

Identification of Anthropogenic Features Through Application of Principal Component Analysis to Hydrochemical Data from the Sines Coastal Aquifer, SW Portugal¹

Paula Galego Fernandes,² Paula Carreira,²
and Manuel Oliveira da Silva³

The Sines coastal sedimentary basin, a tectonic trough with a NE-SW orientation filled with Mesozoic and Cenozoic deposits, has two hydrogeological systems: the Mio-Pliocene and the Jurassic. Both systems are supplying water to the entire region with highly populated and industrialized areas. The aquifers are recharged in the outcrop areas, although the Jurassic aquifer also receives some contribution from the overlying Miocene rocks. The main groundwater flow direction in the Jurassic aquifer is E to W towards the Atlantic Ocean.

In order to understand the influence of anthropogenic activities on the water quality, a statistical principal components analysis (PCA) was performed on the physical–chemical data from groundwater samples in both aquifers. Four components were obtained in the Jurassic and three in the Miocene. The principal components analyses indicate water–rock interactions as the major mechanism responsible for the groundwater solutes (mainly calcium–bicarbonate type), from the calcareous and dolomitic units. Anthropogenic contamination was identified in the influence of the water compositions, corresponding to local polluting inputs. These were magnesium, sulfate, nitrate, iron and copper for PC 3 in the Jurassic and sodium, chloride, nitrate and manganese for PC 2 in the Miocene. These account for about 13% of the system total variance in the Jurassic aquifer and near 23% in the Miocene one.

KEY WORDS: principal component analysis; groundwater evolution; Sines coastal aquifer.

INTRODUCTION

In coastal regions, the problems related to the increase of salinization and pollution in groundwater systems are generally associated with the effects of seawater (seawater intrusion by over-exploitation of the system and by sea-salt-spray

¹Received 3 June 2005; accepted 25 January 2006; Published online: 15 December 2006.

²Departamento Química, Instituto Tecnológico e Nuclear, Estrada Nacional no. 10, Apartado 21, 2686-953 Sacavém, Portugal; e-mail: paulagf@itn.pt

³Centro de Geologia, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, Portugal.

phenomena), and to anthropogenic activities such as domestic, agricultural and industrial wastes. Also, intrinsic properties of aquifers (porous/fractured/karstic media, geological structure, permeability), and external factors, such as climate, which may mitigate or worsen these problems.

The region studied corresponds to the Sines coastal sedimentary basin where two hydrogeological systems were recognized: the Mio-Pliocene and the Jurassic. Both systems supply water to the entire region, which contains highly populated and industrialized areas, although in the remaining region agriculture represents an important activity. Some local losses in the water quality were identified which could be associated to human activities or due to the aquifer matrix lithology.

Water resource assessment and management requires a multidisciplinary approach involving chemical, physical and geological data. The use of naturally occurring parameters is an integral part of studies being undertaken for water resources investigations. Principal component analysis (PCA) is one of the most useful tools to study the influence of the anthropogenic activities in water degradation. The statistical evaluation of the chemical and physical data in time or space should contribute to the qualitative understanding of the processes involved in the occurrence and circulation of water, while a fully quantitative evaluation would require proper conceptual mathematical models to be used for establishing the link between the physical–chemical properties with those of the system parameters.

Principal Component Analysis is an established technique in which the basic aim is to reduce the number of variables in the data set to a smaller number without loss of essential information. PCA transforms the data matrix $[X]$ into a new set of composite variables or principal components based on the variance-covariance matrix of $[X]$. In general, the analysis is based on standardized data, i.e. the correlation matrix. This approach was used in the present study. The principal components are by definition uncorrelated with each other. The first principal component, PC1, explains as much as possible of the total variance of the observations; the second, PC2, explains as much as possible of the residual variance, and so on (Anderson, 1984; Davis, 1984; Melloul and Collin, 1992; Melloul, 1995).

GEOLOGICAL AND HYDROGEOLOGICAL DATA

The Sines sedimentary basin is situated in the Portuguese Atlantic coast and corresponds to a tectonic trough with a NE-SW orientation filled with Mesozoic and Cenozoic deposits (Inverno, 1993) (Fig. 1). The basin deposits have an angular contact with Paleozoic basement rocks in the E and S, with the Sines Subvolcanic Massif at the SW and with Quaternary and Tertiary deposits through the Deixa-o-Resto fault in the west.

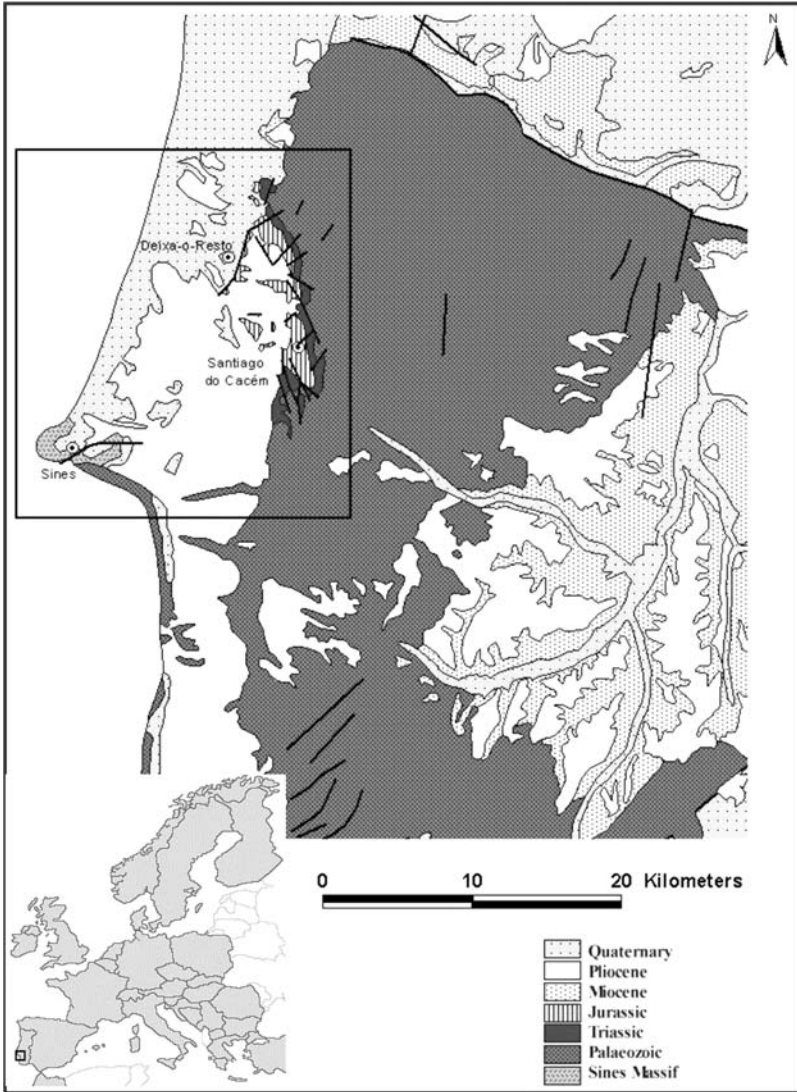


Figure 1. Location of the Sines basin.

The local Mesozoic sequence begins with Triassic deposits consisting of sandstone, evaporite and carbonate series (Grés de Silves Formation) overlain by tuff of the Volcanic-Sedimentary complex.

Carbonate layers with clastic rocks at the top compose the Jurassic sequence. The Liassic is represented by dolomite, dolomitic marl and oolitic limestone with

an average thickness of 100 m. Oolitic limestone, microcrystalline limestone and rare dolomite and marl compose the Dogger, presenting a maximum thickness of 400 m. The Malm with about 600 m is constituted by a sequence of conglomerates, limestone, clay and marl (Manuppela and Moreira, 1989).

Miocene deposits (clayey sandstone, marl and marly limestone), Plio-Pleistocene detrital (sand and clay) and recent alluvial and dune deposits partially cover the Mesozoic formations.

Two hydrogeological systems were identified in the Sines basin: the Mio-Pliocene and the Jurassic (Fig. 2). These hydrogeological systems supply the entire region, in which some areas are highly populated and industrialized. In the Mio-Pliocene system the recharge of the aquifer comes directly from the outcrop areas, while in the Jurassic areas the recharge occurs in the outcrops of carbonate formations and receives some contribution from the overlying Miocene. The main discharge of the systems should be in the continental platform, although some small natural springs occur inland. These two aquifers seem to have a hydrological connection by recharge of the Jurassic aquifer through the Miocene layers.

In the Sines Basin the precipitation varies from 600 mm to 765 mm at Santiago do Cacém and the potential evapotranspiration is about 750 mm/year (Lavaredas and Silva, 1999a). Analyzing the precipitation records shows the influence of the topography in the amount of precipitation over Sines Basin. The values of mean annual precipitation increase to the interior where the altitude is higher (Grândola and Cercal Ridges of mountains).

The main groundwater flow direction of the Jurassic aquifer is E to W towards the Atlantic Ocean. In the central part of the area, near Santo André Lagoon, a concentration of wells supplying water to the city of Santiago do Cacém produces

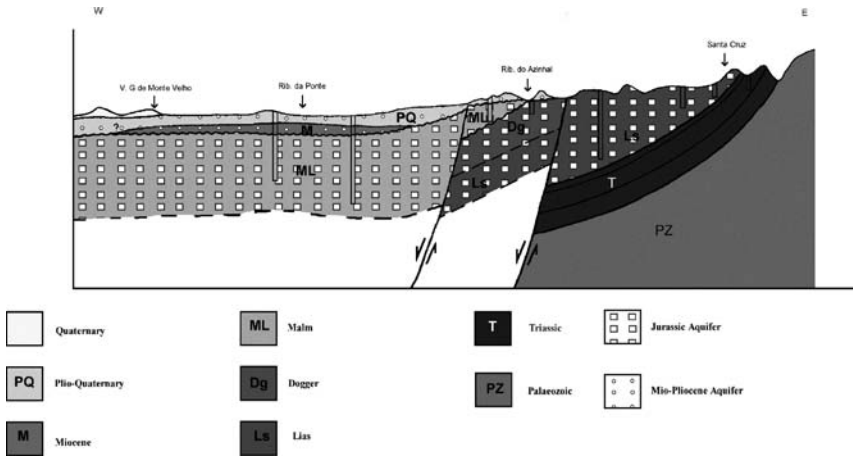


Figure 2. Geologic cross-section (adapted from Inverno, 1993).

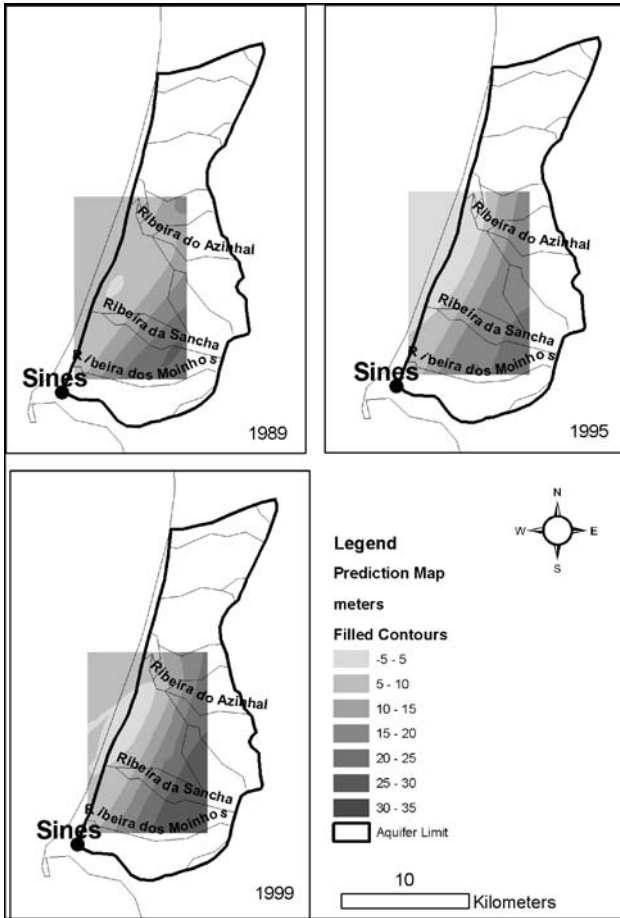


Figure 3. Piezometric map of the Sines basin.

a piezometric depression (Fig. 3). Although the system is not in stress and does not present “important” variations in time from one season to another, this could be due to a recharge contribution from the Tertiary formations to the Jurassic system (Galego Fernandes, 2004).

HYDROCHEMICAL FEATURES

Several hydrogeological and hydrochemical studies of the Sines basin have been performed, in order to evaluate its groundwater resources and to study the

Table 1. Comparison between 1999 and 2001 Campaigns in the Sines basin

	1999				2001			
	Average	Standard deviation	Max	Min	Average	Standard deviation	Max	Min
EC ($\mu\text{S/cm}$)	729.2	258.3	1420	175	708.3	224.9	1179	235
pH	6.9	0.4	7.6	5.7	7.7	0.2	7.9	7.6
Na (mg/l)	37.9	16.1	82.3	0.04	n.a.	n.a.	n.a.	n.a.
Mg (mg/l)	30.8	13.9	57.3	5	n.a.	n.a.	n.a.	n.a.
K (mg/l)	2.8	1.4	7.9	0.8	n.a.	n.a.	n.a.	n.a.
Ca (mg/l)	65.9	28.2	120	5.75	67.0	27.3	101.6	4.8
HCO ₃ (mg/l)	298.6	128.0	566	41	277.6	132.5	518	32
SO ₄ (mg/l)	47.8	44.4	200.0	3.0	47.2	40.6	166.0	11.2
Cl (mg/l)	84.5	44.1	258.0	26.0	73.4	46.0	243.0	26.3
NO ₃ (mg/l)	19.8	15.0	93.4	4.7	26.3	32.8	147.0	0.2
Fe (mg/l)	0.067	0.077	0.400	0.002	n.a.	n.a.	n.a.	n.a.
CO ₂ (mg/l)	22.6	11.6	45.8	4.4	n.a.	n.a.	n.a.	n.a.
Cu (mg/l)	0.008	0.024	0.144	0.001	n.a.	n.a.	n.a.	n.a.
Al (mg/l)	0.083	0.098	0.597	0.007	n.a.	n.a.	n.a.	n.a.
PO ₄ (mg/l)	0.033	0.031	0.130	0.002	n.a.	n.a.	n.a.	n.a.
Mn (mg/l)	0.010	0.020	0.119	0.002	n.a.	n.a.	n.a.	n.a.

Note. Number of cases: $n = 38$ and $n = 20$ for 1999 and 2001 campaign, respectively; n.a.: not analyzed.

chemical characterization and evolution along the flow paths (Lavaredas and Silva 1999a, 1999b; Galego Fernandes, Carreira, and Silva, 1999; Galego Fernandes, Carreira, and Silva, 2001a, 2001b, 2004).

We used physical and chemical data from the Sines basin groundwater system from the most recent and more complete campaigns performed in the region between 1999 and 2001 (cf. Fig. 7 for the well locations) for this study. The average, standard deviation, maximum and minimum values of the data from this period are given in Table 1.

The lithological composition of the aquifer's matrix is dominated by calcareous and dolomitic layers, recognized in the groundwater chemical evolution where the water-rock interaction processes are considered important in the definition of their calcium-bicarbonate type visible in a Piper diagram (Fig. 4), in which the dispersion in the cation (calcium and magnesium) content of the groundwater samples is shown. This could be associated either lithological heterogeneities or with human activities.

Box plots of the chemical concentrations show that bicarbonate, calcium, sulphate, chloride and nitrate have the largest dispersions (Fig. 5). The enrichment in chloride and sodium, 70–200 mg/l and 25–70 mg/l respectively, is present in the waters located near the Moinhos River (in the S of the area).

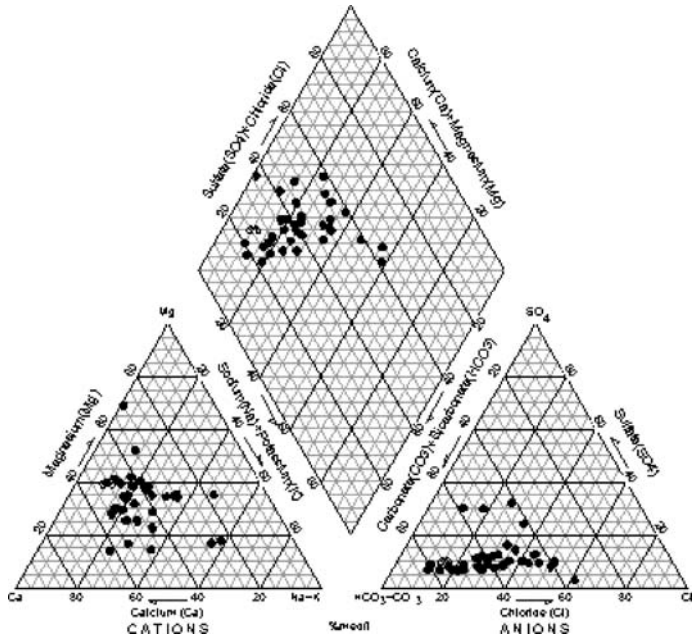


Figure 4. Piper diagram showing the ionic composition of groundwater in the Sines basin for the 1999 campaign.

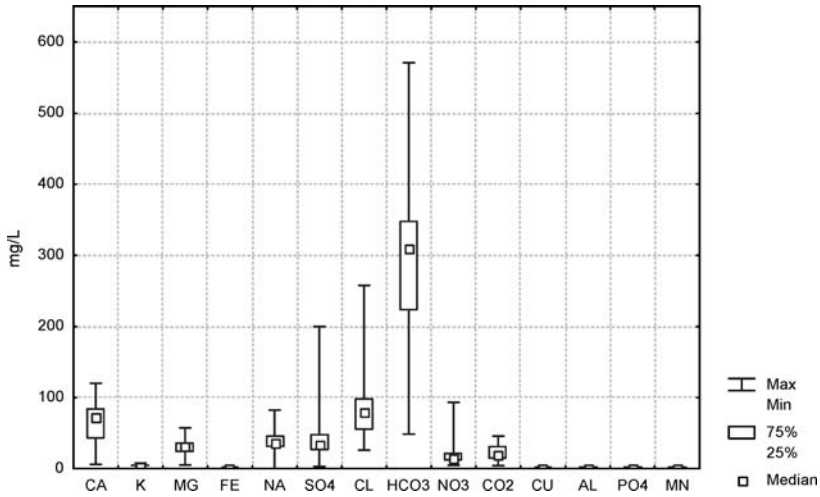


Figure 5. Box plot of chemical analyses in the Sines basin for the 1999 campaign.

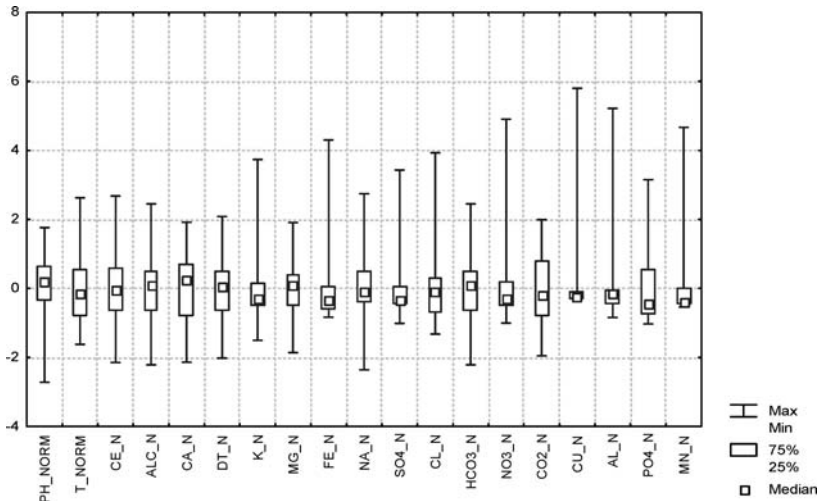


Figure 6. Box plot of standardized parameters in the Sines basin for the 1999 campaign.

The increase in the salt concentration could be associated with water–rock interactions, dissolution of Hetangian evaporites, mixing with actual seawater through the river substratum and/or with ancient seawater trapped in the sediments.

The chloride, nitrate and sulphate content also shows a significant difference between the median and maximum values, mean values are near a quarter of the maximum values, suggesting local contamination inputs to the groundwater system. The wide range of values in the bicarbonate content, from 50–600 mg/l, is the result of the lateral geological variations of the layers conferring to the groundwater different concentrations of bicarbonate. On standardizing the results (Fig. 6), all of the physical and chemical parameters have the same weight in the groundwater evolution and characterization. For some parameters, such as PO_4 , Al, Cu, Mn, SO_4 , NO_3 and Fe, the median is close to the minimum which suggests the presence of an anthropogenic source to the aquifer system, probably indicating pollution input to the groundwater by agricultural or cattle breeding activities.

HYDROGEOCHEMICAL EVOLUTION

PCA was performed on the physical-chemical data in both aquifers using STATISTIC 6 code (Statsoft Inc, 2001). For the Miocene unit, 10 cases and 10 variables were analyzed. The very small number of observations represents the

Table 2. PCA for the Miocene aquifer in the Sines basin

	PC1	PC2	PC3
Ca	0.904	-0.121	-0.093
K	0.069	0.214	0.759
Mg	0.539	0.345	0.656
Na	0.356	0.876	0.254
SO ₄	0.783	0.435	0.160
Cl	0.373	0.685	-0.015
HCO ₃	0.925	0.025	0.229
NO ₃	-0.229	0.718	0.191
Cu	0.002	-0.145	0.890
Mn	-0.136	0.768	-0.535
Total variance (%)	39.46	22.70	17.07

Note. Bold values represent loadings significant at 95% confidence level.

main methodological problem in this study. It is known that PCA requires three or more times the number of cases relative to the number of variables, but this ratio cannot be kept in practice in every case. Errors can be produced in the PCA if we ignore this condition. However, we have decided to start the statistical analysis nevertheless and to verify the results by geological considerations. This concern should be discussed by experts in the field of the application of statistical methods in geology.

Three components with eigenvalues over 1 were obtained (Table 2). The eigenvalues can be used determine an optimum number of principal components associated with the variation in the data. Frequently, all principal components will be selected for which the eigenvalues exceed the unit, but there are also other criteria to determine a “proper” number of PCs.

Despite the extremely small samples/variables ratio we have also used our data from the Miocene aquifer and to compare it with the Jurassic aquifer, but one must be cautious in the generality of the interpretation. It is important to repeat that the use of PCA is just a tool to confirm yet existing geoscientific results and knowledge for these two aquifers.

Component 1 accounts for 40% of the total variance and appears to characterize the water–rock interaction processes. The aquifer matrix dissolution induces high loadings in Ca, SO₄ and HCO₃. PC 2 represents 23% of the total variance and shows high loadings for Na, NO₃ and Mn. Knowing that sample point F12 is located near a pig-farm, increases the odds that the effluents are strongly related to the water solutes (Figs. 7 and 8). PC 3 is dominated by high loadings in K and Cu. However, analyzing the samples in which this factor has a higher score, it could be identified that the increment on the content of waters is only 2 mg/l and 0.002 mg/l in K and Cu, respectively. These small increments lead to an apparent

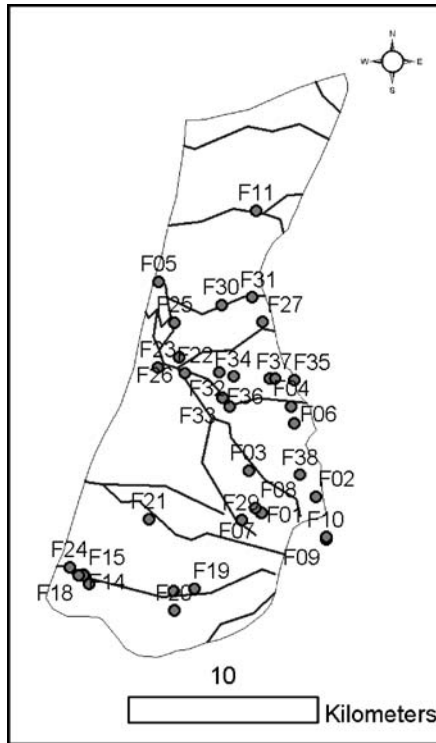


Figure 7. Well location in the Sines basin.

correlation. It seems to be more accurate if we ignore PC 3 and if we consider only the existence of two representative components to characterize this aquifer.

In the Jurassic unit 26 samples and 13 variables were analyzed, and four components with eigenvalues over 1 were obtained (Table 3). In this aquifer we have a sample/variable ratio which is greater than for the Miocene aquifer, although we assume that the number of cases is “in principle” also too small. So, we try to discuss the results with that caution.

Principal component 1 (41% of total variance) shows an increase of bicarbonate and electric conductivity and a decrease of aluminum and iron. This represents the groundwater evolution and the importance of water–rock interaction processes together with loss of the Mesozoic basin influence (pelite). It emphasizes that groundwater, with a higher residence time, have higher values in bicarbonate, calcium and magnesium, related to the lithology of the aquifer. PC 2 (18%) has high loadings for Na, Cl and K. PC 2 most likely represents the dissolution of evaporitic minerals and its contribution to water mineralization (Fig. 9).

Table 3. PCA for the Jurassic aquifer in the Sines basin

	PC1	PC2	PC3	PC4
EC	0.646	0.554	0.391	0.196
Ca	0.642	0.472	0.389	0.004
K	-0.011	0.818	0.136	-0.275
Mg	0.698	0.165	0.523	0.158
Na	0.130	0.915	0.017	0.146
SO ₄	0.173	0.260	0.789	0.077
Cl	0.160	0.888	-0.075	0.195
HCO ₃	0.862	0.094	0.255	0.118
NO ₃	0.536	-0.281	0.418	-0.303
Fe	-0.845	0.027	0.253	-0.066
Cu	-0.027	-0.112	0.816	-0.083
Al	-0.862	-0.189	0.012	0.099
Mn	-0.058	-0.047	-0.005	-0.941
Total variance (%)	40.67	18.00	12.37	8.57

Note. Bold values represent loadings significant at 95% confidence level.

PC 3 shows high loadings in Cu and SO₄. This fact might be the result of agricultural contamination by the use of pesticides. Samples with higher PC scores (F35, F27 and F02) were all located in cultivated fields where the use of fertilizers is more intensive (Fig. 10).

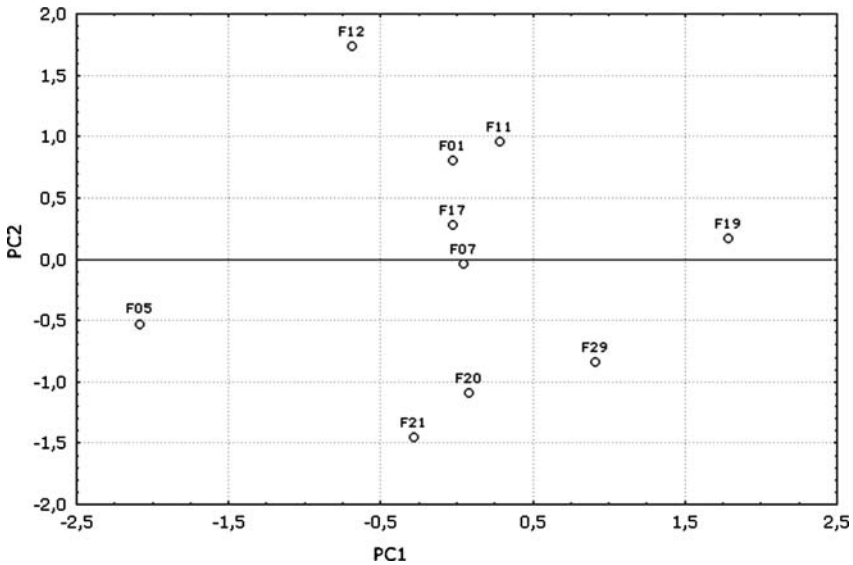


Figure 8. PC 2 versus PC 1 for Miocene aquifer samples.

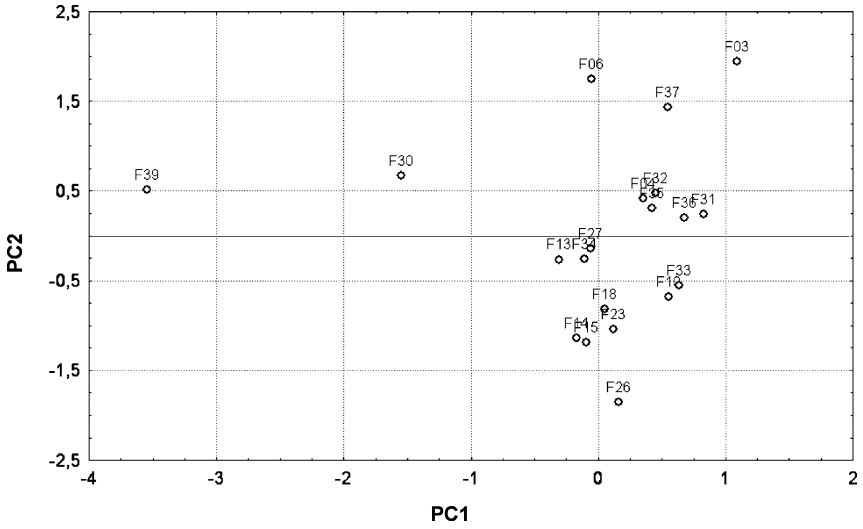


Figure 9. PC 2 versus PC 1 for Jurassic aquifer samples.

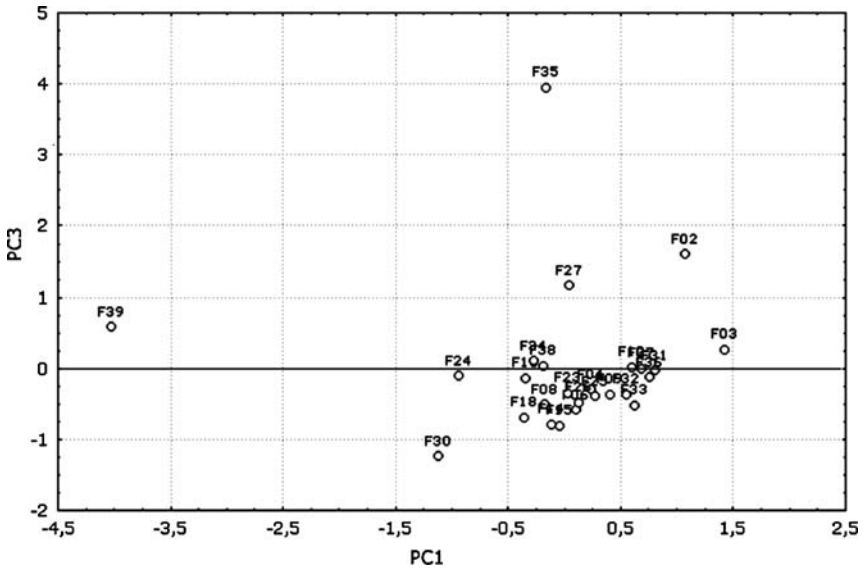


Figure 10. PC 3 versus PC 1 for Jurassic aquifer samples.

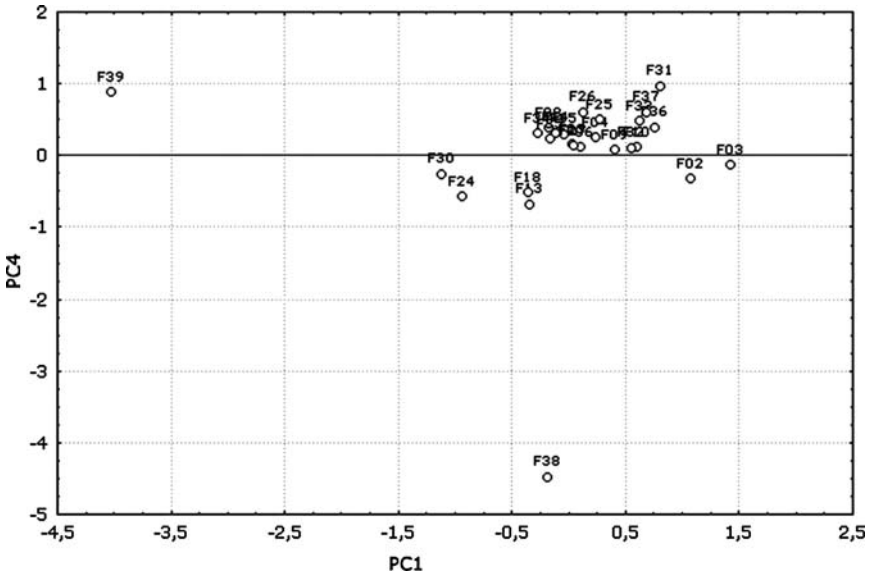


Figure 11. PC 4 versus PC 1 for Jurassic aquifer samples.

PC 4 represents the Jurassic groundwater evolution and the decrease of a Paleozoic influence, as suggested by a low loading of Mn. This results from a decreased influence of iron-magnesium minerals as subordinated constituents of Palaeozoic rocks. It could be seen in Fig. 10 that sample F38 leads to an apparent correlation caused by the 10 times higher mineralization of this sample. Considering this strong outlier, we performed the analysis of PCA omitting F38 and keeping the remaining samples. This new analysis confirmed the correctness of this procedure. Maybe that an elimination of the samples F 35 and F 39 could improve the result again.

Comparing the data of the 1999 and 2001 campaigns (Table 1), the concentrations of relevant chemical species, namely the nitrate concentration, are essentially identical for the water samples of both campaigns, so no significant variation was observed. The distribution of variables like sodium and chloride within Sines allow the evaluation of the water hydrochemical evolution (Fig. 12).

The enrichment in chloride and sodium from 1999 to 2001 is mostly observed near the Moinhos River and Cascalheira River (NE of the area), with an increase of values from 70 to 200 mg/l in chloride and 25 to nearly 70 mg/l in sodium. The increase in the salt content could be accounted for by three different mechanisms: (i) dissolution of evaporitic materials from the Hetangian formation, (ii) seawater mixing with infiltration by the river substratum and (iii) ancient seawater trapped in the sediments during the sedimentary basin formation. Based on chemical and

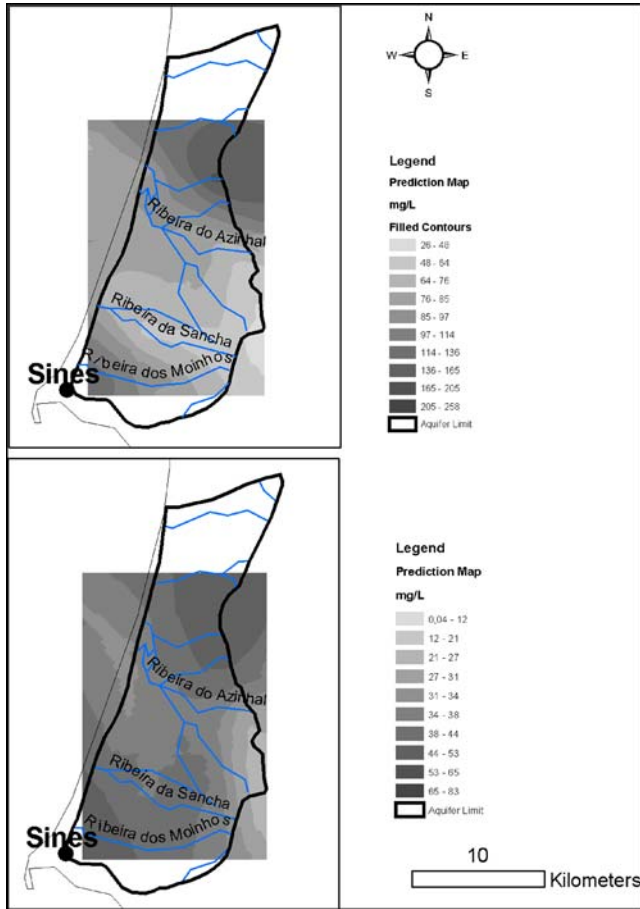


Figure 12. Spatial distribution of sodium and chloride in the Sines basin.

isotopic data the probable justification should be the dissolution of evaporitic materials from the Hetagian formation, as a result of stable isotope composition that do not indicate the mixture with recent or ancient seawater.

CONCLUSION

The hydrogeological system in the Sines basin supplies a highly populated region having both industrial and agricultural activities. The amount of precipitation shows a consistent recharge, leading to an aquifer system that is not in

hydraulic stress and does not present major temporal variations both in quantity and quality.

The results obtained by evaluation of the raw data can be supplemented by using PCA which leads to the interpretation that the water–rock interaction process is the major mechanism responsible for the groundwater solutes, mainly of a calcium-bicarbonate type, invoking the lithological composition of an aquifer matrix dominated by calcareous and dolomitic units. An anthropogenic contamination was identified in both aquifers, corresponding to local pollution inputs (PC 2—Miocene aquifer and PC 3—Jurassic aquifer). This contamination was confirmed *in situ* by the presence of cattle-breeding activities and/or intense agricultural exploitations.

Nowadays for the Sines aquifer, the protection limits for the groundwater quality are respected. However, through the use of qualitative and quantitative evaluations, some nitrate contamination problems are still identified as a result of some agricultural and cattle breeding activities, which lead us to suggests a rigorous control of possible contamination sources, such as effluent management and others.

The main methodological problem is the very small number of observations in this case study which was beyond our control. Consequently, the results of the principal component analysis must be consistent with other geoscientific observations.

ACKNOWLEDGMENTS

This study was carried out within the framework of the research project POCTI/35258/CTA/2000. P. Galego Fernandes acknowledges to FCT (Portugal) the Ph grant SFRH/BD/932/2000.

REFERENCES

- Anderson, T. W., 1984, *An introduction to multivariate statistical analysis*: John Wiley & Sons, New York, 675 p.
- Davis, J. C., 1984, *Statistics and Data Analysis in Geology*: John Wiley & Sons, New York, 646 p.
- Galego Fernandes, P., 2004, *Estudo Hidrogeológico dos sistemas aquíferos Sado-Sines - Evolução Hidrogeoquímica e Isotópica: Dissertação para obtenção do grau de Doutor em Geologia, especialidade Hidrogeologia, FCUL, Departamento de Geologia*, 213 p.
- Galego Fernandes, P., Carreira, P., and Silva, M. O., 2001a, *A Hidrogeoquímica na definição do modelo hidrodinâmico do sistema aquífero da Bacia do Sado (Hydrogeochemistry in the hydrodynamic model definition in the Sado basin aquifer)*: VI Congresso de Geoquímica dos países de Língua Portuguesa—XII Semana de Geoquímica, Faro, p. 435–439.
- Galego Fernandes, P., Carreira, P., and Silva, M. O., 2001b, *Use of hydrochemistry and isotopes as tracers in the Sado-Sines sedimentary basin (S Portugal)*: *in* *New Approaches to Characterising Groundwater Flow*, XXXI IAH Congress, Munich, v. 1, p. 65–70.

- Galego Fernandes, P., Carreira, P., and Silva, M. O., 2004, Identification of anthropogenic inputs through PCA into Sines Coastal Aquifer (SW Portugal): 32nd International Geological Congress, Florence, Abstract volume, publ. Cd-ROM
- Galego Fernandes, P., and Silva, M., 1999, Contribution to the Hydrogeological Characterization of the aquifer system of Sado's Basin": International Association of Hydrogeologists, Bratislava, publ. CD-ROM.
- Inverno, C., 1993, Estratigrafia (Substrato hercínico) (Stratigraphy Hercinic substratum): *in* Notícia explicativa da folha 42C Santiago do Cacém, Carta geológica 1/50 000, serviços geológicos de Portugal.
- Lavaredas, J., and Silva, M., 1999a, Hydrogeochemistry of Sines aquifer system: European Union of Geosciences 10, Strasbourg, v. 4, no. 1, p. 749.
- Lavaredas, J., and Silva, M., 1999b, Sines's aquifer system—Some contribution for its knowledge: International Association of Hydrogeologists, Bratislava, publ. CD-ROM.
- Manuppella, G., and Moreira, J., 1989, Calcários e dolomitos das área de Melides—Santiago do Cacém (Limestones and dolomites in Melides area—Santiago do Cacém): Estudos, Notas e Trabalhos, D.G.G.M., Lisboa, v. 31, p. 79–94.
- Melloul, A., and Collin, M., 1992, The principal components statistical method as a complementary approach to geochemical methods in quality factor identification - application to the Coastal Plain aquifer of Israel: *J. Hydrol.*, v. 140, p. 49–73.
- Melloul, A. J., 1995, Use of principal components analysis for studying deep aquifers with scarce data - Application to the Nubian sandstone aquifer, Egypt and Israel: *Hydrogeol. J.*, v. 3, no. 2, p. 19–39.
- Mackiewicz, A., and Ratajczak, W., 1993, Principal components analysis (PCA): *Comput. Geosci.*, v. 19, no. 3, p. 303–342.
- Statsoft Inc., 2001, *Statistic 6.0*: Tulsa.