*Pol. J. Environ. Stud. Vol. 24, No. 2 (2015), 769-776*
DOI: *10.15244/pjoes/30263*

*Original Research*

# Multivariate Statistical Analysis of Hydrochemical Data for Shallow Ground Water Quality Factor Identification in a Coastal Aquifer

**Qingchun Yang[1]*, Jianing Zhang[1], Yanli Wang[1], Yanna Fang[2], Jordi Delgado Martín[3]**

[1]Key Laboratory of Groundwater Resources and Environment, Ministry of Education, Jilin University, Changchun 130021, P. R. China
[2]Monitoring Center of Geological Environment, Fujian, 350001, P. R. China
[3]Escuela de Ingenieros de Caminos, Universidad de A Coruña, Campus de Elviña, 15192, Spain

*Received: 28 July 2014*
*Accepted: 3 September 2014*

## Abstract

Multivariate statistical techniques, hierarchical cluster analysis (HCA), and principal component analysis (PCA) integrating graphical method (Piper trilinear graphical diagram) were applied to the factor identification of ground water quality in a coastal aquifer, Fujian province, South China. Ground water samples were collected at 12 sites in January (dry season) and July 2011 (wet season). Eleven ground water quality parameters (pH, TH, TDS, $Ca^{2+}$, $Mg^{2+}$, $Na^{+}$, $Cl^{-}$, $SO_4^{2-}$, $HCO_3^{-}$, $NO_3^{-}$, Mn) were selected in order to perform multivariate statistical analysis. During both the past-monsoon and the summer seasons, PCA results revealed the existence of three significant principal components revealing how processes like salinization, water-rock interaction, and anthropogenic pollution influence ground water quality. Three factors which together explain 90.3% and 83.3% of the total variance in the summer and post-monsoon dataset were retained and interpreted. Cluster analysis using the Ward method with squared Euclidean distance measure was performed, which indicated the distribution of the studied wells according to their water quality. Water samples from 12 wells were clustered into three distinct groups to depict different hydrochemical facies. The results proved that multivariate analysis methods like HCA and PCA could be useful for evaluating ground water pollution and identifying ground water hydrochemistry.

**Keywords:** hydrochemistry, principle component, hierarchical cluster analysis, ground water quality, coastal aquifer

## Introduction

Economic paradigms and sustainable socioeconomic development of every community depends much on the sustainability of the available water resources. Ground water quality is a very sensitive issue that transcends national boundaries. It is influenced by many factors, including atmospheric chemistry, the underlying geology, vegetation, and anthropogenic activities. Chemical composition of ground water in coastal regions differs broadly depending on diverse geo-hydrology, hydrometeorology, topography, drainage, and other artificial conditions imposed [1]. For decades, research has been focusing on hydrochemical analysis in various ways. The Piper diagram has been applied broadly to investigate the ground water facies of ground water samples for further research, such as revealing the evolution of phreatic water and understanding the hydrochemical characteristics as well as the formation

*e-mail: qyang@udc.es

mechanism of the ground water [2, 3]. However, it is difficult to study the inherent relationship and the interaction of these values by the common methods, some more sophisticated data analysis techniques are required to interpret ground water quality effectively. Statistical analysis, based on the theory of statistics, can describe water quality factors macroscopically with the neglect of the evolution mechanism of hydrochemical components. The application of different multivariate statistical approaches such as principal component analysis (PCA), factor analysis (FA), discriminate analysis, and cluster analysis (CA) have been widely used to achieve great efficiency of data compression from the original data and to interpret natural associations between samples or variables, highlighting information that is not available at first glance [4]. Recent studies have confirmed the usefulness of these techniques which have been employed to extract critical information from hydrochemical datasets with respect to ground water quality assessment in several studies [5-8]. The combined use of principal component analysis (PCA) and cluster analysis enabled the classification of water samples into distinct groups on the basis of their hydrochemical characteristics.

PCA is a statistical data reduction tool that can be used to aggregate the effects of many variables into a small subset of factors [9]. This method can be used to interpret observed relationships among variables, to yield simpler relationships that provide insight into the underlying structure of the variables, to assess controls on ground water composition, and to evaluate the spatial distribution of the studied pollutants [10]. It assesses the associations between variables as it indicates the participation of individual chemicals among several factors of influence [11-14]. Cluster analysis is a multivariate statistical method aiming to perform classification by assigning observation to a group to make the variable more or less homogeneous and distinct from other groups [15]. Hierarchical cluster analysis, as one branch of cluster analysis, has the advantage of not demanding any prior knowledge of the number of clusters that the non-hierarchical method does [16]. As one of its widespread uses in the field of hydrochemistry, the HCA had been applied successfully in earlier studies to identify the chemical relationships between the water samples by grouping these water samples according to their different chemical characteristics [17-20].

The focus of the present paper lies in:

1) Piper diagram was practiced to find out the chemical type of the studied wells
2) PCA to identify the controlling processes of ground water quality and to interpret the relative importance of the chemical variables
3) HCA to regroup the monitored wells in the study area

## Experimental Procedures

The study area, Dongshan, is a county of far southern Fujian Province (left part of Fig. 1), People's Republic of China, facing Taiwan and next to Hong Kong and Macao. It is considered an important port for international trade and trade with Taiwan and is under the administration of Zhangzhou City and comprises 44 islands for a total area of 194 square kilometers. It lies between 117º17′ E-117º35′ E longitude, 23º33′ N-23º47′ N latitude, consisting of Dongshan Island and the remaining 44 small islands (Fig. 1). The total length of coastline is around 200 m. It is influenced under the subtropical marine monsoon climate. The annual average temperature is about 20.9ºC and varies between 13.1 in January and 27.3 in July. Annual average rainfall is about 1224. 9 mm, most of which occurs during May and September. A typical feature in the study area is frequent typhoons during July and September. Due to the topography, the water body is not well developed within Dongshan town, surface water is scare, and ground water has become a dependent source of water supply and servers in many aspects. The selected ground water quality data of

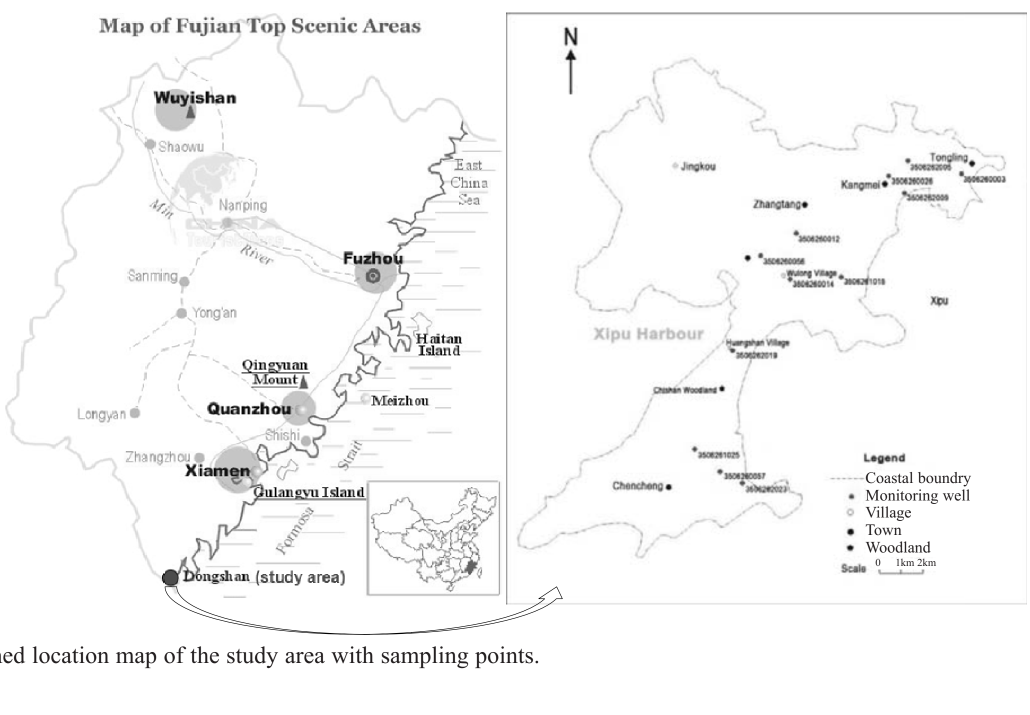


Fig. 1. Outlined location map of the study area with sampling points.

Legend

12 boreholes (right part of Fig. 1) was obtained by sampling the water and monitoring the ion concentrations with an ion chromatograph according to the standard methods for examination of ground water and wastewater at Dongshan hydrological station, which covers pH, TH, TDS, $Ca^{2+}$, $Mg^{2+}$, $Na^{+}$, $Cl^{-}$, $SO_4^{2-}$, $HCO_3^{-}$, $NO_3^{-}$, and Mn.

## Major Ion Balance Analysis

Ion balance analysis is the most fundamental item in quality control of chemical analysis. According to Freeze and Cherry [21], the percent charge balance error (%CBE) is calculated as:

$$\%CBE = \frac{\sum z \cdot m_c - \sum z \cdot m_a}{\sum z \cdot m_c + \sum z \cdot m_a} \cdot 100\%$$

In this equation, z is the absolute value of an ion charge, and mc and ma are the molality of cationic and anionic species, respectively. In this study, major ions like $Na^{+}$, $K^{+}$, $Ca^{2+}$, $Mg^{2+}$, $Cl^{-}$, $SO_4^{2-}$, $HCO_3^{-}$, $PO_4^{2-}$, and $NO_3^{-}$ are considered in the charge balance error calculation for each sampling period. The CEB ranges from 0.07% to 2.97%, and with an average value of 1.67%. Therefore, the percentage error of charge balance is acceptable in the present study.

## Piper Diagram

The piper diagram is a graphical method of ground water classification using a trilinear diagram that contains the concentration of major cations ($Ca^{2+}$, $Mg^{2+}$, $Na^{+}$, $K^{+}$) and anions ($Cl^{-}$, $SO_4^{2-}$, $HCO_3^{-}$, $CO_3^{2-}$). The cations and anions are shown by separate ternary plots. The apexes of the cation plot are calcium, magnesium, and sodium plus potassium. The apexes of the anion plot are sulfate, chloride, and carbonate, plus hydrogen carbonate anions. The two ternary plots are then projected onto a diamond. In the diagram, a water sample is presented with three points, one in each triangle and one in the diamond. The software package Rockware Aq.QA [22] was used to develop the piper diagrams for the classification of water type in this paper.

## Principle Component Analysis (PCA)

The method of principal components is a special case of the more general method of factor analysis. The aim of PCA is construction of new variables called principal components out of a set of existing original variables. The new variables are a linear combination of the existing variables. The PCA is performed to reduce the large data set of variables into a few factors called the principal components, which can be interpreted to reveal underlying data structure. The basic is to classify the variables according to the correlation, i.e., put the variables with high correlation together, whereas the variables with low correlation come from different classes. The aim is to describe the raw data by using the sum of a special factor and a linear function of the fewest common factors on the purpose of interpreting the correlation and simplifying the dimensions. Three steps were followed in carrying out the PCA: computation of correlation matrix by determining the correlation coefficient, estimation of the factor loadings, and factor rotation and interpretation. For the initial factor extraction, Kaiser's [23] eigenvalue criterion of retaining only those components/ factors whose eigenvalues are greater than 1 was adopted. Other criteria used to determine the appropriate number of components to retain include scree plot, variance, and residuals. Kaiser's varimax rotation, an orthogonal rotation procedure that produces a set of component loadings having the maximum variance of the squares of the loadings is used in conducting the principal component analysis to make the factor solutions more interpretable without altering the underlying mathematical structure [24].

## Hierarchical Cluster Analysis (HCA)

Cluster analysis is a statistical method for classification. It groups samples by linking inter-sample similarities and illustrates the overall similarity of variables in the data set [25]. Typical cluster models include connectivity models, centroid models, and distribution models and so on. Hierarchical cluster analysis is the major method for finding a relatively homogeneous cluster of cases based on measured characteristics. It starts with each case as a separate cluster, i.e. there are as many clusters as cases, and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left.

The software package SPSS version 21 was applied for statistical analyses. The techniques contained descriptive statistics, factor analysis, and hierarchical cluster analysis.

# Results and Discussions

The univariate descriptive statistical overview of the hyrochemical data of groundwater sampled is presented in Table 1. The distributions of the water quality parameters were assessed by determining minimum, maximum, mean and standard deviation for each of the 11 variables. We can observe that there was a similar trend of hydrochemical parameters in summer and post-monsoon. The coefficient of variation (CV) is employed to measure the variability of these parameters since some of them have different units. CV eliminates the unit of measurement from the standard deviation of a series of numbers by dividing it by the mean of this series of numbers [12].

## Principle Component Analysis

PCA was performed on the standardized data set of 11 water quality parameters. Table 2 shows the loading for varimax rotated factor matrix, the eigenvalues, the percentage of variance, and the cumulative percentage of the rotated variance associated with each other. The factor loading matrix was rotated in the approach of varimax orthogonal rotation.

In the summer data set, three significant PCs explain 90.28% of the total variation in the hydrochemistry. Most of

Table 1. Descriptive statistics for ground water samples in 2011.

| Variables | Post-monsoon | | | | | | Summer | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Min. | Max. | Median | Mean | SD | CV | Min. | Max. | Median | Mean | SD | CV |
| pH | 4.47 | 7.18 | 6.60 | 6.41 | 0.75 | 0.1171 | 5.41 | 7.02 | 6.49 | 6.42 | 0.46 | 0.0722 |
| TH | 37.47 | 453.21 | 176.23 | 195.77 | 117.74 | 0.6014 | 59.8 | 783.9 | 254.80 | 266.71 | 192.82 | 0.7230 |
| TDS | 109.9 | 1443.6 | 434.30 | 533.34 | 396.07 | 0.7426 | 144.1 | 3689.0 | 521.20 | 835.60 | 982.28 | 1.1755 |
| $Ca^{2+}$ | 7.37 | 116.10 | 42.05 | 49.68 | 34.57 | 0.6959 | 14.13 | 111.10 | 56.91 | 59.57 | 29.02 | 0.4872 |
| $Mg^{2+}$ | 3.57 | 39.47 | 11.84 | 16.92 | 12.37 | 0.7313 | 5.65 | 149.20 | 14.28 | 28.35 | 39.64 | 1.3985 |
| $Na^{+}$ | 14.68 | 353.40 | 44.31 | 95.29 | 115.01 | 1.2069 | 17.52 | 1025.00 | 65.24 | 177.13 | 290.81 | 1.6418 |
| $Cl^{-}$ | 20.4 | 720.0 | 71.98 | 174.30 | 230.20 | 1.3207 | 30.6 | 2180.0 | 118.95 | 364.46 | 621.21 | 1.7045 |
| $SO_4^{2-}$ | 6 | 140 | 75.00 | 74.25 | 46.81 | 0.6305 | 20 | 200 | 74.55 | 85.13 | 55.41 | 0.6509 |
| $HCO_3^{-}$ | 9.1 | 193.7 | 43.90 | 62.53 | 54.67 | 0.8742 | 9.0 | 197.4 | 48.21 | 59.14 | 54.62 | 0.9235 |
| $NO_3^{-}$ | 0.5 | 296.0 | 6.35 | 55.02 | 98.47 | 1.7898 | 0.20 | 274.10 | 22.59 | 54.75 | 85.90 | 1.5690 |
| Mn | 0.01 | 0.55 | 0.05 | 0.15 | 0.19 | 1.2385 | 0.01 | 1.05 | 0.14 | 0.29 | 0.36 | 1.2682 |

Table 2. Rotated component matrix of standardized water quality data set in 2011.

| Variables | Post-monsoon | | | Summer | | |
|---|---|---|---|---|---|---|
| | Factor 1 | Factor 2 | Factor 3 | Factor 1 | Factor 2 | Factor 3 |
| pH | 0.116 | -0.106 | 0.886 | 0.056 | 0.944 | -0.255 |
| TH | 0.714 | 0.695 | -0.011 | 0.951 | -0.105 | 0.268 |
| TDS | 0.927 | 0.310 | -0.008 | 0.988 | -0.106 | 0.026 |
| $Ca^{2+}$ | 0.443 | 0.853 | -0.069 | 0.354 | -0.071 | 0.878 |
| $Mg^{2+}$ | 0.906 | 0.181 | 0.110 | 0.967 | -0.089 | -0.072 |
| $Na^{+}$ | 0.957 | -0.053 | 0.000 | 0.983 | -0.084 | -0.099 |
| $Cl^{-}$ | 0.957 | -0.031 | -0.052 | 0.982 | -0.111 | -0.100 |
| $SO_4^{2-}$ | 0.830 | 0.073 | -0.124 | 0.824 | 0.004 | 0.298 |
| $HCO_3^{-}$ | 0.007 | 0.148 | 0.740 | -0.221 | 0.868 | 0.133 |
| $NO_3^{-}$ | -0.260 | 0.917 | -0.045 | -0.254 | -0.307 | 0.790 |
| Mn | 0.379 | 0.317 | -0.792 | 0.137 | -0.820 | 0.338 |
| Eigenvalues | 5.515 | 2.323 | 2.156 | 5.686 | 2.466 | 1.778 |
| % Variance | 45.961 | 19.355 | 17.963 | 51.691 | 22.417 | 16.167 |
| %Cumulative | 45.961 | 65.316 | 83.279 | 51.691 | 74.108 | 90.275 |

the variance is contained in factor 1 (51.69%), which is associated with the variables TH, TDS, $Mg^{2+}$, $Na^{+}$, $Cl^{-}$, and $SO_4^{2-}$. Factor 1 could be interpreted as the salinization factor. Salinization refers to an increase in the concentration of total dissolved solids in water and can often be detected by an increase in chloride, which was proportionally correlated to the cations like sodium and magnesium. Factor 2 represents 22.42% of the total variation in the hydrochemistry and has high loadings for pH, $HCO_3^{-}$, and Mn. It was obvious that Mn was negative, which meant that its concentration was reduced. In the meantime, factor 2 had a strong positive loading on pH and $HCO_3^{-}$, which indicated that the concentration of $HCO_3^{-}$ increased with increases of pH values in a certain range and is assumed to be indicative of water-rock interaction. The variables $Ca^{2+}$ and $NO_3^{-}$ contribute most strongly to the third component, which explains 16.17% of the total variance. The strong positive correlation of calcium might due to the weathering and the presence of calcium in the rocks of the study area. What's more, the moderate positive correlation of nitrate showed anthropogenic pollution, such as the use of nitrogenous fertilizers in the rural part.

Table 3. Mean values and standard deviation of water quality parameters of water samples post-monsoon in 2011 (wells are representative with the last four numbers).

| Variables Wells | Cluster 1 | | Cluster 2 | | Cluster 3 | |
|---|---|---|---|---|---|---|
| | 0003, 0012, 0014, 0056, 0057, 1025, 2019, 2023 | | 26, 2005 | | 1015, 2009 | |
| | Mean | SD | Mean | SD | Mean | SD |
| pH | 6.53 | 0.48 | 5.80 | 1.88 | 6.53 | 0.37 |
| TH | 136.93 | 77.38 | 280.22 | 19.30 | 346.69 | 150.64 |
| TDS | 334.10 | 166.81 | 566.25 | 27.65 | 1297.40 | 206.76 |
| $Ca^{2+}$ | 33.81 | 20.49 | 89.36 | 16.61 | 73.45 | 60.32 |
| $Mg^{2+}$ | 12.07 | 7.85 | 13.74 | 5.20 | 39.47 | 0.01 |
| $Na^{+}$ | 51.73 | 32.42 | 30.95 | 9.39 | 333.85 | 27.65 |
| $Cl^{-}$ | 83.70 | 57.04 | 58.00 | 10.89 | 653.00 | 94.75 |
| $SO_4^{2-}$ | 62.63 | 43.40 | 55.00 | 21.21 | 140.00 | 0.00 |
| $HCO_3^{-}$ | 76.79 | 62.90 | 30.25 | 4.31 | 37.80 | 14.99 |
| $NO_3^{-}$ | 15.99 | 31.04 | 255.45 | 57.35 | 10.73 | 13.26 |
| Mn | 0.09 | 0.13 | 0.28 | 0.38 | 0.28 | 0.21 |

In the post-monsoon data set, there were still three principle factors extracted from the water quality data set with eigenvalues no less than 1. Factor 1, accounting for 45.96% of variability, is strongly loaded with TDS, $Mg^{2+}$, $Na^{+}$ and $Cl^{-}$, and moderately loaded with TH and $SO_4^{2-}$. This factor could be treated as a process of ion exchange because of the changing hydraulic and climatic conditions. Factor 2 accounted for nearly 19.36% of variance and loaded three parameters: moderately correlated TH and strongly correlated $Ca^{2+}$ and $NO_3^{-}$. TH was a measure of the total concentration of the calcium and magnesium. As shown in Table 1, the max value of TH has been above the permissible limit of the quality standard for ground water in China (≥550 mg/L). This phenomenon in the study area might be imputed to the geological formation of the study area, the improper disposal of wastewater, and the distribution of agricultural activities; hence some measures should be taken like the water softener process and adequate fertilizer. Factor 3 accounted for about 17.96% of variability, including pH, $HCO_3^{-}$, and Mn. This factor might be related to mineral weathering. The value of pH had a negative effect on the precipitation of local minerals.

## Hierarchical Cluster Analysis

Cluster analysis was applied to combine the wells in the study area into homogenous groups due to their ground water quality. In this study, Ward linkage method with squared Euclidean distance was used to group the studied wells into clusters. Cluster analysis suggests three groups of groundwater (Fig. 2, Tables 3 and 4). According to Fig. 2(a), in post-monsoon season cluster 1 is composed of 8 wells and concerns the largest proportion of the total water sam-

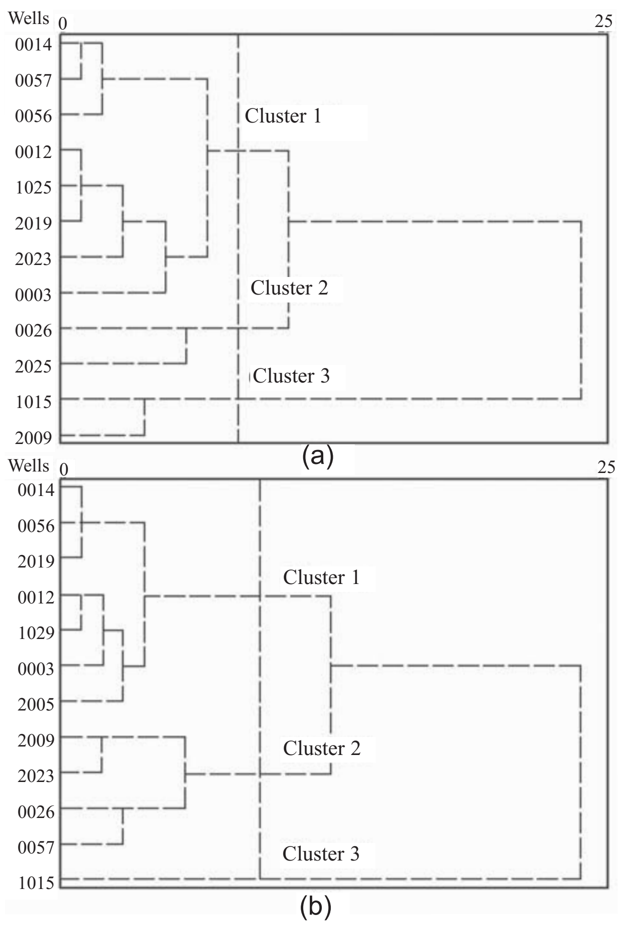


Fig. 2. Dendrogram of the hierarchical cluster analysis of the ground water quality parameters of the wells in the shallow aquifer using Ward method in (a) post-monsoon and (b) summer.

Table 4. Mean values and standard deviation of water quality parameters of water samples in summer in 2011 (wells are representative with the last four numbers).

| Variables Wells | Cluster 1 | | Cluster 2 | | Cluster 3 | |
|---|---|---|---|---|---|---|
| | 0003, 0012, 0014, 0056, 1025, 2005, 2019 | | 0026, 0057, 2009, 2023 | | 1015 | |
| | Mean | SD | Mean | SD | Mean | SD |
| pH | 6.65 | 0.31 | 5.98 | 0.46 | 6.58 | - |
| TH | 161.61 | 73.46 | 321.33 | 81.39 | 783.9 | - |
| TDS | 376.04 | 156.20 | 926.48 | 520.90 | 3689 | - |
| $Ca^{2+}$ | 44.83 | 23.69 | 83.40 | 26.08 | 67.42 | - |
| $Mg^{2+}$ | 11.78 | 7.23 | 27.13 | 12.10 | 149.2 | - |
| $Na^{+}$ | 54.98 | 43.47 | 178.93 | 178.51 | 1025 | - |
| $Cl^{-}$ | 94.28 | 79.01 | 383.39 | 364.65 | 2180 | - |
| $SO_4^{2-}$ | 55.00 | 29.86 | 109.15 | 46.78 | 200 | - |
| $HCO_3^{-}$ | 80.72 | 61.35 | 33.15 | 26.15 | 12.05 | - |
| $NO_3^{-}$ | 44.48 | 65.70 | 86.36 | 125.91 | 0.2 | - |
| Mn | 0.06 | 0.06 | 0.68 | 0.39 | 0.34 | - |

ples, with a value of nearly 66%. Meanwhile, cluster 1 had two sub-clusters. Both cluster 2 and cluster 3 include two wells accounting for about 16.7% of the total samples. Table 3 summarizes the mean values and standard deviations of water quality parameters. One can see that cluster 1 had the highest bicarbonate concentrations compared with other two clusters, with the mean value of 76.79 mg/L. Also, cluster 1 had the lowest salinity represented by the other ion concentrations in the table.

Compared with the quality standard for ground water, these studied wells had good water quality for both industrial production and drinking. Cluster 2 had the highest mean values of $Ca^{2+}$, $NO_3^{-}$, and Mn. It should be noted that the concentration of nitrate exceeded the limit of the quality standard for ground water in China, which had a ceiling of 20 mg/L for drinking. Cluster 3 was considered as suffering most from the salinization since these wells of this cluster owning the maximum of TDS as well as the

Fig. 3. Location distribution of the studied wells with the classification according to the cluster analysis in (a) post-monsoon and (b) summer.

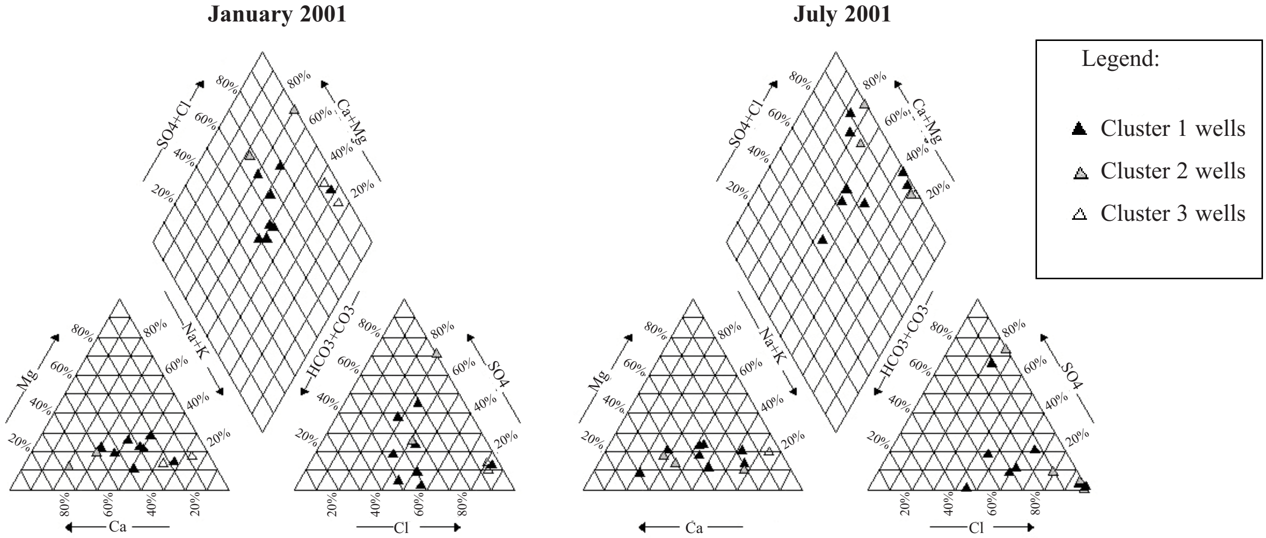


Fig. 4. Piper trilinear diagram of major ions in (a) post-monsoon and (b) summer 2011.

highest concentration of cations like $Na^+$ and $Mg^{2+}$ and the anions like $SO_4^{2-}$ and $Cl^-$. Moreover, the concentrations of TDS, TH, and $Cl^-$ of these wells were definitely more than the threshold values for drinking or direct use. Fig. 3(a) showed the location of the studied wells and their classification based on the cluster analysis.

According to Fig. 2(b) and Table 4, although the clusters had the similar trend of the concentration of each water quality parameter compared with the statistics of the post-monsoon, the distribution of the studied wells had a slight change. Fig. 3(b) showed that the wells (2005, 2009, 0057, and 2023) had been classified differently. It might be thought of as a change caused by spatial and temporal variation. As these observation wells were located on the shallow aquifer in a simple geological unit, the two wells near each other might have a near relation of hydraulic connection, which means the water quality varied accordingly. On the other hand, it can be seen that the water quality of the wells becomes worse from the pre-monsoon to summer, which might be attributed to anthropogenic pollution and mineral dissolution.

Plots of hydrochemical data on Piper [26] trilinear diagram (Fig. 4) show the distribution of chemical constitiuents (cations and anions) in groundwater within the area. From this plot, one can observe that the blended water types exist within the area, for the cations the water samples of Ca-Mg-type and Na-K-type show similar dominance. While for the anion type, $SO_4$-Cl-type of water predominated both in the pre-monsoon and the summer, this phenomenon is more notable (accounting for more than 70% of the total samples) in the summer, which is supposed to result from the decrease of the water table in summer and the subsequent seawater intrusion since the samples are proximate to the sea. Still, the further study should be done to interpret the possible hydrological processes affecting ground water quality in the study area.

## Conclusions

Multivariate statistical techniques including principal component analysis and hierarchical cluster analysis have successfully been used to derive information from the data set about the possible influences of the environment on groundwater quality, and also to identify natural groupings in the set of data. These methods are important to avoid misinterpretation of environmental monitoring data due to uncertainties. This study has employed standard multivariate statistical techniques and traditional graphical methods to identify the factors or sources responsible for water quality variations and hydrochemical characterization in a coastal aquifer in South China. Interpretation of analytical data showed that the abundance of the major ions is as follows: $Na+>Ca^{2+}>Mg^{2+}$ and $Cl^->SO_4^{2-}>HCO_3^-$. PCA produced three factors according to the correlation coefficients and principal component. PCA transforms the hydrochemical variables pH, TH, TDS, $Ca^{2+}$, $Mg^{2+}$, $Na^+$, $Cl^-$, $SO4^{2-}$, $HCO_3^-$, $NO_3^-$, and Mn into three orthogonal principal components, accounting for 90.3% and 83.3% of the total variance in the summer and post-monsoon dataset. Based on similarities in water quality characteristics, HCA grouped the 12 sampling sites into 3 clusters. The multivariate statistical techniques used in this study served as an effective tool for the analysis and interpretation of water samples, identifying water quality factors and understanding temporal and spatial variations in water quality for effective water quality management.

## Acknowledgements

This research was financially supported by National Natural Science Foundation of China (Grant No. 41402202) and Specialized Research Fund for the Doctoral Program of Higher Education (20130061120084).

## References


1. KIM J. H., KIM R. H., LEE J., CHEONG T. J., YUM B. W., CHANG N. W. Multivariate statistical analysis to identify the major factors governing ground water quality in the coastal area of Kimje, South Korea. Hydrological Processes. **19**, (6), 1261, **2005**.
2. LU X.R., ZHOU A.G., WANG M.T., YANG L., LU H. Characteristic analysis of phreatic water equality evolution by Piper diagram in Huaihe drainage area, Jiangsu province. Geotechnical Investigation & Surveying. **2**, 42, **2010**.
3. WANG X.X., WANG W.K., WANG A.F., ZHAO J.L., XIE H.L., WANG X.D. Hydrochemical characteristics and formation mechanism of river water and groundwater along the downstream Luanhe River, northeastern China. Hydrogeology & Engineering Geology. **41**, (1), 25, **2014**.
4. BONANSEA M., LEDESMA C., RODRIGUEA C., PINOTTI L. Water quality assessment using multivariate statistical techniques in Río Tercero Reservoir, Argentina. Hydrology Research, in press, **2014**.
5. PAOPATHEODOROU G., LAMBRAKIS N., PANAGOPOULOS G. Application of multivariate statistical procedures to the hydrochemical study of a coastal aquifer: an example from Crete, Greece. Hydrol. Process. **21**, 1482, **2007**.
6. BELKHIRI L., BOUDOUKHA A., MOUNI L., BAOUZ T. Multivariate statistical characterization of groundwater quality in Ain Azel plain, Algeria. African Journal of Environmental Science and Technology. **4**, (8), 526, **2010**.
7. RAO Y. R. S., KESHARI A. K., GOSAIN A. K. Evaluation of regional groundwater quality using PCA and geostitistics in the urban coastal aquifer, East Coast of India. International Journal of Environment and Waste Management. **5**, (1-2), 163, **2010**.
8. RAO N.S. Spatial control of groundwater contamination, using principal component analysis. Journal of Earth System Science. **123**, (4), 715, **2014**.
9. HAMILTON L.C. Regression with graphics: A second course in applied statistics (Vol. 1, No. 1). Belmont, CA: Duxbury Press, **1992**.
10. LIU C.W., LIN K.H., KUO Y.M. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci. Total Environ. **313**, 77, **2003**.
11. MANIKANDAN K., KANNAN P., SANKAR M. Evaluation and Management of Groundwater in Coastal Regions, Earth Science India, **5**, (I), 1, **2012**.
12. ABDI H. Coefficient of variation. Encyclopedia of Research Design. SAGE Publications, Inc., Thousand Oaks, CA, **169**, **2010**.
13. MOHAPATRA P.K., VIJAY R., PUJARI P.R., SUNDARAY S.K., MOHANYT B.P. Determination of processes affecting groundwater quality in the coastal aquifer beneath Puri city, India: a multivariate statistical approach. Water Sci. Technol. **64**, (4), 809, **2011**.
14. VENKATESH U., VIVEKANAND H., HERANDEZ E.A. Assessment of groundwater water quality in central and southern Gulf Coast aquifer, TX using principal component analysis. Environ Earth Sci., **2013**. DOI: 10.1007/s12665-013-2896-8.
15. HUSSAIN M., AHMED S.M., ABDERRAHMAN W. Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia. J. Environ. Manage. **86**, (1), 297, **2008**.
16. REGHUNATH R., MURTHY T. R., RAGHAVAN B.R. The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka. Water Res. **36**, (10), 2437, **2002**.
17. PANDA U.C., SUNDARAY S. K., RATH P., NAYAK B.B., BHATTA D. Application of factor and cluster analysis for characterization of river and estuarine water systems – A case study: Mahanadi River (India). J. Hydrol. **331**, (3), 434, **2006**.
18. TINA H., NIELS O.J., BRUCE B.Y. Investigation of hydrochemical characteristics of groundwater from the Cretaceous-Eocene limestone aquifer in southern Ghana and southern Togo using hierarchical cluster analysis. Hydrogeol. J. **15**, 977, **2007**.
19. GOLZAR H., SEILIM R., LUTFUN-NESSA., SAMSUDDIN A. Factor and Cluster Analysis of Water Quality Data of the Groundwater Wells of Kushtia, Bangladesh: Implication for Arsenic Enrchment and Mobilizaion. Journal Geological Society of India. **8**, 377, **2013**.
20. JOSHUA O.O., ELISAMA U.T., YINUSA A.A. Application of Sequential Analysis and Geographic Information Systems for Hydrochemical Evolution Survey, Shagari Environ, Southwestern Nigeria. American International Journal of Contemporary Research. **3**, (3), 38, **2013**.
21. Rockware. Inc, Aq.QA© and Prairie City Computing©, Golden, Colorado, USA. Electronic version are available at www.rockware.com.
22. FREEZE R. A., CHERRY J. A. Groundwater. Prentice-Hall INC., Englewood Cliffs NJ, **1979**.
23. KAISER H.F. The varimax criterion for analytical rotation in factor analysis. Psychometrika. **23**, 187, **1958**.
24. MERTLER C. A., VANNATTA R.A. Advanced and Multivariate statistical methods: Practical application and interpretation (3rd ed), Pyrczak, Glendale, CA. pp. 120, **2005**.
25. MASSART D.L., KAUFMAN L. The interpretation of analytical chemical data by the use of cluster analysis, Wiley, New York, **1983**.
26. PIPER A.M. A graphic procedure in the geochemical interpretation of water-analyses. Transactions, American Geophysical Union. **25**, 914, **1944**.