ISSN : 0974 - 7451 Volume 5 Issue 6

Environmental Science
An Indian Journal

Trade Science Inc.

Current Research Papers

ESAIJ, 5(6), 2010 [354-358]

Exploratory data analysis of a large groundwater quality dataset of Bhavnagar city by using Kohonen's self organizing map

H.J.Jani[1*], Shriji Kurup[1], Ratna Trivedi[2], P.M.Chatrabhuji[1], P.N.Bhatt[1]
[1]Department of Analytical Chemistry, Bhavnagar University, Bhavnagar, Gujarat, (INDIA)
[2]Department of Microbiology, Shree Ramkrishna Institute of Applies Sciences,
M.T.B. College Campus, Athwalines, Surat, Gujarat (INDIA)
E-mail : drhitendra75@yahoo.com, drratnatrivedi@gmail.com
Received: 9th October, 2010 ; Accepted: 19th October, 2010

ABSTRACT

In groundwater quality research, researchers are often confronted with large, multidimensional datasets. An exploratory data analysis is often carried out, aiming at summarizing the available data, extracting useful information and formulating hypothesis for further research. The techniques traditionally used in this phase of research are basic statistics (i.e. mean, median, standard deviation) along with graphical techniques (i.e. scatter plots, histograms, time series plots). Also multivariate statistical techniques, like principal component analysis or discriminate analysis are used to explore multidimensional datasets. Kohonen's Self-Organizing Map is an artificial neural network technique, developed by Kohonen in the early '80s, especially designed for 2D visualizing and analyzing large, multidimensional datasets[3].

KEYWORDS

Data analysis;
Multidimensional datasets;
Kohonen's self organizing map.

INTRODUCTION

The safe portable water is absolutely essential for healthy living. Ground water is ultimate and most suitable fresh water resource for human consumption in both urban as well as rural areas. The importance of ground water for existence of human society cannot be overemphasized. There are several states in India where more than 90% population are dependent on groundwater for drinking and other purpose. Ground water is also frequently using as the alternative source for agricultural and industrial sector.

The wells are generally considered as the worst type of ground water sources in the term of physio-chemical contamination due to the lack of concrete plinth and surrounding drainage system[4]. Over burden of the population pressure, unplanned urbanization, unrestricted exploration and dumping of the polluted water at inappropriate place enhance the infiltration of harmful compounds to the ground water.

Study area

The technique is applied to a groundwater quality dataset collected from Bhavnagar city groundwater. For the present study, the city was divided into three main sections and four major zones. Four source water types were included in this analysis: (1) The Northwest-North-Northeast Zone [NW-

N-NE], (2) The Northeast-East-Southeast Zone [NE-E-SE] (3) The Southeast-South-Southwest [SE-S-SW] (4) The Southwest-West-Northwest [SW-W-NW].

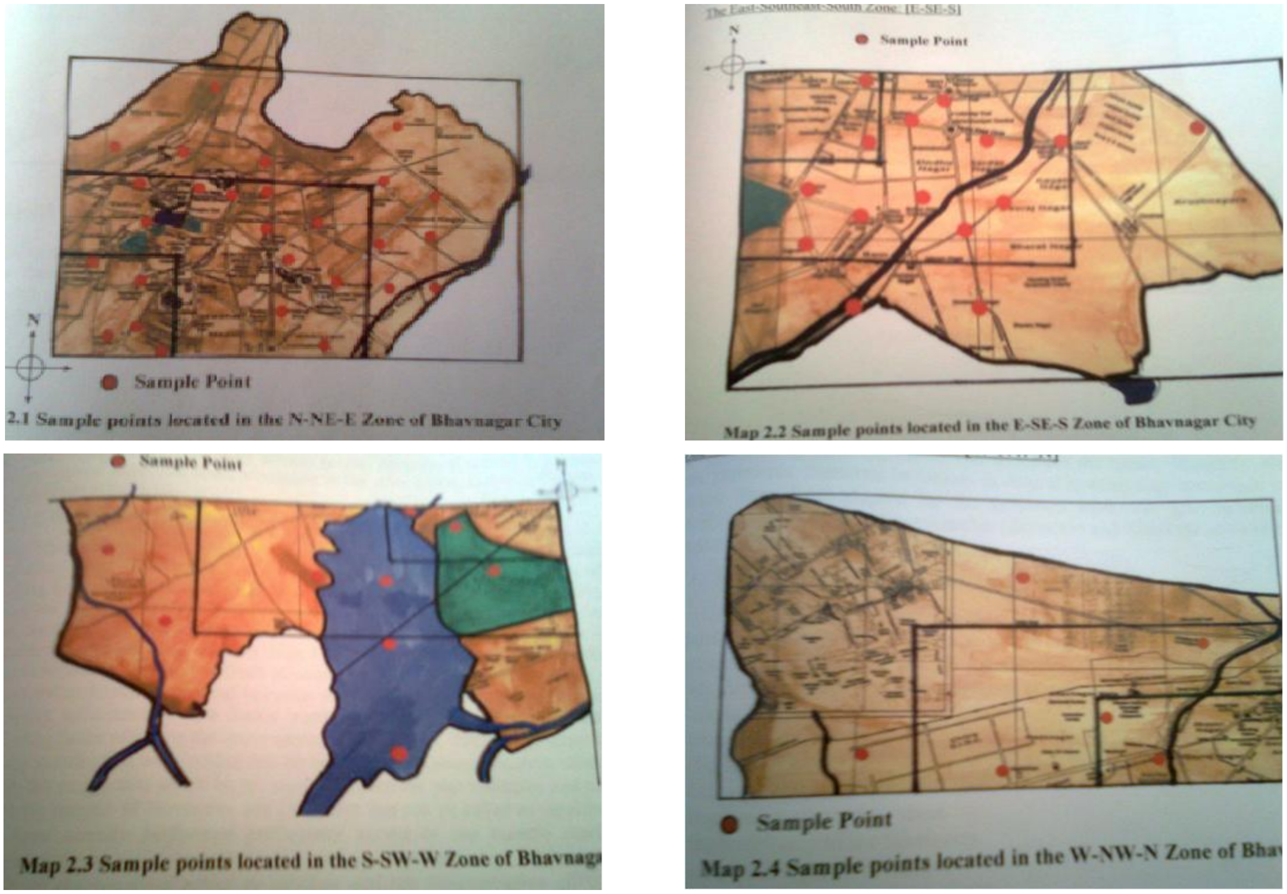


Figure 1 : Self organizing map in hydrogeology of study area

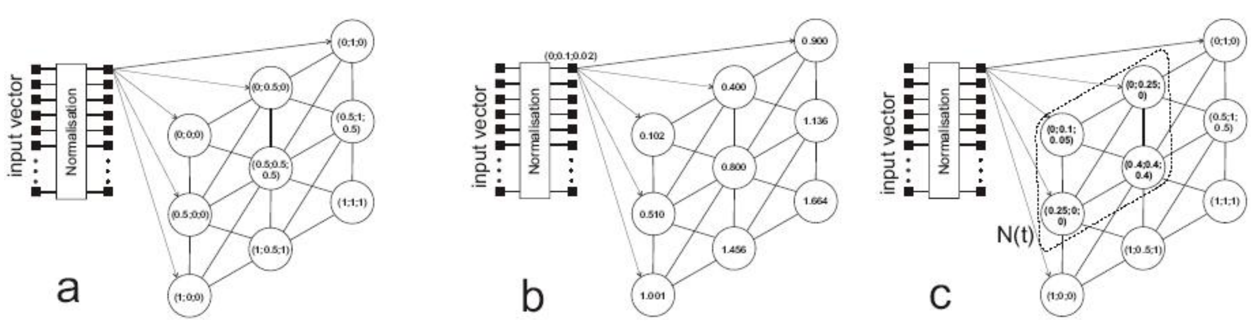
Kohonen's self-organizing map

Artificial neural networks are computer algorithms, based on the functioning of the nervous system of the human brain, capable of learning from data and generalizing.

The SOM neural network is designed to classify data according to their similarity by a learning process. The classification is done by plotting n-dimensional data onto a two-dimensional grid of neurons in such a manner that similar observations are plotted in each others neighbourhood.

The network architecture (Figure 2) consists of an input layer and a layer of neurons, which are fully interconnected. Each of the input vectors is also connected to each of the neurons. The learning process can be divided in six steps[3]:

Step 1: Create m x n matrix (m rows of samples, n variables) of the data set. Since similarity is calculated



Figuur 2 : Kohonen's self organizing map algorithm

based on Euclidean distance, a normalization per variable is carried out.

Step 2: Each neuron is randomly assigned a reference vector with a length equal to the input vectors (n) (Figure 2a)

Step 3: An input vector is shown to the network, the euclidean distance between input vector X and all of the reference vectors M is calculated. (Figure 2b).

Step 4: Choose the Best Matching Unit (BMU), the neuron with the smallest euclidean distance to the input vector (upper left neuron in figure 2b)

Step 5: Update the reference vectors of the BMU and the surrounding neurons so that the euclidean distance to the input vector decreases. (Figure 2c)

Step 6: Repeat steps 3-5 until a predefined maximum of iterations is reached

After step 6 the input vectors are assigned to their best matching units and the result can be visualized, for instance by using Sammon's Mapping (Figure 3)

RESULTS

The studied data set consisted of 390 water samples taken in 2004, of which 20 chemical and physical parameters have been measured. During SOM-analysis depth below ground surface of the sampling point was included. Only one analysis per sampling point was retained and samples containing more than 10% missing values were omitted. This resulted in a data set of 3675 observations on 175 rows and 21 columns. Normalization of the dataset was done by rescaling al variables between 0 and 1.

The results of the SOM-analysis are visualized using Sammon's Mapping (Figure 3). Sammon's mapping projects the neurons in a two-dimensional space in such a mannerthat the distance between two neurons in two dimensions is equal to the distance of the two neurons in the multidimensional space spanned by their reference vectors.

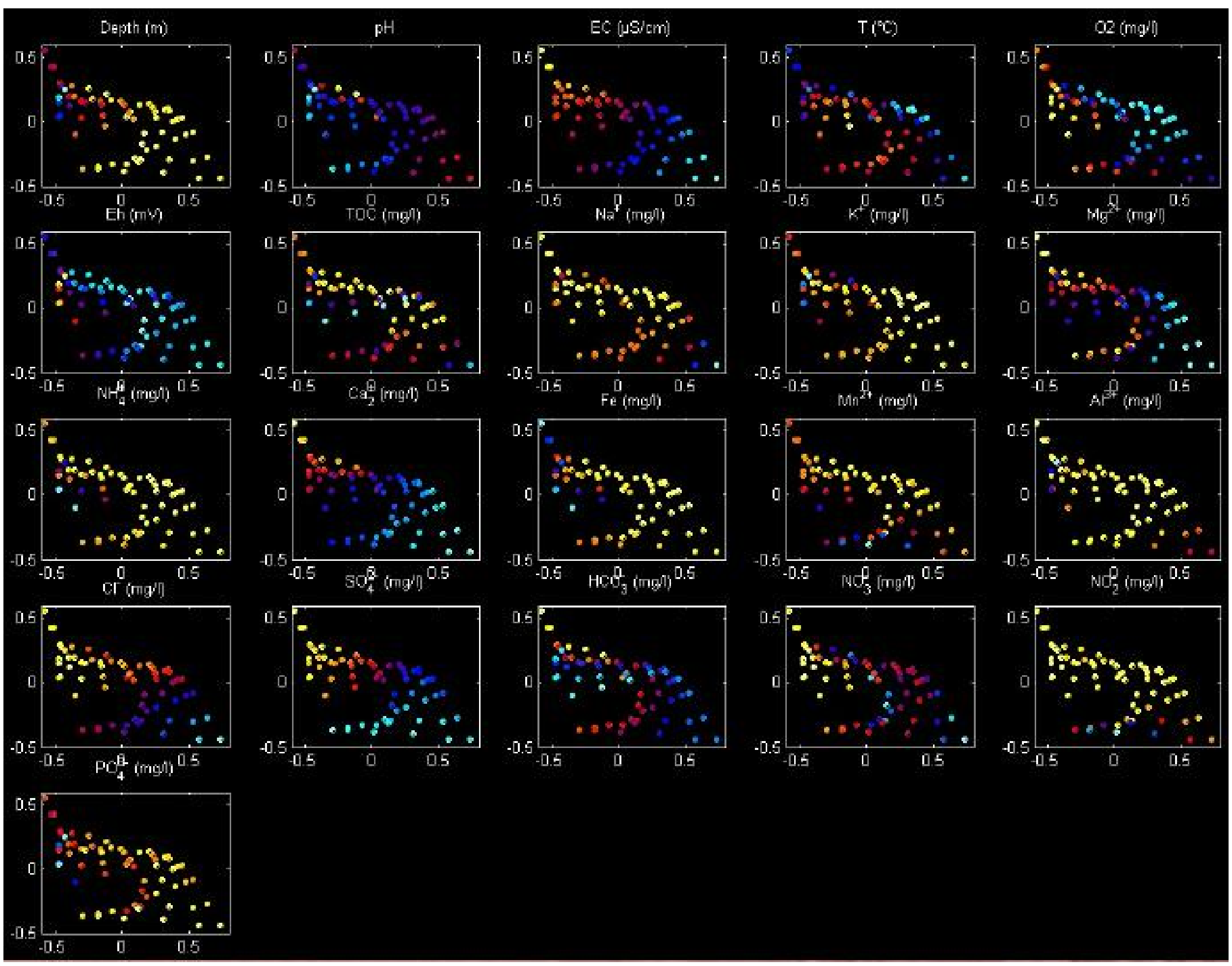


Figure 3 : SOM-analysis by Sammon's mapping

The value of the reference vector for each parameter is indicated by means of a color scale (low values in bleu, high values in red). By comparing these different maps visually one can delineate groups of neurons having similar characteristics and investigate relationships between parameters.

The group of neurons in the upper left corner for instance is characterised by a great depth below ground surface, low concentrations of major cations and anions (Ca, Na, Mg, Cl and SO_4) together with low values of oxygen and Eh. Concentrations of Fe, Mn, Al and K, NH_4 and PO_4 on the other hand are relatively high for this group.

Looking at the graphs as a whole, following relationships between parameters can be observed:

- Oxygen concentration decreases with increasing depth, although shallow anaerobic points exist
- Eh and NO_3 increase with increasing oxygen concentration
- Fe, Mn, Al, TOC and NO_2 increase with decreasing oxygen, although elevated NO_2 levels are restricted to shallow, anaerobic points
- Major cations and anions (Ca, Na, Mg, Cl, SO_4 and HCO_3) together with EC have the same distribution in which elevated concentrations occur at shallow depths. It has to be remarked however that a region of elevated HCO_3 concentrations at greater depth exists.
- K, NH_4 and PO_4 increase with increasing depth.[1]

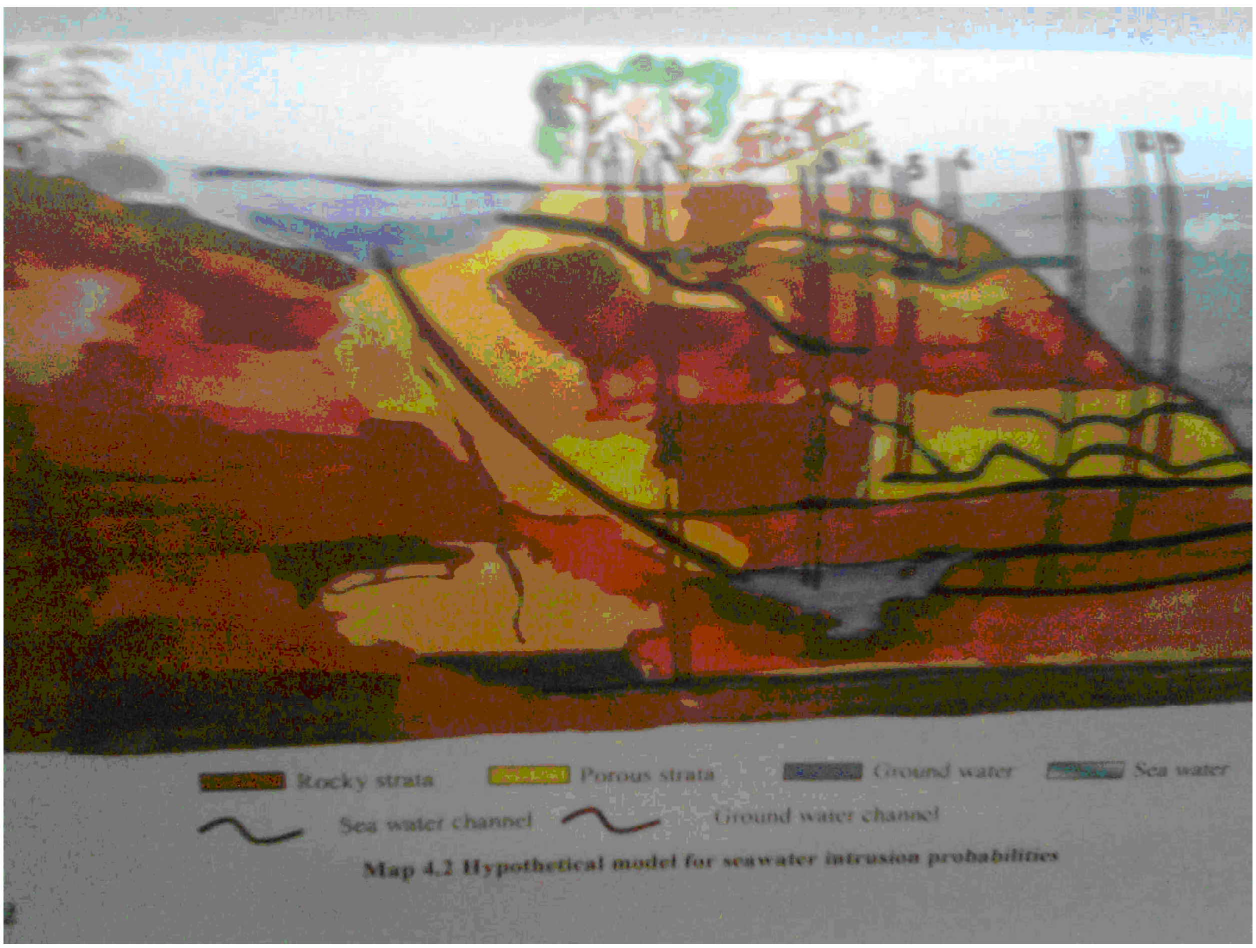


Figure 4 : Hypothetical model for seawater intrusion probabilities

DISCUSSION

Based upon the results of the exploratory data analysis and the geological knowledge of the groundwater system following hypothesis about the hydrogeochemistry of the groundwater in the Dijle basin can be formulated (Figure 4):

- Due to the presence of oxygen in the shallow, phreatic aquifer, no Fe, Al, Mn, NH_4, NO_2 or organic carbon is present in solution. Since oxygen inhibits denitrification, nitrate stays present in the shallow, phreatic aquifer.

- Nitrite only occurs in shallow, anaerobic wells, an indication of local denitrification of nitrate.
- Oxygen concentration decreases with depth, allowing Fe, Al and Mn to be present in solution
- The decrease of Ca and Mg and the increase of NH_4, K and Na with depth can be related to cation exchange on clays where Ca and Mg are exchanged for NH_4, K and Na.
- Elevated phosphate concentrations in depth are probably linked to the presence of phosphate nodules in the Cretaceous aquifer
- Very high concentrations of Na at shallow depth only occur in two wells and is possibly related to anthropogenic influence

ACKNOWLEDGEMENT

The authors express their sincere thanks to Bhavnagar University for their research support and Department of Ocean Development (DoD), Govt. of India for their financial support

REFERENCES

[1] APHA; Standard Methods for the Examination of Water and Wastewater. APHA-AWWA-WPCF, Washington D.C., **(1998)**.

[2] M.Gulinck, W.Loy; Bull.Soc.Belge Géol., Paléont., Hydrol., **80(1-2)**, 77-83 **(1971)**.

[3] T.Kohonen; Self-Organizing Maps. Springer, Berlin, 35-56 **(1995)**.

[4] World Health Organization (WHO); 'Guidelines for Drinking Water Quality' (2nd Edition). Surveillance and Control of Community Supplies. Geneva, Switzerland, **3**, **(1997)**.