Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Environmental Pollution 147 (2007) 771–780

ENVIRONMENTAL
POLLUTION

www.elsevier.com/locate/envpol

Multivariate statistical evaluation of trace elements in groundwater in a coastal area in Shenzhen, China

Kouping Chen [a,*], Jiu J. Jiao [a], Jianmin Huang [b], Runqiu Huang [b]

[a] *Department of Earth Sciences, The University of Hong Kong, Pokfulam Road, Hong Kong, China*
[b] *Department of Environment and Civil Engineering, Chengdu University of Technology, Chengdu, China*

Received 17 December 2005; received in revised form 28 August 2006; accepted 2 September 2006

Multivariate statistical analysis was used to investigate relationships among trace elements and factors controlling trace element distribution in groundwater.

Abstract

Multivariate statistical techniques are efficient ways to display complex relationships among many objects. An attempt was made to study the data of trace elements in groundwater using multivariate statistical techniques such as principal component analysis (PCA), Q-mode factor analysis and cluster analysis. The original matrix consisted of 17 trace elements estimated from 55 groundwater samples colleted in 27 wells located in a coastal area in Shenzhen, China. PCA results show that trace elements of V, Cr, As, Mo, W, and U with greatest positive loadings typically occur as soluble oxyanions in oxidizing waters, while Mn and Co with greatest negative loadings are generally more soluble within oxygen depleted groundwater. Cluster analyses demonstrate that most groundwater samples collected from the same well in the study area during summer and winter still fall into the same group. This study also demonstrates the usefulness of multivariate statistical analysis in hydrochemical studies.

Keywords: Correlation matrix; Principal component analysis; Q-mode factor analysis; Hierarchical cluster analysis; Trace element; Groundwater

1. Introduction

Complex processes control the distribution of trace elements in groundwater, which typically has a large range of chemical composition (Hem, 1970; Drever, 1982; Appelo and Postma, 1993). The trace element composition of groundwater depends not only on natural factors such as the lithology of the aquifer, the quality of recharge waters and the types of interaction between water and aquifer, but also on human activities, which can alter these fragile groundwater systems, either by polluting them or by changing the hydrological cycle (Helena et al., 2000). Sophisticated data analysis techniques are required to effectively interpret trace element data in groundwater.

Shenzhen is a coastal city located in southern China (Fig. 1). The scale and pace of development in Shenzhen are unprecedented in the past decades. Before 1979, Shenzhen was only a small township with a population of less than 20,000 engaged mainly in agricultural activities (Lam, 1986). But now Shenzhen has been developed into a modern industrial, residential and commercial urban conglomeration. Such intense human activities have caused a mounting pressure on the environment, including groundwater systems. Almost little data were available about the trace element composition of groundwater in Shenzhen before. A groundwater monitoring program was recently established to investigate the trace element composition in a coastal area in Shenzhen.

The present study area falls in the alluvial plain in Shenzhen. The Sand River separates the study area into two sides,

---

* Corresponding author. Tel.: +852 2857 8250; fax: +852 2517 6912.
E-mail address: kpchen@hkusua.hku.hk (K. Chen).

0269-7491/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envpol.2006.09.002

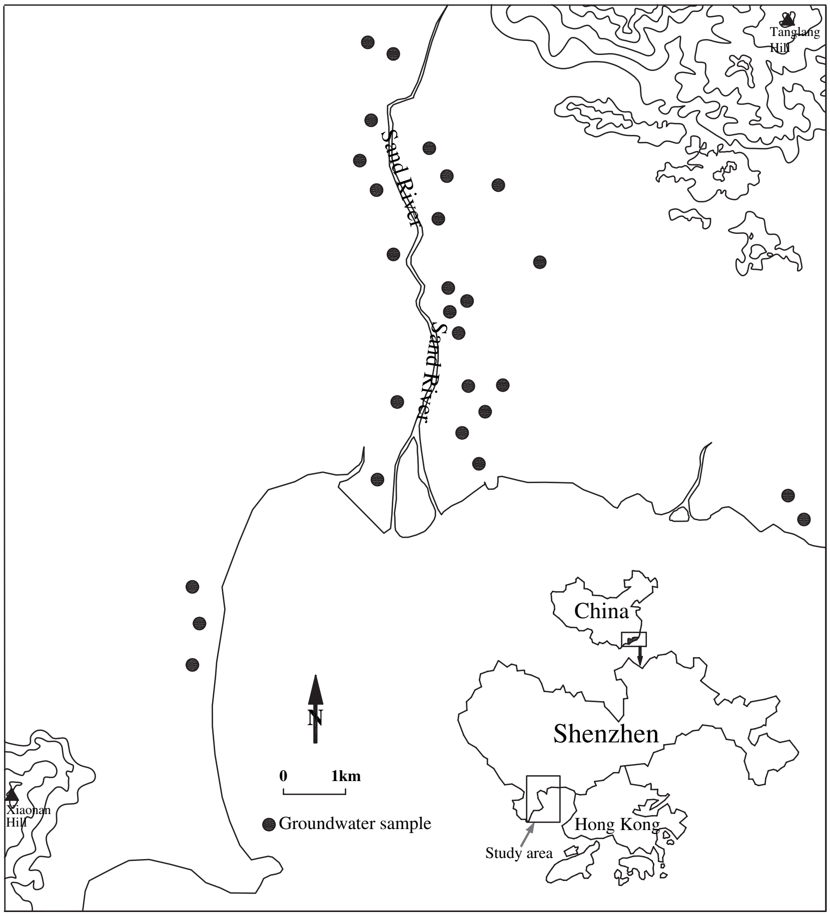

Fig. 1. Sketch map of the study area and locations of groundwater samples.

the west and the east (Fig. 1). The bedrock of this area is predominantly granite. The dominant superficial soil is decomposed granite. Groundwater samples were collected from 27 wells in the study area and some of the wells were sampled multiple times.

The univariate statistical analysis has been generally used to treat trace element data in groundwater (Helena et al., 2000). The simplicity of the univariate statistical analysis is obvious and likewise the fallacy of reductionism could be apparent (Ashley and Lloyd, 1978). In order to avoid this problem in our study, multivariate analysis such as principle component analysis (PCA), Q-mode factor analysis and cluster analysis is used to explain the correlation amongst a large number of variables in terms of a small number of underlying factors without losing much information (Jackson, 1991; Meglen, 1992). The intention underlying the use of multivariate analysis is to achieve great efficiency of data compression from the original data, and to gain some information useful in the interpretation of the environmental geochemical origin. This method can also help indicate natural associations between samples and/or variables (Wenning and Erickson, 1994) thus highlight the information not available at first glance. This multivariate treatment of environmental data is widely successfully used to interpret relationships among variables so that the environmental system could be better managed (Andrade et al., 1992; Aruga et al., 1995; Vega et al., 1998; Tao, 1998; Gangopadhyay et al., 2001; Liu et al., 2003). However, multivariate analysis method has not achieved a comprehensive diffusion in the groundwater studies to date (Aruga et al., 1993; Giammanco et al., 1998; Helena et al., 2000). In this study, multivariate statistical techniques were used to interpret the trace element data in groundwater samples collected from a coastal area in Shenzhen, China.

The goals of the present study were to (1) determine natural associations between groundwater samples and/or variables; (2) investigate the temporal evolution of groundwater composition between two surveys; (3) demonstrate the usefulness of the statistical analysis to improve the understanding of the groundwater composition.

2. Brief review of three multivariate statistical techniques used in this study

Multivariate statistical techniques can help to simplify and organize large data sets to provide meaningful insight (Laaksoharju et al., 1999). In the present study, three multivariate statistical techniques were used to evaluate the concentrations of trace elements in groundwater samples. The statistical software package SPSS11.5 for windows (SPSS Inc., Chicago, IL, USA, 2002) was used for the multivariate statistical calculations. The original data matrix consisted of 17 columns indicating each of the trace elements and 55 rows describing the individual samples.

2.1. Principal components analysis

Principal component analysis (PCA) is a multivariate statistical technique used for data reduction and for deciphering patterns within large sets of data (Wold et al., 1987; Farnham et al., 2003). Principal components are nothing more than the eigenvectors of a variance–covariance or a correlation matrix of the original data matrix. By themselves they may provide significant insight into the structure of the matrix not available at first glance. PCA results vary considerately depending on whether the covariance or correlation matrix is used when large differences exist in the standard deviation of the variables (Davis, 1986). When the correlation matrix is used, each variable is normalized to unit variance and therefore contributes equally. The concentrations of 17 trace elements studied in the groundwater samples of the research area vary differently and the PCA is therefore applied to the correlation matrix for the present study. Thus the raw data matrix was first centered about zero by subtracting the means from each column and then dividing each of the values within each column by the column standard deviation. The eigenvectors of the correlation matrix are principal components and each original observation is converted to what is called principal component score by projecting it onto the principal axes. The elements of the eigenvectors that are used to compute the scores of the observations are called principal component loadings. Because the correlation matrix is symmetrical, the eigenvectors are mutually orthogonal. Typically, the raw data matrix can be reduced to two or three principal component loadings that account for the majority of the variance. The first principal component loading explains the most variance and each subsequent component explains progressively less. As a result, a small number of factors usually account for approximately the same amount of information as the much larger set of the original observations do. The PC loadings can be examined to provide further insight into the processes that are responsible for the similarities in the trace element concentrations in the groundwater samples. PC scores of each groundwater samples are plotted together to investigate similarities between them.

2.2. Q-mode factor analysis

When evaluating groundwater geochemistry, the relative composition of the trace elements in a sample is often as important as their absolute concentrations (Farnham et al., 2003). The attention of Q-mode factor analysis is devoted to interpret the inter-object relationships in a data set rather than the inter-variable relationships explored by the principal component analysis. A similarity matrix that consists of coefficients of proportional similarity between samples is first established in Q-mode analysis. The most widely used similarity measure in Q-mode factor analysis is the cosine θ coefficient of proportional similarity (Davis, 1986),

$$\text{cosine } \theta_{ij}\frac{\sum_{k=1}^{m} x_{ik}x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2 \sum_{k=1}^{m} x_{jk}^2}}$$

This expresses the similarity between object i and object j by regarding each as a vector defined in m-dimensional space. The value of cosine θ ranges from $+1$ for two collinear vectors to 0 for two orthogonal vectors. Since cosine θ measures only the angular similarity, it is sensitive only to the relative proportions of the variables and not to their absolute magnitudes. This means that a concentrated sample exhibits the same similarity matrix to one that is more dilute if the relative proportions of the trace element are the same. Therefore in Q-mode factor analysis, the Q-mode factor loadings rather than the scores are plotted if we wish to see the relationships between samples.

2.3. Hierarchical cluster analysis

Cluster analysis comprises a series of multivariate methods which are used to find true groups of data. In clustering, the objects are grouped such that similar objects fall into the same class (Danielsson et al., 1999). Hierarchical cluster analysis is the most widely applied techniques in the earth sciences and is used in this study. Hierarchical clustering joins the most similar observations, and then successively the next most similar observations. The levels of similarity at which observations are merged are used to construct a dendrogram. Some measure of similarity must be computed between every pair of objects. In this study, a standardized m-space Euclidian distance (Davis, 1986), d_{ij} is used:

$$d_{ij} = \sqrt{\frac{\sum_{k=1}^{m}\left(X_{ik} - X_{jk}\right)^2}{m}}$$

where X_{ik} denotes the kth variable measured on object i and X_{jk} is the kth variable measured on object j.

A low distance shows the two objects are similar or “close together”, whereas a large distance indicates dissimilarity.

3. Methods

3.1. Collection of groundwater samples

A total of 55 groundwater samples were collected from Shenzhen over two 15-day periods from 17 June to 2 July (summer, wet season) and 15 November to 30 November, 2004 (winter, dry season) from 27 wells. The wells sampled were listed in Table 1, along with the sample collection date. Locations of these wells were depicted in Fig. 1. Samples for laboratory trace element analysis were collected into a 125 ml polyethylene narrow-mouth bottle with screw cap. Special care was taken to avoid contamination during sampling for dissolved trace element. Before sample collection, the bottle was rinsed at least three times with groundwater filtered through 0.45 μm mixed cellulose ester membrane (Advantec MFS, USA) directly. After collection, each sample was immediately acidified to $pH < 2$ with ultrapure nitric acid (Fluka, Buchs, Switzerland) and then stored at approximately 4 °C before analysis. Water

Table 1
Trace element concentrations in groundwater samples collected in Shenzhen

Well	Sample data	Code	Concentration (μg/l)																
			V	Cr	Mn	Co	Ni	Cu	Zn	Ga	Ge	As	Mo	Cd	Cs	Ba	W	Pb	U
SW-XWC-1	17-Jun-04	1WXWC1	1.712	1.32	999.6	7.553	4.604	3.77	14.02	0.388	0.066	3.74	2.064	0.069	5.226	132.1	0.254	1.414	1.705
SW-XWC-2	17-Jun-04	2WXWC1	1.468	0.431	303.6	5.455	2.808	1.16	8.337	0.139	0.046	0.702	3.872	0.069	4.521	52.72	0.034	0.602	1.051
SW-CGC-1	17-Jun-04	1WCGC1	0.51	<	276.1	1.155	0.851	<	1.965	0.582	0.029	0.013	0.137	0.143	0.294	107.7	0.02	44.4	0.143
SW-CGC-2	17-Jun-04	2WCGC1	0.975	<	275.3	1.362	0.47	0.418	34.08	0.429	0.049	0.399	0.242	0.149	0.686	101.8	0.101	62.84	0.195
SW-CGC-3	17-Jun-04	3WCGC1	1.147	1.067	7479	9.641	1.803	0.523	6.308	2.521	0.034	1.205	0.508	0.023	2.494	143.4	0.09	0.527	0.169
SW-RWC-1	18-Jun-04	1WRWC1	0.646	<	19.39	0.232	0.872	<	1.864	<	0.019	0.42	0.337	<	1.325	8.549	0.013	<	0.071
SW-WGC-1	18-Jun-04	1WWGC1	0.805	<	12720	6.277	2.273	0.778	<	4.092	0.034	0.709	5.989	0.038	0.312	269.6	0.089	0.207	9.004
SW-DCC-1	18-Jun-04	1WDCC1	0.59	<	114.8	0.264	0.713	0.373	10.73	0.057	0.015	0.148	0.283	<	1.928	23.29	<	<	0.234
NS-WXC-1	23-Jun-04	1NWXC1	0.912	2.674	1613	2.609	4.948	0.967	5.484	0.351	0.071	7.157	4.397	0.025	0.5	14.35	0.104	0.198	0.415
NS-WXC-2	23-Jun-04	2NWXC1	3.811	4.062	927.6	1.076	6.691	2.493	8.278	0.22	0.082	9.481	2.684	0.06	0.389	11.94	0.591	0.25	1.604
NS-WXC-3	23-Jun-04	3NWXC1	9.771	3.196	2224	1.524	6.68	4.059	10.69	1.036	0.197	16.96	11.18	0.072	1.026	81.03	6.882	1.616	1.486
SE-ZGC-1	2-Jul-04	1EZGC1	7.545	<	104.4	1.028	2.135	0.434	577.6	0.036	0.037	1.358	1.689	<	0.087	62.56	0.119	<	1.543
SE-ZGC-2	2-Jul-04	2EZGC1	1.242	0.855	155.4	0.504	2.368	1.655	10.06	0.042	<	0.345	2.205	0.002	0.13	130.2	0.045	<	4.084
SE-XWC-1	2-Jul-04	1EXWC1	4.115	<	253.2	0.87	1.433	1.625	12.2	0.071	<	0.795	2.445	0.078	0.871	131.5	0.158	0.11	1.868
SE-GQC-1	2-Jul-04	1EGQC1	0.495	<	253.6	6.989	0.936	<	<	0.123	<	0.264	0.19	0.075	0.855	98.46	<	8.099	0.398
SE-LJC-1	3-Jul-04	1ELJC1	3.252	<	74.05	0.375	1.076	0.968	<	<	0.007	0.974	11.56	0.048	0.559	93.05	0.033	0.116	5.41
SE-XTC-1	3-Jul-04	1EXTC1	1.801	<	111.1	0.316	1.308	<	<	0.07	0.04	7.134	5.04	0.006	0.097	20.1	0.143	<	0.273
SE-SHS-1	3-Jul-04	1ESHS1	0.982	<	719.8	0.355	2.249	<	3.342	0.137	0.08	1.032	0.705	<	1.194	168	0.059	<	0.196
SE-XBS-1	3-Jul-04	1EXBS1	2.295	<	856.8	0.81	2.259	0.518	<	0.213	0.073	19.71	10.29	0.038	0.583	10.56	0.212	0.103	0.941
SE-TTC-1	3-Jul-04	1ETTC1	1.984	<	55.45	0.236	1.046	<	1.48	0.049	<	0.571	0.677	0.009	3.181	27.63	0.017	<	0.094
SE-TTC-2	3-Jul-04	2ETTC1	2.234	<	50.84	0.253	0.446	<	<	<	<	0.421	0.24	<	1.752	15.08	0.063	<	0.093
SE-BSZ-1	4-Jul-04	1EBSZ1	1.717	<	1096	1.461	2.435	0.682	10.25	<	0.053	11.94	3.282	0.018	0.288	38.67	0.132	0.231	0.531
SE-BSZ-2	4-Jul-04	2EBSZ1	3.011	<	1795	0.996	2.712	<	<	0.555	0.041	9.829	6.449	0.011	0.37	30.92	0.571	0.075	0.911
SE-BSZ-3	4-Jul-04	3EBSZ1	2.876	<	73.26	0.214	0.735	<	<	0.041	<	2.824	2.341	<	0.042	4.862	0.058	<	0.074
SZL-JDG-1	4-Jul-04	1SZL1	0.936	<	61.2	0.355	1.174	2.1	29.08	0.093	0.081	0.59	8.435	0.022	0.527	26.94	0.712	0.899	0.14
SZL-JDG-2	4-Jul-04	2SZL1	1.409	<	658.1	1.487	12.37	<	3.453	0.142	0.01	0.377	0.391	0.151	0.065	169.7	0.051	12.07	0.071
SE-SERS-1	5-Jul-04	1ESER1	21.85	0.981	184.4	0.919	3.799	1.137	<	1.671	1.021	16.26	9.158	0.018	2.938	99.79	3.275	5.081	5.169
NS-WXC-2	18-Sep-04	2NWXC3	15.12	1.323	520.1	0.419	<	2.333	14.01	0.24	0.065	11.52	3.723	0.007	0.222	10.18	0.316	0.719	0.779
NS-WXC-3	18-Sep-04	3NWXC3	10.51	0.975	813.5	0.427	<	2.445	7.608	0.372	0.041	5.253	2.986	0.024	0.327	30.6	0.279	0.334	0.562
SW-XWC-1	24-Nov-04	1WXWC2	2.276	3.581	1292.00	2.685	5.964	0.636	3.103	0.885	0.223	16.540	2.723	<	8.197	96.900	2.704	1.190	3.482
SW-XWC-2	24-Nov-04	2WXWC2	1.593	2.188	180.600	4.239	3.595	1.562	12.860	0.212	0.049	0.856	<	<	5.557	51.180	0.144	0.781	1.058
SW-CGC-1	25-Nov-04	1WCGC2	0.337	0.546	128.200	0.845	0.616	0.491	7.651	0.645	0.070	0.143	<	0.023	0.259	49.610	0.046	29.570	0.133
SW-CGC-2	25-Nov-04	2WCGC2	0.612	1.175	170.400	1.131	0.109	0.966	65.490	0.992	0.183	0.343	<	0.024	0.590	65.120	0.064	50.900	0.188
SW-CGC-3	25-Nov-04	3WCGC2	0.958	2.392	6027.00	7.192	2.767	1.073	7.901	2.837	0.047	1.748	<	<	2.216	131.500	0.082	2.042	0.271
SW-RW-1	25-Nov-04	1WRWC2	0.713	0.755	22.080	0.264	1.188	0.638	6.684	0.098	0.033	0.375	<	<	1.655	10.600	0.031	0.138	0.087
SW-DC-1	25-Nov-04	1WDCC2	0.647	1.215	142.300	0.313	1.194	1.113	25.310	0.190	0.048	0.436	<	<	2.471	26.080	0.016	0.093	0.446
SE-ZGC-1	24-Nov-04	1EZGC2	8.766	1.390	35.690	0.877	4.438	0.886	585.200	0.039	0.032	1.432	<	<	0.115	69.330	0.100	0.169	2.715
SE-ZGC-2	24-Nov-04	2EZGC2	0.857	0.730	53.570	0.417	1.767	1.629	51.780	0.041	-0.006	0.301	<	<	0.077	86.960	0.009	0.121	6.053
SE-XWC-1	24-Nov-04	1EXWC2	2.117	0.082	306.500	0.888	2.197	2.184	22.290	0.142	0.005	0.795	<	0.048	0.995	141.600	0.090	0.471	1.139
SE-GQC-1	24-Nov-04	1EGQC2	0.377	0.512	156.000	2.539	<	0.386	18.240	0.141	0.017	0.322	<	0.008	0.311	72.880	0.039	1.273	0.161
SE-LJC-1	23-Nov-04	1ELJC2	1.993	1.648	106.200	0.515	1.879	1.882	10.450	0.056	0.005	0.780	8.752	<	0.896	102.600	0.037	0.459	17.010
SE-XTC-1	23-Nov-04	1EXTC2	1.123	0.570	219.300	0.380	1.628	0.715	3.811	0.124	0.045	14.150	<	<	0.098	18.710	0.189	0.124	0.255
SE-SHS-1	23-Nov-04	1ESHS2	0.587	0.763	349.400	0.400	2.746	0.734	2.135	0.149	0.036	0.949	<	<	1.725	143.200	0.074	0.116	0.233
SE-XBS-1	23-Nov-04	1EXBS2	1.179	1.504	510.500	0.700	1.659	0.541	2.626	0.257	0.048	46.250	<	<	0.517	5.300	0.178	0.158	0.452
SE-TTC-1	23-Nov-04	1ETTC2	0.855	1.579	48.840	0.270	0.662	0.753	7.220	0.293	0.032	0.504	<	<	3.420	25.920	<	0.332	0.145
SE-TTC-2	23-Nov-04	2ETTC2	0.893	1.224	54.430	0.286	0.565	0.542	2.599	0.076	-0.008	0.088	<	<	1.624	13.410	<	0.170	0.084

SE-BSZ-1	21-Nov-04	1EBSZ2	0.783	1.616	600.100	1.004	2.704	0.625	4.623	0.283	0.029	3.043	<	<	0.275	17.460	0.126	0.263	2.607
SE-BSZ-2	21-Nov-04	2EBSZ2	1.105	0.906	669.500	0.630	2.094	0.586	2.070	0.631	0.048	10.490	<	<	0.362	24.350	1.658	0.202	0.894
SE-BSZ-3	21-Nov-04	3EBSZ2	3.161	0.562	141.900	0.379	0.847	0.971	3.355	0.093	0.008	3.780	<	<	0.048	7.291	0.115	0.227	0.345
SZL-JDG-1	21-Nov-04	1SZL2	0.841	1.043	369.700	0.423	<	<	1.342	0.165	0.107	1.671	1.184	<	1.012	67.440	0.912	1.352	0.066
SZL-JDG-2	21-Nov-04	2SZL2	0.346	<	122.300	0.479	4.552	<	11.750	0.071	0.005	0.223	<	1.825	0.041	102.300	0.033	37.750	0.031
SE-SERS-1	21-Nov-04	1ESER2	6.957	1.204	102.500	0.392	0.321	<	0.737	0.181	0.661	29.220	<	<	0.216	15.400	5.500	0.279	1.873
NS-WXC-1	15-Nov-04	1NWXC2	0.895	1.011	1417.00	1.956	1.606	0.572	60.360	0.654	0.018	2.598	<	<	0.486	21.250	0.235	0.101	0.860
NS-WXC-2	15-Nov-04	2NWXC2	4.05	0.89	381.00	0.80	0.97	1.05	2.45	0.19	0.02	12.81	<	<	0.21	6.20	0.70	0.11	0.79
NS-WXC-3	15-Nov-04	3NWXC2	2.40	1.13	406.60	0.99	0.54	1.48	3.40	0.18	0.02	4.73	<	<	0.36	21.31	0.16	0.12	0.51

<: Below detection limit.

temperature, pH, electric conductivity and the dissolved oxygen were all measured in the field with portable electronic instruments.

3.2. Analysis of trace elements in groundwater samples

Collected samples were analyzed at The University of Hong Kong and all trace element concentrations were determined by Inductively Coupled Plasma-Mass Spectrometer (ICP-MS) (Model VG EXCELL). The precision of the values obtained by ICP-MS was determined as the percentage of the relative standard deviation (RSD) of the three consecutive measurements of the standard reference solution and was listed in Table 2. The drift of all elements analyzed was <5%.

In the quality assurance program, both the international standard reference material (SRM 1640) and the synthetic solutions were used for each batch of 10-sample analysis in order to verify the accuracy in the analytical procedures applied. Three replications of each analysis were performed and the mean values were used for calculations. The recovery of the standard reference material SRM 1640 (Trace Elements in Natural Water, National Institute of Standards and Technology) was greater than 90% in all trace elements analyzed. The methods employed for the determination of the other species, not certified in the reference materials, were instead checked by preparing synthetic aqueous solutions. Table 2 showed the results of the analytical determinations carried out on the standard reference solutions.

Table 1 summarizes the concentrations of 17 trace elements analyzed in groundwater samples. All statistical calculations were based on the data listed. For the purpose of constructing all plots and statistic calculation, data below analytical detection limits were set to a value of half of the detection limit.

4. Results and discussion

4.1. Seasonal variation of groundwater composition

The concentrations of trace elements (V, Cr, Mn, Co, Ni, Cu, Zn, Ga, Ge, As, Mo, Cd, Cs, Ba, W, Pb, U) in the groundwater samples were reported in Table 1. All statistical analyses were based on this data set. Table 3 summarized the measured variables, the limits of detection, and the mean and standard deviations found in two surveys during the wet and dry seasons in 2004. The concentrations of V, Mn, Co, Ni, Cu, Ga, Ge, Mo, Ba, W, Pb displayed relatively high values during the summer and very low values during the winter, however, the concentrations of Cr, Zn, As, Cd, Cs and U were lower in groundwater samples in the summer than in the winter.

Table 2
Accuracy (as relative error) and precision (as relative standard deviation) data for the method applied

Species	Solution	Certified/reference value (μg/l)	Measured value (μg/l)	Relative error (%)	RSD (%)
V	SRM 1640	12.99 ± 0.37	13.42	3.31	2.59
Cr	SRM 1640	38.6 ± 1.6	39.13	1.37	2.58
Mn	SRM 1640	121.5 ± 1.1	124	2.06	0.33
Co	SRM 1640	20.28 ± 0.31	22.74	12.13	0.692
Ni	SRM 1640[a]	27.4 ± 0.8	29.77	8.65	0.273
Cu	SRM 1640[a]	85.2 ± 1.2	89.42	4.95	1.38
Zn	SRM 1640[a]	53.2 ± 1.1	52.37	1.56	2.29
As	SRM 1640	26.67 ± 0.41	28.17	5.62	2.45
Mo	SRM 1640	46.75 ± 0.26	41.82	10.55	3.99
Cd	SRM 1640	22.79 ± 0.96	22.63	0.70	1.67
Ba	SRM 1640	148.0 ± 2.2	143.4	3.11	1.6
Pb	SRM 1640	27.89 ± 0.14	27.92	0.11	1.98

[a] Values for reference.

Table 3
Detection limits, mean value and standard deviation in each survey

Variables	Detection limit (μg/l)	Survey 1		Survey 2	
		Mean (μg/l)	Standard deviation	Mean (μg/l)	Standard deviation
V	0.033	2.97	4.33	1.79	2.01
Cr	0.42	0.56	1.31	1.16	0.72
Mn	0.03	1239.07	2719.5	538.99	1173.3
Co	0.018	2.01	2.65	1.19	1.54
Ni	0.16	2.64	2.60	1.86	1.49
Cu	0.37	0.88	1.13	0.85	0.48
Zn	0.66	27.75	134.0	35.59	113.5
Ga	0.026	0.48	0.97	0.37	0.57
Ge	0.01	0.08	0.22	0.07	0.13
As	0.036	4.27	5.91	5.95	10.74
Mo	0.061	3.58	3.68	0.49	3.99
Cd	0.001	0.04	0.05	0.07	0.80
Cs	0.001	1.19	1.37	1.30	1.89
Ba	0.087	76.83	65.57	53.61	44.36
W	0.008	0.51	1.47	0.51	1.22
Pb	0.067	5.14	17.31	4.94	13.06
U	0.0008	1.40	2.13	1.61	3.43

4.2. Statistical analysis

4.2.1. Correlation between variables

In this study, the principal component analysis was based on the eigenanalysis of the correlation matrix. The close inspection of correlation matrix was useful because it can point out associations between variables that can show the overall coherence of the data set and indicate the participation of the individual chemical parameters in several influence factors, a fact which commonly occurred in hydrochemistry (Helena et al., 2000). Instead of eliminating trace element concentrations below detection limits of some groundwater samples, a value corresponding to half of the detection limit of the element was assigned to that sample. Table 4 was the correlation matrix of the 17 trace element variables. Only those with correlation values higher than 0.50 were considered. Inspection of Table 3 and the correlation matrix showed that the trace elements such as Cr, Zn, As, Cd, Cs and U which displayed lower concentration in the summer than in the winter were closely correlated, as can be seen from the correlation matrix (Cr and As, $r = 0.813$; Cr and Cs, $r = -0.841$; Zn and Cd, $r = 0.776$; Zn and Cs, $r = 0.560$; Cd and U, $R = 0.903$). One interpretation of these observations was that these trace elements in groundwater had similar hydrochemical characteristics in the study area. Relatively more reducing groundwater exhibited greater concentrations of Mn, Co and Ba, and relatively more oxidizing groundwater was characterized by greater concentrations of Cr, As, and U. Trace elements such as Cr, As, and U typically occurred as soluble oxyanions in oxidizing waters, whereas in reducing environments U(VI) was reduced to U(IV) and precipitated as a solid phase (Farnham et al., 2003). Mn is very soluble in low pH (reducing) waters. The redox sensitive element, As, was relatively more soluble in oxidized groundwater in which it existed as oxyanion AsO_4^{2-} or $H_2AsO_4^-$. However, in reducing waters, arsenic was easily to be incorporated in insoluble minerals (Langmuir,

Table 4
Pearson correlation coefficients for the 17 trace elements

	V	Cr	Mn	Co	Ni	Cu	Zn	Ga	Ge	As	Mo	Cd	Cs	Ba	W	Pb	U
V	1.000																
Cr	*0.555*	1.000															
Mn	−0.112	−0.225	1.000														
Co	−0.572	*−0.834*	0.587	1.000													
Ni	*0.676*	*0.892*	−0.539	−0.866	1.000												
Cu	*0.703*	0.392	−0.419	−0.342	*0.705*	1.000											
Zn	0.336	−0.073	−0.387	0.093	0.306	*0.864*	1.000										
Ga	0.028	−0.232	*0.984*	0.574	−0.496	−0.294	−0.274	1.000									
Ge	0.965	0.588	−0.203	−0.638	0.743	0.723	0.335	−0.075	1.000								
As	0.900	0.813	−0.217	−0.784	0.874	0.650	0.183	−0.132	0.947	1.000							
Mo	0.878	0.415	−0.291	−0.643	0.602	0.549	0.199	−0.167	0.938	0.841	1.000						
Cd	0.522	0.064	*−0.663*	−0.296	0.438	0.758	*0.776*	−0.536	0.463	0.319	0.442	1.000					
Cs	−0.391	*−0.841*	−0.118	0.704	−0.563	0.080	*0.560*	−0.075	−0.427	−0.656	−0.359	0.373	1.000				
Ba	−0.049	−0.577	0.603	0.829	−0.516	0.152	0.435	0.672	−0.122	−0.319	−0.230	−0.019	0.594	1.000			
W	*0.971*	0.429	−0.028	−0.475	0.565	0.648	0.306	0.117	0.969	0.862	0.924	0.430	−0.339	0.048	1.000		
Pb	*0.621*	−0.093	−0.090	0.086	0.243	0.817	0.833	0.067	0.628	0.401	0.549	0.630	0.381	0.564	0.682	1.000	
U	0.497	0.351	−0.674	−0.402	0.667	0.871	0.817	−0.581	0.467	0.435	0.327	*0.903*	0.186	−0.089	0.359	0.553	1.000

Figures in italics indicate absolute values greater than 0.5.

1997; Welch and Lico, 1998; Farnham et al., 2003). So the fact that the concentrations of Cr, As and U in groundwater during the winter were higher comparing to summer may be due to the high dissolved oxygen content in groundwater in winter. And those elements such as Mn, Co and Ba which had higher concentration in the summer may be caused by the depletion of oxygen in groundwater.

4.2.2. Principal component analysis

In order to obtain detailed statistical information, more robust statistical methods such as PCA, Q-mode, and cluster analysis were applied to the data set to shed some light on the origin of the elements under study. The PCA was carried out by diagonalization of the correlation matrix, so the problem of different numerical ranges of the original variables was avoided, since all variables were scaled to variance unit and contributed equally. Table 5 summarized the PCA results including the loadings and the eigenvalues of each PC. There were several criteria to identify the number of PCs to be retained in order to understand the underlying data structure (Jackson, 1991). In the present study, factors with eigenvalues greater than 1 were taken into account. Following this rule, three independent factors were extracted, which explained 93.2% of the total variance. The first one was responsible for 50.3% of the total variance and was best represented by V, Cr, Mn, Co, Ni, Cu, Ge, As, Mo, Cd, W, Pb and U. PC 2 explained 24.5% of the total variance and was mainly participated by Cr, Zn, Co, Cd, Cs, Ba and Pb. Additional 18.4% of the total variance was explained in PC 3 and Mn and Ga gave the most contribution. Evaluation of the PC loadings (Table 5) showed that most of the trace elements with greatest positive PC 1 loadings typically occurred as soluble oxyanions in oxidizing waters, whereas two of the trace elements with greatest negative PC 1 loadings (Mn −0.52; Co −0.77) were generally more soluble within oxygen depleted groundwater (Hem, 1989). The solubility of Mn, as Mn^{2+}, was very high in low pH (reducing) waters, and much lower in oxidizing waters because manganese precipitates as Mn(IV)-oxide scavenging other trace elements such as Co, Pb, Zn, Cu and Ni from solution in more oxidizing waters (Hem, 1989; Farnham et al., 2003). Trace elements with the greatest negative PC 1 loadings (Fig. 2) were those expected to be more abundant in more reducing groundwater. V, Cr, As, Mo, W and U having greatest positive PC 1 loadings exhibited more soluble in oxidizing alkaline environments (Hem, 1989). For example, vanadium tended to occur in solution as oxyanion HVO_4^{2-} in oxidizing, moderate to high pH waters, but under reducing conditions vanadium could exist as the oxycation $V(OH)_2^+$ which could show strong adsorption to some aquifer materials (Collier, 1984; Domagalski et al., 1990). As, the redox sensitive element, was commonly more soluble in oxidized groundwater occurring as oxyanion AsO_4^{2-} or $H_2AsO_4^-$. However, in reducing waters, arsenic tended to be incorporated in insoluble minerals (Langmuir, 1997; Welch and Lico, 1998). Mo was highly mobile in oxidizing waters and predominant as MoO_4^{2-} in groundwater. All these may in some extent demonstrated that the PC 1 reflected the oxidizing/reducing conditions within groundwater.

Figs. 3 and 4 show the score plot for the first two PCs, explaining 74.8% of the total variance. Two main zones of each survey could be classified through the diagram of the scores for PC 1 versus PC 2. In both surveys, groundwater samples (except sample 1WXWC2) from wells in the west of the Sand River were in zone 2. Also, comparison of Figs. 3 and 4 indicated that groundwater samples collected in the summer (first survey) displayed more variations than samples collected in the winter.

Table 5
Principal component loadings

Variable	PC 1	PC 2	PC 3
V	*0.88*	0.03	0.42
Cr	*0.68*	−0.60	0.02
Mn	*−0.52*	−0.03	0.82
Co	*−0.77*	0.59	0.24
Ni	*0.90*	−0.28	−0.16
Cu	*0.83*	0.48	−0.02
Zn	0.48	0.82	−0.18
Ga	−0.41	0.07	0.89
Ge	*0.92*	−0.02	0.37
As	0.91	−0.27	0.29
Mo	*0.83*	−0.06	0.29
Cd	*0.66*	0.55	−0.37
Cs	−0.36	0.87	−0.29
Ba	−0.31	0.77	0.53
W	*0.82*	0.08	0.53
Pb	*0.56*	0.75	0.30
U	*0.73*	0.42	−0.43
Eigenvalues	8.55	4.17	3.12
% Variance explained	50.3	24.5	18.4
% Cumulative variance	50.3	74.8	93.2

Figures in italics indicate absolute values greater than 0.5.

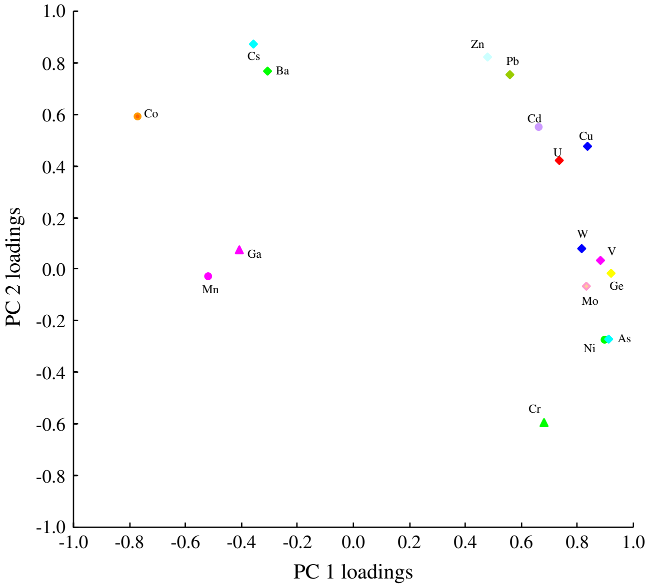

Fig. 2. Contribution of each element to the PC loadings obtained by the principal component analysis.

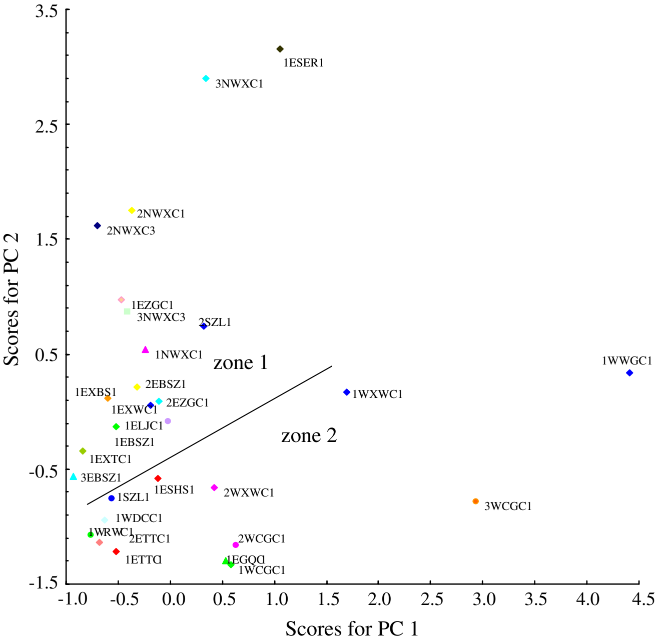

Fig. 3. Principal component scores for the groundwater samples of the first survey.

4.2.3. Q-mode factor analysis

Q-mode factor analysis of the 55 groundwater samples was carried out. The analysis generated three factors which together accounted for 99.9% of the total variance (Table 6). The three factors obtained in this way were rotated using the Varimax procedure (Knudson et al., 1977), which could be more easily interpreted. The first factor (which explains 66.0% of the total variance) was considered as major factor controlling the relative proportions of trace element existing in the groundwater samples and had the high loadings of

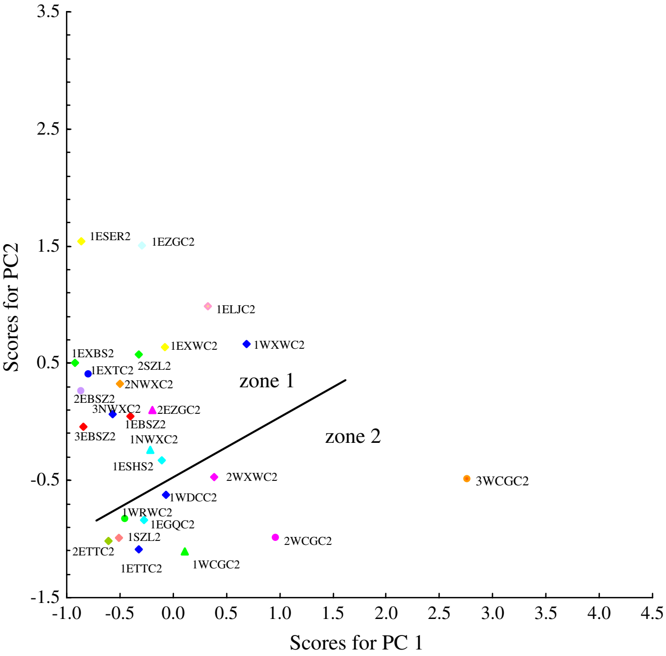

Fig. 4. Principal component scores for groundwater samples of the second survey.

Table 6
Varimax rotated Q-mode factor loading matrix

	Factor 1	Factor 2	Factor 3
1WXWC1	0.899	0.438	0.019
2WXWC1	0.881	0.472	0.027
1WCGC1	0.756	0.654	0.011
2WCGC1	0.769	0.639	0.011
3WCGC1	0.941	0.336	0.021
1WRWC1	0.732	0.679	0.019
1WWGC1	0.941	0.338	0.021
1WDCC1	0.866	0.500	0.022
1NWXC1	0.945	0.326	0.018
2NWXC1	0.945	0.327	0.012
3NWXC1	0.937	0.350	0.013
1EZGC1	*0.622*	*0.781*	−0.007
2EZGC1	*0.464*	*0.886*	0.000
1EXWC1	0.673	0.739	0.005
1EGQC1	0.759	0.650	0.019
1ELJC1	*0.237*	*0.971*	−0.008
1EXTC1	0.882	0.470	−0.043
1ESHS1	0.848	0.530	0.014
1EXBS1	0.946	0.324	0.001
1ETTC1	0.692	0.720	0.019
2ETTC1	0.819	0.572	0.018
1EBSZ1	0.937	0.349	0.011
2EBSZ1	0.943	0.333	0.016
3EBSZ1	0.930	0.366	−0.017
1SZL1	0.726	0.687	0.004
2SZL1	0.835	0.551	0.013
1ESER1	0.670	0.736	−0.069
2NWXC3	0.945	0.327	−0.001
3NWXC3	0.936	0.350	0.014
1WXWC2	0.923	0.384	0.011
2WXWC2	0.826	0.563	0.030
1WCGC2	0.758	0.652	0.011
2WCGC2	0.762	0.648	0.011
3WCGC2	0.941	0.339	0.021
1WRWC2	0.706	0.705	0.024
1WDCC2	0.876	0.482	0.021
1EZGC2	*−0.004*	*0.996*	−0.033
2EZGC2	*0.097*	*0.994*	−0.003
1EXWC2	0.710	0.704	0.007
1EGQC2	0.707	0.706	0.011
1ELJC2	*0.383*	*0.918*	0.021
1EXTC2	0.923	0.383	−0.040
1ESHS2	0.744	0.668	0.008
1EXBS2	0.950	0.307	−0.061
1ETTC2	0.669	0.740	0.025
2ETTC2	0.845	0.534	0.026
1EBSZ2	0.939	0.344	0.017
2EBSZ2	0.937	0.349	0.007
3EBSZ2	0.934	0.357	−0.005
1SZL2	0.875	0.484	0.013
2SZL2	0.465	0.885	−0.004
1ESER2	0.892	0.373	−0.253
1NWXC2	0.943	0.332	0.020
2NWXC2	0.946	0.324	−0.009
3NWXC2	0.931	0.364	0.010
Eigenvalue	36.3	18.6	0.1
Percentage of variance	66.0	33.7	0.2
Cumulative percentage	66.0	99.7	99.9

Figures in italics correspond to samples displaying low loadings in Factor 1 but high loadings in Factor 2.
Rotation method: Varimax with Kaiser normalization.

almost all the samples except samples from wells SE-LJC-1, SE-ZGC-1 and SE-ZGC-2. On the other hand, groundwater samples from the three wells SE-LJC-1, SE-ZGC-1 and SE-ZGC-2 all had high loadings in the second factor. Fig. 5 showed the factor loadings of all groundwater samples for the first two factors. As mentioned earlier, the Q-mode factor analysis described the relative proportions of these trace elements in groundwater samples. Therefore, the relative proportions of trace elements in these groundwater samples were controlled completely by the three factors which together explain 99.9% of the total variance. Factor 1 could represent the type of soil through which groundwater flows. Factor 2 may be related to the reducing/oxidizing conditions within groundwater. And the relative proportions of trace elements in most groundwater samples were controlled more by the types of soils through which the water flowed than the reducing/oxidizing conditions within groundwater. However, the relative proportions of trace elements in groundwater samples from the three wells SE-LJC-1, SE-ZGC-1 and SE-ZGC-2 may be controlled more by the reducing/oxidizing conditions. Table 6 listed the three loadings, eigenvalues of each loading, percentage of variance and the cumulative percentage of variance of the three factors. In Q-mode factor analysis, matrix of coefficients of proportional similarity was row normalized but not column normalized, so the first few factors typically described the variance of the trace elements with the greatest concentrations. Fig. 6 showed the score plot for the first two factors of the 17 trace elements. In this figure, all the trace elements except Mn were clustered together, which may indicate the dominance of Mn in the Q-mode factor analysis. From Table 1, the element of Mn exhibited higher concentration in the 55 groundwater samples when compared to all other trace elements studied, which confirmed the result of the Q-mode factor analysis.

4.2.4. Cluster analysis

The result of the hierarchical cluster analysis was given as a dendrogram (Fig. 7). As can be seen from this figure, the samples collected from the same well during different surveys, except for wells SE-BSZ-1, SZL-JDG-1, SZL-JDG-2 and SE-XTC-1, were clustered together. Therefore, most groundwater samples collected from the same well during different periods in the study area still fell into the same type, which can indicate that the seasonal variability did not affect the groundwater much. However, groundwater samples colleted from the four wells SE-BSZ-1, SZL-JDG-1, SZL-JDG-2 and SE-XTC-1 during different times were not clustered together. This may demonstrate that the groundwater characters in these four wells change very significantly with seasons.

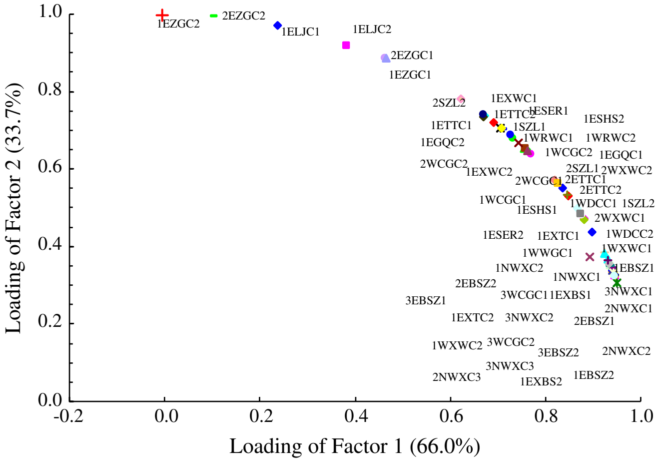

Fig. 5. Plot of the Q-mode factor loadings.

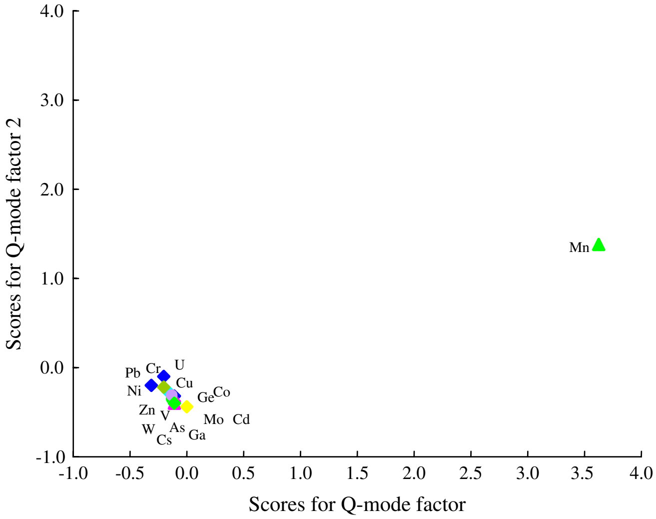

Fig. 6. Plot of Q-mode factor scores.

5. Conclusion

The study showed that the analysis of hydrochemical data using the multivariate statistical techniques such as principal component analysis, Q-mode factor analysis and cluster analysis can give some information not available at first glance. The results of the PCA allowed the reduction of the original data matrix to three important PCs explaining 93.2% of the total variance. Trace elements (V, Cr, As, Mo, W, U) with greatest positive loadings typically occurred as soluble oxyanions in oxidizing waters, while Mn and Co with greatest negative loadings were generally more soluble within oxygen depleted groundwater. PC 1 in some extent reflected the oxidizing/reducing conditions within the groundwater and controlled some redox sensitive trace elements. Q-mode factor analyses demonstrated that the relative proportions of trace elements in groundwater were mostly controlled by the soil type through which the water flowed. Cluster analyses investigated the variation of groundwater in the two surveys and the results demonstrated that most groundwater samples collected from the same well in the study area during the summer and winter still fell into the same group. This study also demonstrated the usefulness of the multivariate statistical analysis in hydrochemical studies.

Acknowledgements

The study was partially supported by the "Two Bases" Project of National Natural Science Foundation of China, the

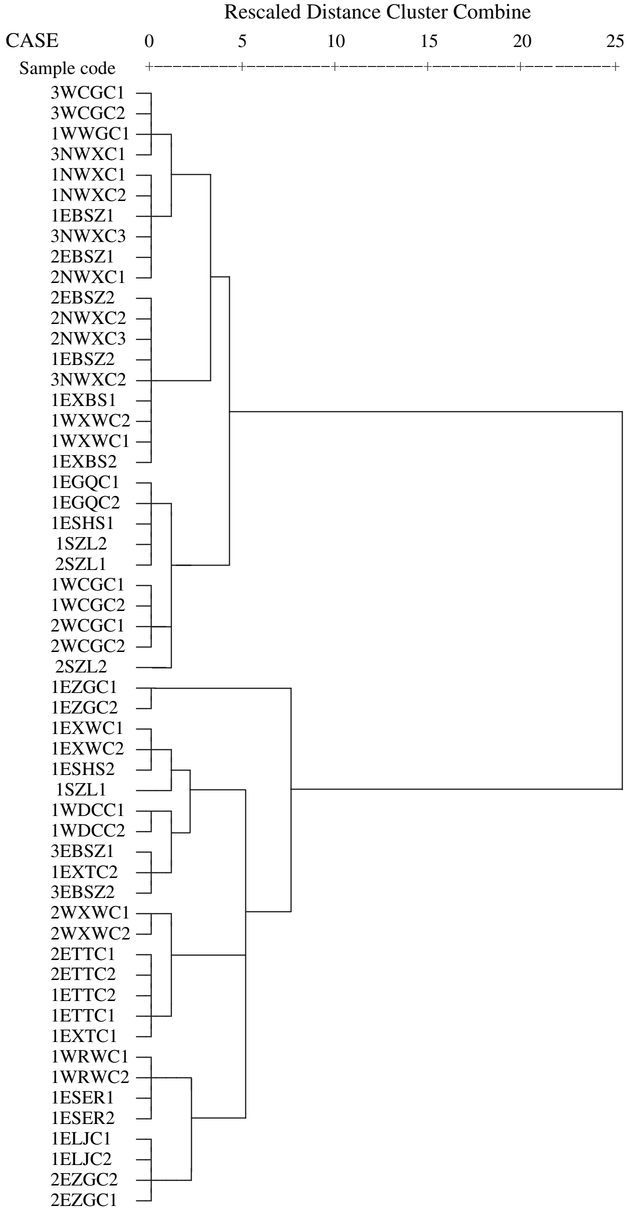

Fig. 7. Dendorgram of the hierarchical cluster analysis using the Ward method.

Research Grants Council of the Hong Kong Special Administrative Region (HKU 7105/02P), and Committee on Research and Conference Grants (CRCG) at the University of Hong Kong. Field assistance in collecting data in Shenzhen from local geotechnical companies including Shenzhen Research and Design Institute, China Academy of Railway Sciences, Shenzhen Gongkan Geotechnical Engineering Co. Ltd, and Shenzhen Geotechnical Engineering Company are highly appreciated. We thank Mr. Guoping Ding, Haipeng Guo and Chiman Leung for their contributions to this research.

References

Andrade, J.M., Padra, D., Muniategui, S., 1992. Multivariate analysis of environmental data for two hydrographic basins. Analytical Letter 25, 379–399.

Appelo, C.A.J., Postma, D., 1993. Geochemistry, Groundwater and Pollution. Balkema, Rotterdam.

Aruga, R., Gastaldi, D., Negro, G., Ostacoli, G., 1995. Pollution of a river basin and its evolution with time studied by multivariate statistical analysis. Analytica Chimica Acta 310, 15–25.

Aruga, R., Negro, G., Ostacoli, G., 1993. Multivariate data analysis applied to the investigation of river pollution. Fresenius' Journal of Analytical Chemistry 346, 968–975.

Ashley, R.P., Lloyd, J.W., 1978. An example of the use of factor analysis and cluster analysis in groundwater chemistry interpretation. Journal of Hydrology 39, 355–364.

Collier, R.W., 1984. Particulate and dissolved vanadium in the north Pacific-ocean. Nature 309, 441–444.

Danielsson, A., Cato, I., Carman, R., Rahm, L., 1999. Spatial clustering of metals in the sediments of the Skagerrak/Kattegat. Applied Geochemistry 14, 689–706.

Davis, J.C., 1986. Statistics and Data Analysis in Geology, second ed., pp. 563–565.

Domagalski, J.L., Eugster, H.P., Jones, B.F., 1990. Fluid–mineral interactions: a tribute to H.P. Eugster. The Geochemical Society Special Publication 2, 315–353.

Drever, J., 1982. The Geochemistry of Natural Waters. Prentice-Hall, Englewood Cliffs, NJ.

Farnham, I.M., Johannesson, K.H., Singh, A.K., Hodge, V.F., Stetzenbach, K.J., 2003. Factor analytical approaches for evaluating groundwater trace element chemistry data. Analytica Chimica Acta 490, 123–138.

Gangopadhyay, S., Gupta, A.S., Nachabe, M.H., 2001. Evaluation of groundwater monitoring network by principal component analysis. Ground Water 39 (2), 181–191.

Giammanco, S., Ottaviani, M., Valenza, M., Veschetti, E., Principio, E., Giammanco, G., Pignato, S., 1998. Major and trace elements geochemistry in the groundwaters of a volcanic area: Mount Etna (Sicily, Italy). Water Research 32, 19–30.

Helena, B., Pardo, B., Vega, M., Barrado, E., Fernandez, J.M., Fernandez, L., 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. Water Research 34 (3), 807–816.

Hem, J.D., 1970. Study and interpretation of the chemical characteristics of natural waters. Geological Survey Water-Supply Paper 1473.

Hem, J.D., 1989. U.S. Geological Survey Water-Supply Paper 2254, 263 pp.

Jackson, J.E., 1991. A User's Guide to Principal Components. Wiley, New York.

Knudson, E.J., Duewer, D.L., Christian, G.D., Larson, T.V., 1977. Application of factor analysis to the study of rain chemistry in the Puget Sound region. In: Kowalski, B.R. (Ed.), Chemometric: Theory and Application. ACS Symposium Series, Washington, DC, pp. 80–116.

Laaksoharju, M., Skarman, C., Skarman, E., 1999. Multivariate mixing and mass balance (M3) calculation, a new tool for decoding hydrogeochemical information. Applied Geochemistry 14, 861–871.

Lam, K.C., 1986. Environment and development in Chinese special economic zones: the case of Shenzhen. Science of the Total Environment 55, 147–156.

Langmuir, D., 1997. Aqueous Environmental Geochemistry. Prentice-Hall, New Jersey, 600 pp.

Liu, W.X., Li, X.D., Shen, Z.G., Wang, D.C., Wai, O.W.H., Li, Y.S., 2003. Multivariate statistical study of heavy metal enrichment in sediments of the Pearl River Estuary. Environmental Pollution 121, 377–388.

Meglen, R.R., 1992. Examining large databases: a chemometric approach using Principal Component Analysis. Marine Chemistry 39, 217–237.

Tao, S., 1998. Factor score mapping of soil trace element contents for the Shenzhen area. Water, Air and Soil Pollution 102, 415–425.

Vega, M., Pardo, R., Barrado, E., Deban, L., 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research 32, 3581–3592.

Welch, A.H., Lico, M.S., 1998. Factors controlling As and U in shallow groundwater, southern Carson Desert, Nevada. Applied Geochemistry 13, 521–539.

Wenning, R.J., Erickson, G.A., 1994. Interpretation and analysis of complex environmental data using chemometric methods. Trends in Analytical Chemistry 13, 446–457.

Wold, S., Esbensen, K., Geladi, P., 1987. Chemometric and Intelligent Laboratory Systems 2, 27–52.