ISSN: 2447-3359

REVISTA DE GEOCIÊNCIAS DO NORDESTE

Northeast Geosciences Journal

v. 11, nº 1 (2025)

https://doi.org/10.21680/2447-3359.2025v11n1ID35605

Use of Multivariate Statistical Analysis in the Characterization of Hydrogeochemical Heterogeneity of Aquifer Systems, a Methodological Review

Uso da Análise Estatística Multivariada na Caracterização de Heterogeneidades Hidrogeoquímicas dos Sistemas Aquíferos, uma Revisão Metodológica

Deize Elle Ribeiro Moitinho[1]; Natanael da Silva Barbosa[2]; Maria da Conceição Rabelo Gomes[3]; Sizenando Bispo-Silva [4]; Cristovaldo Bispo dos Santos[5]; Ludmila Amorim Gomes [6]

[1] Institute of Geosciences, Federal University of Bahia (UFBA), Salvador, BA, Brazil. Bahia Water and Sanitation Engineering Company (CERB), Salvador, BA, Brazil. E-mail: deize.moitinho@cerb.ba.gov.br
ORCID: https://orcid.org/0000-0002-7061-6914

[2] Institute of Geosciences, Federal University of Bahia (UFBA), Brazil. E-mail: ndbarbosa@ufba.br
ORCID: https://orcid.org/0000-0003-0923-6266

[3] Pará State University (UEPA), Brazil. E-mail: conceicaorabelo@yahoo.com.br
ORCID: https://orcid.org/0000-0001-7841-4201

[4] Petrobrás, Rio de Janeiro, Brazil. E-mail: sizenando@petrobras.com.br
ORCID: https://orcid.org/0000-0001-6049-7774

[5] Institute of Geosciences, Federal University of Bahia (UFBA), Salvador, BA, Brazil. E-mail: bispo@ufba.br
ORCID: https://orcid.org/0000-0002-1448-3643

[6] Institute of Geosciences, Federal University of Bahia (UFBA), Salvador, BA, Brazil. E-mail: ludimilla.amorim@ufba.br
ORCID: https://orcid.org/0000-0002-1651-3738

Abstract: There is a growing concern regarding underground water resources, driving the development of more refined methods to assess changes in these resources, whether they arise from natural processes or direct and indirect interventions in the environment, of a physical or chemical nature. This study addresses the main techniques of the Multivariate Statistical Analysis (MSA) method applied to this purpose. MSA has been widely employed in characterizing aquifers and in studies related to contamination from mining activities, urban or rural interventions (agricultural and livestock), and in identifying processes of marine intrusion resulting in salinization. Additionally, this approach was used to segment the territory into areas with different qualities of groundwater. Forty articles were analyzed, with 80% of these originating from countries in Asia, Africa, and North America. During this evaluation, it was also possible to examine the relative applicability and complementarity of Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) to achieve the scientific objective, i.e., the investigation of natural and/or anthropogenic processes in groundwater. Both methods, when compared to conventional geochemical grouping or other multivariate statistical techniques, stood out as the most suitable and common in characterizing aquifer heterogeneities. We conclude that the applicability of this statistical approach is universal, as the employed techniques are independent of lithology type, as they are linked to the hydrogeochemical nuances addressed in the studies.Based on articles published between 1965 and 2022, it was possible to recognize in MSA tools a crucial instrument in the management and development of underground water resources, addressing both current and future demands..

Keywords: Multivariate Statistical Analysis; Water Resources; Methodological Review.

Resumo: Cresce a preocupação em relação aos recursos hídricos subterrâneos, impulsionando o desenvolvimento de métodos mais refinados para avaliar as mudanças nesses recursos, sejam elas decorrentes de processos naturais, intervenções diretas ou indiretas no meio, de natureza física ou química. Este estudo aborda as principais técnicas do método de Análise Estatística Multivariada (AEM) aplicadas a esse propósito. A AEM tem sido amplamente empregada na caracterização de aquíferos e em estudos relacionados à contaminação proveniente de atividades mineiras, intervenções urbanas ou rurais (agropecuárias) e na identificação de processos de intrusão marinha resultando em salinização. Além disso, essa abordagem foi utilizada para segmentar o território em áreas com diferentes qualidades das águas subterrâneas. Foram analisados 40 artigos, sendo 80% destes provenientes de países da Ásia, África e América do Norte. Durante essa avaliação, foi possível também examinar a aplicabilidade relativa e a complementaridade da Análise de Cluster Hierárquica (HCA) e da Análise de Componentes Principais (PCA) para atingir o objetivo científico, ou seja, a investigação de processos naturais e/ou antrópicos nas águas subterrâneas. Ambos os métodos, quando comparados ao agrupamento geoquímico convencional ou a outras técnicas da estatística multivariada, destacaram-se como os mais adequados e comuns na caracterização das heterogeneidades dos aquíferos. Concluímos que a aplicabilidade dessa abordagem estatística é universal, pois as técnicas empregadas independem do tipo de litologia, uma vez que estão vinculadas às nuances hidrogeoquímicas abordadas nos estudos. Com base nos artigos publicados entre 1965 e 2022, foi possível reconhecer nas ferramentas da AEM um instrumento crucial na gestão e desenvolvimento dos recursos hídricos subterrâneos, atendendo tanto às demandas presentes quanto às futuras.

Palavras-chave: Análise Estatística Multivariada; Recursos Hídricos; Revisão Bibliográfica.

Recebido: 13/03/2024; Aceito: 21/02/2025; Publicado: 28/05/2025.

1. Introduction

Underground water resources are among the main natural elements exploited by humanity. The sustainable management of these resources emerges as a crucial issue for all countries, especially developing ones (Zghibi et al., 2014; Rakotondrabe et al., 2018). Globally, 97.5% of water is saline, distributed in oceans, leaving only 2.5% as freshwater. Of this amount, 68.9% is frozen, leaving 31% as liquid freshwater, with 4% coming from surface sources, while the largest portion is composed of groundwater (Revenga et al., 2000; MMA-SRHU, 2007). The multiple uses of this water serve various purposes: economic (industry, agriculture, mining), energy generation, and sustenance of living organisms.

Aquifers stand out partly due to their lower cost compared to surface waters, which are more vulnerable to pollution and depletion. Unlike surface waters, underground reservoirs offer perennial flow, purity, and protection, reinforcing the need for effective management policies to ensure their sustainable use (Wanda et al., 2011; Narany et al., 2014; Ahmed et al., 2017). Consequently, countries must collect representative data to monitor and anticipate changes in these resources.

Numerous recent hydrogeochemical studies corroborate that salinity is among the main causes of water quality degradation in coastal aquifers subjected to arid climates (Zghibi et al., 2014; Slama et al., 2017; Ahmed et al., 2018; Sae-Ju et al., 2019). Variations in groundwater quality frequently result from the regional and geological configuration of the subsurface context (Melloul & Collin 1992; Shin et al., 2020).

In recent decades, national capacities for hydrogeological monitoring and quality data production have declined significantly, impacting decision-making (Keita & Zhonghua et al., 2017). Addressing this issue requires a deeper understanding of aquifer heterogeneity, influenced by lithological-structural factors (fissured, karstic, or porous) and environmental elements such as climate, recharge areas, evaporation, and rock-water interactions. Additionally, assessing land use, identifying contamination from human activities, and analyzing past and future geochemical characteristics are crucial (Narany et al., 2014; Ravikumar et al., 2015; Kumar Das et al., 2022).

Recently, hydrologists have increasingly relied on mathematical and statistical techniques to reduce numerous variables to a manageable set of uncorrelated dimensions (Seyan et al., 1985; Farnham et al., 2000). Given the vast number of influencing factors, analyzing them all is impractical. However, Multivariate Statistical Analysis (MSA) provides a powerful tool for simultaneously processing multiple variables, helping identify the most relevant elements for further study.

The pioneer in applying MSA in hydrology was Wallis in 1965 in a comparative work between common techniques of the time such as Regression Analysis contrasted with the application of MSA using Principal Component Analysis (PCA) and Cluster analysis with Varimax rotation. Two years later, Dawdy & Feth (1967) performed the first application on hydrogeochemical data adopting the method in a study on groundwater quality in Mojave, California.

Since then, this tool has seen increasing use across various fields. As state-of-the-art research provides an efficient tool for understanding what has been studied within a specific field (Roma Romanowiski & Ens, 2006), this work corresponds to a review based on global-scale research covering the last five decades. The objective is to reach the main theoretical references cited in the literature on multivariate statistical analysis to understand how MSA is applied and what are the main tools for studying geochemical heterogeneities (i.e., natural evolution of aquifers, salinization processes, anthropogenic contamination, contamination from mineral deposits related to mining activities or natural background) of aquifer systems.

2. Methodology

2.1 Methodology of Principal Component Analysis (PCA)

The methodology of Principal Component Analysis (PCA) is a specific case of the broader method of factor analysis. PCA focuses on constructing new variables called principal components from an existing set of original variables (Shin et al., 2020; Odat et al., 2015; Dawdy & Feth, 1967). These new elements are derived through a linear combination of the original variables.

The application of PCA aims at reducing dimensions across multiple parameters by categorizing those most relevant that explain the variance of the phenomenon under study. When interpreted, these components reveal underlying data structures. The basic approach consists of grouping dimensions based on correlation—i.e., gathering highly correlated variables while those with low correlation are separated into distinct classes (Güler et al., 2002; PingHeng et al., 2010; Sae-Ju et al., 2019). The primary goal is to describe raw data using the sum of one special factor in a linear function with as few common factors as possible thus simplifying dimensions.

With principal components in hand, analyses are conducted in Q mode (when investigating similarities between individuals or objects), considering all measured variables simultaneously for each individual and R mode (exploring similarities between variables). The technique is based on analyzing eigenvalues from a correlation or covariance matrix with results expressed in tables/matrices (Table 1) (Melloul & Collin, 1992; Güler et al., 2012; Shin et al., 2020).

Table 1 – Tablet with Main Components.

Variant	**Main Component**			
	PC1	**PC2**	**PC3**	**PC4**
pH	-0,182	0,036	**0,952**	0,045
EC	**0,972**	0,099	-0,144	-0,027
TDS	**0,972**	0,099	-0,144	-0,027
TH	**0,991**	0,045	-0,026	-0,027
CaH	**0,967**	0,076	0,080	-0,066
TA	**0,994**	0,004	-0,044	-0,042
Ca	**0,967**	0,076	0,256	-0.197
Mg	**0,970**	0,015	-0,123	0,010
Na	0,132	0,762	0,256	-0.197
K	**0,937**	0,044	-0,141	-0,007
HCO_3	**0,994**	0,004	-0,044	-0,042
F	-0,035	**0,659**	-0,114	0,210
Cl	**0,981**	-0,007	-0,119	-0,066
SO_4	**0,962**	0,118	-0,068	0,003
PO_4	-0,055	-0,021	0,044	**0,967**
NO_3	0,076	**0,748**	-0,023	-0,066
Eigen value	10,49	1,63	1,10	1,04
Proportion%	65,55	10,17	6,88	6,52
Cumulative %	65,55	75,71	82,59	89,11

Source: Ravikumar & Somashekar (2015).

Ravikumar & Somashekar (2017) employed Principal Component Analysis (PCA) to estimate aquifer quality in the Varahi River Basin, Karnataka, India. PCA was chosen to manage elements that vary across different orders of magnitude. The analysis focused on the correlation matrix, normalizing each variable to unit variance to ensure equal contribution. Results were graphically represented based on the principal component table (Table 1) and illustrated in Figure 1.

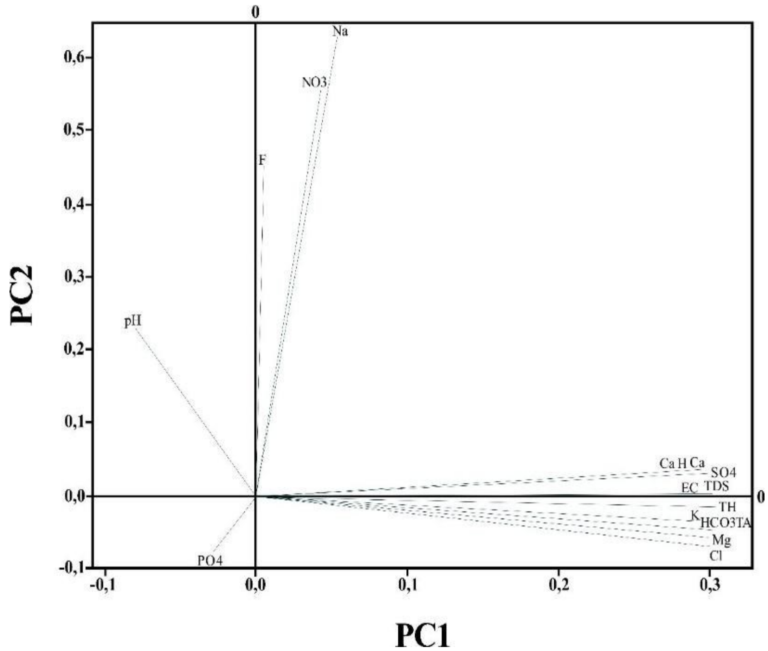

Figure 1 – Graphical representation of main components displayed e Table 1.
Source: Modified from Ravikumar & Somashekar (2015).

The summarized information from the PCA graph is crucial for identifying relationships patterns similarities extreme chemical composition enabling subsequent statistical modeling. Analyzing the loading graph (Fig.2) for the first two PCs (PC1 and PC2), all physicochemical parameters are observed distributed across upper right and lower right quadrants respectively. Lines connecting variables passing through origin on factor loading graphs indicate each variable's relative contribution. This visual representation is crucial for a deeper understanding of data obtained during aquifer quality assessment. This section summarizes methodology used clarifying PCA's application process within aquifer quality analysis preparing for subsequent comprehension regarding results obtained through this method.

2.2 Methodology - Factor Analysis (FA)

In factor analysis, factors are rotated using Kaiser's varimax rotation to enhance interpretability. Proposed by Kaiser (1958), this method emphasizes significant factor loadings essential for analysis. Data preparation involves standardizing raw data, eliminating unit differences, and calculating the covariance matrix (Farnham et al., 2003; Che et al., 2021; Tessema et al., 2012). Eigenvalues and the variance proportions associated with each factor are then computed.

The number of factors to be extracted is determined by summing proportions and calculating cumulative eigenvalues. Various methods, such as the Scree Plot and Kaiser's rule, guide this decision, with most studies utilizing one or both approaches.

2.2.1 *Scree Plot*

The Scree Plot is an effective tool for identifying relevant factors, highlighting abrupt changes in eigenvalues that indicate a significant change from a steep curve to a more gradual descent. In Figure 3, this pattern shows a notable slope change after the first two factors. Eigenvalues decrease and fall below one between factors four and five, suggesting that a four-component solution may be the most appropriate, capturing the best range of variation (Ravikumar & Somashekar, 2015; Yang et al., 2015).

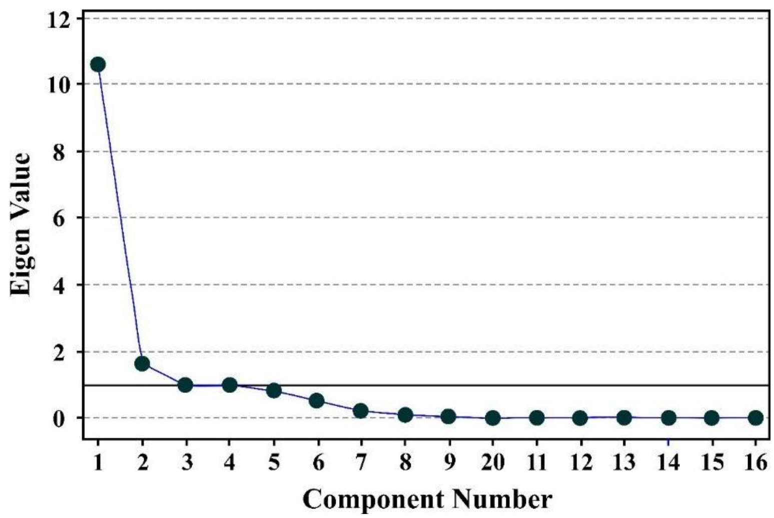

Figure 2 – Scree Plot of Eigenvalues, used to define the amount of components, most representative components highlighted by the solid line.
Source: Adapted from Ravikumar & Somashekar (2015).

2.2.2 *Kaiser's Rule*

In factor analysis, the number of factors selected is based on the elements that explain the most variance, typically around seventy percent of the dataset's variance. Varimax rotation is applied using eigenvalues greater than one, making the results interpretable often guided by Kaiser's criterion (1958). Factor loadings are categorized as "strong" (>0.75), "moderate" (0.75–0.50), and "weak" (0.50–0.30) (Liu et al., 2003). The classification of factor loadings is based on the correlation matrix of standardized variables, aiming to simplify interpretation and maximize the relationship between them by applying Kaiser's Varimax rotation (1958). In this scheme, the first component (F1) typically reflects the most significant variations, while the last component explains less of the data (Usunoff & Guzman, 1989; Helena, 2000).

2.3 Methodology – Cluster Analysis

Cluster analysis is a statistical grouping method divided into hierarchical non-hierarchical categories. Hierarchical algorithms generate dendrograms representing similarity among samples while non-hierarchical algorithms directly form groups. Hierarchical method widely used groups samples based similarities illustrating overall variable similarity within dataset often employing Euclidean distance (Davis ,1986). Among various techniques, the average linkage method and Ward's method are prominent (Seyhan & Keet, 1981).

Hierarchical Cluster Analysis (HCA) is distinguished as the primary method for identifying homogeneous sets based measured characteristics. It begins each case as separate cluster merging sequentially until only one group remains grounded theory similar parameters originate same source (Bhuiyan,et.al ,2016). This technique widely applied Earth sciences (Davis ,1986) utilizes similarity levels constructing dendrogram (Güler et.al ,2002; Cloutier et.al ,2008 ; Güler et.al ,2012 ; Moya et.al , 2015).

Hierarchical clustering calculates the similarity between pairs of objects by measuring the distances between points in a multidimensional space, where smaller distances indicate greater similarity, and larger distances indicate less similarity (Tessema et al., 2012). In Figure 4, the smallest distances highlight similarities among samples grouped by their origins, such as Cluster A (low pollution), which is subdivided into A-1 (urban groundwater) and A-2 (surface waters and springs subject to natural mineralization). Cluster B represents downstream waters from gold mining sites.

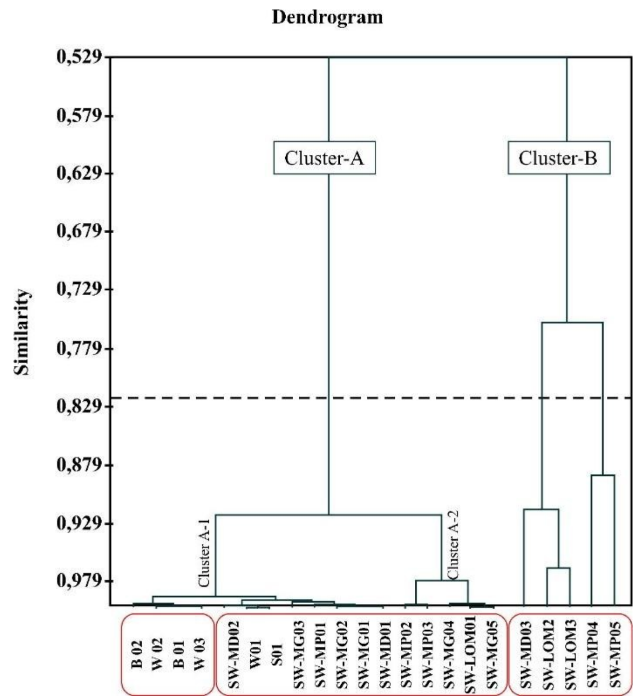

Figure 4 – Dendrogram developed on the identification from different contamination's sites.
Source: Adapted from Rakotondrabe et.al. (2018).

2.4 Correlation matrix (MC)

Correlation matrix serves essential tool for identifying associations among variables especially hydrochemistry where individual chemical parameters may play diverse roles (Orakwe Chukwuma ,2015; Rakotondrabe et.al ,2018; Helena et.al ,2000).

In correlation analysis, we estimate the relationships between pairs of variables using the Pearson correlation coefficient, which reflects the strength and direction of the relationship. The results are presented in a matrix (Table 2), where values ranging from -1 to +1 are considered, with values greater than |0.5| deemed statistically significant (Schot Vander Wal, 1992; Odat, 2015; Armed et al., 2017; Zhang et al., 2018).

Table 2 – Correlation matrix.

	Ec	*Tc*	*pH*	*TDS*	*Na*	*K*	*Ca*	*Mg*	*Cl*
Ec	*1,00*								
Tc	*0,09*	*1,00*							
pH	*-0,25*	*-0,27*	*1,00*						
TDS	*0,99*	*0,09*	*-0,25*	*1,00*					
Na	*0,94*	*0,09*	*0,17*	*0,95*	*1,00*				
K	*0,70*	*0,05*	*-0,20*	*0,71*	*0,72*	*1,00*			
Ca	*0,96*	*0,06*	*0,29*	*0,96*	*0,86*	*0,69*	*1,00*		
Mg	*0,95*	*0,07*	*-0,24*	*0,94*	*0,82*	*0,66*	*0,92*	*1,00*	
Cl	*0,99*	*0,06*	*-0,23*	*0,99*	*0,94*	*0,73*	*0,95*	*0,94*	*1,00*

Source: Modified From Odat (2015).

3. Results and discussions

When analyzing a wide range of parameters in samples, multivariate statistics emerges as an essential tool for the efficient analysis of these data. Collecting multiple measurements provides a comprehensive understanding, allowing for the simultaneous consideration of the variability of diverse properties (Landim, 2011). Among the most widely used methods are: Principal Component Analysis, Cluster Analysis, Factor Analysis (including Rotation and Scree Plot), and the Correlation Matrix.

Forty articles were examined regarding the techniques used and their objectives. The distribution was performed based on the themes in which Multivariate Statistical Analysis (MSA) was applied, with the most frequent methodologies highlighted. Figure 5 illustrates the techniques employed in the referenced research, revealing the simultaneous use of some of them. Notably, the predominance of Principal Component Analysis (PCA) is observed in the theme of hydrogeochemistry characterization/evolution. The results for each theme addressed by MSA are presented, highlighting the techniques employed and their frequent simultaneous use.

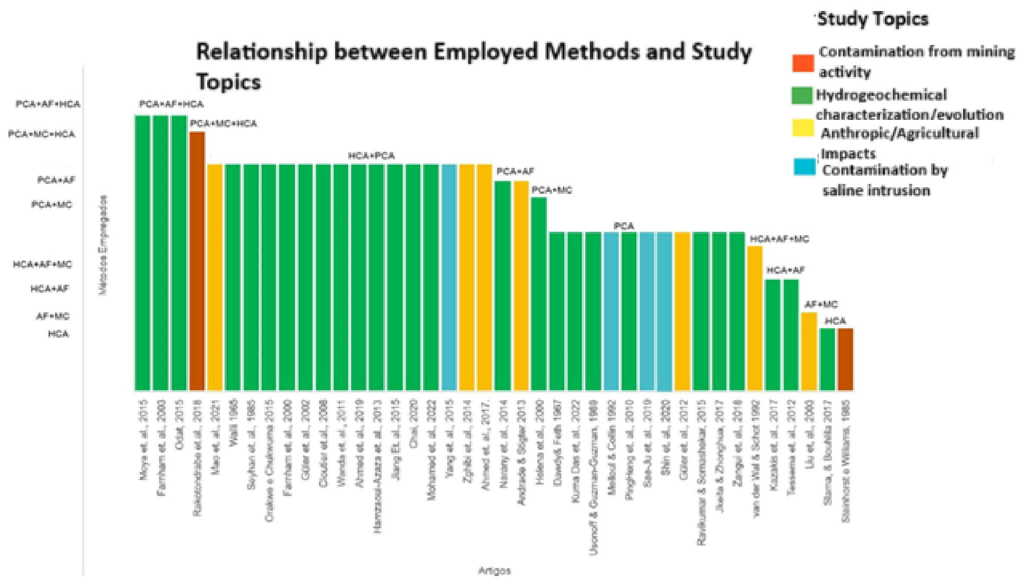

Figure 5 – distribution of applied techniques.
Source: Authors (2024).

3.1 Characterization and/or Evolution Hydrogeochemistry

In a study conducted in Kangding County, southwest China, Zhang et al. (2018) employed Principal Component Analysis (PCA) along with a Scree plot to identify three main components, as represented in Figure 6. By integrating information from other tools, it was possible to infer that thermal water follows deep circulation through the Xianshuihe fault zone, while groundwater flows through fractures, recharging the thermal water. These analyses resulted in the construction of a conceptual hydrological model, providing a deeper understanding of the natural water system.

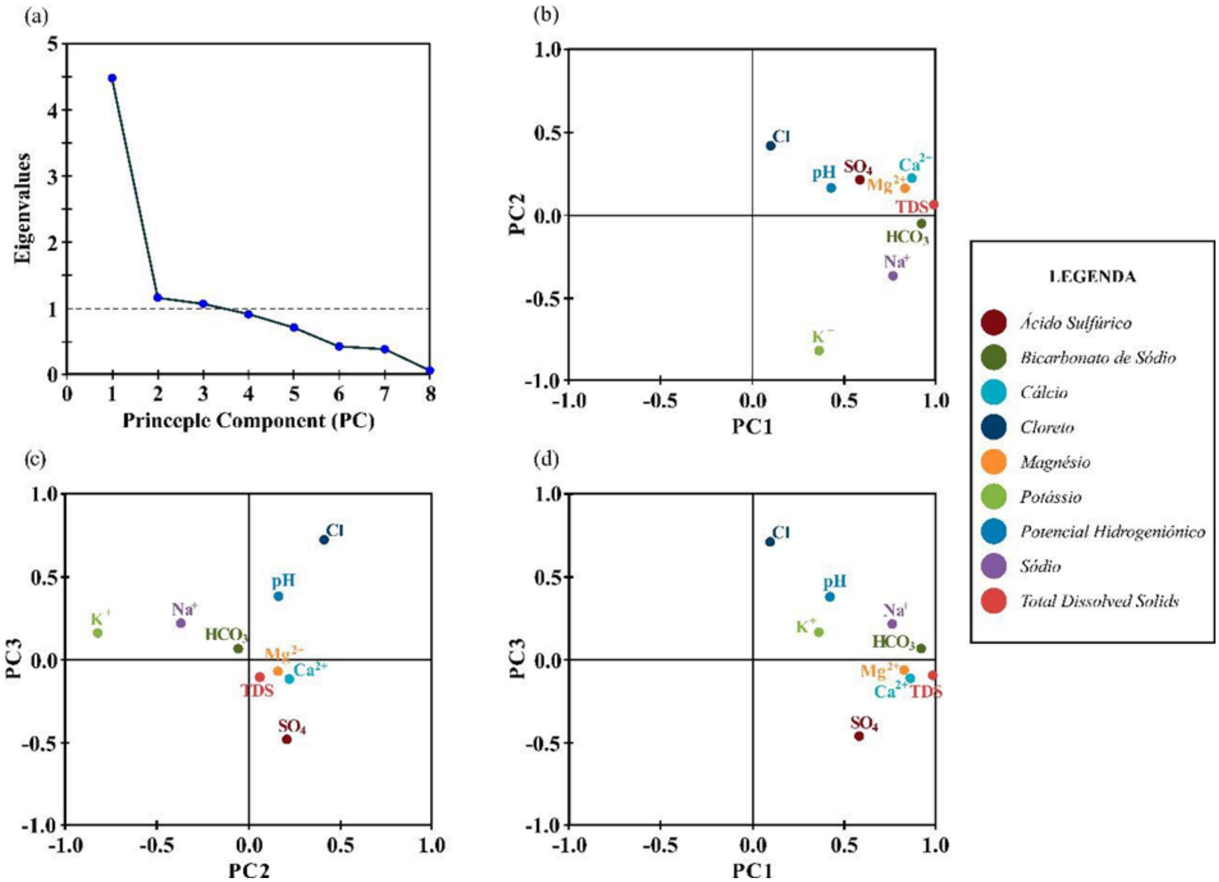

Figure 6 – Scree plot graphics and PCA Diagrams.
Source: Adapted from Zhang et al (2018).

3.2 Investigation Anthropogenic Urban/Agricultural Impacts Groundwater

In a study conducted in the Sylhet region of Bangladesh, Ahmed et al. (2017) investigated the hydrogeochemical characteristics of groundwater, analyzing potential influences from surface waters. They used the Pearson Correlation Matrix (CM), Principal Component Analysis (PCA), and Hierarchical Cluster Analysis (HCA) to assess the controlling factors. Significant correlations were observed, such as K+SO4- (0.838) and K+–NO3 (0.543), indicating the influence of excessive agricultural practices in the study area, associated with high precipitation throughout the discharge period.

PCA revealed six principal components, accounting for 76.16% of the total variance. PC1, PC3, PC4, and PC6 were indicative of Geogenic/Natural Factors, while PC2 and PC5 highlighted Anthropogenic Factors, related to the use of pesticides and fertilizers. The hierarchical cluster analysis (Figure 7) confirmed these results, identifying four distinct clusters. Clusters 1 and 3 reflect weathering and natural factors, cluster 2 indicates both natural and anthropogenic influences, and cluster 4 is exclusively associated with anthropogenic factors.

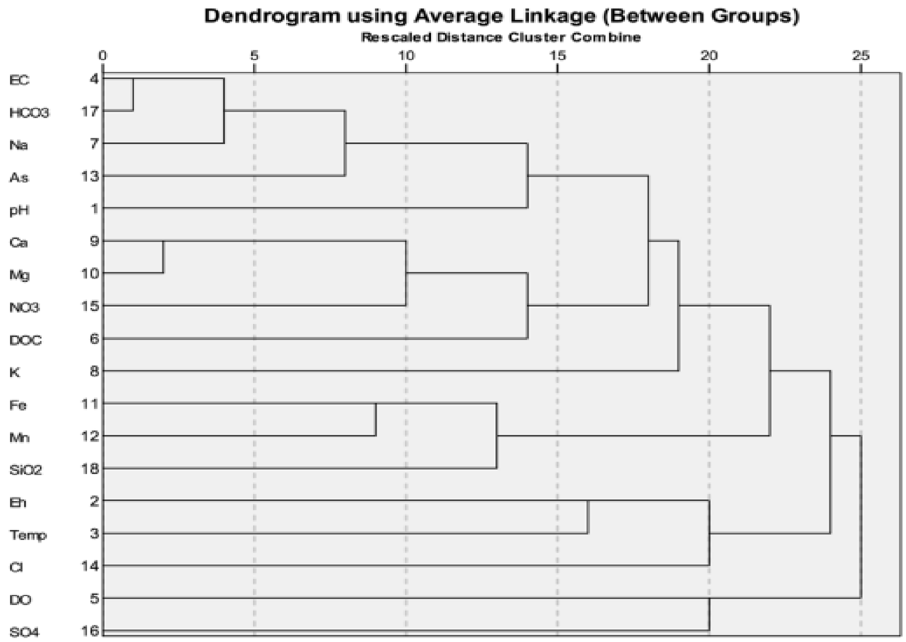

Figure 7 – HCA Diagrams.
Source: Ahmed et al. (2017).

3.3 Mining Activity Contamination Analysis

As reported by Rakotondrab et al. (2018), the influence of gold mining activities on water quality in the Mari River basin in Bétaré-Oya, located in eastern Cameroon, was analyzed. Groundwater and surface water samples were collected, and in addition to conventional analysis methods and the Heavy Metal Pollution Index (HPI), all data were subjected to Multivariate Statistical Analysis using PCA, HCA, and CM techniques.

The correlation matrix in this study helped identify associations between variables, assess the consistency of the data set, and determine the contribution of individual chemical parameters to different influencing factors. Similarly, PCA was applied to distinguish and verify these factors (Figures 7a-b), resulting in the classification of five groups: two for major elements and physical parameters (Figure 7-a) and three for heavy metals (Figure 7-b).

In the first group (red circle), which includes turbidity and total suspended solids (TSS), it is suggested that physical pollution may be attributed to erosion from mining and waste. The second group (green circle), containing EC, pH, alkalinity, and major ions (K+, Ca2+, Na+, Cl-, HCO-3, NO-3), represents water mineralization, which may have either natural or anthropogenic origins, as in the case of nitrate ions.

Regarding the groups related to heavy metals, the first group (green circle), composed of Fe and Cr, showed a significant correlation (0.728), suggesting lithogenic origins for these heavy metals. The second group (red circle), consisting of Mn/Cd and Pb, may have a combination of geogenic and anthropogenic sources. The last group (Zn/Cu), in the purple circle, was attributed to anthropogenic sources such as mining and associated activities, representing potential pollution sources in the study area.

The multivariate statistical approach enabled the identification of highly polluted areas within the study region, related to both physical and chemical pollutants.

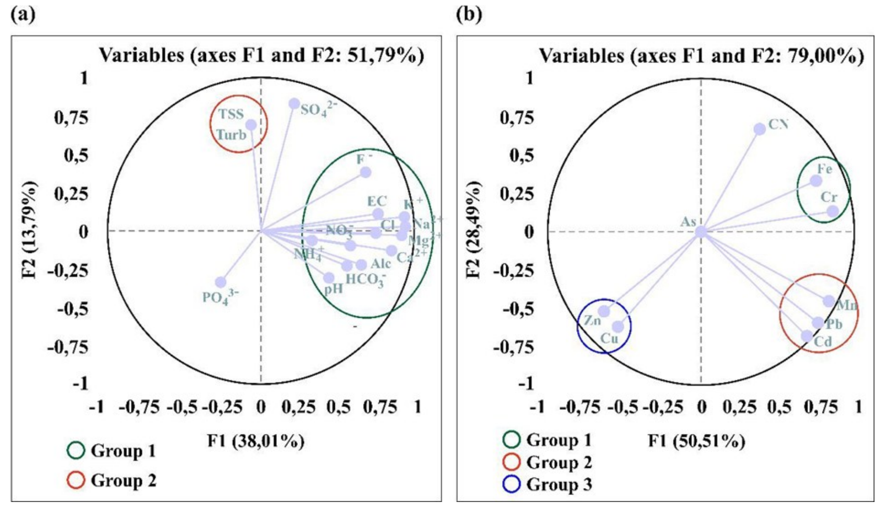

Figure 8 – Analysis of Main Components (PCA) of Major Elements and physical parameters (a) e heavy metals (b). Source: Adapted from Rakotondrab et. al., (2018).

3.4 Aquifer Contamination Analyses for Saltwater Intrusion

Shin et al. (2020) investigated a data set from an archipelago off the southwest coast of Korea, near the Muanna Peninsula, using Principal Component Analysis (PCA) to understand the impact of seawater on groundwater. Seventy-four groundwater samples were collected and classified based on water type and the molar ratio of Cl-/HCO-3. Thirty-six samples were classified as Ca2+-Cl-, and thirty-two as Na+-Cl- (representing 91.9%), indicating that they were influenced by seawater. The classification of samples based on the molar ratio of Cl-/HCO-3 showed that forty out of seventy-four samples had a molar ratio of 2.8 or higher, indicating a strong influence of seawater on groundwater.

Principal Component Analysis revealed the influence of oceanic water in the first component, which explained 54.1% of the variance. These results showed that saltwater intrusion affects the region's water, with samples exhibiting a significant proportion of seawater mixing.

3.5 Complex Studies - Simultaneous Hydrogeochemical Heterogeneities

Kazakis (2017) applied multivariate statistical techniques to hydrochemical data from three distinct regions in northern Greece, each characterized by unique hydrogeological, climatological, and land-use conditions (Figure 8). The study employed the Pearson correlation coefficient, Factor Analysis (FA) in R-mode, and Hierarchical Cluster Analysis (HCA) to assess groundwater quality and hydrochemical processes in these regions.

In an innovative approach, the data from all three regions were analyzed collectively, without presupposing similarities in hydrochemical processes. Factor analysis, following Varimax rotation, identified two factors explaining 70.6% of the total variance.

The findings revealed that contamination in the first region was linked to industrial and agricultural activities; in the second region, seawater intrusion and geothermal fluids were the primary sources of contamination; while in the third region, anthropogenic impacts—particularly the intensive use of pesticides—were the main contributors.

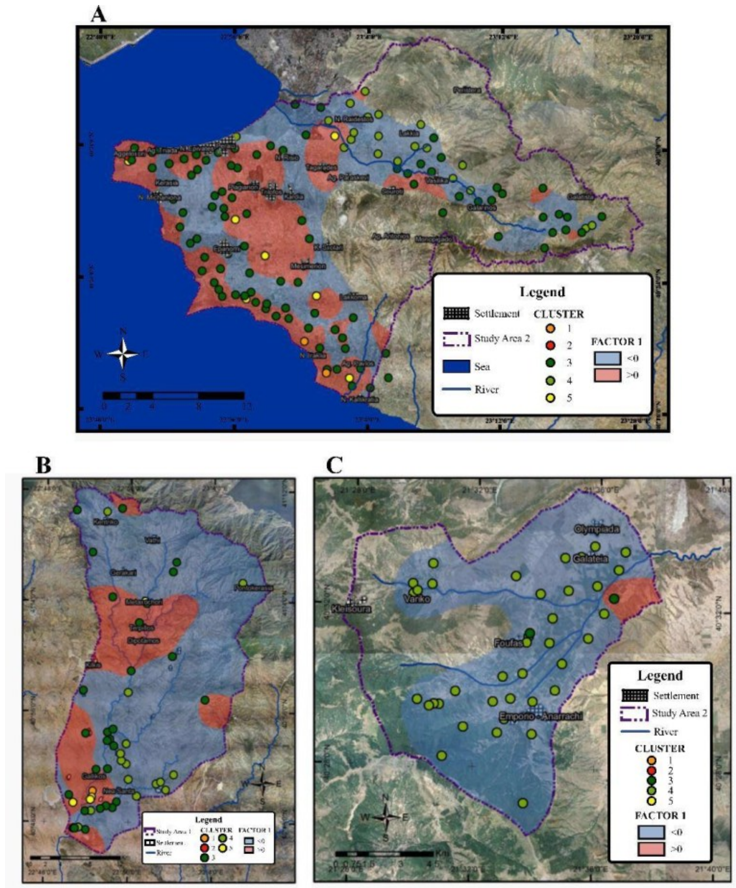

Figure 8 – The results of Cluster Analysis *application in 3 different sites of Greece, with different land use e sobre different aquifers.*
Source: Adapted map from Kazakis (2017).

4. Conclusions

Multivariate Statistical Analysis (MSA) is an essential tool for handling vast amounts of data, providing integration and efficient interpretation to represent results. This methodology enables a quantitative approach and favors the classification of aquifer samples by analyzing correlations among variables, such as physicochemical parameters, and evaluating the heterogeneities of sampling sites and waters, among other relevant applications. The success of MSA in hydrogeochemical studies is evident in its ability to identify key water factors, understand temporal and spatial variations, and delineate sites based on the processes to which they have been subjected. Such information is central for the effective management of water quality.

It is crucial to emphasize that proper utilization of MSA requires a solid understanding of the techniques and their limitations. While MSA does not establish cause-effect relationships, it offers valuable insights into relationships that can be inferred.

Based on the research analyzed, we conclude that the multivariate statistical approach is universally applicable, regardless of lithology type, with the main tools employed being Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA), often applied together.

References

AHMED, N.; BODRUD-DOZA, MD.; ISLAM ,S. M. D.; CHOUDHRY, M. A.; MUHIB, MD. I.; ZAHID, A.; HOSSAIN, S.; MONIRUZZAMAN, MD.; DEB, N.; BHUIYAN, MD. A. Q. 2019. *Hydrogeochemical evaluation and statistical analysis of groundwater of Sylhet,* north-eastern Bangladesh. Acta Geochim 38, 440-455). https://doi.org/10.1007/s11631-018-0303-6

AHMED, A. H., RAYALEH, W. E., ZGHIBI, A., & OUDDANE, B. 2017. *Assessment of chemical quality of groundwater in coastal volcano-sedimentary aquifer of Djibouti*, Horn of Africa. Journal of African Earth Sciences, 131, 284–300. doi:10.1016/j.jafrearsci.2017.04.010

ANDRADE, A. I. A. S. S. & STIGTER, T. Y.. 2013. *The distribution of arsenic in shallow alluvial groundwater under agricultural land in central Portugal: insights from multivariate geostatistical modeling*. Sci Total Environ. DOI: 10.1016/j.scitotenv.2013.01.033

BHUIYAN M.A.H, BODRUD-DOZA M, ISLAM A.T, RAKIB M.A, RAHMAN M.S, RAMANATHAN A.L.. 2016. *Assessment of groundwater quality of Lakshimpur district of Bangladesh using water quality indices, geostatistical methods, and multivariate analysis*. Environ Earth Sci 75:1020. https://doi.org/10.1007/s12665-016-5823-y

CHE, Q.; SU, X.; WANG, S.; ZHENG, S.; LI, Y. 2021. *Hydrochemical Characteristics and Evolution of Groundwater in the Alluvial Plain (Anqing Section) of the LowerYangtze River Basin: Multivariate Statistical and Inversion Model Analyses.* Water, 13, 2403. https://doi.org/10.3390/w13172403

CHAI, Y.; XIAO, C.; LI, M., LIANG, X. 2020. *Hydrogeochemical Characteristics and Groundwater Quality Evaluation Based on Multivariate Statistical Analysis*. Revista Water, 12(10):2792, DOI:10.3390/w12102792

CLOUTIER, V.; LEFEBVRE, R.; THERRIEN, R.; SAVARD, M. M. 2008. *Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system*. Journal of Hydrology 353, 294– 313.

DAVIS, J.C., 1986. *Statistics and Data Analysis in Geology*. John Wiley & Sons Inc., New York.

DAWDY, D.R. AND FETH, J.H., 1967. *Application of factor analysis in the study of chemistry of groundwater quality, Mojave River Valley, California*. Water Resour. Res., 3(2): 505-510.

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*. Analytical Chimica Acta 490, 123–138.

FARNHAM, I.M., STETZENBACH, K.J., SINGH, A.K., JOHANNESSON, K.H., 2000. *Deciphering groundwater flow systems in Oasis Valley, Nevada, using trace element chemistry, multivariate statistics, and geographical information system*. Mathematical Geology 32, 943–968.

GÜLER, C., THYNE, G.D., MCCRAY, J.E., TURNER, K.A., 2002. *Evaluation of graphical and multivariate statistical methods for classification of water chemistry data*. Hydrogeol. J.10:455–474. http://dx.doi.org/10.1007/s10040-002-0196-6.

GÜLER, C., KURT, M.A., ALPASLAN, M., AKBULUT, C., 2012. *Assessment of the impact of anthropogenic activities on the groundwater hydrology and chemistry in Tarsus coastal plain (Mersin, SE Turkey) using fuzzy clustering, multivariate statistics and GIS techniques*. J. Hydrol. 414, 435–451.

HELENA, B., PARDO, B., VEGA, M., BARRADO, E., FERNANDE, J.M., FERNANDEZ, L., 2000. *Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis*. Water Res. 34 (3), 807–816.

JIANG, Y.; GUOA, H.; JIAA, Y.; CAOA, Y.; HUB, C. 2020. *Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia*. Chemie der Erde 75 (2015) 197–205.

KAISER H., F. 1958. *The varimax criteria for analytical rotation in factor analysis*. Psychometrika, 23(3):187 –200.

KAZAKIS, N., MATTAS, C., PAVLOU, A., PATRIKAKI,O., VOUDOURIS, K. 2017. *Multivariate statistical analysis for the assessment of groundwater quality under different hydrogeological regime*s. Environ Earth Sci, 76:349 DOI 10.1007/s12665-017-6665-y.

KEITA, S., & ZHONGHUA, T. 2017. *The assessment of processes controlling the spatial distribution of hydrogeochemical groundwater types in Mali using multivariate statistics*. Journal of African Earth Sciences, 134, 573–589. doi:10.1016/j.jafrearsci.2017.07.023

KUMAR DAS,T.; AHMED, S.; HOSSEN, A.; RAHAMAN,MD.H.; RAHMAN, M., M. 2022. *Multivariate statistics and hydrogeochemistry of deep groundwater at southwestern part of Bangladesh*, Heliyon, Volume 8, Issue 10, e11206, ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2022.e11206.

LIU CW, LIN KH, KUO YM. 2003. *Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan.* Sci. Total Environ. 2003;313(1):77–89. http://dx.doi.org/10.1016/S0048-9697(02)00683-6.

MAO, H.; WANG, G.; RAO, Z.; LIAO,F.; SHI,Z.; HUANG,X.; CHEN, X.; YANG, Y. 2021. *Deciphering spatial pattern of groundwater chemistry and nitrogen pollution in Poyang Lake Basin (eastern China) using self-organizing map and multivariate statistics,* Journal of Cleaner Production, Volume 329, 129697, ISSN 0959-6526, https://doi.org/10.1016/j.jclepro.2021.129697.

MELLOUL, A., COLLIN, M., 1992. *The 'principal components' statistical method as a complementary approach to geochemical methods in water quality factor identification; application to the Coastal Plain aquifer of Israel*. Journal of Hydrology 140, 49–73.

MOHAMED,N. A.; WACHEMO, A. C.; KARUPPANNAN, S. & DURAISAMY,K. 2022. *Spatio-temporal variation of groundwater hydrochemistry and suitability for drinking and irrigation in Arba Minch Town, Ethiopia: An integrated approach using water quality index, multivariate statistics, and GIS, Urban Climate*, Volume 46,2022,101338, ISSN 2212-0955, https://doi.org/10.1016/j.uclim.2022.101338.

MOYA, C.E., RAIBER, M., TAULIS, M., COX, M.E., 2015. *Hydrochemical evolution and ground-water flow processes in the Galilee and Eromanga basins, Great Artesian Basin, Australia: a multivariate statistical approach*. Sci. Total Environ. 508:411–426. http://dx.doi.org/10.1016/j.scitotenv.2014.11.099.

NARANY, S. T., RAMLI, M. F., ARIS, A. Z., SULAIMAN, W. N. A., & FAKHARIAN, K. 2014. *Spatiotemporal variation of groundwater quality using integrated multivariate statistical and geostatistical approaches in Amol–Babol Plain, Iran. Environmental Monitoring and Assessment*, 186(9), 5797–5815. doi:10.1007/s10661-014-3820-8

ODAT, S. 2015. Cluster and Factor Analysis of Groundwater in Mafraq Area, Jordan, Current World Environment Vol. 10(2), 422-431.

ORAKWE, L. C., CHUKWUMA, E.C. 2015. *Multivariate Analysis Of Ground Water Characteristics Of Geological Formations Of Enugu State Of Nigeria.* International Journal Of Scientific & Technology Research /Volume 4, Issue 08, August.

PINGHENG, Y., DAOXIAN, Y., WENHAO, Y., YINGLUN, K., PENG, Y., E QIUFANG H. 2010. *Formations of groundwater hydrogeochemistry in a karst system during storm events as revealed by PCA*. Chinese Sci Bull May 2010 Vol.55 No.14: 1412−1422 doi: 10.1007/s11434−010−0083−9

RAKOTONDRABE, F., NGOUPAYOU, J. R. N., MFONKA, Z., RASOLOMANANA, E. H., ABOLO, A. J. N. E AKO, A. A. 2018. *Water quality assessment in the Bétaré-Oya gold mining area (East-Cameroon): Multivariate Statistical Analysis approach*. Science of the Total Environment 610–611 (2018) 831–844

REVENGA, C. 2000. *Will the be enough water? In: Mock G. Ed. "Pilot Analysis of Global Ecosystems: Freshwater Systems.* EarthTrends World Resources Institute. p. 1-6.

RAVIKUMAR, P.; SOMASHEKAR, R.K; PRAKASH, K.L. 2015. *A comparative study on usage of Durov and Piper diagrams to interpret hydrochemical processes in groundwater from SRLIS river basin, Karnataka, India.* www.elixirpublishers.com. Elixir Earth Sci. 80 (2015) 31073-31077.

ROMANOWISKI, JOANA PAULIN & ENS, ROMILDA TEODORA (2006). *As pesquisas denominadas do tipo "estado da arte" em educação.* In Diálogo Educacional. Curitiba, v.6, n. 19, p. 37-50, set./dez., 2006.

SAS - CARY, N. C., 1982. SAS Institute Inc. SAS User's Guide.

SAE-JU, J., CHOTPANTARAT, S., & THITIMAKORN, T. (2019). *Hydrochemical, geophysical and multivariate statistical investigation of the seawater intrusion in the coastal aquifer at Prachuap-Khiri-Khan Province, Thailand.* Journal of Asian Earth Sciences, 104165. doi:10.1016/j.jseaes.2019.104165

SCHOT, P.P., VAN DER WAL, J., 1992. *Human impact on regional groundwater composition through intervention in natural flow patterns and changes in land use*. Journal of Hydrology 134, 297–313

SEYHAN, E. AND KEET, B., 1981. *Multivariate statistical analysis (part 1), Application to hydromopho-metrica! data (case-study; Ahr River basin, Bolzano, Italy).* Aqua-VU series A, No. 8. Rodopi, Amsterdam.

SEYHAN, E., VAN DE GRIEND, A.A. AND ENGCLEN, G.B., 1985. *Multivariate analysis and interpretation of the hydrochemistry of a dolomite reef aquifer, Northern Italy.* Water Resour. Res., 21: 1010-1024.

SHIN, K.; KOH,D-C.; JUNG, H.;AND LEE, J. 2020. *The Hydrogeochemical Characteristics of Groundwater Subjected to Seawater Intrusion in the Archipelago,* Korea. Water, 12, 1542; doi:10.3390/w12061542.

SLAMA, F., & BOUHLILA, R. 2017. *Multivariate statistical analysis and hydrogeochemical modelling of seawater-freshwater mixing along selected flow paths: Case of Korba coastal aquifer Tunisia*. Estuarine, Coastal and Shelf Science, 198, 636–647. doi:10.1016/j.ecss.2016.10.005

STEINHORST, R.K., WILLIAMS, R.E., 1985. *Discrimination of groundwater sources using cluster analysis, MANOVA, canonical analysis and discriminant analysis*. Water Resources Research 21, 1149– 1156.

TESSEMA, S. G., MIRGORODSKY ,D., MERTEN , D., HILDEBRANDT , A., ATTINGER S., BU ̈CHEL, G. 2012. Hydrochemical characterization of groundwater system of biodiversity experimental field: Jena, Germany. Environ Earth Sci. DOI: 10.1007/s12665-012-2171-4

USUNOFF, E.J., GUZMA ́N-GUZMA ́N, A., 1989. *Multivariate analysis in hydrochemistry: an example of the use of factor and correspondence analyses*. Ground Water 27, 27–34.

WANDA, E.; MONJEREZI, M.; MWATSETEZA, J. F.; KAZEMBE, L. N. 2011. *Hydrogeochemical appraisal of groundwater quality from weathered basement aquifers in Northern Malawi*. Physics and Chemistry of the Earth, Volume 36, Issue 14-15, p. 1197-1207.

WALLIS, J. R. 1965. *Multivariate statistical methods in hydrology--a comparison using data of known functional relationship.*Water Resources Res., 1(4), 447-462.

YANG, Q.; ZHANG, J.; WANG, Y.; FANG, Y; & MARTÍN, J. D. 2015. *Multivariate Statistical Analysis of Hydrochemical Data for Shallow Ground Water Quality Factor Identification in a Coastal Aquifer*. Pol. J. Environ. Stud. Vol. 24, No. 2 (2015), 769-77.

ZHANG, Q.; GUO, J.; YU, L. E ZHAO, R. 2018. *Hydrochemical Characteristics and Multivariate Statistical Analysis of Natural Water System: A Case Study in Kangding County, Southwestern China*; Water, 10, 80; doi:10.3390/w10010080

ZGHIBI, A., MERZOUGUI, A., ZOUHRI, L., & TARHOUNI, J. 2014. *Understanding groundwater chemistry using multivariate statistics techniques to the study of contamination in the Korba unconfined aquifer system of Cap-Bon (North-east of Tunisia).* Journal of African Earth Sciences, 89, 1–15. doi:10.1016/j.jafrearsci.2013.