

# Assessment of Groundwater Quality by Means of Self-Organizing Maps: Application in a Semiarid Area

## FRANCISCO SÁNCHEZ-MARTOS

Department of Hydrogeology  
Almería University  
04120 Almería, Spain

## PEDRO A. AGUILERA\*

Department of Ecology  
Almería University  
04120 Almería, Spain

## ANTONIA GARRIDO-FRENICH

Department of Analytical Chemistry  
Almería University  
04120 Almería, Spain

## JOSÉ A. TORRES

Department of Computer Science  
Almería University  
04120 Almería, Spain

## ANTONIO PULIDO-BOSCH

Department of Hydrogeology  
Almería University  
04120 Almería, Spain

**ABSTRACT** / The Kohonen neural network was applied to hydrochemical data from the Detritic Aquifer of the Lower Andarax, situated in a semiarid zone in the southeast of Spain. An activation map was obtained for each of the sampling points, in which the spatial distribution of the activated neurons indicated different water qualities. To extract the information contained in the activation maps, they were divided into nine quadrats. Cartesian coordinates were assigned to each quadrant ( $x$ ,  $y$ ), and for each sampling point, three derived variables were selected, which were assigned the values  $x$  and  $y$  of the corresponding quadrat. A classification was defined based on this simple matrix system which allows an easy and rapid means of evaluating the water quality. This assessment highlights the various processes that affect groundwater quality. The method generates output that is easier to interpret than from traditional statistical methods. The information is extracted from the activation maps without significant loss of information. The method is proposed for assessing water quality in hydro-geochemically complex areas, where large numbers of observations are made.

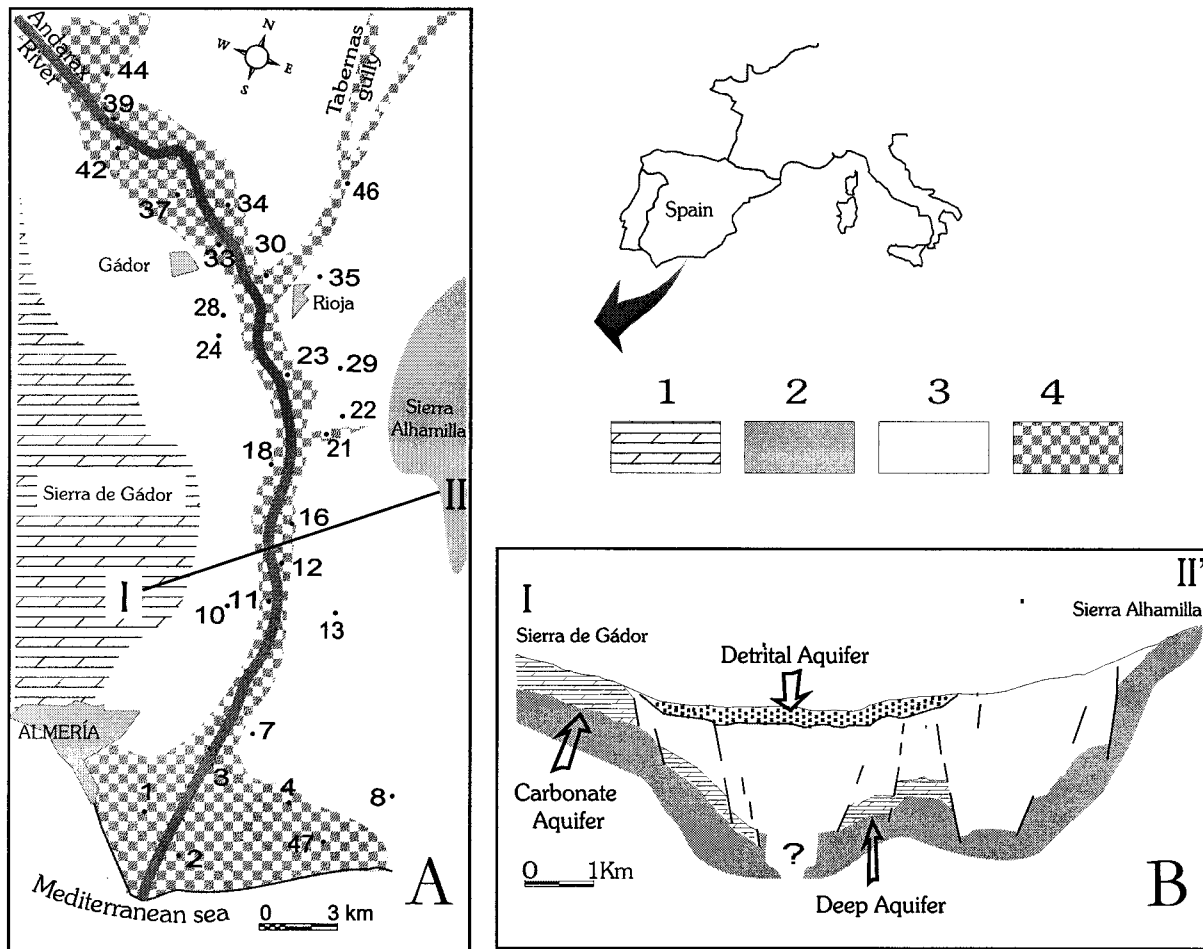
Salinization of groundwater is a problem of great interest to regional and national administrations, particularly in semiarid regions. The problem is magnified in coastal areas, where human activity encourages over-exploitation of the aquifers, marine intrusion, and agricultural pollution (Candela and others 1991, Appelo and Postma 1993, Carter 1996). In semiarid areas, other natural processes frequently take place, such as the dissolution of evaporite rocks and the action of brines, which progressively salinize the groundwater (Vengosh and Rosenthal 1994, Sánchez-Martos and others 1999). As a result of these various processes, the water's physical and chemical characteristics are transformed and the quality may fall below that desirable for a particular use (e.g., domestic water supply or crop irrigation). It is therefore necessary to assess water qual-

ity to adequately plan and manage these groundwater resources. Such an assessment must allow the processes that affect water quality to be identified.

Studies aimed at water quality assessment involve large quantities of data. The data are normally processed using multivariate statistical techniques (Davis 1986), which are suitable for groundwater quality assessment (Cerón and others 2000, Helena and others 2000, López-Chicano and others 2001) and pollution studies (Burkart and others 1999). These methods contribute to the decision-making process involved in managing the water resources (Karydis 1992). Recently, artificial neural networks (ANN) have become the focus of interest across many scientific disciplines (Lek and Guegan 2000, Foody, 1999, Li and others 1999, Schleiter and others 1999). They have been applied to groundwater, both in pollution studies (Gumrah and others 2000) and in flow simulation studies (Hamed and Hassan 2000). Although the use of ANN has spread rapidly, there are few cases in which ANN has been used as a complement to the traditional multivariate statistical methods.

**KEY WORDS:** Self-organizing maps; Multivariate analysis; Groundwater quality; Semiarid area

\*Author to whom correspondence should be addressed.



**Figure 1.** (A) Study area and location of sampling points: 1, limestones and dolomites; 2, metapelite rocks; 3, conglomerates, silty marls, and sandy marls; 4, quaternary alluvial and deltaic deposits. (B) Hydrogeological sections of the Andarax Valley.

Much attention has focused on feed-forward neural networks (Mastrorillo and others 1997) with considerably less effort directed at exploiting the potential of the Kohonen or self-organizing feature map networks. This neural network organizes the data by similarity. The output is a low, typically two-dimensional, array in which similar samples are clustered together (Foody 1999). In this paper the Kohonen neural network (KNN) has been applied to hydrochemical data collected from the Detritic Aquifer of the Lower Andarax. The information generated by KNN has been extracted using a system proposed by Aguilera and others (2001). The method results in a simple matrix that forms the basis of a classification for the rapid assessment of groundwater quality. It utilizes the full information content and generates output that is easy to interpret. The method is proposed as a complement to traditional multivariate techniques.

## Methodology

### Hydrogeological Setting

The valley of the Lower Andarax River lies in a coastal zone in the southeast of Spain. The area is semiarid (precipitation <300 mm/yr) and supports intensive agricultural activity. The valley is flanked by the Sierra Alhamilla and the Sierra Nevada, where pelitic metasediments outcrop. Its west edge corresponds to the Sierra de Gádor, which is a limestone–dolomite massif with outcrops of phyllites (Figure 1). The depression is filled by postorogenic detrital deposits of diverse lithology (marls, sandy loams, sands and conglomerates) with numerous gypsiferous evaporite intercalations. Based on the geological characteristics of the area, the following hydrogeological units have been defined: Detrital Aquifer, Carbonate Aquifer, and Deep Aquifer (Pulido-Bosch and others 1992, Sánchez-

Martos 1997). The Detrital Aquifer runs the entire length of the valley and includes Quaternary alluvial and deltaic deposits, together with deltaic sandy-silt conglomerates dating from the Pliocene. The impermeable base of the aquifer corresponds to Miocene marls and to Pliocene sandy marls in the southernmost part. The thickness of the aquifer varies from 200 m in the NW to 20–40 m next to the coast. The geology of the area is highly complex: there is intense tectonic activity (Bousquet and Phillip 1976, Sanz de Galdeano and others 1985) and a highly diverse lithology, which includes frequent changes of facies (Voermans and Baena 1983). As a result, there are sharp changes of permeability, both across the aquifer and over its depth. The Detrital Aquifer has a close hydraulic connection with the River Andarax, which is the principal source of recharge. Groundwater levels show sharp fluctuations, with clear seasonal recovery. Piezometric levels vary from 190 m asl., to below sea level in the coastal band at certain times of the year.

The physicochemical characteristics of the groundwaters are variable in space and time as a result of the functioning of the aquifer itself and the diversity of its lithology. They show wide thermal variations and contain certain indicators of thermalism associated with deep flows. The dominant facies is calcium sulfate, while some points close to the coast have a chloride facies. The saline content increases from north to south, with values of between 1000 and 8000 mg/liter. To the north of Rioja lies an intermediate section, where there is elevated salinity (close to 4000 mg/liter), and a high  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ , and  $\text{Na}^+$  content. The delta area provides the most saline waters (3500–10,000 mg/liter) with mixed chloride-sulfate facies. Study of the hydrogeochemical evolution of the Detrital Aquifer of the Lower Andarax is difficult because of the diversity of its groundwaters, which feature thermal processes, marine intrusion, and high sulfate and boron contents (Sánchez-Martos 1997, Sánchez-Martos and others 1999, Sánchez-Martos and Pulido-Bosch 1999).

#### Hydrogeochemical Data

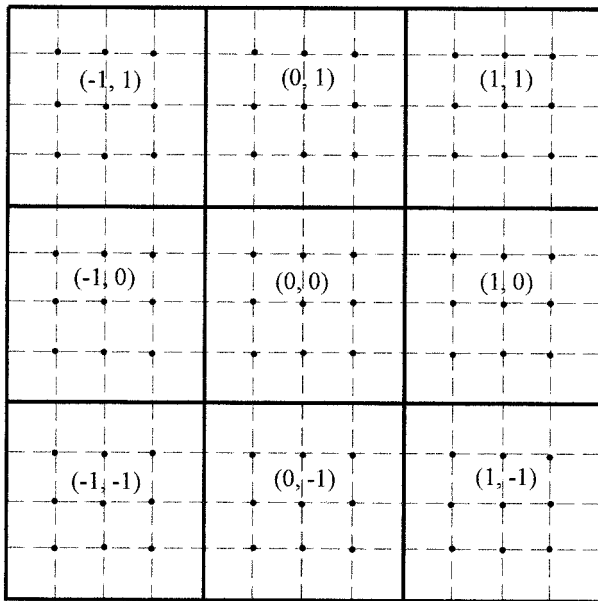
Hydrogeochemical information for the Lower Andarax basin was obtained from a sampling network in the Detrital Aquifer comprising 28 boreholes. Sampling was carried out two times a year over a three-year period. Temperature was directly determined in the borehole using a WTW field conductivity meter (model LFT 91).  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ , and  $\text{NO}_3^-$  anions were analyzed using ion chromatography with a Dionex chromatograph (model DX-100).  $\text{HCO}_3^-$  was analyzed by potentiometric determination in a Methrom Titroprocessor (model 686). The concentrations of

the cations  $\text{Na}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ , and  $\text{K}^+$  were measured by inductively coupled plasma analysis (ICP-OES) using a Leeman Labs atomic emission spectrophotometer (model PS-1000). The balance error did not exceed 5%. Total dissolved solids (TDS) were used to estimate the salinity of the water.

#### Kohonen Neural Network (KNN)

Artificial neural networks (ANN) are computer techniques that attempt to simulate the function and decision-making processes of the human brain (Eberhart and Dobbins 1990). They may be broadly classified into networks that learn in a supervised or unsupervised way (Beale and Jackson 1990, Haykin 1994, Bishop 1995). In supervised learning the network is given many different examples of a particular problem, including the desired solutions. In unsupervised learning the desired outputs are not given, and the network learns to classify the examples by recognizing different patterns.

The Kohonen neural network is a type of neural network designed for unsupervised pattern recognition tasks (Kohonen 1982, 1989, Zurada 1992). The goal of KNN is to map similar objects on the same or neighboring neurons. As such, it is a clustering method in which the objects or samples are distributed over a toroidal map (usually a square grid of  $n \times n$  cells). To enable a visual appreciation, the neurons are arranged in two dimensions. The network comprises  $n \times n \times p$  weights, where  $p$  is the number of input variables and each  $p$  dimensional vector is a neuron. During training  $n$  objects are presented to the network—one at a time—on a fixed number of occasions; each object is assigned to the cell for which the distance between each input object and the neuron is the minimum. The weights of the cell to which an object is assigned and the topologically nearest cells are modified in such a way as to reproduce the input object. Different neighborhood functions can be used, such as block function, triangular function, Gaussian-bell function or Mexican-hat shaped function (Melssen and others 1994). The principle is always that units closer to the winning unit are adapted most. When the network is trained, similar objects fall into the same or proximate cells. The maximum and minimum correction parameters for the weights were chosen as 0.5 and 0.05, respectively. Before training, all KNN weights were initialized by randomly assigning a weight in the minimum to maximum range of the input values, and the number of iterations was fixed at 1500 in all cases. The output of the process is a map of activated neurons.



**Figure 2.** Division into quadrats of the activation map. (●) indicates the location of the neurons.

#### Method for Extracting Information from Activation Maps

Aguilera and others (2001) developed a methodology to extract all the information contained within the activation maps. This methodology assumes that the spatial distribution of the activated neurons reflects water quality and that neurons corresponding to similar water quality types are placed close to one another on the activation map. Thus, the method follows the principle of the topographic map that similar outputs are placed close to one another on the activation map (Kohonen 1990). The method was developed assuming that the activation maps are defined in a two-dimensional plane. The plane is divided into nine quadrats (Figure 2), each composed of nine neurons. Using Cartesian coordinates to define the plane, each quadrat is assigned two values ( $x$ ,  $y$ ). In this way, for each sampling station, a series of derived variables can be selected, which assume the ( $x$ ,  $y$ ) coordinate of the corresponding quadrat. These values are indicative of the spatial placement of the activated neurons.

#### Results and Discussion

The mean concentrations of raw data are given in Table 1. A principal components analysis of the field data was done (168 observations  $\times$  9 variables). The first and second components were selected, account-

ing for 65% of the variance. The plane defined by the first two components enables an interpretation of the hydrogeochemical evolution (Figure 3). The  $\text{SO}_4^{2-}$ ,  $\text{Cl}^-$ ,  $\text{Na}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ , and  $\text{K}^+$  ions are related to the first component, which represents general saline enrichment, with less salinization along the negative semiaxis and more on the positive semiaxis. The second component indicates influence of sulfate (positive semiaxis) and chloride, i.e. marine, influence (negative semiaxis). Figure 3 shows the distribution of the 168 samples in the plane defined by the first two components. It is difficult to distinguish waters of different quality in the set of samples as a whole. Only a few of the samples, which are associated with high salinity, are clearly differentiated. The clustering of points at lower salinity (negative semiaxis of the first component) makes it difficult to distinguish different groups of water quality.

Results from the cluster analysis (Figure 4), demonstrate the same problem: it is difficult to identify groups of sampling points that correspond to similar water quality. This is basically due to the large number of samples. Reducing the number of samples would result in a classification that would be easier to interpret, but in this way information needed to define the various types of groundwater quality would be lost. It would also be possible to use the mean values for the variables, such that there would be a single sample for each variable. However, if the variables fall over a wide range, their means could deviate and lead to a poor interpretation of water quality.

For these reasons, it is first necessary to use all of the samples and all the data from the monitoring programs in determining the various classes of groundwater quality. Secondly, a mathematical technique must be used that conserves all the information and allows methods to be developed that result in easily interpretable graphical outputs. KNN could be used as a tool in assessing groundwater quality.

The output from applying KNN is an activation map. Figures 5 and 6 show that each sampling point is represented by an activation map of  $9 \times 9$  neurons. The position of the activated neurons on each map is indicative of different water qualities. Thus, activation of a particular neuron on the map depends on the water quality, which is determined by the combination of physico-chemical variables. For example, the activation maps for sampling points 21 and 22 indicate that they have similar water quality, since the arrangement of activated neurons is similar. These two points show a slight thermal anomaly and their waters are of good quality (Table 1). In contrast, the activation maps for sites 37 and 46 are quite different

Table 1. Temperature  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ ,  $\text{HCO}_3^-$ ,  $\text{NO}_3^-$ ,  $\text{Na}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Ca}^{2+}$ , and total dissolved solids (TDS)<sup>a</sup>

S	Temp. (°C)	$\text{Cl}^-$	$\text{SO}_4^{2-}$	$\text{HCO}_3^-$	$\text{NO}_3^-$	$\text{Na}^+$	$\text{Mg}^{2+}$	$\text{Ca}^{2+}$	TDS
21	26.0 ± 1.3	316.8 ± 51.7	587.3 ± 60.3	281.1 ± 15.0	0.0 ± 0.0	243.9 ± 34.3	89.8 ± 12.8	104.5 ± 20.6	1492.7 ± 47.9
22	23.4 ± 1.9	356.0 ± 59.3	625.2 ± 43.1	289.3 ± 34.7	4.8 ± 7.8	239.3 ± 14.7	109.4 ± 14.8	115.3 ± 19.3	1609.2 ± 82.4
28	24.1 ± 0.1	217.7 ± 15.4	448.7 ± 37.7	262.5 ± 41.4	6.6 ± 3.0	119.8 ± 9.0	86.9 ± 5.1	122.5 ± 33.2	1143.0 ± 80.8
33	25.2 ± 0.5	133.3 ± 28.2	440.0 ± 18.8	267.7 ± 19.6	11.4 ± 5.1	147.2 ± 41.2	52.4 ± 16.9	121.8 ± 33.5	1044.8 ± 47.1
34	19.7 ± 2.1	119.6 ± 27.4	502.8 ± 125.6	270.8 ± 17.5	23.3 ± 3.6	118.6 ± 46.6	69.4 ± 8.3	151.2 ± 62.2	1046.3 ± 58.2
37	21.3 ± 1.24	87.0 ± 9.2	372.7 ± 33.8	263.6 ± 8.8	10.9 ± 2.4	82.6 ± 31.0	56.6 ± 7.9	105.9 ± 46.8	850.4 ± 49.1
44	21.8 ± 1.0	133.7 ± 6.7	489.5 ± 38.4	189.0 ± 25.0	13.6 ± 5.6	111.6 ± 27.9	54.3 ± 10.9	148.1 ± 34.5	1050.0 ± 39.6
11	19.9 ± 0.6	470.0 ± 62.1	1299.8 ± 195.6	318.3 ± 53.9	81.0 ± 14.4	349.2 ± 79.1	184.2 ± 25.0	282.1 ± 100.8	2835.4 ± 338.1
12	19.4 ± 0.6	478.8 ± 88.0	1246.3 ± 142.6	351.3 ± 25.5	85.8 ± 16.0	348.3 ± 17.2	158.6 ± 16.5	294.3 ± 24.8	2897.8 ± 117.2
18	20.4 ± 0.5	401.5 ± 72.2	1255.8 ± 235.8	369.3 ± 77.4	72.8 ± 24.9	279.6 ± 50.7	182.3 ± 42.4	264.4 ± 121.9	2741.5 ± 469.6
30	21.0 ± 1.3	323.7 ± 46.9	987.2 ± 204.6	332.9 ± 38.5	63.2 ± 11.9	257.5 ± 34.1	141.6 ± 37.7	207.2 ± 65.0	2788.7 ± 933.2
3	20.4 ± 1.3	346.8 ± 53.7	938.2 ± 149.6	335.2 ± 22.4	35.1 ± 9.8	228.4 ± 23.0	169.8 ± 70.7	196.6 ± 91.2	2089.4 ± 269.4
39	18.8 ± 1.5	114.8 ± 7.8	664.7 ± 77.9	229.5 ± 59.5	25.4 ± 4.6	94.9 ± 17.7	87.6 ± 10.3	178.9 ± 57.7	1285.3 ± 138.9
42	19.3 ± 0.7	166.3 ± 30.9	826.8 ± 26.6	287.5 ± 42.8	29.6 ± 5.1	132.1 ± 34.4	102.2 ± 19.2	196.5 ± 59.9	1602.7 ± 77.5
7	22.3 ± 0.5	609.2 ± 62.6	1096.8 ± 140.1	278.5 ± 20.1	30.4 ± 7.2	336.7 ± 37.4	161.2 ± 13.5	311.7 ± 50.6	2695.7 ± 173.2
23	20.4 ± 0.8	713.2 ± 89.1	1401.5 ± 88.6	288.9 ± 55.7	52.2 ± 14.0	505.3 ± 91.2	164.9 ± 29.0	282.3 ± 107.9	3195.2 ± 242.3
24	21.4 ± 1.2	537.2 ± 50.0	870.5 ± 74.1	327.5 ± 21.9	40.0 ± 12.4	308.7 ± 48.5	155.9 ± 14.5	196.4 ± 42.2	2268.6 ± 91.9
10	21.5 ± 0.6	603.0 ± 55.2	1780.7 ± 217.3	380.0 ± 12.4	55.5 ± 6.3	466.6 ± 23.8	248.5 ± 25.8	342.1 ± 51.0	3706.0 ± 304.1
16	20.7 ± 0.5	658.7 ± 189.6	1810.2 ± 228.3	396.9 ± 216.6	163.5 ± 83.2	506.4 ± 101.4	238.6 ± 30.8	394.3 ± 54.9	3996.0 ± 373.7
35	22.5 ± 1.1	523.3 ± 45.9	1921.8 ± 135.0	197.5 ± 49.4	31.9 ± 18.8	285.7 ± 132.2	213.5 ± 57.9	446.8 ± 123.4	3528.0 ± 179.5
46	20.4 ± 0.9	420.8 ± 58.0	2224.3 ± 219.5	247.3 ± 17.0	1.1 ± 1.7	407.2 ± 56.4	225.2 ± 27.4	431.9 ± 88.2	3850.6 ± 251.2
1	21.7 ± 1.0	1557.3 ± 215.0	2466.2 ± 257.8	548.6 ± 88.5	222.6 ± 47.6	952.2 ± 148.4	432.9 ± 65.9	471.9 ± 51.6	6540.7 ± 611.5
47	22.1 ± 0.4	1047.8 ± 82.0	1160.2 ± 138.4	344.5 ± 16.8	93.2 ± 21.2	559.5 ± 45.5	224.9 ± 23.6	275.1 ± 33.3	3547.7 ± 179.0
2	19.5 ± 0.6	1626.2 ± 286.4	2080.7 ± 76.9	427.7 ± 19.9	71.9 ± 19.4	739.0 ± 90.0	459.2 ± 52.5	511.1 ± 102.1	5711.3 ± 567.4
8	22.8 ± 1.0	4406.3 ± 164.0	1914.0 ± 59.4	242.7 ± 21.0	143.3 ± 58.7	2472.2 ± 131.6	415.3 ± 70.0	368.6 ± 54.0	9877.6 ± 239.3
4	23.7 ± 0.8	1679.8 ± 330.2	789.3 ± 87.8	253.8 ± 34.7	29.6 ± 7.2	774.0 ± 96.5	239.2 ± 29.0	217.7 ± 110.4	3871.7 ± 446.9
13	24.6 ± 1.9	577.8 ± 61.4	730.0 ± 63.5	257.2 ± 29.4	1.1 ± 2.0	324.6 ± 59.8	113.3 ± 13.8	180.4 ± 40.3	2067.2 ± 150.5
29	22.7 ± 1.1	1096.8 ± 121.7	1023.8 ± 224.2	107.3 ± 105.3	9.8 ± 12.1	673.0 ± 57.4	188.5 ± 28.2	117.1 ± 114.9	3983.9 ± 759.8

<sup>a</sup>Concentrations are expressed in mg/liter (mean ± SD). S: sampling point.

from one another, and this suggests they have very different water qualities. Sampling point 37 is directly affected by the inflow of surface water from the river Andarax, while point 46 is associated with the Tabernas gully. Water from point 46 is saline (3900 mg/liter) and has high concentrations of  $\text{SO}_4^{2-}$  (2224 mg/liter) and  $\text{Ca}^{2+}$  (431 mg/liter), whereas point 37 is less saline (850 mg/liter), although its facies is also calcium sulfate.

The information contained in the activation maps can not be used to assess the groundwater quality. To achieve this aim, it is necessary to extract the information using the system described in the Methods section. The variables selected for extracting the information from the activation maps were: (1) quadrat containing the greatest activation, (2) quadrat containing the neuron with the highest frequency of activation, and (3) quadrat containing the centroid of the activation map. The values obtained for these variables are shown in Table 2. Each variable comprises two values, and each observation is defined by six values. Cluster analysis was then applied to the data matrix to identify groups of similar sampling sites. Euclidean distance was used as a measure of similarity and Ward's method as the clustering strategy. From the dendrogram obtained (Figure 7), five water quality groups are differentiated.

Group I contains sampling points 21, 22, 28, 33, 34, 37, and 44. Salinity at these points is the lowest in the study area, oscillating between 850 mg/liter and 1609 mg/liter. The low salinity is directly related to the surface water of the river Andarax, which forms the principal recharge of the detritic aquifer (34, 37, and 44). Points 33 and 34 indicate the influence of direct recharge associated with water from the River Andarax together with the deep flow from the Carbonate Aquifer towards the Detritic Aquifer (Pulido Bosch and others 1992). These two sampling points are low in  $\text{NO}_3^-$  (<11.4 mg/liter) and the water temperature here is the highest measured in the Detritic Aquifer (>24°C). Water from points 21 and 22 has lower salinity (<1600 mg/liter). These sampling points exhibit a thermal anomaly (23.4°C and 26°C) and their  $\text{NO}_3^-$  content is very low (<4.8 mg/liter). The origin of this water is related to some deep flow associated with the edge of the Sierra Alhamilla at the confluence of a fracture zone, which lifts the substratum (Sánchez-Martos and others 1999).

Groups II and III show water with a strong sulfate component and variable salinity. Sampling points 11, 12, 18, and 30 in group II lie in the central part of the valley, downstream of the confluence of the river Andarax and the Tabernas gully (Figure 1). Their

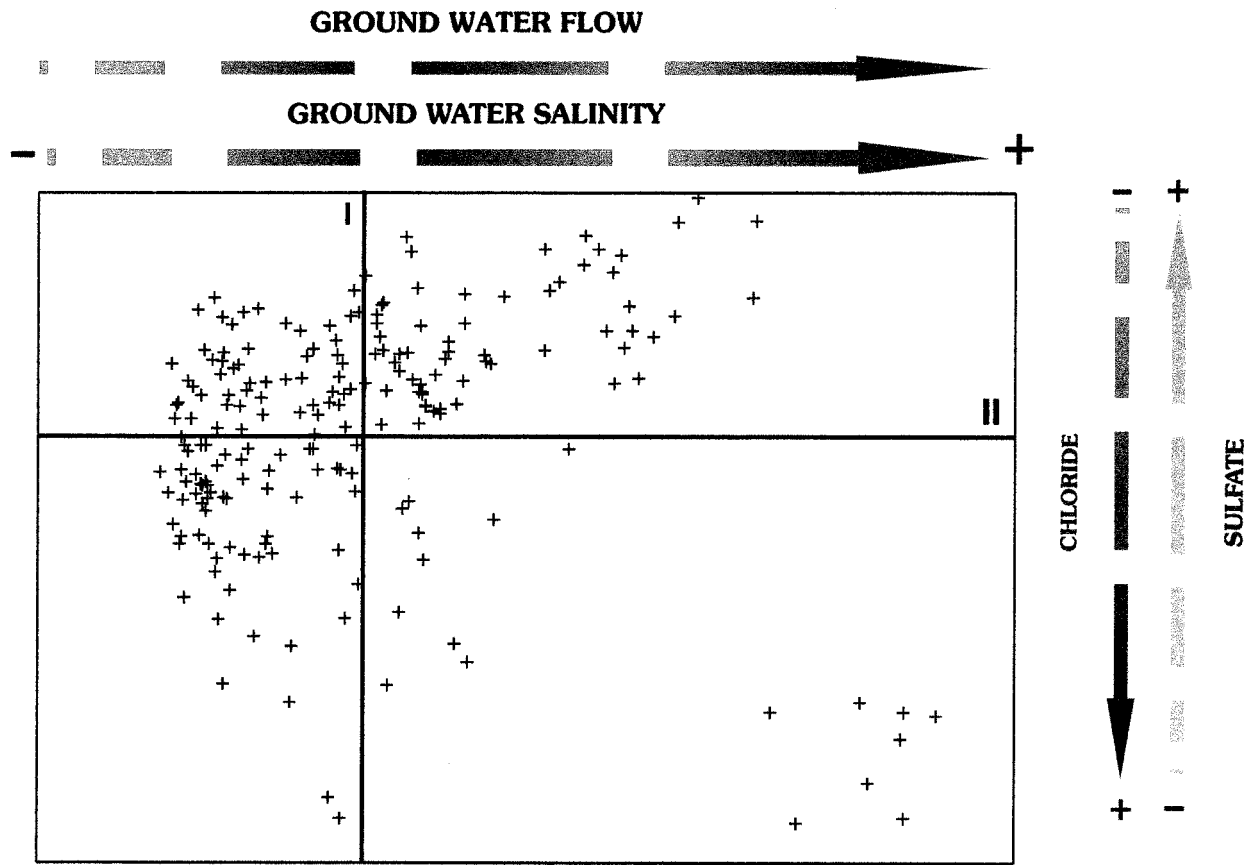


Figure 3. Principal component analysis: distribution of the samples in the plane defined by the first two components.

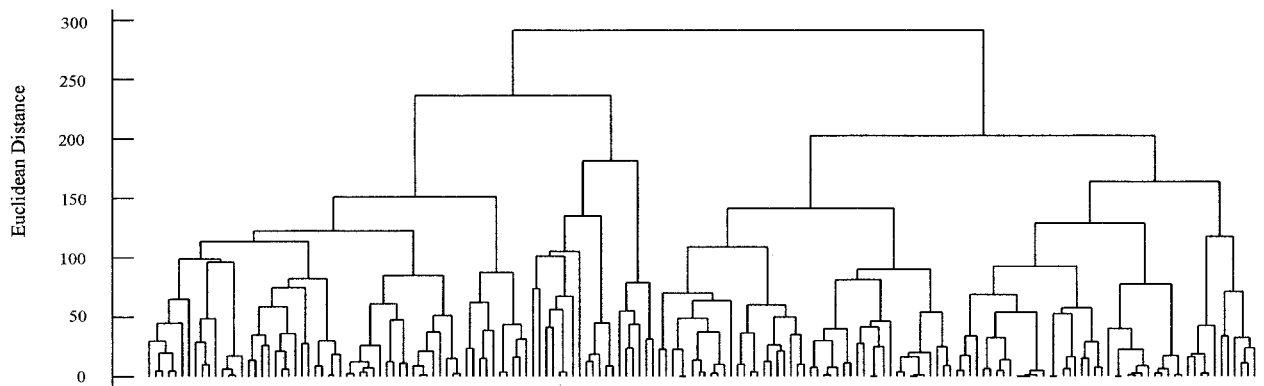
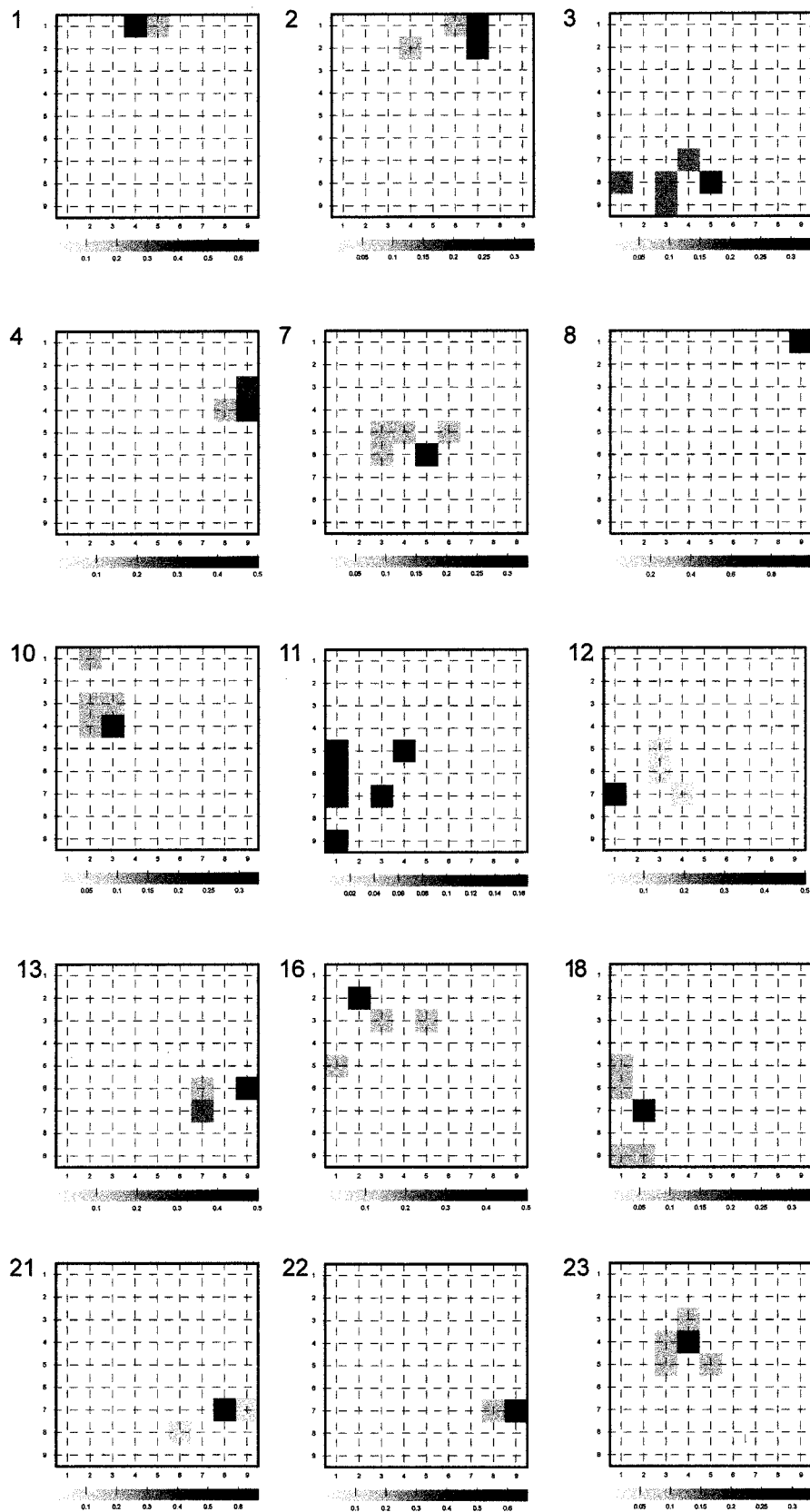


Figure 4. Dendrogram obtained from the raw data.

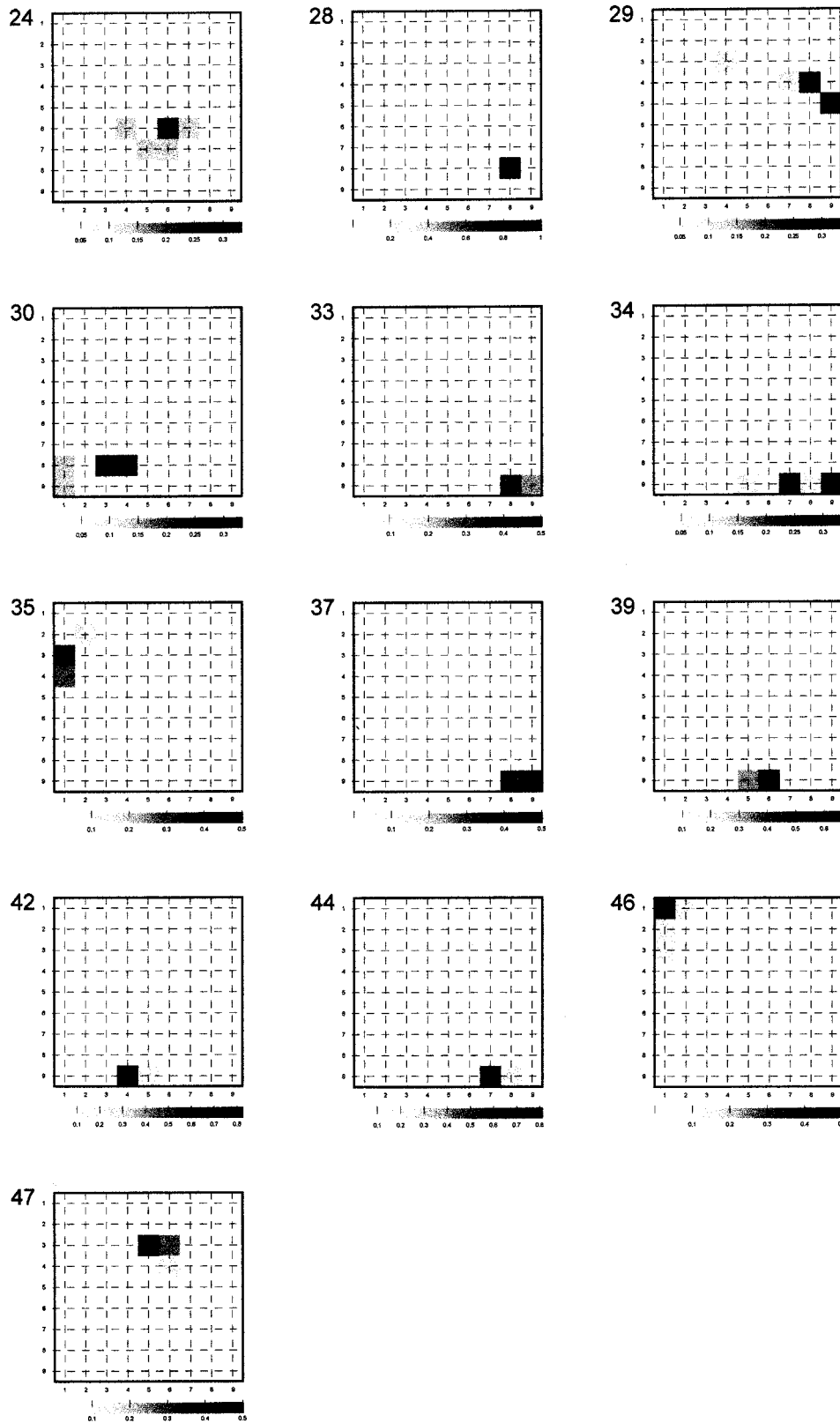
location determines their water chemistry: salinity varies little, from 2741 mg/liter to 2897 mg/liter, and the  $SO_4^{2-}$  content is triple that of  $Cl^-$ . Compared to group I, the markedly higher salinity and  $SO_4^{2-}$  can be attributed to the influence of groundwater flow through the alluvial deposits of the Tab-

ernas gully, whose waters are rich in  $SO_4^{2-}$  (Sánchez-Martos 1997).

Water in group III (sampling points 3, 39, 42, 7, 23, and 24) shows variable salinity of between 1285 mg/liter and 3195 mg/liter. The most saline sampling points (7, 23, and 24) are slightly separated from the



**Figure 5.** Activation maps for sampling points 1, 2, 3, 4, 7, 8, 10, 11, 12, 13, 16, 18, 21, 22, and 23, obtained by applying KNN. The color intensity indicates the frequency of activation that ranges from 0 to 1.



**Figure 6.** Activation maps for sampling points 24, 28, 29, 30, 33, 34, 35, 37, 39, 42, 44, 46, and 47, obtained by applying KNN. The color intensity indicates the frequency of activation that ranges from 0 to 1.

Table 2. Values obtained applying quadrat system to activation maps for each sampling point

S <sup>a</sup>	Quadrat with greatest activation		Quadrat with the neuron most-intensely activated		Quadrat containing the centroid	
21	1	-1	1	-1	1	-1
22	1	-1	1	-1	1	-1
28	1	-1	1	-1	1	-1
33	1	-1	1	-1	1	-1
34	1	-1	1	-1	1	-1
37	1	-1	1	-1	1	-1
44	1	-1	1	-1	1	-1
11	-1	0	-1	-1	-1	-1
12	-1	-1	-1	-1	-1	-1
18	-1	-1	-1	-1	-1	-1
30	-1	-1	-1	-1	-1	-1
3	0	-1	0	-1	0	-1
39	0	-1	0	-1	0	-1
42	0	-1	0	-1	0	-1
7	0	0	0	0	0	0
23	0	0	0	0	0	0
24	0	0	0	-1	0	0
10	-1	0	-1	0	-1	1
16	-1	1	-1	1	-1	1
35	-1	1	-1	1	-1	1
46	-1	1	-1	1	-1	1
1	0	1	0	1	0	1
47	0	1	0	1	0	1
2	1	1	1	1	0	1
4	1	0	1	0	1	0
8	1	1	1	1	1	1
13	1	0	1	0	1	0
29	1	0	1	0	1	0

<sup>a</sup>S: sampling point.

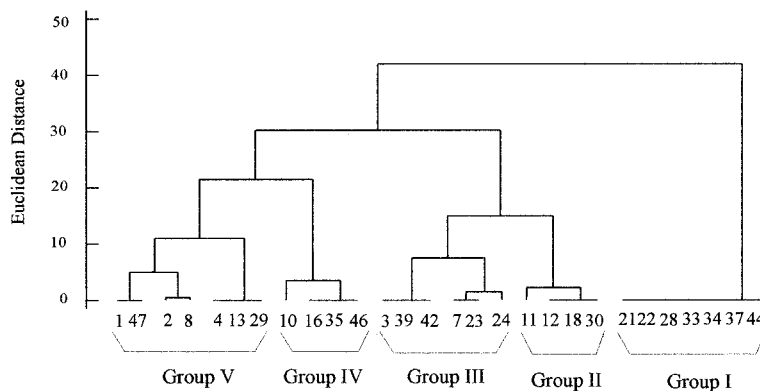
alluvial materials and very close to the marls. This means that there is a less connection with the preferential flowpath, across the alluvial deposits, and shows the influence on the groundwater of the marly materials, with some evaporite intercalations. The less saline points (3, 39, and 42) are located over the alluvial strata, where there is a preferential flow, which determines the lower salinity of the water. Sampling points 39 and 42 fall geographically within group I. Their salinity (1285 mg/liter and 1602 mg/liter) is similar to group I, but the  $\text{SO}_4^{2-}$  (664 mg/liter and 826 mg/liter) and  $\text{Ca}^{2+}$  (178 mg/liter and 196 mg/liter) contents are elevated. These points tap permeable levels at greater depth, close to the impermeable basement. The basement is composed of marly Miocene deposits containing some evaporites, and this explains their slightly higher salinity and elevated  $\text{SO}_4^{2-}$  and  $\text{Ca}^{2+}$ .

Groups IV and V contain the most saline waters in the Detritic Aquifer. Salinity oscillates between 2067

mg/liter and 9877 mg/liter. The waters of group IV (sampling points 10, 16, 35, and 46) have a sulfate facies, with  $\text{SO}_4^{2-}$  content varying between 1780 mg/liter and 2224 mg/liter, together with a narrow range of salinity (3528 mg/liter to 3996 mg/liter). This water type represents the direct influence of the Miocene deposits with their highly soluble gypsiferous intercalations, which form the main source of the  $\text{SO}_4^{2-}$ . These boreholes are situated directly over the very low permeability Miocene marls, which is the reason for their limited exploitation. This rock type is distributed along the length of the valley, so that the points within group IV are widely dispersed (Figure 1).

Group V (sampling points 1, 47, 2, 8, 4, 13, and 29) contains groundwaters showing a wide range of salinity (2067 mg/liter to 9877 mg/liter). Points 4, 13, and 29 have the lowest salinity (2067 mg/liter to 3983 mg/liter), and these lie far way from the groundwater flow that takes place over the alluvial deposits. The presence of Miocene rocks of marine origin in this area, together with Pliocene river delta deposits, may explain the sulfate-chloride nature of these waters, even though the sampling points are located far from the sea. Points 4, 13, and 29 lie in an area of little agricultural activity (as indicated by their low  $\text{NO}_3^-$  content of less than 29 mg/liter). The remaining points in group V (1, 2, 8, and 47) are highly saline (>3500 mg/liter), with concentrations of  $\text{Cl}^-$  (1004–4006 mg/liter) and  $\text{NO}_3^-$  (70 mg/liter to 222 mg/liter) being the highest of those measured in the Detritic Aquifer. These sites are situated on the delta of the River Andarax, where the agricultural activity is significant and where marine intrusion also occurs. These factors lead to increases in  $\text{NO}_3^-$  and  $\text{Cl}^-$ , respectively.

The use of KNN provides a rapid and simple method of groundwater quality assessment. Using this method, it has been possible to differentiate the various processes that determine water quality. In comparison with traditional multivariate statistical methods, such as PCA or cluster analysis, it accepts raw data as input and generates an output (in the form of a classification) that is easier to interpret. It is able to summarize the information without losing some of the information that defines water quality. For this reason, the method could be developed for different regions, with the aim of providing an early system for managers of groundwater quality. It may be of particular interest in areas which have a complex hydrogeochemistry, in which there is a marked interplay of processes, both natural and anthropogenic, contributing to the decline in groundwater quality.



**Figure 7.** Dendrogram obtained from the activation maps.

## Acknowledgments

This article was compiled within the framework of project HID99-0597-CO2, financed by the Spanish CICYT (Interministerial Commission of Science and Technology). We are grateful to reviewers for his comments and suggestions which have contributed to improving the manuscript.

## Literature Cited

- Aguilera, P. A., A. Garrido Frenich, J. A. Torres, H. Castro, J. L. Martínez Vidal, and M. Canton. 2001. Application of the Kohonen neural network in coastal water quality management: methodological development for the assessment and prediction of water quality. *Water Research* 35:4053–4062.
- Appelo, C. A. J., and D. Postma. 1993. Groundwater contamination and pollution. Balkema, Rotterdam, 536 pp.
- Beale, R., and J. Jackson. 1990. Neural computing: an introduction. Adam Hilger.
- Bishop, C. M. 1995. Neural networks for pattern recognition. Oxford University Press, Oxford.
- Bousquet, J. C., and H. Phillip. 1976. Observations tectoniques y microtectoniques sur la distension Plio-Pléistocènes anciens dans l'Est des Cordillères Bétiques (Espagne Meridionale). *Cuadernos de Geología de la Universidad de Granada* 7:57–67.
- Burkart, M. R., D. W. Kolpin, and D. E. James. 1999. Assessing groundwater vulnerability to agricultural contamination in the Midwest US. *Water Science and Technology* 39:103–112.
- Candela, L., M. B. Gómez, L. Puga, L. Rebollo, and F. Villaroya eds. 1991. Acuífero sobreexplotado. XXII International Congress. IAH. Vol I, 580 pp.
- Carter, L. W. 1996. Nitrates in groundwater. Lewis Publishers, BocaRaton, Florida, 264 pp.
- Cerón, J. C., R. Jiménez-Espinosa, and A. Pulido-Bosch. 2000. Numerical analysis of hidrogeochemical data: a case study (Alto Guadalentín, southeast Spain). *Applied Geochemistry* 15:1053–1067.
- Davis, J. C. 1986. Statistic and data analysis in geology. John Wiley, New York.
- Eberhart, R. C., and R. W. Dobbins. 1990. Neural networks PC tools: a practical guide. Academic Press, San Diego.
- Foody, G. M. 1999. Applications of the self-organising feature map neural network in community data analysis. *Ecological Modelling* 120:97–107.
- Gumrah, F., B. Ox, B. Guler, and S. Evin. 2000. The application of artificial neural networks for the prediction of water quality of polluted aquifer. *Water, Air and Soil Pollution* 119: 275–294.
- Hamed, K. H., and A. E. Hassan. 2000. Accuracy of neural network approximators in simulation-optimization. *Journal of Water Resources Planning & Management* 126:48–56.
- Haykin, S. 1994. Neural networks: a comprehensive foundation. Macmillan, New York.
- Helena, B., R. Pardo, M. Vega, E. Barrado, J. M. Fernández, and L. Fernández. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research* 34: 807–816.
- Karydis, M. 1992. Scaling methods in assessing environmental quality: a methodological approach to eutrophication. *Environmental Monitoring and Assessment* 22:123–136.
- Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43:59–69.
- Kohonen, T. 1989. Self-Organization and associative memory. Springer-Verlag, Berlin.
- Kohonen, T. 1990. The self-organizing map. *Proceedings of the IEEE*, pp. 1464–1480.
- Lek, S., and J. F. Guegan (eds.). 2000. Artificial neural networks: application to ecology and evolution. Springer-Verlag, Berlin, 262 pp.
- Li, Y., J. H. Jiang, Z. P. Chen, C. J. Xu, and R. Q. Yu. 1999. A new method based on counter propagation network algorithm for chemical pattern recognition. *Analytica Chimica Acta* 388:161–170.
- López Chicano, M., M. Bouamama, A. Vallejos, and A. Pulido Bosch. 2001. Factors which determine the hydrogeochemical behaviour of karstic springs. A case study from the Betic Cordilleras, Spain. *Applied Geochemistry* 16:1179–1192.
- Mastrorillo, S., S. Lek, F. Dauba, A. Belaud. 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology* 38:237–246.

- Melssen, W. J., J. R. M. Smits, L. M. C. Buydens, and G. Kateman. 1994. Using artificial neural network for solving chemical problems. II. Kohonen self-organizing feature maps and Hopfield networks. *Chemometrics and Intelligent Laboratory Systems* 23:267–291.
- Pulido-Bosch, A., F. Sánchez-Martos, J. L. Martínez-Vidal, and F. Navarrete. 1992. Groundwater problems in a semiarid area (Low Andarax river, Almería Spain). *Environmental Geology and Water Science* 20:195–204.
- Sánchez-Martos, F. 1997. Estudio hidrogeoquímico del Bajo Andarax (Almería). PhD dissertation. University of Granada, Granada, Spain, 290 pp.
- Sánchez-Martos F. and Pulido-Bosch A. 1999. Boron and the origin of salinization in an aquifer in the south-east of Spain. *C.R. Acad. Sci. Paris, Sciences de la Terre et des Planets* 328:751–757.
- Sánchez-Martos, F., A. Pulido-Bosch, A. and J. M. Calaforra-Chordi. 1999. Hydrogeochemical processes in an arid region of Europe (Almería, SE Spain). *Applied Geochemistry* 14:735–745.
- Sanz de Galdeano, C., J. Rodríguez Fernández, and A. C. López Garrido. 1985. A strike-slip fault corridor within the Alpujarra Mountains (Betic Cordilleras, Spain). *Geologische Rundschau* 74:641–675.
- Schleiter, I. M., D. Borchardt, R. Wagner, T. Dapper, K. D. Schmidt, H. H. Schmidt, and H. Werner. 1999. Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks. *Ecological Modelling* 120:271–286.
- Vengosh, A., and E. Rosenthal. 1994. Saline groundwater in Israel: its bearing on the water crisis in the country. *Journal of Hydrology* 156:389–430.
- Voermans, F. M., and J. Baena. 1983. Memoria y Hoja Geológica n° 1045 (Almería). MAGNA-IGME, 52 pp.
- Zurada, J. M. 1992. Introduction to artificial neural systems. West Publishing Company, New York.