Variation of Groundwater Quality in Seawater Intrusion Area Using Cluster and Multivariate Factor Analysis

Yi-Chu Huang[1]*, Chou-Ping Yang[2], Yu-Chun Lee[1], Ping-Kuei Tang[1], Wen-Ming Hsu[1]
[1]Department of Environmental Science and Engineering
[2]Center for Teaching Excellence
National Pingtung University of Science and Technology
Pingtung, Taiwan
ych@mail.npust.edu.tw

Tien-Nien Wu
Department of Environmental Engineering
Kun Shan University of Technology
Tainan County, Taiwan

***Abstract*- Numerous monitoring data and water quality index obtained from 76 regional shallow-layer monitoring wells are simplified and classified by applying the multivariate statistical methods such as factor and cluster analysis to search for the interrelation between the water quality parameters, factors representing the characteristics as well as possible pollution sources of groundwater quality in Pingtung Champaign, Taiwan. Four principal factors were recognized in concerned area using principal component analysis (PCA) including salinization factor, mineralization factor, inorganic factor, and inorganic reduction factor. All four factors can interpret 81.0% variances of the integrated groundwater characteristics. In addition four clusters were classified according to the similar and dissimilar characteristics of water quality of monitoring wells in Pingtung Champaign. The results showed that the groundwater quality of hinterland was better than that of coastal area. Some coastal areas have already been affected by the seawater intrusion resulting in aquifer salinization. Multivariate statistical methods provided by this study can not only reduce the harassment of the missing items of monitoring water quality, but also refer as a management alternative for groundwater resources.**

Keywords: seawater intrusion; salinization; cluster analysis, principal component analysis

I. INTRODUCTION

Either naturally occurring processes or human activities may have a significant impact on the water quality of aquifer that may limit its use. Although in Pingtung plain groundwater resource is abundant, the unregulated exploitation of groundwater resources would result in undesirable adverse impacts such as land subsidence, saline water intrusion, and aquifer salination, especially the groundwater over-pumping problems for aquaculture along the costal regions of study area. Several indexes and their correlation are necessary to determine the regional water quality characteristics of aquifer. The multivariate statistical analysis is utilized to process the problems of multi-factors and multi-characteristics. It has been applied to study the interrelation and correlation of variables in the fields of environmental science, hydrology, geology, etc. [1-7].

The objective of this study is to utilize the factor and cluster analysis to understand the factors affecting the groundwater quality and discriminate their influence area. Multivariate statistical methods provided by this study can not only reduce the harassment of the missing items of monitoring water quality, but also refer as a management alternative for the groundwater resources.

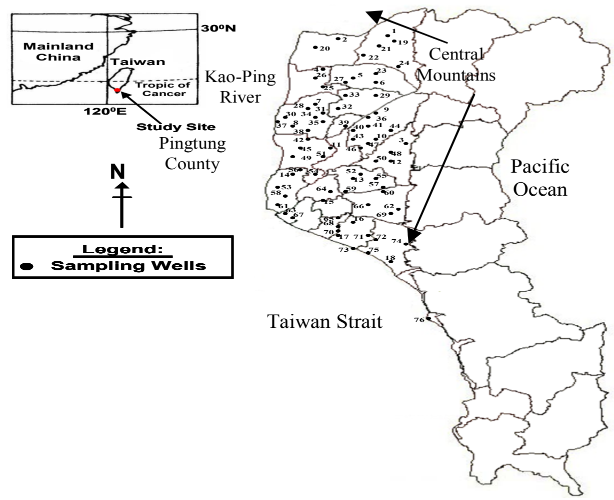

Figure 1. Study area with the location of sampling wells.

II. MATERIALS AND METHODS

A. Study Area

Pingtung County is a strip-shaped region that belongs to the Pingtung Champaign groundwater sub-regions, one of the major groundwater resources in Taiwan. The area of Pingtung County situated in the tropic region of southern Taiwan is approximately 2,775 km^2 and extends 112 km from north to south and 47 km from east to west (Fig. 1). The area is enclosed by the Kao-Ping River to the west, the central mountains to the north and east, and the Taiwan Strait to the south. The Kao-Ping, Tung-Kung, and Lin-Bian are the three major rivers that flow through the area. Annual precipitation concentrates from May to September and is around 2,700 mm at average.

Agriculture and aquaculture are the primary sources of revenue in Pingtung area. To boost earnings, many farmers have transformed their croplands into aquafarms. Over the past 30 years, a large quantity of groundwater has been overly extracted from the shallow layer aquifer to mix with seawater to supply those aquafarms. The over-pumping behaviors have caused serious land subsidence, seawater intrusion, flooding and deterioration of the surrounding environment [4]. The most serious land subsidence and seawater intrusion area occurred at the Donggang, Linbian, Jiadung, and Fangliau townships of Pingtung County. The most recent case, the rainfall of the typhoon Morakot on August 8 of 2009 reached 2,000 mm in two days resulting in serious disasters including stream swollen, instantaneous inundation and seawater flooding in Pingtung region. Vertical infiltration from the aquafarms and lateral intrusion of seawater are the main causes of local shallow groundwater salinization along the coastal area of Pingtung.

Regional groundwater recharges are mainly from the infiltration of the western Kao-Ping River as well as the northern and eastern mountains of the Pingtung Plain. The hydro-geological stratification from top to bottom can be divided into four aquifers and three aquitards within depth of 220 m in the groundwater sub-regions of Pingtung Champaign. In general, the thickness of aquifers is large and extends over the whole region. Aquitard is sandwiched between the aquifers within 220 m below the ground. Monitoring wells whose permeability coefficients ranging from 1.0×10^{-4} to 9.9×10^{-4} m s^{-1}, from 1.0×10^{-5} to 9.9×10^{-5} m s^{-1}, and greater than 1.0×10^{-3} m s^{-1} are classified to good-class (68.5% of monitoring wells), medium (14.2%), and excellent level permeability (13.4%), respectively. The thickness of the first aquifer covering the entire region ranges from 23.5 to 83.5 m with an average thickness of about 49.9 m. In this study, the periodically sampled groundwater quality data are mainly from the first aquifer.

B. Groundwater Data

There are 76 monitoring wells set up on the shallow aquifer and periodically surveyed since 2003 (Fig. 1). These wells with depth ranging from 8 to 30 m were designed to monitor the groundwater quality with an emphasis on aquifer salinization and potential groundwater contamination in the concerned region. From 2003 to 2008, groundwater quality including total hardness (TH), total dissolved solids (TDS), chloride (Cl^-), sulfate (SO_4^{2-}), total organic carbon (TOC), nitrate nitrogen (NO_3-N), nitrite nitrogen (NO_2-N), ammonia nitrogen (NH_3-N), iron (Fe) and manganese (Mn) were selected as their more complete geochemical data [8].

C. Factor Analysis

The multivariate statistical process of environmental data is widely used to characterize and assess the surface water and groundwater quality, and it is useful for evidencing temporal and spatial variations caused by natural and human factors linked to seasonality [4, 9-10]. With the aid of multivariate statistical techniques, the groundwater data can be simplified, organized and classified to bring about useful meanings [5].

Factor analysis such as principal component analysis (PCA) is known as a powerful technique for reduction of variables dimension by providing the correlation among measured chemical variables and their multivariate patterns based upon eigenanalysis of the correlation or covariance matrix that may be help to classify the original large sets of data [9-10]. The geochemical interpretation of determined factors gives insights into the dominant processes, which may command the distribution of hydro-chemical variables [2].

As the measurement scales and numerical ranges of the original variables vary widely, all raw data should be first standardized by Z-score mode to exclude the effects of different units of measurement, and make the data dimensionless. Standardization is to make each variable within the original data matrix subtract the column mean and then is divided by the column standard deviation. The variances/co-variances and correlation coefficients measures how well the variance of each constituent can be explained by relationships with each of the others. Then, eigenvalues and eigenvectors were calculated for the covariance matrix and the data were also transformed into factors [2, 4-5].

PCA is based on the diagonalization of the correlation matrix that can give the overall coherence of the data set. By utilizing PCA, the original p-dimensional standardized data matrix is transformed into an m-dimensional PC matrix with less degree of freedom, i.e. $p<m$. Considering the correlations present in the original data, PCs can reduce the overall complexity of the data and still reserve inherent inter-dependencies. Only factors with eigenvalues greater than 1 are preserved for further processed [11]. Generally speaking, the first few PCs explain the majority of the variance within the original dataset, then the first one accounts for the most variance and each subsequent PC explains progressively less. The factor loadings are responsible for the correlations between PCs and selected variables, and those with the greatest positive and negative loadings make the largest contribution. As a result, the loadings can offer more information to track the sources that are responsible for the similarities of collected samples in groundwater quality [2].

D. Cluster Analysis

Cluster analysis (CA) is a statistical tool to classify the true groups of data according to their similarities to each other. All variables were also standardized by Z-score mode before being subjected to CA. A short Euclidean distant implies the high similarity between the measured objects. In clustering, the distinct groups can reveal either the interaction among the variables (R-mode) or the interrelation among the samples (Q-mode) [4-5]. Two types of CA methods, hierarchical cluster analysis and nonhierarchical cluster analysis, have been performed in two-step procedures in this study. Ward's clustering procedure commonly used in hierarchical method of cluster analysis was employed to identify the number of clusters. Next, the nonhierarchical cluster analysis was utilized to obtain the correct classified observations by K-means method. Several previous works have successfully demonstrated the application of two-step clustering procedures to group monitoring wells according to the identified mechanisms affecting groundwater quality in southwest coast area of Taiwan [2, 4-5]. A multivariate analysis including PCA and CA was conducted by the computer package named Statistical Product and Service Solutions (SPSS-12.0) in this study. In addition, a comprehensive description of the multivariate analysis method and its application can be found in textbook on statistics [12-13].

Ⅲ. RESULTS AND DISSCUSSION

A. Factor Analysis

PCA is employed to both variables and sampling groundwater data collected from 76 wells in Pingtung Champaign from 2003 to 2008. The Kaiser-Meyer-Oklin (KMO) test conducted on the correlation matrix shows a calculated value KMO = 0.74 greater than the acceptable value 0.6, thus meaning that PCA can successfully reduce the dimensionality of the original data set. The Bartlett's sphericity test with a calculated $\chi 2$ = 979.29 (P = 0.0 <0.01) implies the concerned correlation matrix are suitable for further factor analysis.

The rotated factor loading, eigenvalues, percentages of variance, and cumulative percentages of variance associated with each factor for PCA are summarized in Table 1. The highlighted variables have their absolute values of loadings greater than 0.7 for the reason that it is an indicator of the participation of the variables in each PC, i.e. they represent the most important information on which the interpretation of factors is based. Four factors with respective eigenvalues greater than unity were identified which account for 80.98% of the total variance in the original data set.

In Table 1, Factor 1 shows that seawater intrusion affecting the studied aquifer was identified, and it is reasonable to observe a strong positive correlation of TDS with SO_4^{2-}, Cl^-, and TH. Factor 1 explains for 43.28% of the total variance and is characterized by very strong positive loading. Even though calcium and magnesium were not measured in this study, these cations were expected to be present for the electro-neutrality of saline water and the TH, thus the presence of these cations and anions in groundwater results in the enhancement of TDS. In addition, TDS represented the abundant mineral elements in the groundwater and TH increases with TDS. Therefore, Factor 1 is accordingly defined as the salinization factor and Cl^- serves as an indicator of seawater intrusion in the studied aquifer. According to Huang's (2005) research showed that the over-pumping had led to the deterioration of the groundwater due to seawater intrusion [4]. In the coastal area of Pingtung plain, the shallow aquifer has suffered serious salinization where the concentrations of SO_4^{2-}, Cl^-, TDS, and TH far exceeding the agricultural water quality standards regulated by Taiwan because farmers have overly extracted groundwater from the aquifer under their aquafarms.

Factor 2 accounts for 15.94% of the total variance and is mainly associated with very high loadings of TOC and Fe (Table 1). TOC is attributed to the infiltration from improper handle of pasturage and municipal wastewater. High levels of Fe and Mn in the groundwater are closely related to the dissolution processes of the geological features and formation in the study area. Therefore Factor 2 is characterized as the organic mineral factor. Fig. 2 demonstrates the PC score of each sampling data for PC 1 and PC 2.

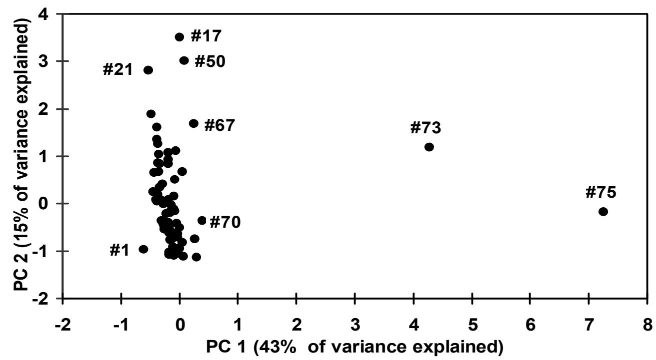

Figure 2. PC score of each sampling well for PC 1 and PC 2.

Factor 3 and Factor 4 are related to 11.62% and 10.15% of the total variance of NO_3-N and NO_2-N, respectively (Table 1). The single dominant variable, NO_3-N and NO_2-N, presented in respective Factor 3 and Factor 4 is most likely contributed to the leakage of pasturage and industrial wastewater. NH_3-N can be transformed into NO_2-N and then NO_3-N during a long period of time, so NO_3-N is treated as a long-term indicator of water polluted by municipal wastewater. Municipal wastewater is commonly treated through septic tanks in which NH_3-N is bio-transformed to NO_3-N through nitrification. Hence, Factor 3 and Factor 4 are characterized as the inorganic factor and inorganic-reduction factor.

Table 1. Variance of rotated R-mode factor loading matrix in PCA

Variables	Factor 1	Factor 2	Factor 3	Factor 4
SO_4^{2-}	**0.989**	0.052	-0.035	0.019
Cl^-	**0.988**	0.082	0.031	-0.023
TDS	**0.971**	0.112	0.087	-0.011
TH	**0.986**	0.106	0.047	0.009
TOC	-0.100	**0.827**	-0.026	-0.105
Fe	0.198	**0.760**	-0.076	0.334
NO_3-N	0.029	-0.083	**0.874**	0.021
NO_2-N	-0.021	0.015	-0.002	**0.968**
NH_3-N	0.271	**0.551**	0.479	-0.171
Mn	0.361	0.413	-0.310	-0.034
Eigenvalue	4.328	1.593	1.162	1.015
Variance (%)	43.28	15.93	11.62	10.15
Cumulative variance (%)	43.28	59.21	70.83	80.98

B. Cluster Analysis

As the relative compositions of the constituents in a sample are often as important as their absolute concentrations, CA is utilized to classify the similarity among samples. The hierarchical method is consuetudinary CA, which has been successfully demonstrated in hydro-geochemical studies [4-5, 14]. With Ward's clustering procedure, the data set is first categorized into four groups in a dendrogram. Following hierarchical CA, the K-means method is utilized to reclassify the data set on the basis of the similarity between clusters. The output of nonhierarchical CA with the chemical data for each cluster is given in Table 2.

There are 47 monitoring wells identified within Cluster 1, which has the lowest loading scores with respect to TDS, TH, TOC, NH_3-N, SO_4^{2-}, Cl^-, Fe, and Mn (Table 2). The groundwater quality in cluster 1 is better as compared with the other clusters. The recognized 26 sampling wells located on the inland of the study area in cluster 2 with high values of TOC and TH implies the aquifers are affected by industrial and municipal wastewaters (Table 2). There are only one and two monitoring wells identified in Cluster 3 and Cluster 4, respectively. Those wells located on the coastal region of the Pingtung plain resulting in noticeable seawater intrusion phenomenon. In addition, over-pumping may be one of factors led to the deterioration of the groundwater quality. Therefore, high level of TDS, SO_4^{2-}, Cl^-, Fe, and Mn are observed in cluster 3 and 4 whose groundwater quality was the worst in the study region. The relative position of each identified cluster is shown in Fig. 3.

Table 2. Chemical properties of each cluster classified by the K-Mean method

Variable	Unit	Cluster 1	Cluster 2	Cluster 3	Cluster 4
TDS	mg/L (mean)	501.8	708.94	4180	23187.5
	mg/L (range)	51.5~2110	172~2950	4180	21900~24475
TH	mg/L (mean)	337.9	420.1	1035.5	5580
	mg/L (range)	63.1~791.5	119.5~1102	1035.5	4095~7065
TOC	mg/L (mean)	0.52	1.64	0.55	1.08
	mg/L (range)	0~2	0.6~3.95	0.55	0.5~1.65
NH_3-N	mg/L (mean)	0.065	0.395	0.79	0.795
	mg/L (range)	0~0.28	0~2.4	0.79	0.71~0.88
NO_3-N	mg/L (mean)	4.39	2.2	236.14	1.11
	mg/L (range)	0~11.85	0~11.7	236.14	0.21~2.01
NO_2-N	mg/L (mean)	0.006	0.0098	0	0
	mg/L (range)	0~0.08	0~0.255	0	0
Cl^-	mg/L (mean)	36.7	140.9	1300	12425
	mg/L (range)	2.22~627	5.11~1557	1300	9000~15850
SO_4^{2-}	mg/L (mean)	74.4	84.44	175	1795
	mg/L (range)	15.2~225	3.33~243	175	1305~2285
Fe^{2+}	mg/L (mean)	0.8	5.45	1.16	7.69
	mg/L (range)	0~3.46	0.075~11.85	1.16	6.04~9.34
Mn^{2+}	mg/L (mean)	0.38	1	0.67	3.15
	mg/L (range)	0~2.29	0~11.06	0.67	1.62~4.68
Number of Sampling Wells		1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 23, 24, 25, 27, 28, 29, 30, 31, 34, 35, 37, 42, 43, 44, 45, 46, 49, 52, 53, 55, 56, 57, 58, 59, 60, 61, 69, 74, 76	4, 17, 21, 22, 26, 32, 33, 36, 38, 39, 40, 41, 47, 48, 50, 51, 54, 62, 63, 64, 65, 66, 67, 68, 71, 72	70	73, 75

IV. CONCLUSIONS

This study has successfully demonstrated the utility of multivariate statistical tools including factor and cluster analysis to characterize the groundwater quality of Pingtung Champaign from 2003 to 2008. Four principal factors were recognized using PCA. The first salinization factor accounts for 43.28% of the total variance and includes the variables of SO_4^{2-}, Cl^-, TDS, and TH. The second organic mineral factor specifies 43.28% and contains the variables of TOC and Fe. The third inorganic factor comprises the variable of NO_3-N. The fourth inorganic reduction factor encloses NO_2-N. All four factors can interpret 80.96% variances of the integrated groundwater characteristics. Four clusters of monitoring wells were classified according to the similar and dissimilar characteristics of groundwater quality in Pingtung Champaign. The results showed that the groundwater quality of hinterland was better than that of coastal area. Some coastal areas have already been affected by the seawater intrusion resulting in aquifer salinization. With the aid of statistical techniques, it is predictable to be aware of the underlying processes and the distribution of sources that might affect the concerned groundwater quality. Furthermore, it can offer the requisite information for the authority to pursue the sustainable approaches on groundwater management and contamination prevention. In the future, this study will utilize the artificial neural network (ANN) models coupling with the results of multivariate statistical methods to predict the groundwater quality and further to understand the characteristics of groundwater quality in the Pingtung Champaign.

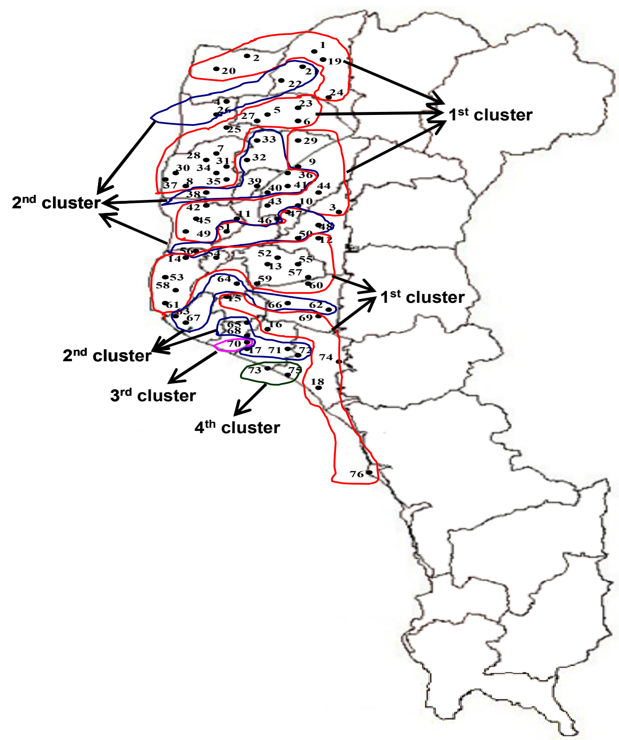

Figure 3. Relative position of each identified cluster.

REFERENCES

[1] M. M. Morales, P. Marti, A. Llopis, L. Campos, and S. Sagrado, “An environmental study by factor analysis of surface seawaters in the Gulf of Valencia (Western Mediterranean),” Analytica Chimica Acta, Vol. 394, No. 1, pp. 109-117, 1999.

[2] C. W., Liu, K. H. Lin, and Y. M. Kuo, “Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan,” The Science of the Total Environment, Vol. 313, No. 1-3, pp. 77-89, 2003.

[3] K. P. Singh, A. Malik, D. Mohan, and S. Sinha, “Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)- a case study,” Water Research, Vol. 38 No. 18, pp. 3980-3992, 2004.

[4] Y. C. Huang, T. N. Wu, and P. J. Cheng, “Characterization of groundwater quality by multivariate statistical analysis: an example from Kaohsiung County, Taiwan.” In Water and Wastewater

Management for Developing Countries, eds K Mathew and I Nhapi, Water and Environment Management Series, IWA Publishing, London, 2005.
[5] T. N. Wu, Y. C. Huang, M. S. Lee, and C. M. Kao, "Source identification of groundwater pollution with the aid of multivariate statistical analysis," Water Science and Technology: Water Supply, Vol. 5, No. 6, pp. 281-288, 2005.
[6] K. Chen, J. J. Jiao, J. Huang, and R. Huang, "Multivariate statistical evaluation of trace elements in groundwater in a coastal area in Shenzhen, China," Environment Pollution, Vol. 147, pp. 771-780, 2007.
[7] V. A. Tsihrintzis, C. S. Akratos, and J. N. E. Papaspyros, "An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands," Chemical Engineering Journal, Vol. 143, pp. 96-110, 2008.
[8] Pingtung County Environmental Protection Bureau, Soil and groundwater database, Website: http://163.29.224.196/soilandgw/, 2009.
[9] B. Helena, R. Pardo, M. Vega, E. Barrado, J. M Fernandez, and L. Fernandez, "Temporal evolution of groundwater composition in an alluvial aquifer Pisuerga River, Spain by principal component analysis," Water Research, Vol. 34, No. 3, pp. 807-816, 2000.
[10] H. F. Kaiser, "The varimax criteria for analytical rotation in factor analysis," Psychometrika, Vol. 23, No. 3, pp. 187-200, 1958.
[11] R. A. Reyment and K. H. Joreskog, Applied Factor Analysis in the Natural Science, Cambridge University Press, p. 371, 1993, New York, NY, USA.
[12] J. C. Davis, Statistics and Data Analysis in Geology, 2nd ed., John Wiley and Sons, p. 656, 1987, New York, NY, USA.
[13] J. E. Jackson, A User's Guide to Principal Components. John Wiley and Sons, 1991, New York, NY, USA.
[14] R. Reghunath, M. T. R. Sreedhara, and B. R. Raghavan, "The utility of multivariate statistical techniques in hydro-geochemical studies: an example from Karnataka, India," Water Research, Vol. 36, pp. 2437-2442, 2002.