7], among the most widely used programs for statistical analysis in social science. In addition to statistical analysis, data management (case selection, file reshaping, creating derived data) and data documentation (a metadata dictionary is stored with the data) are features of the base software. This includes: Descriptive statistics(Cross tabulation, Frequencies, Descriptives, Explore, Descriptive Ratio Statistics); Bivariate statistics(Means, t-test, ANOVA, Correlation, Nonparametric tests); Prediction for numerical outcomes(Linear regression);Prediction for identifying groups(Factor analysis, cluster analysis,Discriminant analysis). SPSS datasets have a 2-dimensional table structure where the rows typically represent cases and the columns represent measurements, as explained in previous section. This data format is shown in figure 12.

The screenshot shows the SPSS data editor window titled '*Sin título2 [Conjunto_de_datos1] - Editor de datos SPSS'. The window displays a data table with the following columns: sampleid, Analysis_date, stationid, X, Y, NH4, HCO3, Ca, and Cl. The data is organized into rows representing individual cases. The table is displayed in a grid format with a scroll bar on the right side. The status bar at the bottom indicates 'SPSS El procesador está preparado'.

| sampleid | Analysis_date | stationid | X | Y | NH4 | HCO3 | Ca | Cl |
|----------|---------------|-----------|--------|---------|------|--------|--------|---------|
| 1 | 25-MAY-1995 | 17068-1 | 498420 | 4657920 | 0 | 430700 | 170500 | 35100 |
| 2 | 25-MAY-1995 | 17068-2 | 499900 | 4656900 | 0 | 309400 | 105700 | 57900 |
| 3 | 28-MAY-1996 | 17004-1 | 508905 | 4661748 | 0 | 259000 | 96700 | 67800 |
| 4 | 28-MAY-1996 | 17004-7 | 506246 | 4661877 | 0 | 339000 | 186600 | 97400 |
| 5 | 28-MAY-1996 | 17004-8 | 508389 | 4661102 | 250 | 294000 | 102100 | 39500 |
| 6 | 28-MAY-1996 | 17085-1 | 500847 | 4658107 | 0 | 433000 | 171300 | 61200 |
| 7 | 28-MAY-1996 | 17211-1 | 503800 | 4656431 | 0 | 416000 | 171000 | 71700 |
| 8 | 29-MAY-1996 | 17070-1 | 507320 | 4653731 | 4290 | 481000 | 233200 | 191500 |
| 9 | 29-MAY-1996 | 17124-7 | 512226 | 4650095 | 870 | 362000 | 88300 | 76300 |
| 10 | 29-MAY-1996 | 17191-1 | 505825 | 4653381 | 0 | 363000 | 135200 | 68400 |
| 11 | 29-MAY-1996 | 17195-2 | 506308 | 4655973 | 0 | 427000 | 140700 | 53500 |
| 12 | 29-MAY-1996 | 17199-1 | 510961 | 4654068 | 0 | 454000 | 185700 | 115300 |
| 13 | 29-MAY-1996 | 17199-14 | 515721 | 4651507 | 1190 | 506000 | 224800 | 2724100 |
| 14 | 29-MAY-1996 | 17199-20 | 512379 | 4651813 | 0 | 546000 | 150900 | 127800 |
| 15 | 29-MAY-1996 | 17199-3 | 513615 | 4653236 | 0 | 376000 | 138100 | 72500 |
| 16 | 29-MAY-1996 | 17199-31 | 515662 | 4654884 | 3520 | 791000 | 238900 | 868000 |
| 17 | 29-MAY-1996 | 17199-7 | 513862 | 4652342 | 1480 | 632000 | 127800 | 238900 |
| 18 | 30-MAY-1996 | 17081-6 | 510684 | 4652802 | 0 | 410000 | 136700 | 93900 |
| 19 | 30-MAY-1996 | 17195-1 | 504993 | 465295 | 0 | 413000 | 176200 | 73900 |
| 20 | 30-MAY-1996 | 17195-3 | 506549 | 4655840 | 0 | 325000 | 117500 | 41200 |
| 21 | 05-JUN-1996 | 17068-3 | 501341 | 4655515 | 0 | 351000 | 156400 | 55600 |
| 22 | 13-JAN-1998 | 17195-1 | 504993 | 465295 | 100 | 404000 | 169400 | 73700 |
| 23 | 13-JAN-1998 | 17211-1 | 503800 | 4656431 | 100 | 416000 | 174800 | 72500 |
| 24 | 15-JAN-1998 | 17070-1 | 507320 | 4653731 | 1590 | 278000 | 104300 | 64200 |
| 25 | 15-JAN-1998 | 17081-6 | 510684 | 4652802 | 570 | 389000 | 136800 | 81400 |
| 26 | 15-JAN-1998 | 17085-1 | 500847 | 4658107 | 100 | 422000 | 157800 | 58000 |
| 27 | 15-JAN-1998 | 17124-7 | 512226 | 4650095 | 100 | 613000 | 174400 | 268500 |
| 28 | 15-JAN-1998 | 17191-1 | 505825 | 4653381 | 100 | 362000 | 133600 | 65200 |
| 29 | 15-JAN-1998 | 17195-3 | 506549 | 4655840 | 100 | 297000 | 122800 | 48000 |
| 30 | 15-JAN-1998 | 17199-1 | 510961 | 4654068 | 100 | 435000 | 175600 | 108100 |
| 31 | 15-JAN-1998 | 17199-14 | 515721 | 4651507 | 2640 | 635000 | 280700 | 3504100 |
| 32 | 15-JAN-1998 | 17199-20 | 512379 | 4651813 | 100 | 553000 | 231700 | 177900 |
| 33 | 15-JAN-1998 | 17199-3 | 513615 | 4653236 | 100 | 378000 | 159100 | 66900 |
| 34 | 15-JAN-1998 | 17199-31 | 515662 | 4654884 | 1880 | 671000 | 158900 | 433700 |
| 35 | 15-JAN-1998 | 17199-7 | 513862 | 4652342 | 100 | 628000 | 129800 | 248500 |

Fig. 12: SPSS working environment.

- ArcGIS.

ArcGIS describes itself as a complete system for authoring, serving, and using geographic information. It is an integrated collection of GIS software products for building and deploying a complete GIS. Among its many functions, arcGIS has a tool for transforming multivariate data: Principal component analysis(PCA). PCA is used to transform the data attributes in a multiband raster from the input multivariate attribute space to a new multivariate attribute space whose axes are rotated with respect to the original space. The result of a PCA is a multiband raster with the same number of bands as the original raster (one band per axis in the new multivariate space). The first principal component will have the greatest variance, the second will show the second most variance not described by the first, and so forth. The Principal Component function requires an input multiband raster to be identified, the number of principal components into which to transform the data, the name of the statistics output file, and the name of the output raster. The output raster will contain the same number of bands as the specified number of components. Each band will depict each component. More information about this software can be looked up at the esri web page. Its interface is shown in figure 13.

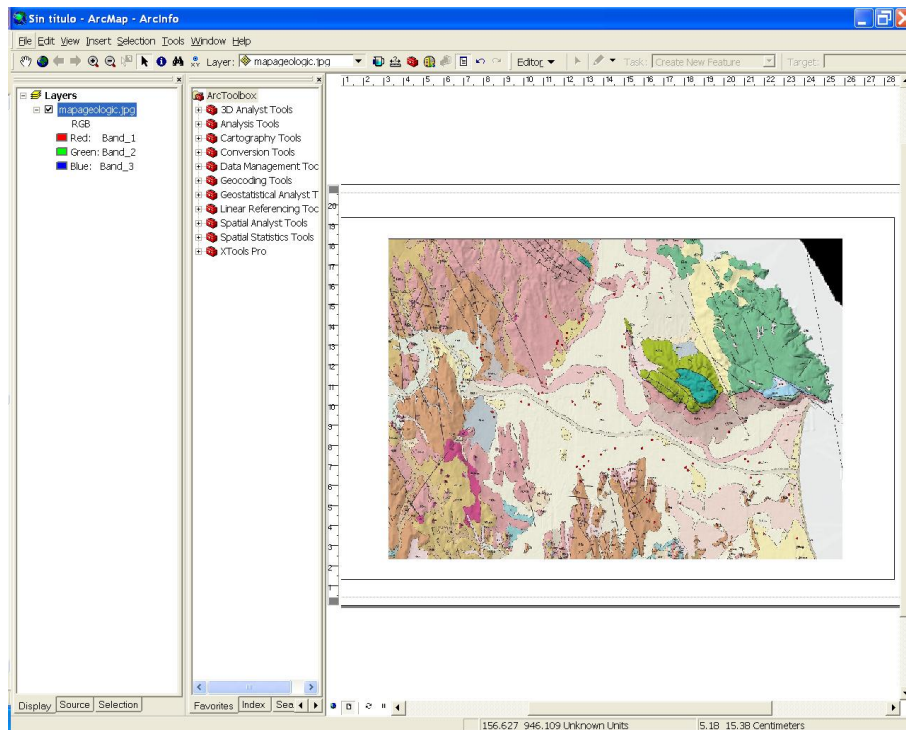


Fig. 13: ArcGIS interface.

- R.

R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form. R defines itself in its web page [11] as a language and environment for statistical computing and graphics. It provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible as it allows users to add additional functionality by defining new functions. The R environment is an integrated suite of software facilities for data manipulation, calculation and graphical display and it includes:

- an effective data handling and storage facility
- a suite of operators for calculations on arrays, in particular matrices
- a large, coherent, integrated collection of intermediate tools for data analysis
- graphical facilities for data analysis and display either on-screen or on hardcopy
- a programming language which includes conditionals, loops, ser-defined recursive functions and input and output facilities.

Figure 14 displays the interface of this program:

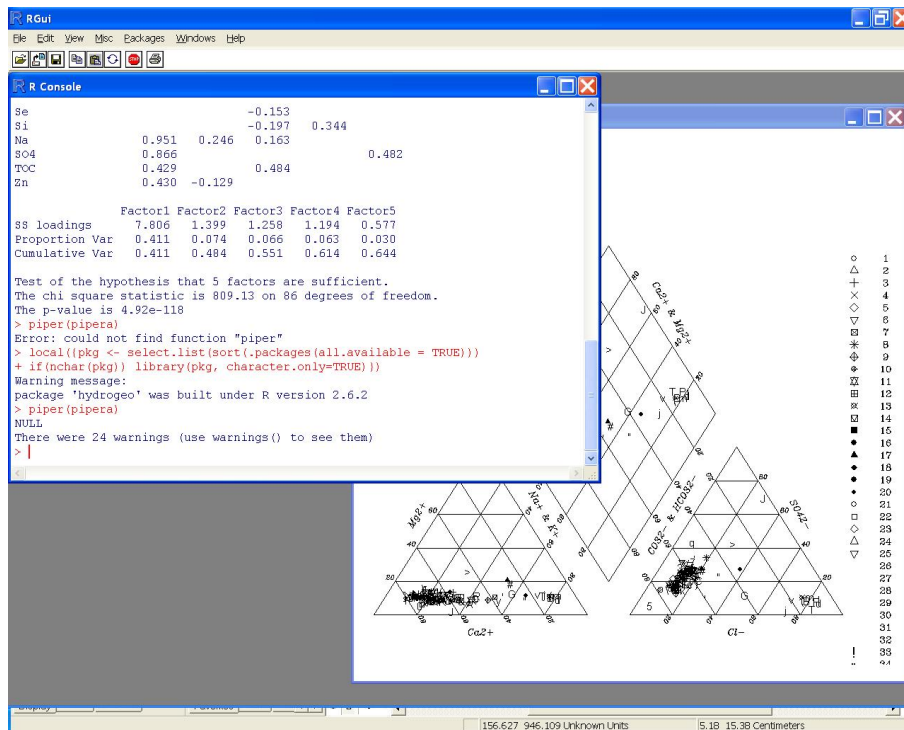


Fig. 14: R interface.

6.3 Classical study techniques for chemical analysis

Chemical hydrogeology is the science which studies groundwater chemicals and its interactions with rock components. To deal with such a complicated task, it uses standardized graphics and diagrams to simplify its study. Those are specially useful when it comes to comparing analysis of the same zone in different moments or places. This way, spatial or temporal relationships between different ions can be more easily highlighted. So, both spatial and temporal analysis have been done.

This chapter pretends to introduce the reader to each of the different used methods to ease its comprehension of the following graphics. As a first approach, a short description of each technique is done and, afterwards, the most important results are shown with their corresponding explanations. These descriptions have been extracted from one of most renowned manuals for groundwater analysis (Custodio et al., 1967)[2] as it has been one of this thesis most important references.

On the other hand, the most limiting factor to choose which graphics and diagrams to use in this analysis has been the possibility to represent it with the previously explained programs and the easiness to perform them. As most of them are more easily interpreted when they are represented on a map, when it has been possible, they have been plotted on the geologic map of our study zone. When this has not been possible, separate graphics are shown for each station or date.

6.3.1 Column or Collins diagrams

It consists of two adjacent columns with a shared base where one represents the major cations and the other one the major anions of one sole analysis. The cation order is usually $Ca - Mg - Na - K$ and the anion order $(CO_3H + CO_3) - CO_4 - Cl - NO_3$. A third column with silicon or any other particular characteristic can be added.

This diagram is not present in most of the referred programs. The only one capable to do a similar kind of plot has been ArcGIS. It offers the option of representing on each point a set of bars proportional to the variables defined by the user. The represented ions in this case have been Ca, Mg, Na, NO_3, HCO_3 and SO_4 . This graphic has only helped to detect extreme values as seen on figure 15. The rest of analysis are represented in too short columns and its interpretation is difficult. If the scale is changed, the highest values turn into exaggeratedly high columns, so this is also an unsatisfying graphic. As this is a not very outstanding result, no further analysis with it has been done.

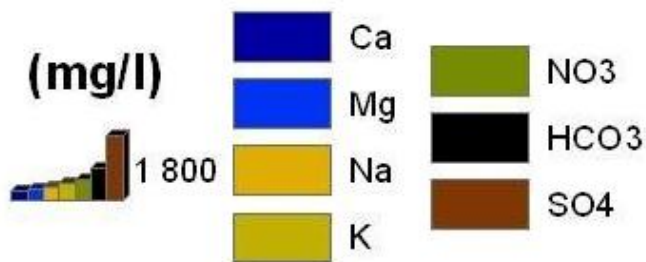
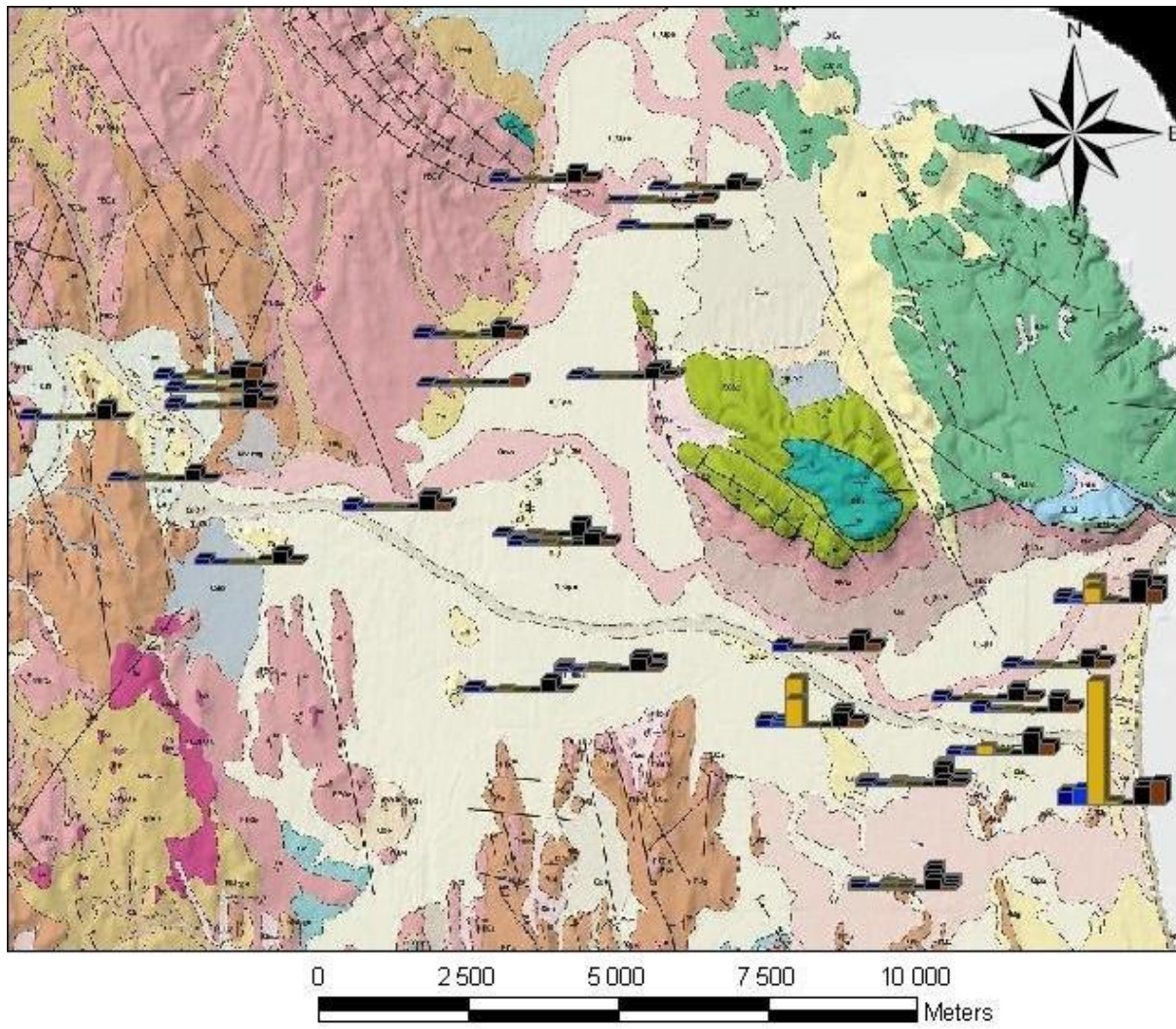


Fig. 15: Column diagram made with ArcGIS

6.3.2 Triangular or Piper diagrams

This kind of diagram is specially appropriate for representing simultaneously three components. It consists of an equilateral triangle where each vertex represents 100% of each component. This way, points are represented as ternary mixtures. Ions must be reduced to %. The most usual way is with % values in meq/l. Usually, the major anions and cations ($CO_3 + (CO_3)$, SO_4 , $Cl + (NO_3)$ and $Na + K$, Ca , Mg) are represented.

There are two sorts of triangular diagrams:

- a) Represent cations and anions on one same diagram, joining two representative points using an oriented segment.
- b) Use of two separate diagrams for representing cations and anions, distributed in different manners, and with a central rhomboidal or square field, where a third point is represented, deduced from those represented by cations and anions. The most used is the Piper diagram, with rhomboidal central field. The main advantage of those graphics is that they can represent a lot of different analysis without confusion. Those waters of similar composition stay grouped into very definite zones. Those diagrams can also reflect mixtures of water, precipitation, dissolution of chemicals, or other modifications.

This diagram is purported to be useful for identifying a number of processes such as mixing, dissolution-precipitation, cation exchange, etc. For example, a point located on the line joining two points can be the result of a mixture of the two waters represented with those two points. If dissolution or precipitation of a salt occurs, the different representative points are found located on the line that passes through the vertex corresponding to the affected anion or cation. A change of base does not produce any change on the anion diagram, while the points found on cation diagram are displaced over a line parallel to the side that joins the vertices of the involved cations. In the case of sulphate reduction, the cation diagram would not change, while the anion diagram would show a displacement of the representative points along a parallel line to the side of the representative vertices of SO_4 and HCO_3 as approximately one meq/l of SO_4 is changed for one meq/l of HCO_3 . The lines that concur on a same vertex represent waters with the same relationships between ions corresponding to the other two vertices.

These graphics have been made preferably with Aquachem(Fig.16) as it only needs the data to be imported in the usual format and the rest of the process is done automatically by the program. First of all, all samples have been represented in one only piper diagram.

As seen in figure 16, this graphic might be too confusing. So, a piper diagram for each year has been done showing clearer results such as in figure 17.

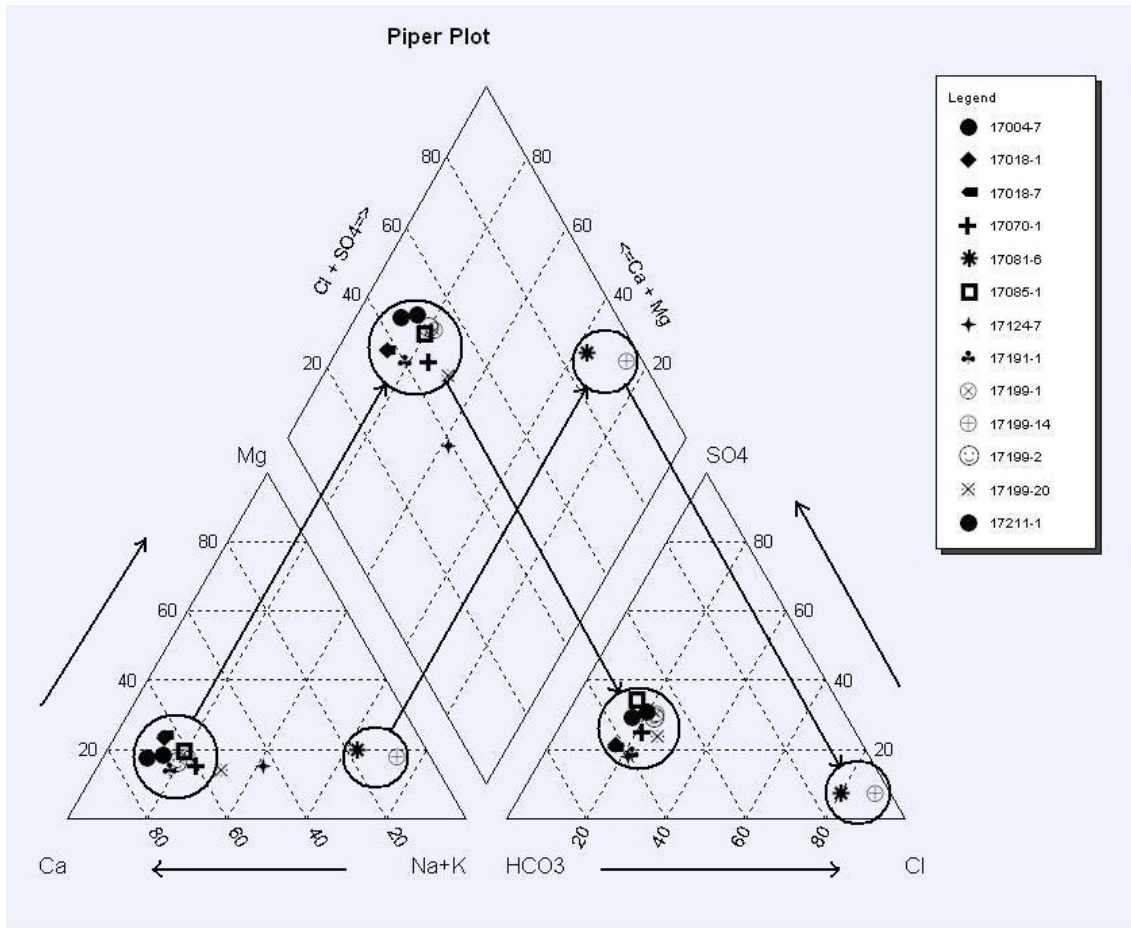


Fig. 17: Piper diagram for year 2004.

The most notable feature of the data displayed here is the horizontal line (about 20% *Mg*) in the cations diagram. This may reflect several processes as those told before. However, the fact that the high *Na* samples coincide with high *Cl* suggests a mixture between *Ca* – *HCO*₃ and *Na*–*Cl* (sea water). Moreover, marine intrusion in this area is known in advance and the extreme values correspond to stations located in areas affected by this process (17081-6 and 17199-14).

On the other hand, piper diagrams can also be done with R. It is necessary to install the hydrogeo package and perform manually much more laborious steps. First of all, data concentrations have to be homogenized to meq/l. After that, the concentrations of Ca , Cl , Mg and SO_4 also have to be converted to % by the user before introducing the data. Only this 4 data % are required as those of $(Na + K)$ and HCO_3 are deducted subtracting from 100%.

All previous steps are done automatically by Aquachem. It also plots the graphic with a better presentation and allows station labeling and other features. So, when it is possible, it is always preferable to use Aquachem for these plots. An example of one diagram done with R can be seen in figure 18:

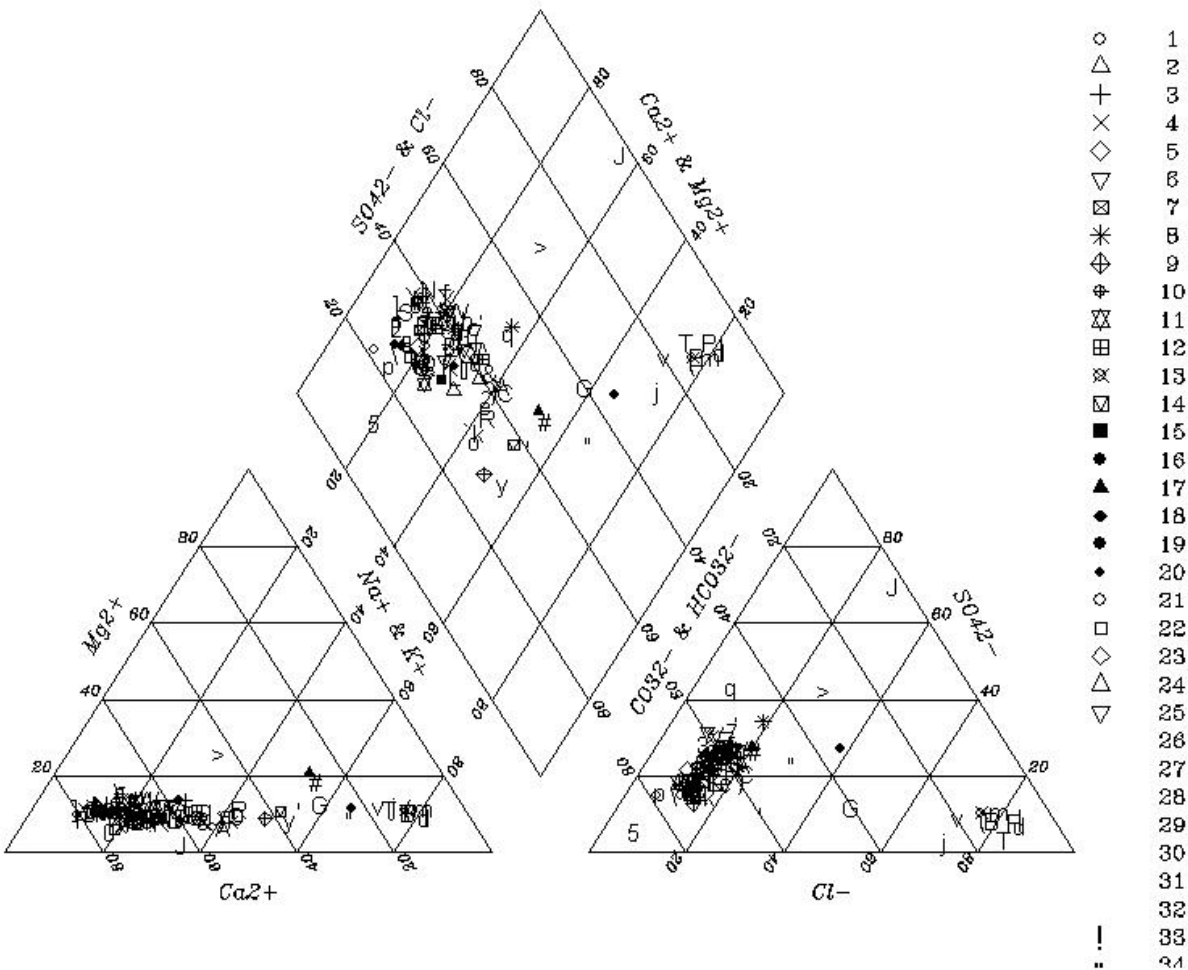


Fig. 18: Piper diagram with R

6.3.3 Circular and Pie diagrams

A circle proportional to dry residue or total dissolved meq/l is drawn, and is divided in sectors proportional to the % of different compounds. If it is done with the ions represented in meq/l, one half of the circle represents cations, and the other, anions. Representation of only cations or only anions can also be performed. These graphics have been mainly done with Aquachem (figure 19) and with ArcGIS (figure 20).

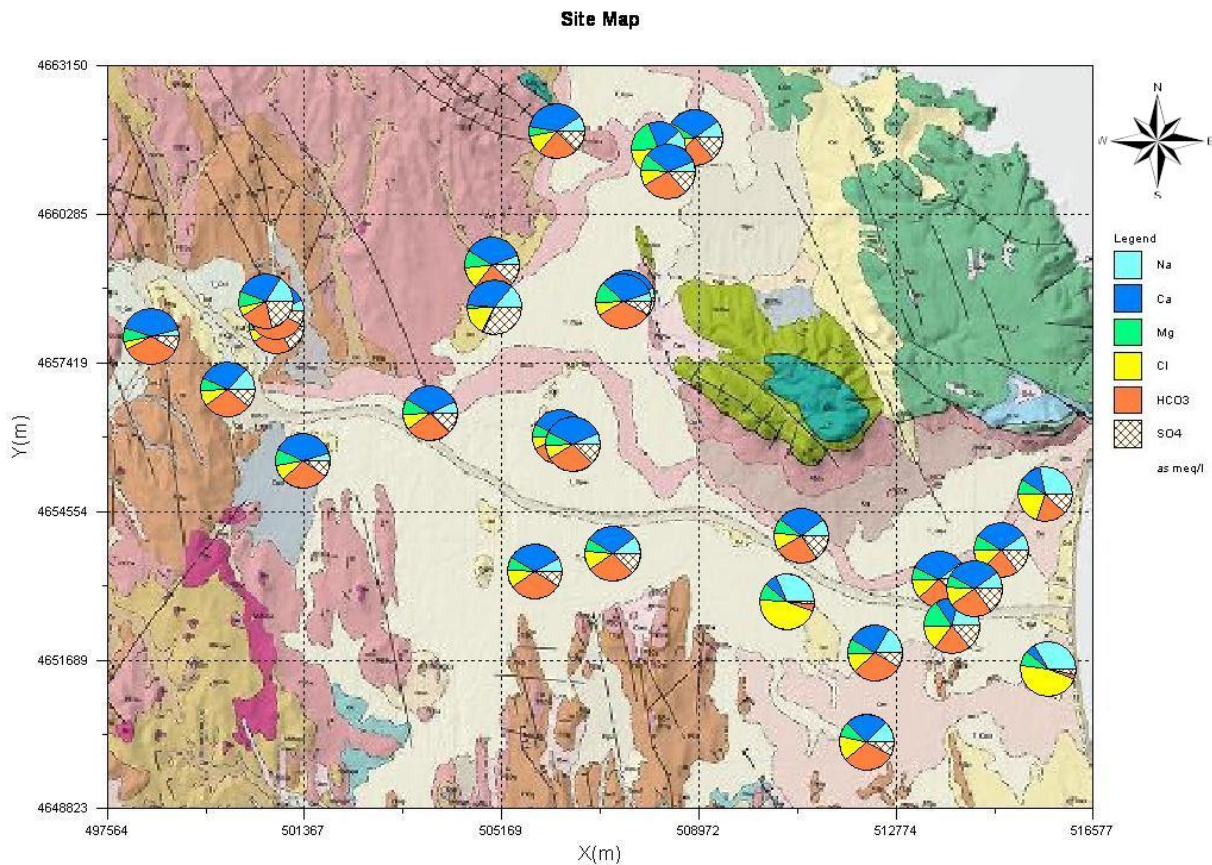


Fig. 19: Pie chart of year 1995-2004 done with Aquachem

These graphs allow a fast visualization of dominant ions. The circles are dominantly dark blue (*Ca*) and orange (*HCO₃*) in the interior, but become increasingly light blue (*Na*) and yellow (*Cl*) towards the coast. This information is similar to the one obtained with the Piper diagram though less quantitative but easier to locate in space.

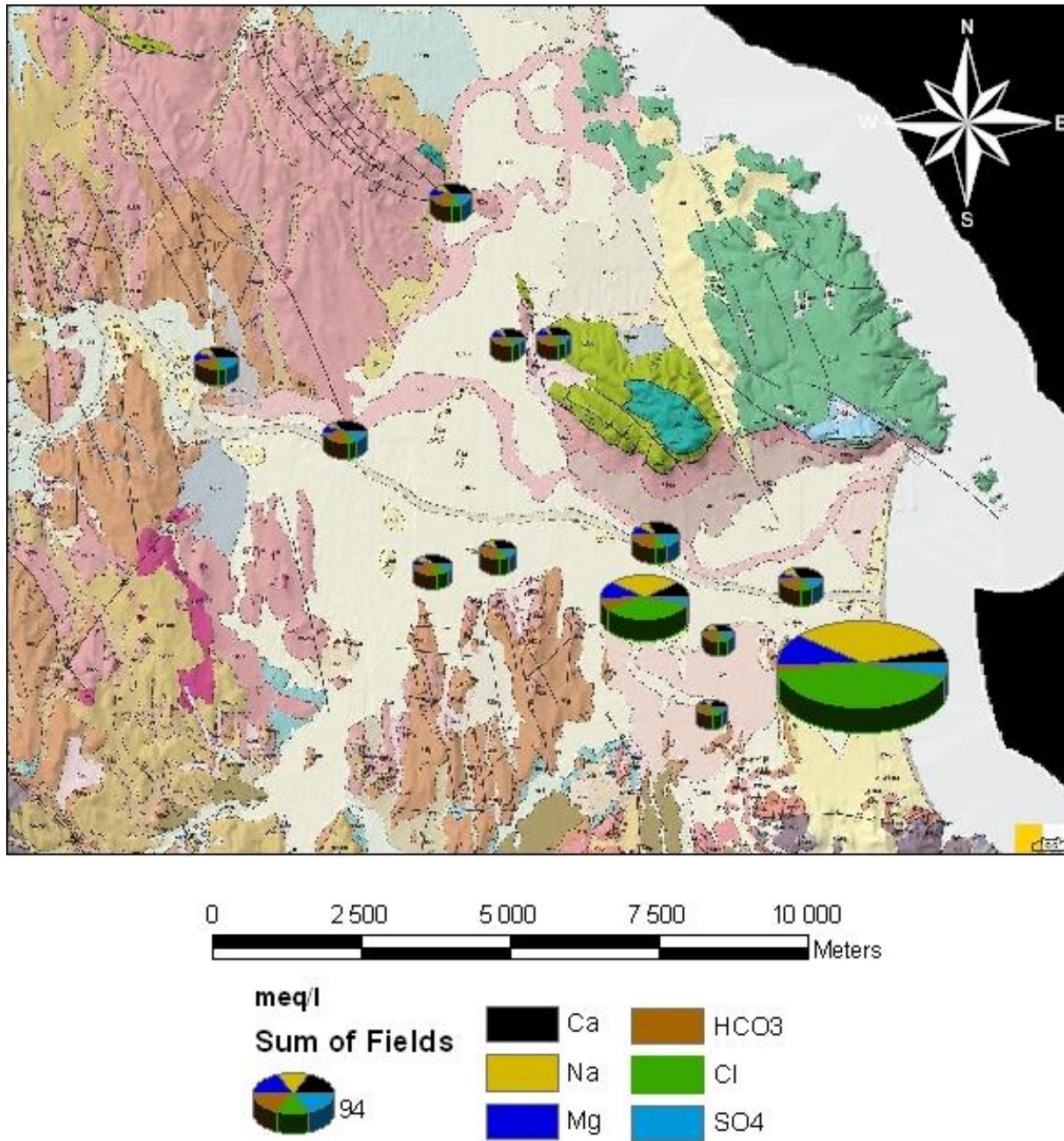


Fig. 20: Pie chart of year 2004 done with ArcGIS

When more of one analysis correspond to a same location, aquachem overlaps the pie representations corresponding to that point. So, a diagram for each different year has also been done for these charts. The lack of this program regarding the pie diagrams is that the size of the circles is unique for all samples, instead of being proportional to the sum of ions. However, this program automatically converts the ions to the required units and also represents them in correct order, allowing a fast generation of these graphics.

On the other hand, for the representation of these diagrams with ArcGIS it is necessary to manually convert all data to meq/l and introduce in the symbology options the specified ions in correct order. Although using this program is more laborious, it has the option to represent this diagrams with proportional size to the sum of the sample ions. This may help to also make quantitative interpretations of the results.

6.3.4 Polygonal, radial and modified Stiff diagrams

They consist in representing segments proportional to each ion on either convergent or parallel half-lines and joining the ends, resulting in a polygon. In both cases, every line represents an ion or the sum of an anion and a cation. The first ones are called radial diagrams, which are scarcely used. The second ones, are the so-called polygonal diagrams, of which the Stiff diagram is the best known. Originally, it was used for waters from oil extraction sites, and it had 4 parallel lines giving place to 4 half-lines for anions, and other 4 half-lines for cations. Generally, for the analysis of groundwater, a disposition of three base lines with the combination of anions and cations ($Na + K, Mg, Ca, Cl + NO_3, SO_4, HCO_3 + CO_3$) is used. These diagrams can only be done with Aquachem. As the scale is automatically adapted to the highest values, only extreme analysis can be correctly seen. If scale is changed to smaller values, then the extreme ones appear exaggeratedly in the chart. So, no different graphics for each year have been done, due to the difficulty to visualize practical results with any of them. However, a radial diagram (figure 21) and a stiff diagram (figure 22) charts are shown next:

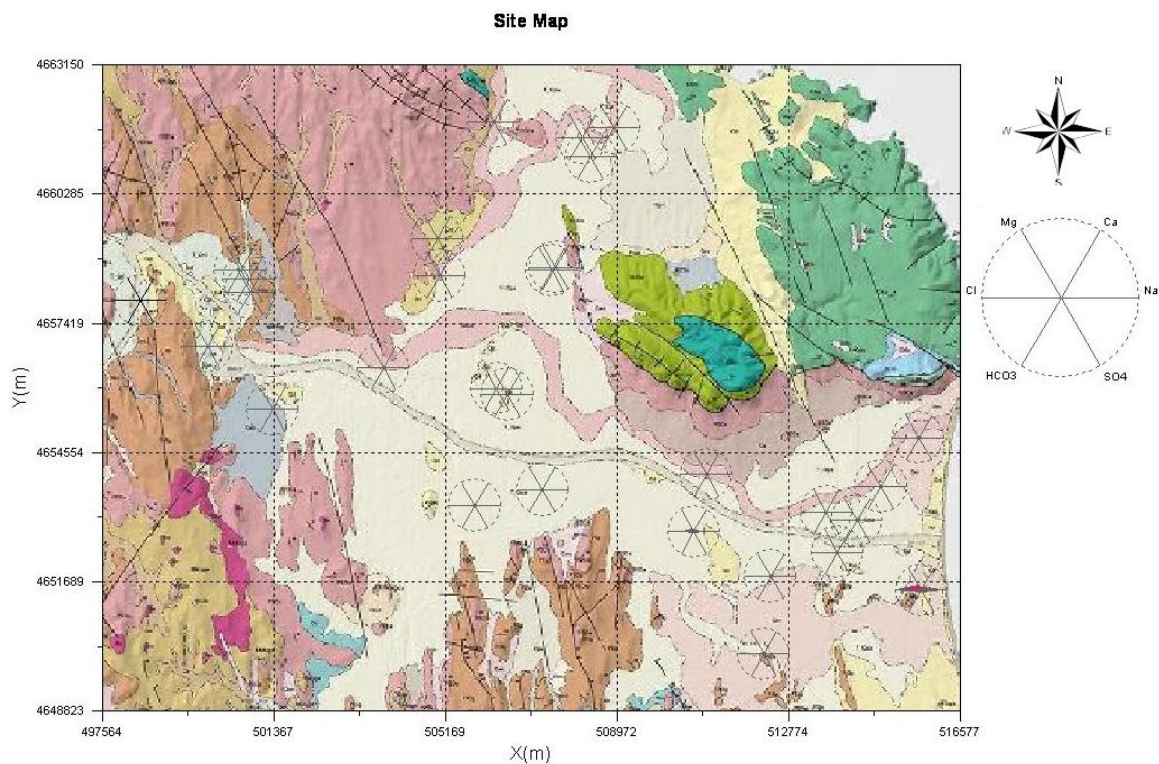


Fig. 21: Radial diagram chart.

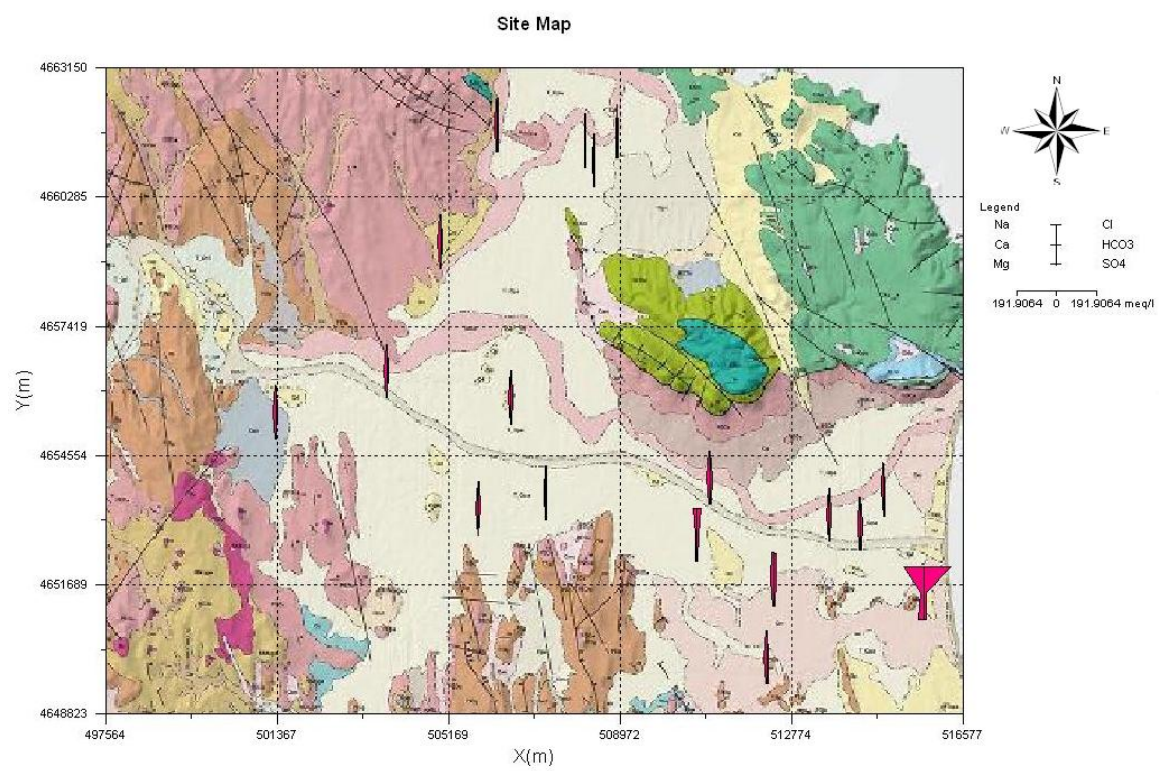


Fig. 22: Stiff chart.

6.3.5 Vertical line and Schoeller-Berkaloff diagrams

Several vertical half-lines are distributed equispacedly, and on each of them an ion is represented by means of a segment proportional to its concentration in meq/l, ppm or %. The different points obtained are subsequently joined by a polyline. A few simultaneous analyses can be represented simultaneously. When the axes are logarithmical one is dealing with Schoeller-Berkaloff diagrams. Absolute concentrations in meq/l are used. Generally the columns represent, from left to right: Ca , Mg , $Na+K$, Cl , SO_4 , HCO_3 . The inclination of the line joining the representative points of the two ions measures its relationship.

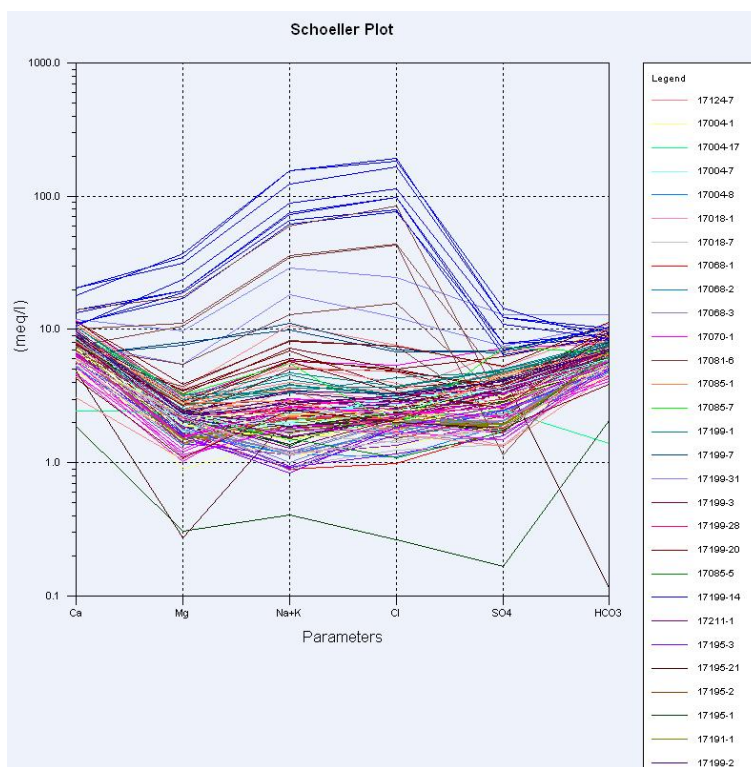


Fig. 23: Schoeller diagram years 1996-2004.

One plot with all samples of all years might be too confusing (Fig.23), so one diagram for each year have been done as such in figure 24.

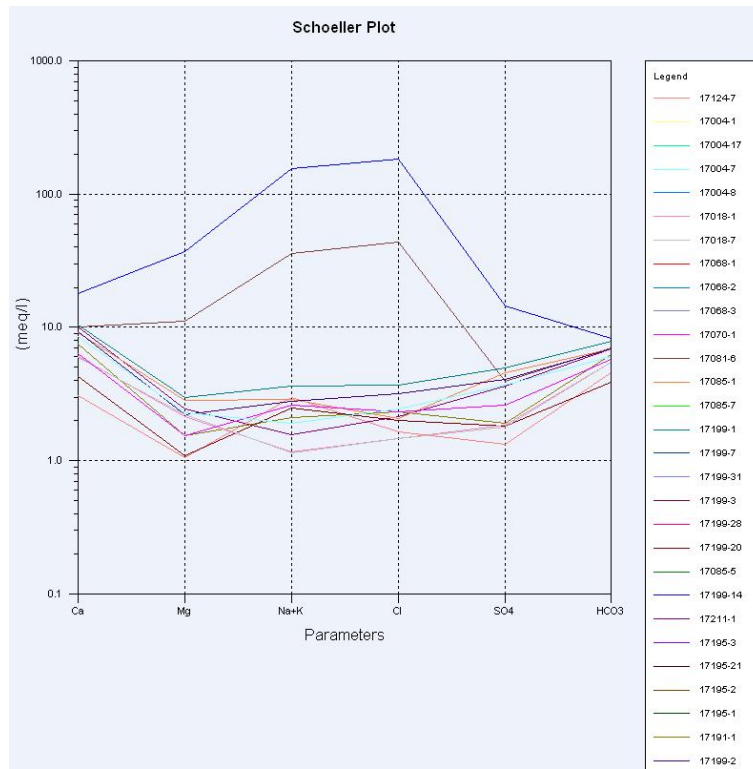


Fig. 24: Schoeller diagram year 2004

As with the previous diagrams, the most outstanding feature of these diagrams is the presence of two groups of waters ($Na - Cl$ and $Ca - HCO_3$). Gaining more information from these graphs requires some expertise. The main problem is, like in the case of Piper diagrams, the difficulty of spatial associations. Information about time evolution can be obtained by succession of pictures which provides a rapid visualization of time changes in the representation.

6.3.6 Scatter plots

These plots represent the value of two ions over two cartesian coordinate axes. They are interesting for comparing analysis performed on a same point at different time or for different water analysis over a same zone. They can be used to highlight relationships between ions. The diagrams can be lineal or double-logarithmic if the variations range is over two orders of magnitude.

For example, representing rCl vs rNa , we get the rNa/rCl relationship. If $rCl \leq rNa$, it can be a sign of mineralization due to silicate attack (granitic terrains) as the content of Cl is small and, on the other hand, the content of Na can be very high. These diagrams have been represented with Aquachem (figure 25) and SPSS (figure 26).

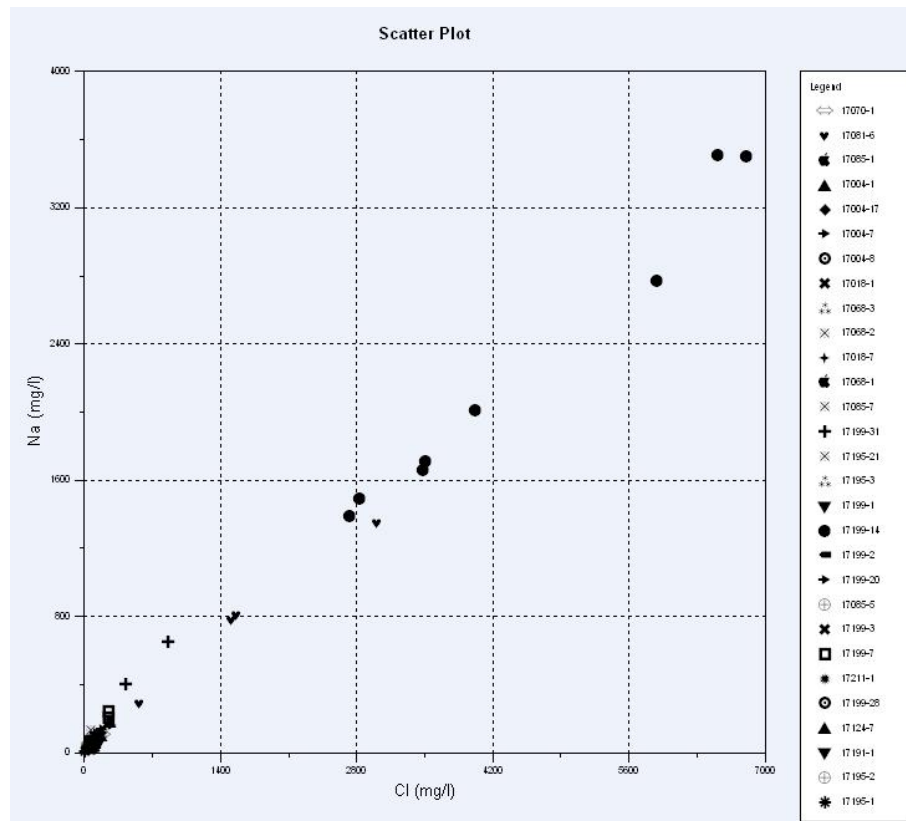


Fig. 25: Na-Cl scatter plot with Aquachem.

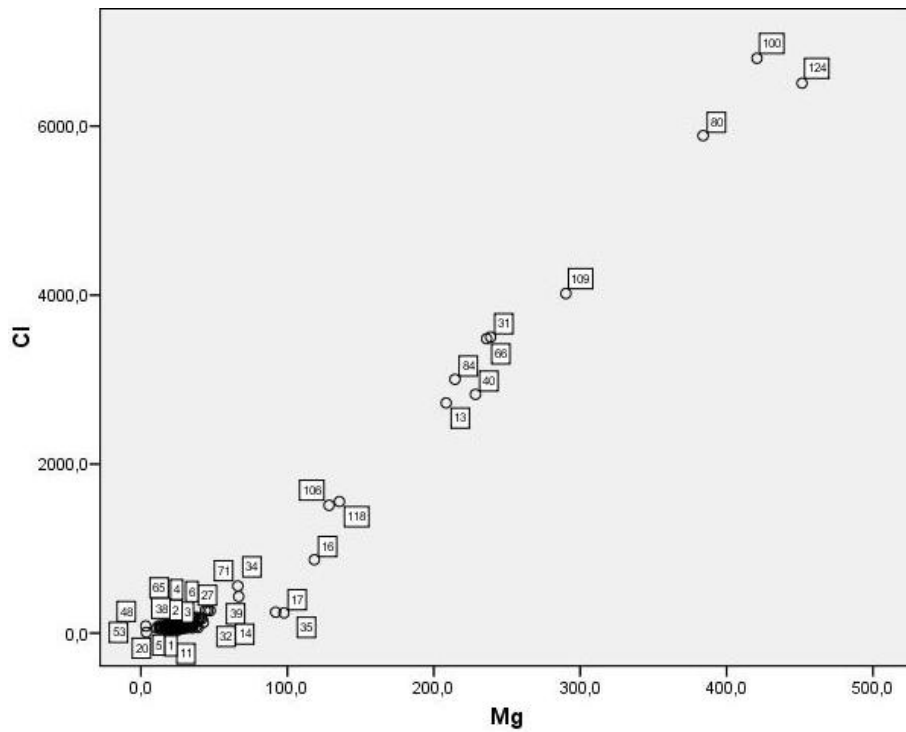


Fig. 26: Mg(mg/l)-Cl(μ g/l) Scatter plot with spss

Labeling and scaling is more easily done with aquachem. However, SPSS has also the option of representing these scatter plots as a matrix, which can provide a better understanding of the ion relationships as can be seen in figure 27.

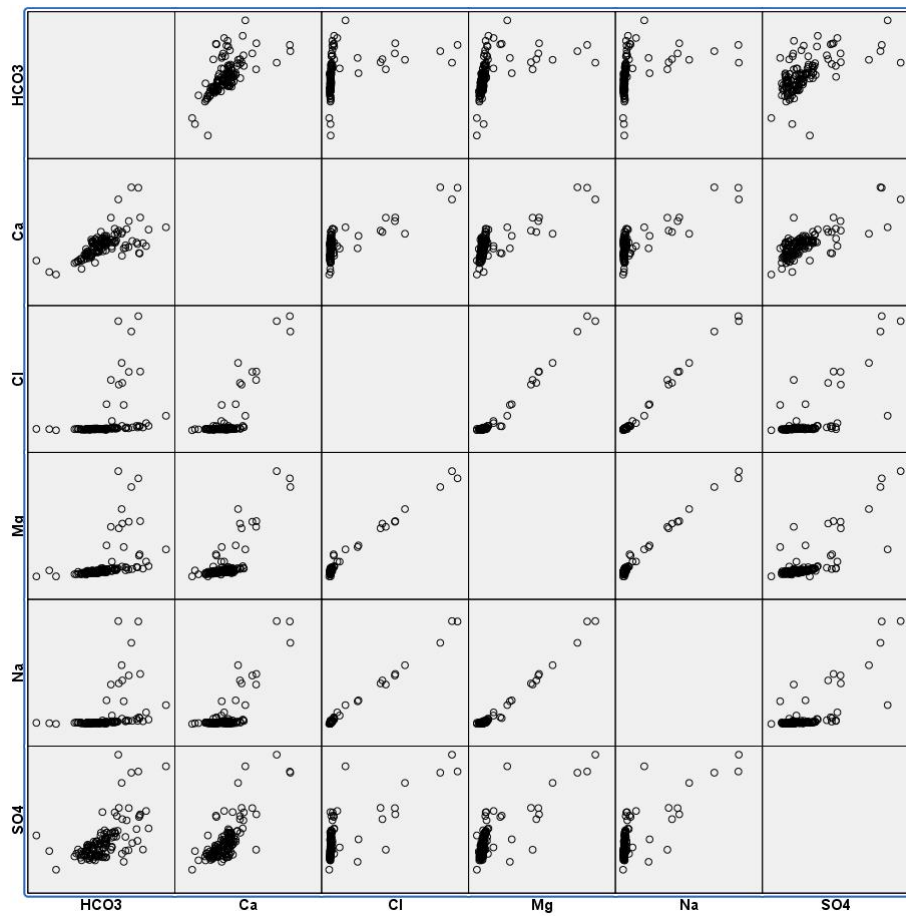


Fig. 27: Scatter plot matrix done with SPSS

In figure 27, simultaneous linear relationships can be observed between Cl , Mg and Na . When these ions are represented against the others, sea water mixing controls the appearance of the graphs. In order to contrast this information, a same matrix plot has been done, but in logarithmic scales.

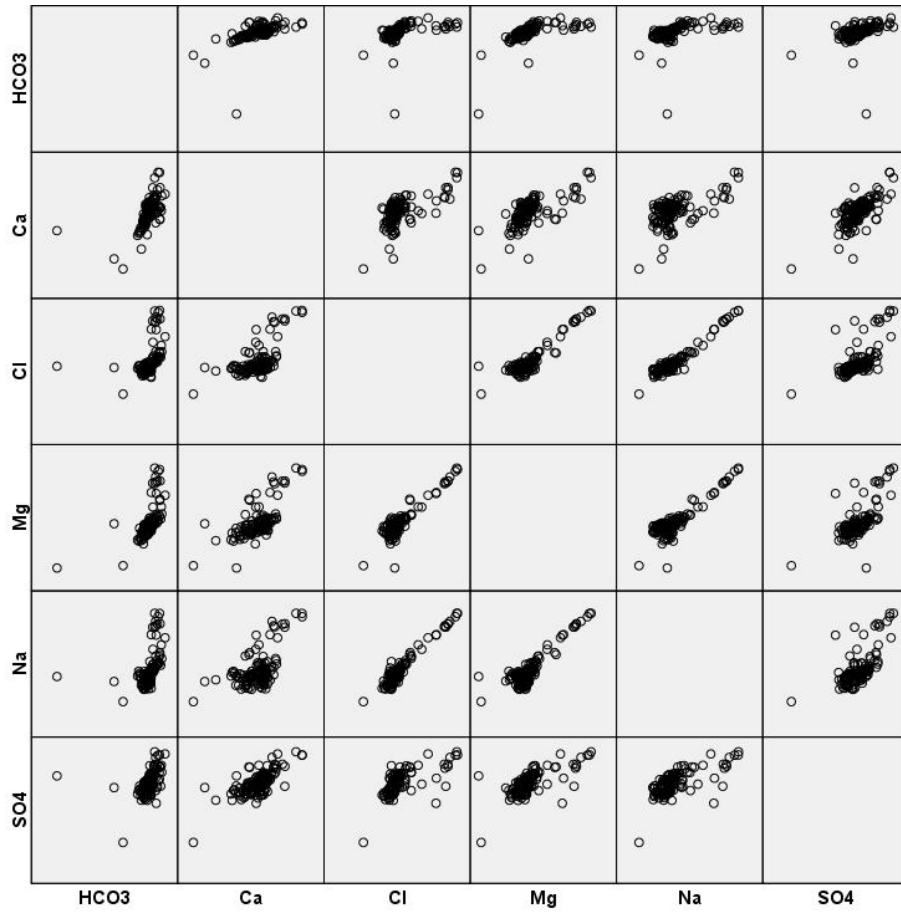


Fig. 28: Scatter plot log-matrix done with SPSS

6.3.7 Frequency diagrams

In case of having a huge amount of chemical data it can be interesting to analyze the frequency on which given values occur. For every range of values, the frequency or the number of samples that correspond to every value, and the representation is made accordingly. So, for instance, if we obtain a definite gaussian function it means that all analyses correspond to waters of a same aquifer. If we obtain two or more aggrupations, it means that we have two or more aquifers or zones. In case of no appearacnce of any aggrupation, it could mean that one is dealing with many different aquifers or with very variate special characteristics. These diagrams have been done with Aquachem (figure 29) and SPSS (figure 30). Only one example with each program is represented as they have not showed any bimodality.

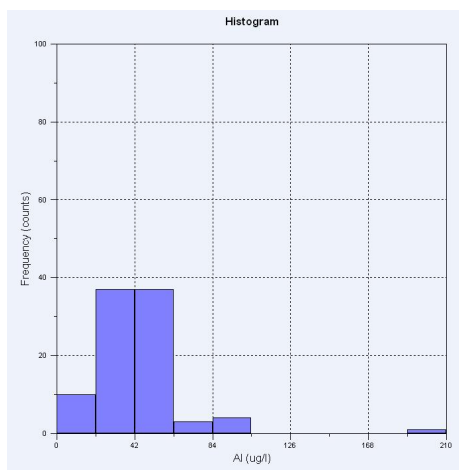


Fig. 29: Al ($\mu\text{g/l}$) histogram made with Aquachem

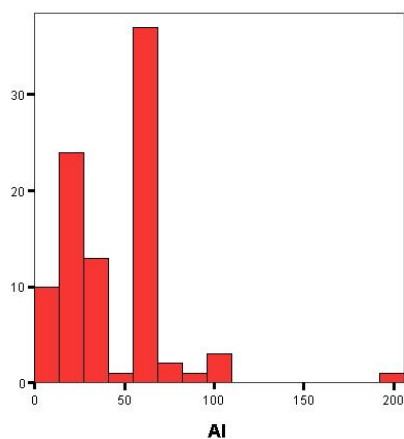


Fig. 30: Al ($\mu\text{g/l}$) histogram made with SPSS in $\mu\text{g/l}$.

6.3.8 Hydrograms

They are graphics which represent the temporal evolution of a given characteristic. It is the most clear form of visualizing temporal variations. If several characteristics are drawn on one same hydrogram, we can see how they influence each other. Aquachem allows us to represent in one graph the evolution of one value in many stations or the evolution of different parameters in one station. Due to the huge amount of stations the first kind of diagrams are too confusing, as can be seen in figure 31:

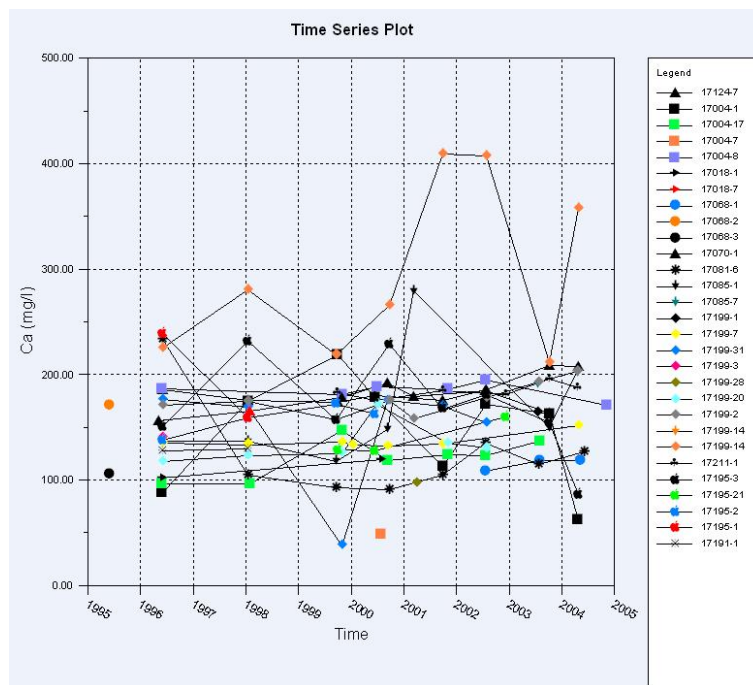


Fig. 31: Ca(mg/l) vs. time

The visualization of multiple parameters in on only station is an easier way to interpret temporal evolution. The obtained graphics are less confusing and allow an easier interpretation of concentration trends. In order to not get confused with false concentrations between different years due to dilution of components by recent rain events, these analysis have been accompanied with their rain hydrograms to understand sudden increases or decreases of concentration. Two examples are shown in figures 32 to 39:

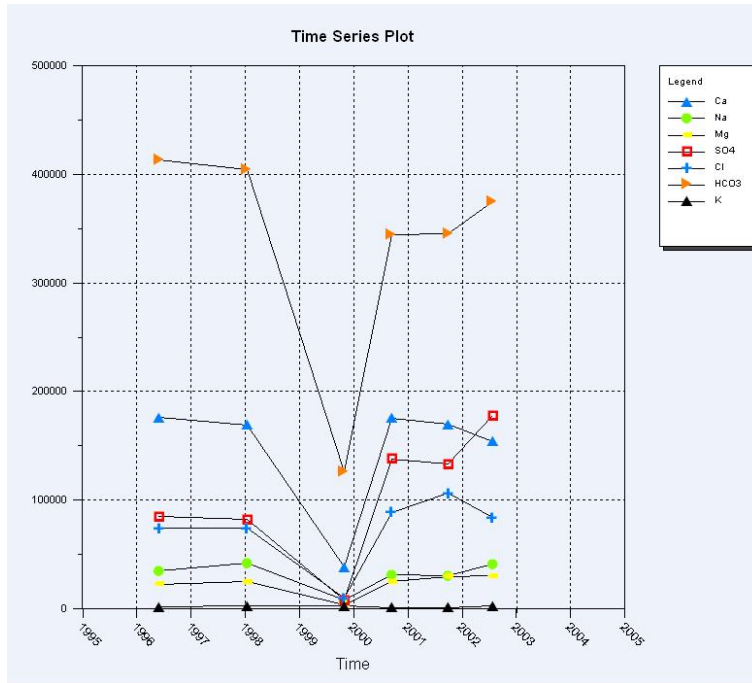


Fig. 32: Multiparameter station 17195-1 vs. Time. Low peak in year 99 analysis due to “wet” period.

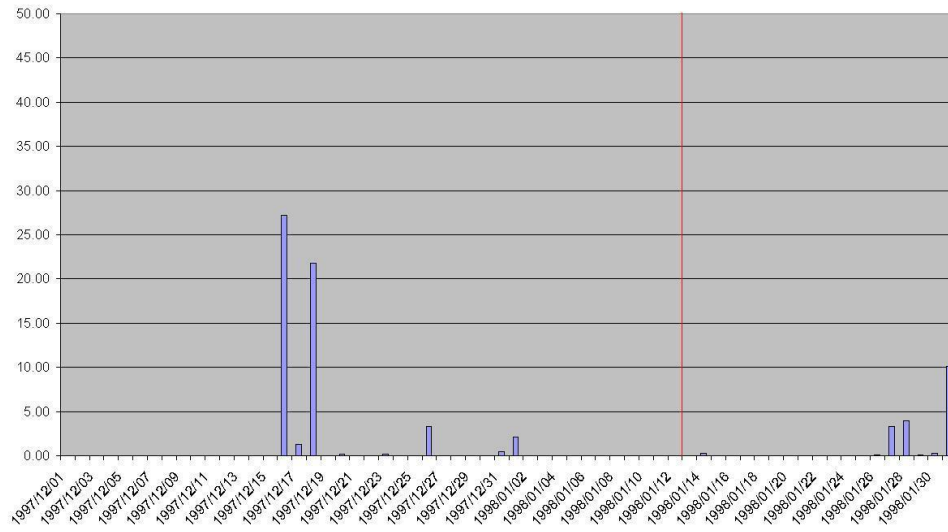


Fig. 33: Accumulated daily rain events for station 17195-1 prior to the 1998 sampling(red line). Dry period, 25 days without rain.

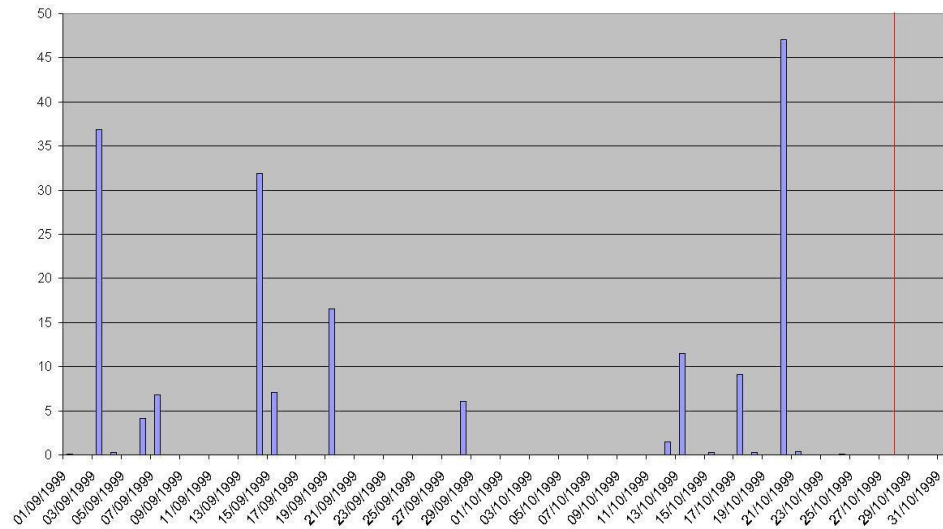


Fig. 34: Accumulated daily rain events for station 17195-1 prior to the 1999 sampling (red line) in mm. Recent rain event, 47,5mm 7 days before.

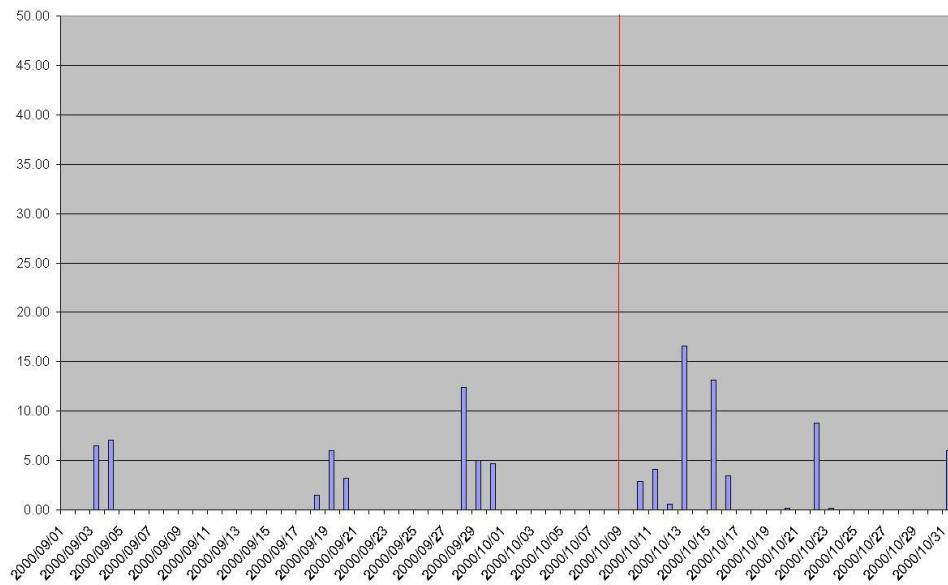


Fig. 35: Accumulated daily rain events for station 17195-1 prior to the 2000 sampling (red line). Dry period, only 12,5mm rain 10 days before.

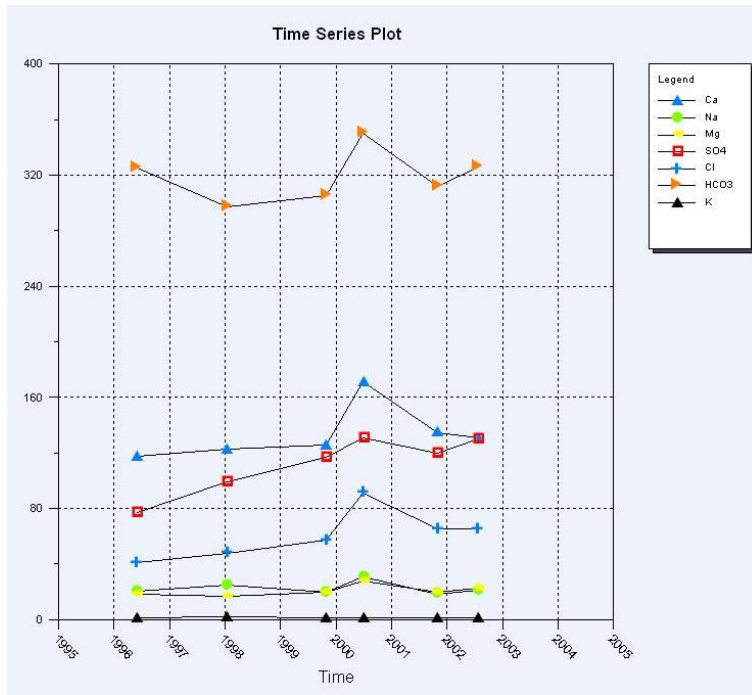


Fig. 36: Multiparameter station 17195-3 vs. Time. High concentration in year 2000 analysis due to “dry” period.

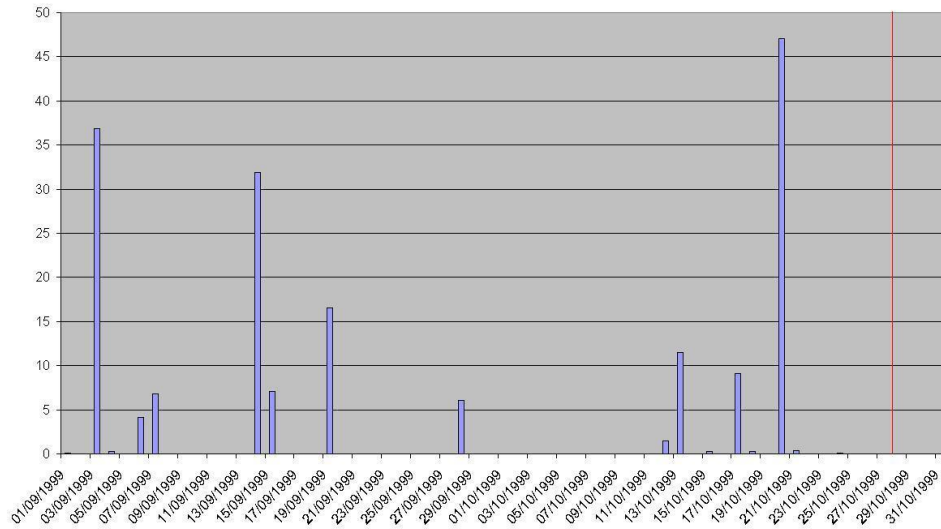


Fig. 37: Accumulated daily rain events for station 17195-1 prior to the 1999 sampling(red line) in mm. Wet period, recent rain event (47,5mm) 7 days before.

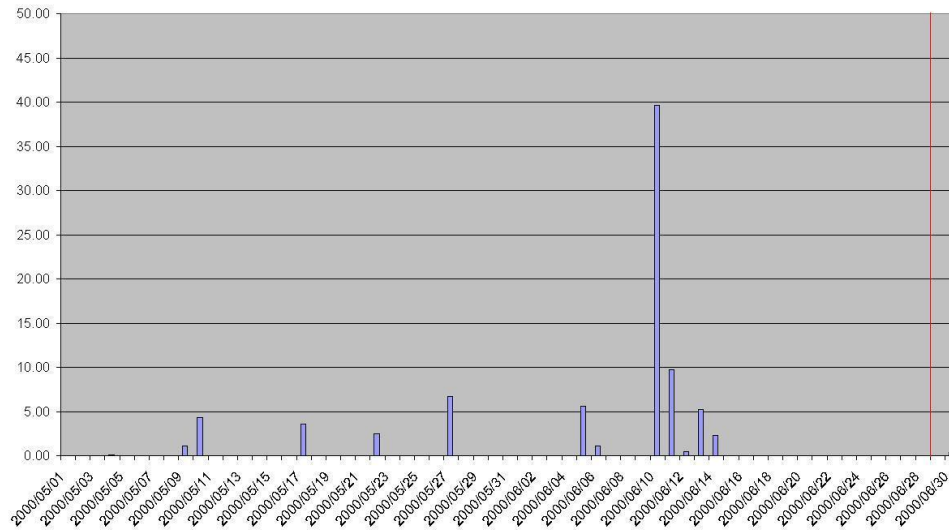


Fig. 38: Rainfall events for station 17195-3 prior to the 2000 sampling(red line) in mm. Dry period, 19days without rain.

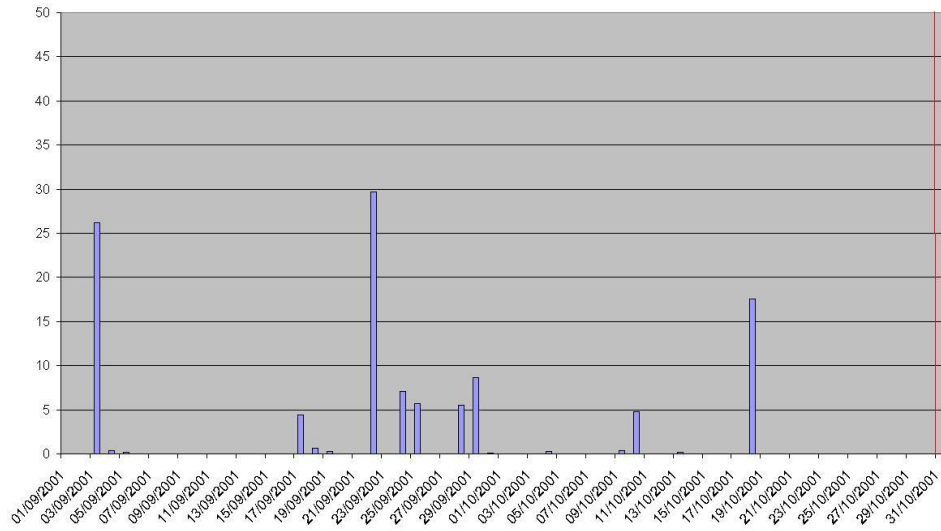


Fig. 39: Rainfall events for station 17195-3 prior to the 2001 sampling(red line) in mm. Wet period, 3 high rain events, last 12 days before the analysis date.

If this last example is also represented in logarithmic scale, temporal evolution of parameters with small values are more easily detected. However, variations in the highest ones are smoother. This is due to the scale change. This can be seen in figure 40:

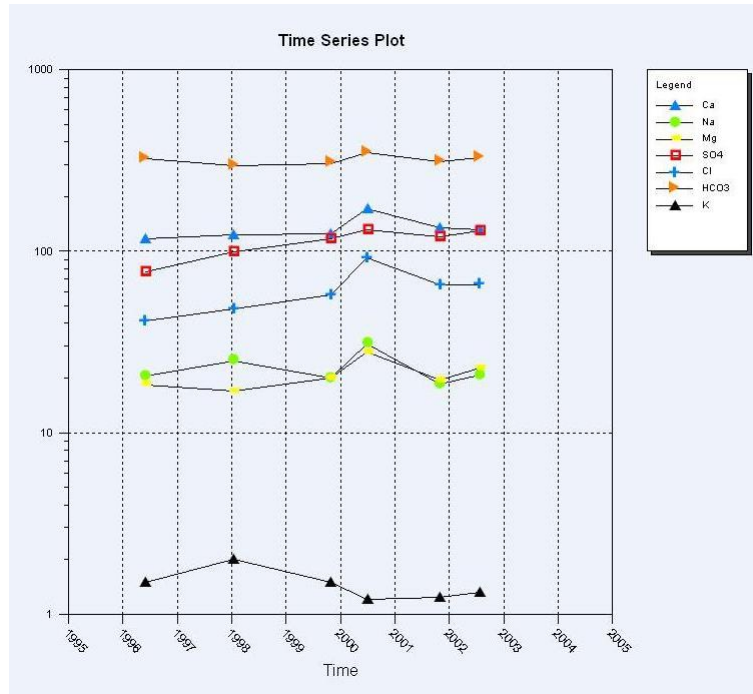


Fig. 40: Multiparameter station 17195-3 vs. Time. Changes in K , Na and Mg are more evident and in HCO_3 and SO_4 smoothened.

6.4 Multivariate statistics techniques.

Prior to the results explanation, this section has summarized the information extracted from Härdle-Simar's Applied Multivariate Statistical Analysis, ?? in order to have a global idea of what these techniques are used for to ease its interpretation.

One of the main objectives in statistics is to better understand and model the underlying processes which might generate a set of data. This is known as statistical inference: we infer properties of the population from which the observations are taken from information contained in a sample. Via decomposition of data matrices by factors we can take a descriptive perspective on how using geometrical approach a way of reducing the dimension of a data matrix can be derived with respect to least-squares criterion. The result will be low dimensional graphical picture of the data matrix. That is the objective is to explain as much variability observed in a large data set in terms of a few factors.

We assume that the data matrix $\chi(n \times p)$ is composed of n observations (or individuals) and p variables. Thus, each row corresponds to an observation and each column to a variable.

$$\chi = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

If we consider each row as a vector $x_i^T = (x_{i1}, \dots, x_{ip}) \in R^p$, we can represent our data matrix as a cloud of n points in R^p such as in figure 41.

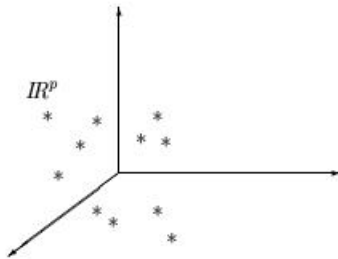


Fig. 41: n -points cloud in p -dimensional space.

When p is large (larger than 2 or 3), we cannot produce interpretable graphs of these clouds of points. Therefore, the aim of the factorial methods to be developed here is two-fold. We shall try to simultaneously approximate the column space $C(X)$ and the row space $C(X_i)$ with smaller subspaces without losing too much information about the variation and structure of the point clouds in both spaces. Ideally, this will provide insights into the structure of χ through graphs in R , R^2 or R^3 . The main focus then is to find the dimension reducing factors.

To project n -points cloud onto a space of lower dimension we begin finding a subspace of dimension 1. Assuming that the data have been previously centered, the problem boils down to finding a straight line F_1 through the origin (Figure 42). The direction of this line can be defined by a unit vector $u_1 \in R^p$. Hence, we are searching for the vector u_1 which gives the best fit of the initial cloud of n points.

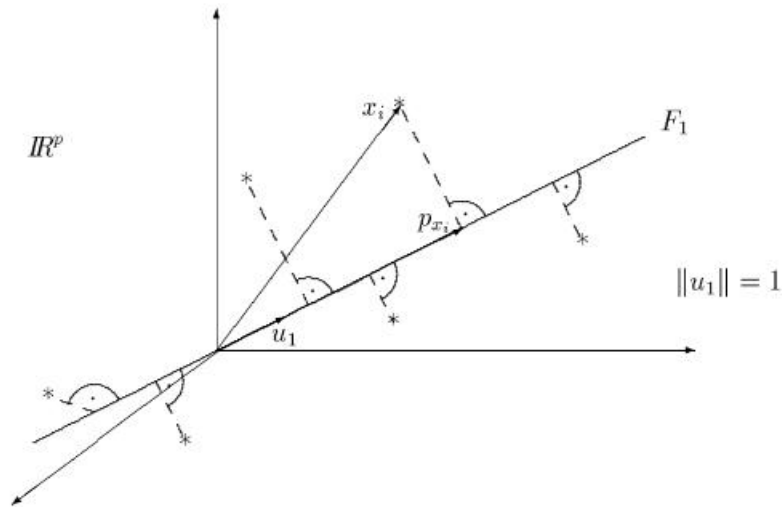


Fig. 42: Reduction of a n point cloud to a 1D subspace.

The representation of the i -th individual $X_i \in R^p$ on this line is obtained by the projection of the corresponding point onto u_1 , the projection point p_{X_i} . The coordinate of x_i on F_1 is given by:

$$p_{x_i} = x_i^T \frac{u_1}{\|u_1\|} = x_i^T u_1.$$

We define the best line F_1 in a least-squares sense: Find $u_1 \in R^p$ which minimizes the sum of the square distances between the original x_i points and their projections p_{X_i} :

$$\sum_{i=1}^n \|x_i - p_{x_i}\|^2.$$

The coordinates of the n individuals on F_1 is called the first factor and u_1 the first factorial axis. The n individuals, x_i , are now represented by a new factorial variable $z_1 = \chi u_1$. This new factorial variable is a linear combination of the original variables (x_1, \dots, x_p) whose coefficients are given by the components of vector u_1 :

$$z_1 = u_{11}x_1 + \dots + u_{p1}x_p.$$

If we want to approximate the n individuals to a 2-dimensional subspace, this space will contain u_1 . The plane will be determined by the best linear fit u_1 and a unit vector u_2 orthogonal to u_1 as in figure 43.

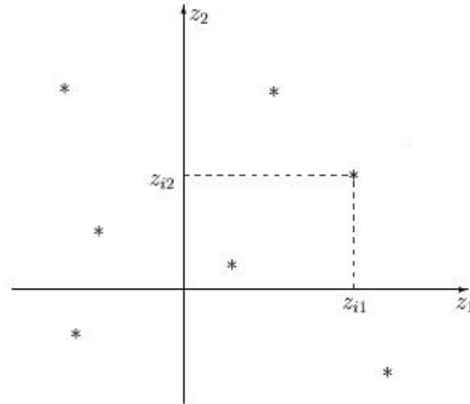


Fig. 43: Reduction of a n point cloud to a 2D subspace.

If we generalize this to q -dimensional subspaces where $q \leq p$, the best subspace is generated by u_1, u_2, \dots, u_q , the orthonormal eigenvectors of $\chi^T \chi$ associated to their corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q$. The coordinates of the n individuals on their k factorial axis will be given by the factorial variables $z_k = \chi u_k$ for $k = 1, \dots, q$. Each factorial variable is a linear combination of the original variables. The representation of the n individuals is obtained by plotting $z_1 = \chi u_1$ vs. $z_2 = \chi u_2$ (and, if necessary, vs. $z_3 = \chi u_3$).

The quality of the factorial representation is usually defined using τ_q which is the percentage of inertia explained by the first q factors:

$$\tau_q = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_q}{\lambda_1 + \lambda_2 + \dots + \lambda_p}.$$

The main idea of the principal components analysis is to find a standardized linear combination which maximizes the variance of the projection of the original point cloud. An example is drawn in figure 44.

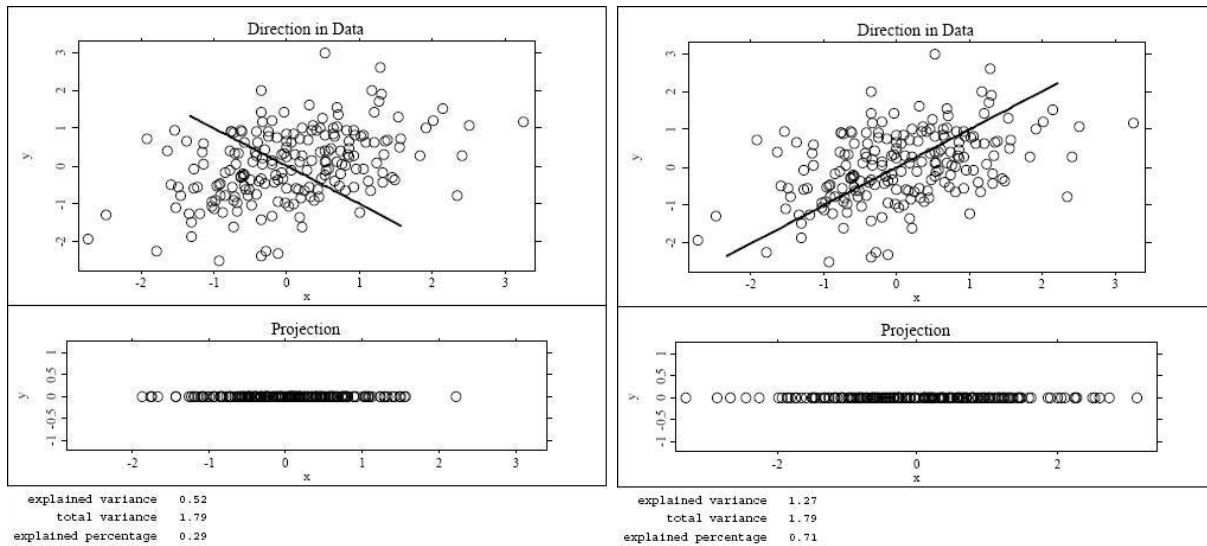


Fig. 44: Projection of n points cloud onto a random line vs. projection of n points cloud onto line that maximizes the variance of the projected points.

As principal components analysis is performed on the covariance matrix it is sensitive to scale changes. So, it should be applied on a data set with approximately the same scale in each variable. If the variables are on different scales, we will have to standardize the variables. This way, each variable will have the same weight to find the solution. If the data set is not normalized, variables with more variance are considered more important.

6.4.1 Factor analysis

In factor analysis the same problem of reducing the dimension of a multivariate random variable data set is addressed but the number of factors is fixed from the start. Each factor is interpreted as a latent characteristic of the individuals revealed by the original variables.

The aim of this technique is also to reduce the dimension of the observed data. The perspective used, however, is different: we assume that there is model, Factor Model, stating that most of the covariances between the elements of χ can be explained by a limited number of latent factors. As this technique is based on the covariance matrix, it should be scale invariant.

The non-uniqueness of solutions will have to be dealt by using techniques which lead to easier interpretations. This will involve (geometric) rotations of the factors. The number of solutions is defined by the degrees of freedom of the system:

$$d = 1/2(p - k)^2 - 1/2(p + k).$$

Where k is the number of fixed factors.

A way of understanding how well explained a variable is in our factor model will be the communalities. A communality is the proportion of a variable's variance explained by a factor structure. This value ranges between 0 and 1. The closer this value is to 1, the better the variable is explained.

There are different methods to estimate k . It is important that the number of factors explains an acceptable cumulated variance. Some of those methods are:

- Maximum likelihood method.
- Method of Principal factors.
- Principal component method.

As the comprehension of these methods is complicated the main idea is to try different hypothesis. So, it is recommended to use different methods, including rotation, and compare which solution results in a more comprehensible factor model(considering loadings, communalities and factor scores).

SPSS performs factorial analysis giving numerical results with the % of variance explained by the factors of the given solution, the components of each factor with or without rotation. This way one can get an idea of which variables can be involved in a latent factor. The process of applying factor analysis has involved many hypothesis as including or excluding variables, conversions to logarithmic scale or homogenization of all parameter units. Each one has given different interpretations.

No best solution can be defined as all factor analysis algorithms are considered equally valid. The criteria to decide the most "correct" solution is to find the one which gives a more understandable explanation of the data set structure. The only common factor in all case studies has been a marine intrusion factor. The output format of results that is given by SPSS can be seen in figures 45 to 47:

Comunalidades

| | Inicial | Extracción |
|---------------|---------|------------|
| NH4 | 1.000 | .733 |
| HCO3 | 1.000 | .770 |
| Ca | 1.000 | .885 |
| Cl | 1.000 | .938 |
| Cond | 1.000 | .823 |
| Meas_hardness | 1.000 | .972 |
| Fe | 1.000 | .684 |
| Mg | 1.000 | .959 |
| Mn | 1.000 | .774 |
| NO3 | 1.000 | .702 |
| NO2 | 1.000 | .765 |
| K | 1.000 | .849 |
| Na | 1.000 | .944 |
| SO4 | 1.000 | .845 |
| TOC | 1.000 | .390 |
| Si | 1.000 | .615 |

Método de extracción: Análisis de Componentes principales.

Fig. 45: Table of communalities. Its closeness to 1 gives an idea of how good a variable is explained by the obtained solution.

Varianza total explicada

| Componente | Autovalores iniciales | | | Sumas de las saturaciones al cuadrado de la extracción | | |
|------------|-----------------------|------------------|-------------|--|------------------|-------------|
| | Total | % de la varianza | % acumulado | Total | % de la varianza | % acumulado |
| 1 | 8.572 | 53.573 | 53.573 | 8.572 | 53.573 | 53.573 |
| 2 | 1.708 | 10.672 | 64.245 | 1.708 | 10.672 | 64.245 |
| 3 | 1.292 | 8.077 | 72.322 | 1.292 | 8.077 | 72.322 |
| 4 | 1.077 | 6.733 | 79.055 | 1.077 | 6.733 | 79.055 |
| 5 | .934 | 5.837 | 84.892 | | | |
| 6 | .690 | 4.314 | 89.206 | | | |
| 7 | .552 | 3.452 | 92.659 | | | |
| 8 | .472 | 2.950 | 95.609 | | | |
| 9 | .265 | 1.654 | 97.263 | | | |
| 10 | .234 | 1.465 | 98.728 | | | |
| 11 | .096 | .601 | 99.328 | | | |
| 12 | .069 | .429 | 99.757 | | | |
| 13 | .036 | .226 | 99.983 | | | |
| 14 | .002 | .016 | 99.999 | | | |
| 15 | .000 | .001 | 100.000 | | | |
| 16 | 1.7E-007 | 1.1E-006 | 100.000 | | | |

Método de extracción: Análisis de Componentes principales.

Fig. 46: % of variance explained by each factor. In this case, only the 4 first factors are considered for the solution.

Matriz de componentes^a

| | Componente | | | |
|---------------|------------|-------|-------|-------|
| | 1 | 2 | 3 | 4 |
| NH4 | .461 | -.418 | -.334 | .484 |
| HCO3 | .552 | .325 | -.596 | .067 |
| Ca | .764 | .539 | .005 | -.110 |
| Cl | .960 | .005 | .118 | -.046 |
| Cond | .786 | -.255 | -.014 | .374 |
| Meas_hardness | .968 | .174 | .058 | -.031 |
| Fe | .643 | -.330 | -.020 | .400 |
| Mg | .976 | .016 | .075 | .002 |
| Mn | .859 | -.128 | -.008 | -.142 |
| NO3 | -.217 | .666 | .306 | .343 |
| NO2 | -.029 | -.511 | .708 | -.046 |
| K | .900 | -.057 | .037 | -.187 |
| Na | .963 | .006 | .114 | -.055 |
| SO4 | .858 | .159 | .163 | -.237 |
| TOC | .602 | -.090 | .072 | -.120 |
| Si | .068 | .419 | .400 | .524 |

Método de extracción: Análisis de componentes principales.
a. 4 componentes extraídos

Fig. 47: Components of the factors. The parameters with highest values are highlighted. So, a marine first factor can be deduced.

R also allows a more limited factor analysis with the maximum likelihood extraction method. As we also need a data matrix with no voids, not all analysis can be used and results are sensibly different as can be seen in figure 48.

```

Loadings:
      Factor1 Factor2 Factor3 Factor4 Factor5
NH4      0.281  0.552      0.256 -0.384
HCO3     0.343  0.140  0.279  0.789  0.273
Ca        0.825      0.484  0.216
CE...lab  0.920  0.334  0.160
Cl        0.951  0.250  0.141
Cu        0.355 -0.136
Meas_hardness_mg 0.949  0.225      0.206
Fe        0.337  0.703  0.281      0.203
Mg        0.938  0.298  0.144
Mn        0.775  0.161  0.329
NO3       -0.566  0.116
NO2      -0.100 -0.190
K         0.799  0.109  0.464  0.158
Se        -0.153
Si        -0.197  0.344
Na        0.951  0.246  0.163
SO4       0.866      0.482
TOC       0.429  0.484
Zn        0.430 -0.129

      Factor1 Factor2 Factor3 Factor4 Factor5
SS loadings  7.806  1.399  1.258  1.194  0.577
Proportion Var 0.411  0.074  0.066  0.063  0.030
Cumulative Var 0.411  0.484  0.551  0.614  0.644

Test of the hypothesis that 5 factors are sufficient.
The chi square statistic is 809.13 on 86 degrees of freedom.
The p-value is 4.92e-118
> |

```

Fig. 48: Factor analysis made with R. First factor clearly continues as marine but others change.

With a more complicated data management, ArcGIS has a Principal Components analysis option to perform on multiple rasters. If data are converted to appropriate format this tool can be used to represent the new factors components on our map. However, results do not look very similar to previous ones as the mathematic algorithm used by this code might be considerably different to the ones used in SPSS or R. An example of the results obtained with ArcGIS are shown in figures 49-52. In figure 49 the big differences in factor loadings with regard to the ones obtained previously can be seen. Figure 50 shows a graphic where apparently samples with marine intrusion are easily detected. But, in following figures, the plotting of second factor (figure 51) and third factor (figure 52) have no evident interpretation.

| # Eigenvalues | | | | | | | |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------|
| 2,598137e+009 | 2,453243e+007 | 6,078373e+006 | 2,114472e+006 | 8,253627e+005 | 1,874155e+005 | 1,445012e+005 | |
| # Eigenvectors | | | | | | | |
| # Input | Layer | | | | | | |
| nh4 | | 0.0002 | 0.0006 | 0.0012 | -0.0038 | 0.0013 | -0.0074 |
| hco3 | | 0.0243 | 0.9023 | 0.3809 | -0.1125 | 0.0094 | 0.0857 |
| ca | | 0.0299 | 0.1749 | 0.0279 | 0.7178 | 0.3360 | -0.3858 |
| cl | | 0.8760 | -0.1345 | 0.1631 | 0.1595 | 0.1970 | 0.2736 |
| fe | | 0.0012 | -0.0003 | 0.0000 | -0.0139 | -0.0005 | -0.0103 |
| mg | | 0.0586 | 0.0623 | -0.0197 | -0.0540 | -0.0037 | 0.5293 |
| mn | | 0.0005 | 0.0004 | 0.0000 | -0.0017 | 0.0025 | -0.0021 |
| no3 | | -0.0040 | 0.0474 | -0.0229 | 0.5447 | -0.8003 | 0.2131 |
| no2 | | 0.0000 | -0.0005 | -0.0017 | 0.0010 | -0.0008 | 0.0023 |
| ph | | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| k | | 0.0134 | 0.0156 | -0.0203 | -0.1822 | -0.0858 | 0.2535 |
| se | | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0002 |
| si | | 0.0000 | 0.0000 | 0.0001 | 0.0003 | 0.0002 | 0.0004 |
| na | | 0.4712 | 0.1319 | -0.1822 | -0.3292 | -0.4167 | -0.5872 |
| so4 | | 0.0739 | 0.3368 | -0.8905 | 0.0624 | 0.1586 | 0.1717 |
| toc | | 0.0034 | -0.0069 | -0.0024 | -0.0377 | -0.0398 | 0.0679 |
| zn | | 0.0000 | 0.0000 | 0.0001 | 0.0002 | 0.0003 | 0.0004 |

Fig. 49: Factor loadings given by ArcGIS.

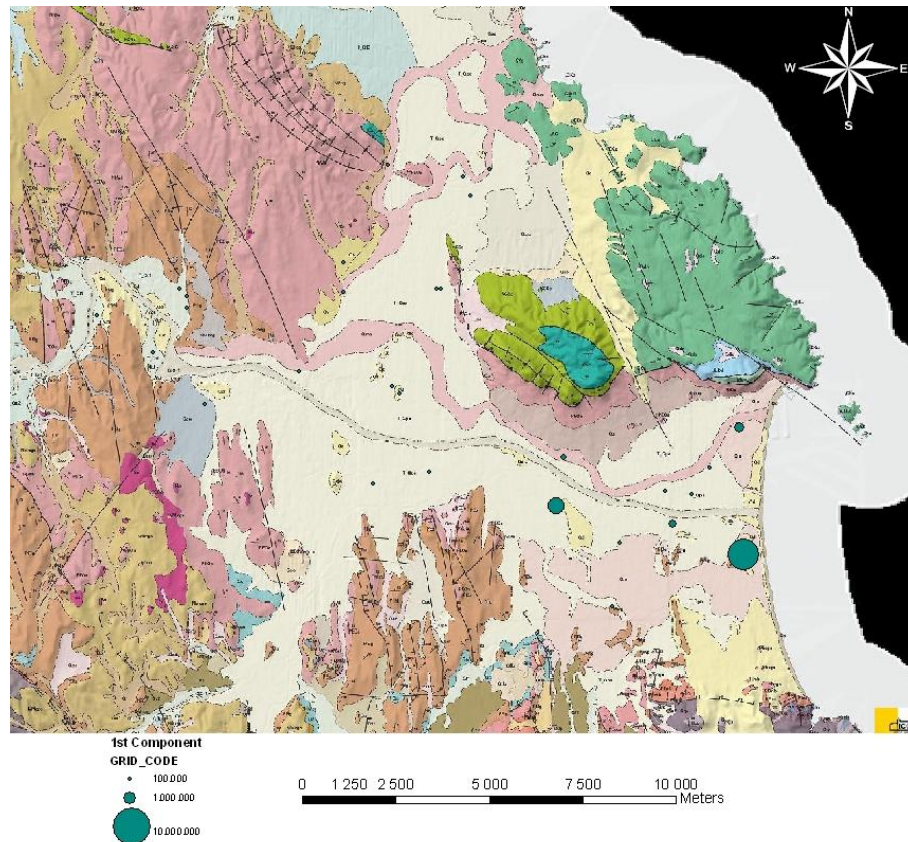


Fig. 50: Representation of first factor with ArcGIS.

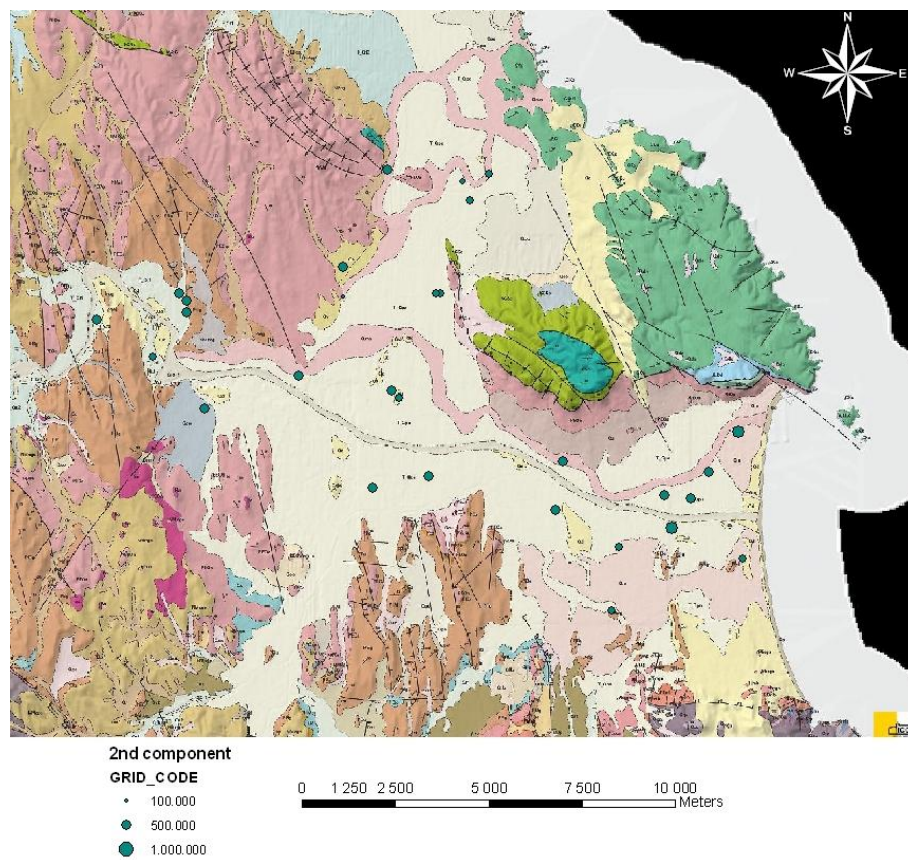


Fig. 51: Representation of second factor with ArcGIS.

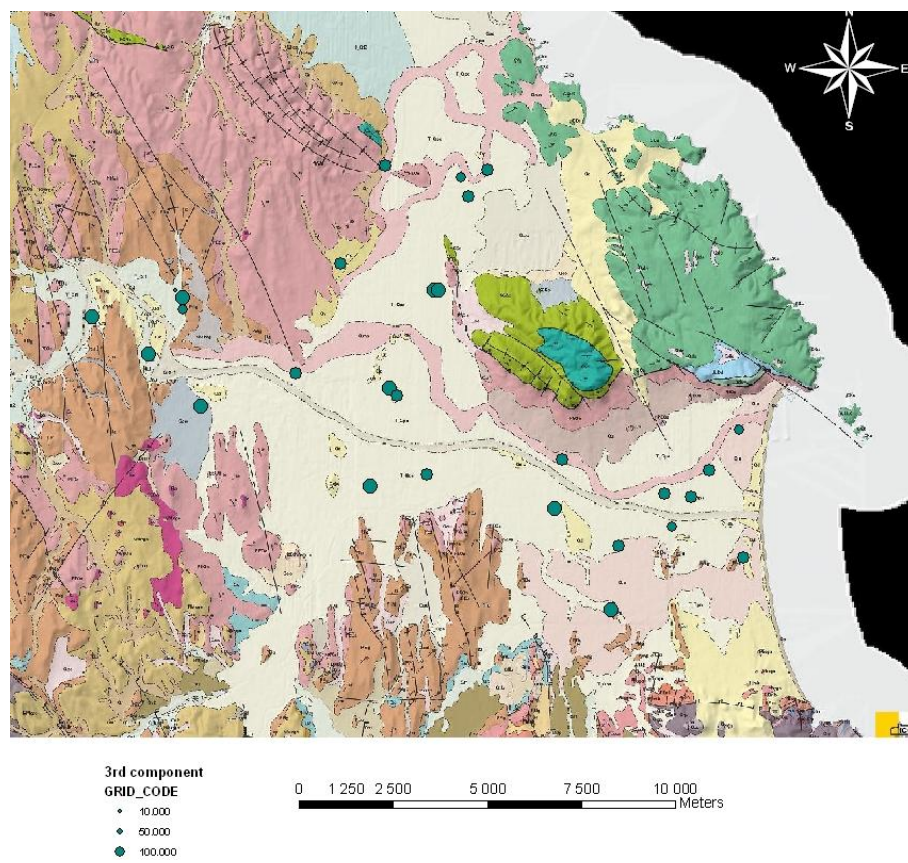


Fig. 52: Representation of third factor with ArcGIS.

6.4.2 Cluster analysis

When considering groups of objects in a multivariate data set, two situations can arise. Given data set containing measurements on individuals, in some cases we want to see if some natural groups or classes of individuals exist, and in other cases, we want to classify the individuals according to set of existing groups. Cluster analysis develops tools and methods concerning the former case, that is, given data matrix containing multivariate measurements on large number of individuals (or objects), the objective is to build some natural subgroups or clusters of individuals. This is done by grouping individuals that are similar according to some appropriate criterion. Once the clusters are obtained, it is generally useful to describe each group using some descriptive tool to create better understanding of the differences that exist among the formulated groups. Decision rules are provided in classifying multivariate observation into one of the known groups. In cluster analysis the criterion chosen to measure the similarity among objects clearly plays an important role.

The best way to understand the aim of cluster analysis is with the graphic description of the obtained clusters, called dendrogram, shown in figure 53:

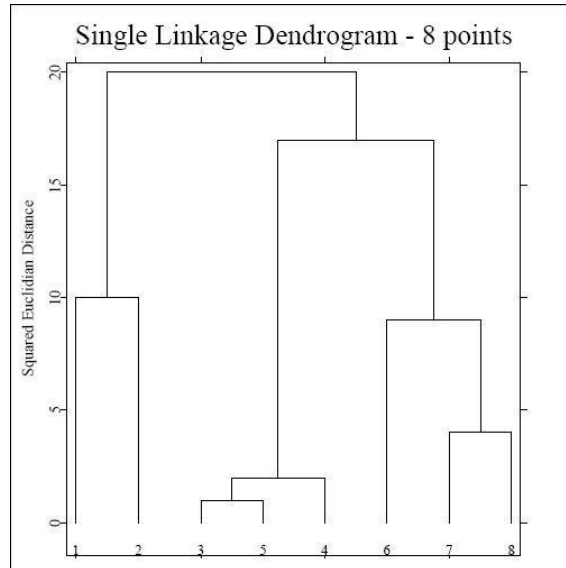


Fig. 53: Cluster analysis dendrogram.

So, as seen in figure 53, cluster analysis constructs groups and subgroups from a multivariate data set. This process is done following two steps: definition of a proximity or distance measure and choice of the group constructing algorithm.

There are two kinds of methods, hierarchical and partitioning algorithms. The first ones

can also be subdivided into agglomerative and splitting procedures. Agglomerative types start from the finest partition possible(each observation is a cluster) and groups them. They are very frequently used and depend on the definition of the distance between clusters. Splitting procedures starts with the coarsest partition possible(one cluster contains all observations) and proceeds splitting the cluster up to smaller sized clusters.

Partitioning algorithms start from a given group definition and exchange elements between groups until a certain score is optimized. The main difference between hierarchical and partitioning algorithms is that when groups are founded in the first methods, this assignment can not be changed. In partitioning algorithms the objects contained in groups may change during the process.

This technique has only been successfully applied with SPSS. R and ArcGIS theoretically have similar functions but it has not been possible to apply them successfully on our data set. So, only results with SPSS will be shown in figures 54 and 55.

The only conclusion that can be drawn from the previous graphics is that there an outstanding sample quite different to the other ones. It corresponds to a sample(17199-14) where saline intrusion had already been detected with other analyzing techniques.

```

***** H I E R A R C H I C A L   C L U S T E R   A N A L Y S I S *****
Dendrogram using Average Linkage (Between Groups)

          C A S E
Label      Num  Rescaled Distance Cluster Combine
-----+-----+-----+-----+-----+
24/07/2002 17195-1  97  |
11/05/2004 17004-7 130  |
21/07/2003 17211-1 115  |
15/04/2004 17211-1 119  |
24/07/2002 17211-1 102  |
19/07/2002 17004-7  91  |
11/10/2001 17085-1  86  |
19/07/2002 17085-1  95  |
10/07/2003 17199-1 108  |
22/04/2004 17199-1 123  |
28/09/2001 17199-1  79  |
28/09/2001 17191-1  77  |
22/04/2004 17191-1 122  |
09/10/2000 17191-1  72  |
19/07/2002 17004-1  90  |
31/07/2003 17004-1 116  |
09/10/2000 17004-1  69  |
31/10/2001 17004-1  87  |
31/07/2003 17018-7 117  |
05/05/2004 17018-1 127  |
09/10/2000 17195-1  73  |
03/10/2001 17068-3  83  |
31/10/2001 17004-7  88  |
09/10/2000 17068-3  70  |
17/07/2003 17085-5 112  |
28/09/2001 17124-7  76  |
22/04/2004 17124-7 121  |
17/07/2003 17085-7 113  |
19/07/2002 17068-3  93  |
22/04/2004 17085-1 120  |

```

Fig. 54: Cluster analysis results, upper half.

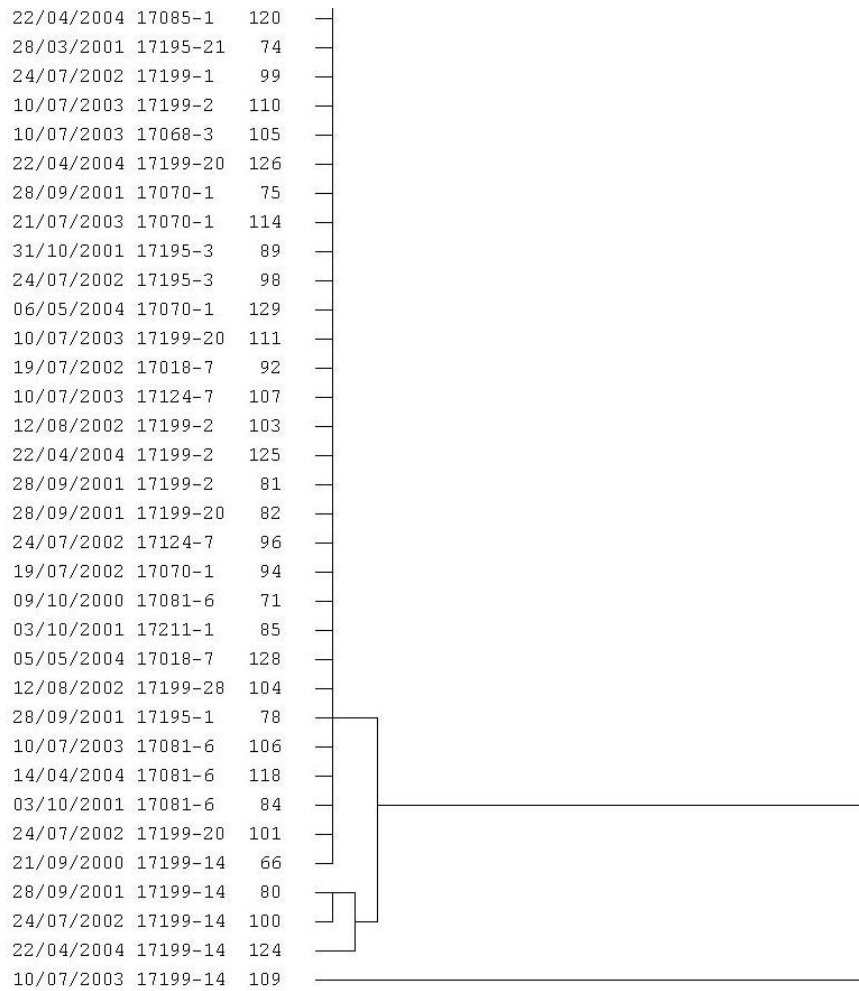


Fig. 55: Cluster analysis resulting dendrogram, lower half.

6.5 Modelling techniques

This section shows two examples of modelling techniques that can be done with aquachem. First of all, an the saturarion indexes of different minerals can be modelled thanks to its PHREEQC environment. Another more briefly commented feature has been sample mixing.

6.5.1 Saturation indexes

In order to obtain a map which displays spatially the saturation indexes of our samples regarding different minerals we use an example of “team work” between two codes. Aquachem calculates these indexes that can be exported to ArcGIS to finally represent them on a geologic map if desired. Figure 56 shows an example of this combination of tools:

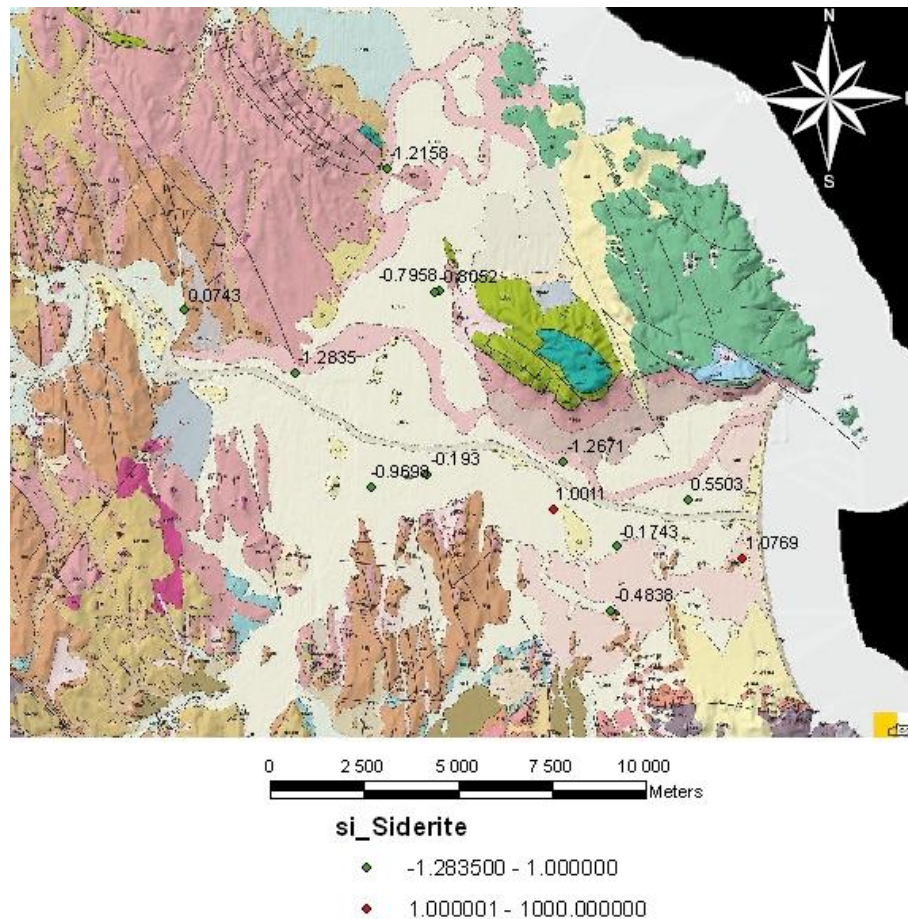


Fig. 56: Saturation index siderite

6.5.2 Sample mixing

This option of Aquachem is not commented by its practical interpretation. It only shows a conceptual mistake that is done by this software when mixing samples. As can be seen in figure 57, a linear combination of the samples pH is done.

| | | | | | | | |
|---|---------|--------|--------|--------|--------|--------|-------|
| Solution 1 | 17081-6 | | | | | | |
| Solution 2 | 17018-1 | | | | | | |
| Percentage of solution 1 in target solution 50%-50% | | | | | | | |
| Solution 1 | 1.0 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.0 |
| Solution 2 | 0.0 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 1.0 |
| Na | 809.7 | 417.04 | 417.04 | 417.04 | 417.04 | 417.04 | 24.38 |
| K | 18.1 | 11.17 | 11.17 | 11.17 | 11.17 | 11.17 | 4.24 |
| Ca | 202.4 | 161.3 | 161.3 | 161.3 | 161.3 | 161.3 | 120.2 |
| Mg | 135.6 | 80.9 | 80.9 | 80.9 | 80.9 | 80.9 | 26.2 |
| Cl | 1557.0 | 804.25 | 804.25 | 804.25 | 804.25 | 804.25 | 51.5 |
| HCO3 | 431.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 332.0 |
| SO4 | 189.0 | 138.5 | 138.5 | 138.5 | 138.5 | 138.5 | 88.0 |
| pH | 7.05 | 7.425 | 7.425 | 7.425 | 7.425 | 7.425 | 7.8 |

Fig. 57: Linear pH mixing

7 Discussion of methods and conclusions.

The objectives of this thesis have been twofold. First, applying and analyzing the results obtained with the different techniques, tools and methods that have been used during this study. One of the final aims of this work has been determining which ones could be more easily interpreted and applied on a generic analysis of a groundwater data set without the need of being an expert hydrogeologist. So, the easiness of interpretation and the degree of explanation given about the processes that might affect the analyzed water samples will be addressed during this chapter.

On the other hand, a second objective has been evaluating the different software programs used to perform the previous steps in terms of how helpful they have been, how laborious to manage and if they have a user friendly environment. Improvement capabilities and cost have also been considered.

- First of all, data treatment and filtering has been a previous step during this work. The actual storage format of the ACA (each row representing one analyzed parameter) is not appropriate to export the data to any of the software tools that have been used. Although Aquachem theoretically accepts this data format, in practice, no successful data import could be done.
- Regarding the obtained results, the classical study techniques for chemical analysis have been the most visual and easy to understand. However, they have not always been useful to explain the chemical status of our samples or to interpret processes. In some cases this has been due to the diagram itself and sometimes to limitations in the software with which it has been represented.
- Radial and Stiff diagrams have not showed interesting results because it has not been possible with the available software to represent the samples with different scales according to their value ranges. Another set of unsuccessful representations have been the frequency diagrams. In this case, as there are not two distinct water bodies, the diagrams have not shown distinct groups. The problem did not lie on the technique itself but on our data samples. Finally, another technique which lead to poor interpretation has been the columnar diagrams, which only served to highlight extreme values that are also seen with other representation techniques.
- More interesting results have been obtained with circular and pie diagrams. These diagrams lead to a fast visualization of the dominant ions as well as their spatial variations

when they are represented on a map. Aquachem, however, does not allow a plot for each sample proportional to the sum of its ions and, therefore, making quantitative interpretations is not possible. Nevertheless, ArcGIS, in a more laborious way can do such a scale variant representation.

- Piper diagrams and scatter plots are techniques that have been useful to represent chemical processes such as water mixing. However, mixing can be easily confused with ion exchange and these techniques have to be interpreted with care. It likely happens with hydrograms of ion concentration evolution. Some cases showed that a high or low peak in concentrations had to be compared with rain events in order to prevent false interpretations. Finally, Schoeller-Berkaloff diagrams might be an interesting tool to interpret water groups or temporal evolution but this needs some expertise.
- Multivariate statistical analysis was a completely different approach to our data sample. Both factor analysis and cluster analysis were applied onto the data set. These techniques have been previously studied in detail in order to better understand their aim.
- Factor analysis has lead to interesting interpretations. However, no unique solution and conclusions can be drawn. This technique requires lots of hypothesis, iterations and data transformations or reductions as considering one parameter or not can lead to substantial changes in the final solutions.
- Cluster analysis has not showed really outstanding results as the only remarkable grouping were the saline samples, which can be easily detected with other methods. However, reworking the distance definition of this technique might improve the results, although it lies outside the scope of this thesis.
- Regarding the evaluated software and its usefulness, the more succesful code has been Aquachem for hydrochemical techniques and SPSS for multivariate statistics. ArcGIS has been considered as an alternative for some diagrams that could not be done with Aquachem. It also has an interesting tool for graphical representation of principal components analysis, which is a first approach to factor analysis. Finally, the R code has been used to get alternative results for some specific techniques such as piper diagrams and factor analysis. Its main interest is the fact that it is an open code and that it can be modified to the user's requirements.

- Additional features of Aquachem are worth commenting. Data unit conversions and parameter order in hydrochemical displays are automatically done. The code also calculates saturation indexes, which might be useful to understand the interactions between water and source rock. However, to be able to use this option a specific and more laborious way of importing data must be done. It has also shown some conceptual errors such as a linear mixing of water samples pH. Finally, its most important graphical lack is that it has no option to define different scales for different sample groups.
- On the other hand, ArcGIS has not been intended to be a complete alternative to Aquachem but has been helpful for some particular graphical representations. Defining parameter order in standardized hydrochemical diagrams is more laborious, but it has shown more efficient graphical display. In the case of principal component analysis, introducing and converting data to an appropriate format has been quite an intensive work, and the results have not been very coherent. However, the option of making a graphical representation would be an interesting tool to combine with other codes used.
- Regarding the statistical codes, SPSS has been the most user-friendly software and the main tool to perform most of the statistical analysis. Data importing and processing has been easily done. Moreover, the comparison of different hypothesis has lead to a more understandable interpretation of the data set structure. It has also been interesting to use its matrix scatter plot option for a fast and easy global understanding of our samples.
- Finally, R has been used in some cases to compare its results with the ones obtained with the previously commented software. The main reason to consider this application is the fact that it is a free and open code that can be conformed to user's requirements. However, the packages that are available at the moment require much more data processing and filtering. It has only been useful for specific techniques so at the moment it can not be considered as a complete alternative to any of the other codes.

8 Future work lines

The future work lines that can be derived from this work will have to deal with three different areas: data formatting, techniques to use and software to apply.

First of all, an appropriate data format will have to be defined in order to improve its export easiness to the different codes that will be used to analyze it.

In some cases the performance of the methodologies employed is case dependent. In others, their usefulness is determined by the data quality and, sometimes, the hypothesis to apply has had influence on the obtained results. To better understand which factors condition the quality of the analysis performed on a generic data set, further study should be done. Therefore, the validation of the methodology and tools proposed here requires the application to other case studies to compare results and extracted interpretations.

After this validation, it might be interesting to define a new code which includes the most interesting features of previously explained software programs. So, Aquachem's modelling and hydrochemical plotting options with the advantages of ArcGIS graphic display such as scale variant legend would be interesting to be implemented in one only tool. Another interesting feature to include would be a multivariate statistical package like the one from SPSS combined with a possibility to display the recalculated factors. Maybe this way a better understanding of multivariate techniques would be achieved.

References

- [1] Härdle, Wolfgang and Simar, Leopold, Applied multivariate statistical analysis, Berlin-New York : Springer, 2003,
- [2] Custodio, Emilio et al., Hidrología subterránea, Barcelona Omega, 1983
- [3] Gould, Stephen Jay, The mismeasure of man, New York-London, W.W.Norton, 1996
- [4] Appelo, C.A.J. and Postma, D., Geochemistry, groundwater and pollution, Rotterdam, A. A. Balkema 2005
- [5] Instituto Geologico y Minero de España, Mapa geológico de España E. 1:50.000, Torroella de Montgrí, Madrid, 1983
- [6] ACA web page, <http://mediambient.gencat.cat/aca/ca/inici.jsp>
- [7] Wikipedia web page, Spss description, <http://en.wikipedia.org/wiki/Spss>
- [8] Geoservei, S.L web page, <http://www.geoservei.com/>
- [9] Institut cartogràfic de Catalunya web page, <http://www.icc.es/web/content/ca/index.html>
- [10] Schlumberger Water Services web page, http://www.swstechnology.com/software_product.php?ID=1
- [11] R-project web page, <http://www.r-project.org/>