An overview of the application of principal component analysis and factor analysis (PCA/FA) in water resource management

Adamu Mustapha

Department of Geography, Faculty of Earth and Environmental Sciences, Kano University of Science and Technology Wudil, Nigeria

Correspondence e-mail: amustapha494@gmail.com, amustapha@kustwudil.edu.my

Abstract— Principal component analysis (PCA) and factor analysis (FA) are multivariate statistical techniques that have been adopted by researchers to provide ease of interpretation of a large data set. An ever increasing literature on the use of PCA/FA tools suggest that the techniques are useful in data reduction and structural simplification, revealing the latent pollution sources. The combined used of these statistical techniques could simplify the process within a convenient size and enable the classification of water samples into distinct groups, sources apportionment, relationship and differences in the parameters used based on hydrochemical characteristics. They reflect more accurately the multivariate nature of natural ecosystem and provide a way to handle large data sets with large number of parameters by summarizing the redundancy and noise from the datasets and provide means of detecting and quantifying truly multivariate patterns of the data sets.

Keywords— Principal component analysis, factor analysis, water pollution sources apportionment

I. INTRODUCTION

Water is the most important natural resource in the world, since without it life cannot exist and industries cannot operate [1]. It is a common and widespread chemical compound in nature, which is a major constituent of all living creatures. It is a fundamental for metabolism and many accompanying processes and constitutes the only habitat for life for many organisms [2]. Water is a finite resource that is very essential for human existence, agriculture and industry by providing a medium for transport, recreation, tourism, worship, food, ecosystem functioning, and a place to experience a serenity of nature [3]. Without any doubt, inadequate quantity and quality of water resources have a serious impact on sustainable development.

The increase in human population and economic activities has grown in scale; the demand for large-scale suppliers of fresh water from various competing end users has increased. Declining in quality and quantity of water resources can be attributed to the water pollution and improper management of the resource [4]. Many regions in the world are simultaneously impacted by urbanization processes, industrial and agricultural activities and many cities in developing countries have been developed without adequate and proper planning, this has led to indiscriminate actions including dumping of wastes in to the water, washing, and bathing in open surface water bodies [5]. In such areas, the environmental compartments are affected by pollutants originating from different potential sources. The deteriorating water quality affects man, animal, and plant life with far reaching consequences [6]. From environmental, economical and or social point of view, it is important to identify these sources and their contribution to the total contamination of the area [7].

The application of PCA/FA in interpretation of complex data sets offers a better understanding of the water quality and ecological status. It allows the identification of the possible factors/sources that influence the water quality variation and offers valuable tools for reliable management of water resources as well as rapid solutions on pollution problems [8].

II. PRINCIPAL COMPONENT ANALYSIS/FACTOR ANALYSIS (PCA/FA)

Principal Component Analysis and factor analysis (PCA/FA) are multivariate statistical approaches that can be used to analyse interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimension by providing an empirical estimate of the structure of variables [9]. PCA is about extracting of set of uncorrelated, mutually independent (orthogonal) and mathematically represented by a linear combination of the parameters of the study to capture the maximum amount of variability of a given data set [10].

The objective of PCA/FA is to reduce the number of variables necessary to describe the observed variation within a datasets [11]. PCA/FA lessens the number of variable's understudy and discovers the structure between the variables (Pejman et al., 2009). This is achieved by forming linear combinations of the variables (components) that describe the distribution of the data [13]. The linear combinations are derived from some measures of association such as correlation or covariance matrix. PCA converts the original dataset which comprises of measurements for each variable measured for each sample and convert them to an equal number of composite variables [14]. The goal is that the first few principal components (PC's) will retain most of the information in all the original variables, thus reducing the practical dimensionality of the dataset [15].

In order to maximize the relationship between the variables under study, the PC's are rotated. Factor rotation is used to facilitate interpretation by providing a simpler factor structure [16]. Factor rotation has been used to extract related variables and infer the processes that control water chemistry. Varimax rotation is commonly applied to the PC's in order to determine factors that can be easily explained in terms of hydro-chemical and anthropogenic process [17]. The goal of rotation is to maximize the variance (variability) of the new variable while minimizing the variance around the new variables [18]. The relevant information is carried out by the first principal component (PCs). PCs are ordered in such a way, that the variance of the first PC (PC1) is the highest; the variance of the second PC (PC2) is the second highest, and so on, whereas the last PC is the lowest in explaining variation of the data sets. The total variance of all the PC's is equal to the total variation of the original variables [19].

PCA does not actually make assumption about the distribution of the data but, instead is a tool that identifies variation and pattern in a dataset. PCA techniques minimize the number of variables under study and at the same time maximize the amount of information in the analysis. The original set of variables is reduced to a much smaller set which account for most of the reliable variance of the initial variable pool [20]. There are two main conditions necessary for PCA/FA: there is need to for a relationship to be existed between the variables. Further, the larger the sample size, especially in relationship to the number of variables, the more reliable the factors [21].

The first principal component score obtained for individual i on component $ci1$ uses weight $wi1 .....,wp1$ in the linear combination as follows:

$$ci1 = yi1wi1 + yi2wi2 + ... + yip + wip$$

The linear combination is chosen, so that the sum of square of $ci1$ is as large as possible subject to the condition that $wi1 +.....+ wi2$ = 1. The second principal component is another linear combination of yip .

FA follows PCA; FA focuses on reducing the contribution of less significant variables to simplify even more of the data structure coming from PCA [22]. It is used to describe the variability among the observed variables in terms of fewer unobserved variables called factors [23]. This can be achieved through the use of factor rotation, by rotating the axis defined by PCA according to well established rules. The basic motivation for using any rotational methods is to achieved a simpler and more meaningful representation of the underlying factors, producing a new group of variables known as varifactor (VFs) [24-25]. The varifactor can be written as follows:

$$yji + fj1xi1 + fj2zi2 + ... + fjmzjm + eij$$

Where y is the measured variable, f is the factor loading, z is the factor score, e is the residual term accounting for errors, ij is the sample numbers and m is the total number of factors.

Before conducting a factor analysis, it is essential to check our sampling adequacy and sphericity to see if it is worth proceeding with the PCA/FA analysis through the use of Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity. KMO test measure the sampling adequacy and determined whether the data in the samples are adequate to correlate and is it suitable for factor analysis [26]. KMO is a measure of sampling adequacy that provides an index between 0 and 1 reflecting the proportion of variance among the variables. This statistical measure indicates the proportion of variance in the variables that might be caused by the underlying factors. A decision of when to stop extracting factors depends on when there is only very little random variability left [27]. High values close to 1.0 generally indicate that factor analysis may be useful in data reduction, if the value is less than 0.50, the results of the factor analysis will probably not be very useful [28]. As a rule of thumb, if the KMO test is 0.5 or higher, factor analysis is valid. The KMO guiding rules were presented in the Table1.

Table 1 Guiding rules for interpreting KMO results after Kaiser's 1974

KMO value	Interpretation
0.90-1.00	Marvellous
0.80-0.89	Meritorious
0.70-0.79	Middling
0.60-0.69	Mediocre
0.50-0.59	Miserable
0.00-0.49	Unacceptable

The Bartlett test of Sphericity, on the other hand, investigates if there is a relationship within the variables worth to be investigated. If no relationship is found, then there is no point in proceeding with the factor analysis. A lower probability value ($p < 0.05$) indicates that it makes sense to continue with the factor analysis.

III. WATER POLLUTION SOURCES APPORTIONMENT USING PRINCIPAL COMPONENT ANALYSIS AND OR FACTOR ANALYSIS (PCA/FA)

An ever increasing literature on the use of PCA in identifying pollution sources suggest that the technique is useful in revealing the latent pollution sources and it is practical in various environmental studies. Factor analysis (FA) imposes a strict underlying structure of fixed common factors. FA follows PCA after rotation and gives the most important factors that bring variation in the data sets by simple clarification of the data structure.

Shrestha and Kazama; Huang et al.; Juahir et al. [21-31] studied spatial variability of surface water quality and source's apportionment and classified the studied water bodies into High pollution site (HP), Moderate pollution site (MP) and Low pollution site (LP). Factor analysis revealed that the pollution levels in the three zones were mainly influenced by natural sources (temperature and river discharge) and anthropogenic sources (industrial, municipal and agricultural run-off). Onojake et al., [32] in their studies, they discovered that Rivers in Delta's state of Nigeria were heavily polluted as a result of industrial discharge and municipal waste (anthropogenic source of pollution). They used PCA/FA to identify the latent factors that explain the chemistry of the surface water; PCA/FA yielded three PCs with more than 82% variance. Vidal et al. [33] used PCA/FA for the study of water resource's contamination due to the use of livestock slurries as fertilizer; PCA/FA revealed in the sampled data is related to dilution of water with strong saline content and redox condition.

Surface water pollution sources identification in the Peri-Urban Interface of Wuxi Taihu Lake area, China was studied by Huang et al. [30]. The most striking result to emerge from the data is that lead (Pb) and nitrogen (N) were related to the influence of urban run-off and domestic wastewater while copper (Cu) and chromium (Cr) were associated to the industrial effluent around the vicinity Taihu Lake. Moreover, Phosphorous (P) and Zinc (Zn) were related to both activities of urban domestic wastewater and effluent from the industries.

Huang et al. [30] applied PCA/FA to identify latent factors of pollution and yielded three potential pollution sources, the output of this study has demonstrated that, water quality variation were influenced primarily by pollution due to industrial waste water, agricultural run-off and domestic activities. PCA/FA proved to be robust at unveiling the sources of variation in the hydrochemistry of the surface water. Yidana [34] uses Kaiser Criterion to extract PCs and used varimax rotation to maximized variation in the datasets and ease interpretation, in their studies, they identified that various sources of pollution influence the hydrochemistry of the surface water and the basin is controlled largely by the weathering of minerals (silicates, carbonates, gypsum and apatile) from the underlying metasediments, little contributions were observed from the decay of organic matter from the heavily forested regions. Ouyang [35] applied PCA/FA to evaluate the effectiveness of the surface water quality monitoring network; PCA/FA revealed that total organic compound, total nitrate and nitrite were the most statistically significant parameters that brought variation in water chemistry in the studied area.

Similarly, Xu et al. [36] in their studies on a Taihu lake region in China, they discovered that, the surface water in the region is progressively susceptible to anthropogenic pollution. This was revealed by PCA/FA, in which it yielded three PCs corresponding to urban residential subsistence, livestock farming and farmland's run-off. PCA/FA was equally applied to analysed data sets of marine water quality and yielded four latent factors of pollution sources: organic/eutrophication, natural, mineral and nutrient/faecal pollution sources, that influence water chemistry of southern Hong Kong [37].

Koklu et al. [38] focused on water quality of Melen River, Turkey; PCA/FA was used and factors were identified which are responsible for the variation in water quality. The study concluded that, all stations under study were affected by agricultural and industrial pollution due to the intensive agricultural activities through application of fertilizer. Li et al. [39] used PCA/FA to detect the latent pollution sources of Songhua River and identified five latent factors: organic, inorganic, petrochemicals, physiochemical and heavy metals pollution sources, they concluded that Songhua River water quality was mainly controlled by domestic waste water and industrial effluent. Surface water quality data for 29 parameters were collected from Esenkara monitoring station along the Porsuk stream, Turkey. FA shows that, river water temperature and dissolved oxygen were the main variable responsible for water quality variation in the area [40].

Similarly, in their studies Varol and Sen [41] used PCA/FA to identify the factors/sources responsible for variation in Behrimaz stream, Turkey, PCA/FA yielded five latent factors amounting to 88.32% of the total variance. The latent factors explained that, the variation of water quality are mainly related to natural (stream discharge, temperature, and soluble minerals) and anthropogenic (nutrient/organic pollutants and agricultural run-off). Similarly, Calijuri et al. [42] used PCA to assess the influence of human activities and natural characteristic in surface and underground water resources in Lagoa santa Karstic Region of Brazil. Five PCs were generated for each resource to explain the latent pollution sources: erosive, anthropogenic, weathering, dilution and domestic sewage influences.

PCA/FA techniques have also been applied by Han et al. [43] in their studies on Nakdong River watershed; they used PCA/FA to identify pollution sources in the study area. They discovered that, anthropogenic pollutants are responsible for high variation in the water quality of the studied area. Equally, Wong [44] studied spatial variability of physiochemical elements in Hong Kong River. PCA/FA was used to identify latent factors or pollution sources; it yielded four components in which the different watercourse studied falls under nutrient and organic pollutants, heavy metal's contamination; other water courses suffered multiple types of pollution or have a low pollution problem. The regional distributions of physiochemical determinants verified that, the marine and anthropogenic sources have strong influences to Rivers/Streams in Hong Kong.

Similarly, Tanriverdi et al. [45] applied PCA/FA in the surface water quality data to assess and examine water quality of Ceyhan River. Three PCs were significantly identified correspond to areas close to cities presented low dissolved oxygen content and high concentration of physiochemical parameters suggesting anthropogenic inputs. The stations in the vicinity of industries have higher pollution due to discharge of waste water from industries and domestic activities.

The PCA/FA technique has been applied in to groundwater pollution sources apportionment. For examples Aris et al. [46] used PCA and identified three main processes associated with the sea water intrusion into groundwater system of small Tropical Island. Equally, Omor-Irabor et al. [47] concluded that PCA proved to be a powerful tool in revealing the sources apportionment and explained the various sources of underground pollution: soil groundwater interaction, domestic and agricultural pollution, industrial pollution and atmospheric and vehicular pollution. Cloutier et al. [48] found that the factors responsible for controlling groundwater geochemistry as revealed by PCA/FA are groundwater mixing with sea water and solute diffusion from marine clay acquitard, salinity, dissolution of carbonate rocks and hardness.

Application of PCA by Morell et al. [49] showed that the chemical characterization of groundwater is the result of three components: intruding on seawater, freshwater and rainfall infiltration and saline water with a characteristic sulphate-calcium-magnesium facies, derived from bordering aquifers. Liu et al. [50] applied factor analysis to the assessment of groundwater in a Blackfoot disease area in Taiwan: a two-factor model was revealed and explained over 77 % of the total groundwater quality variation as a result of seawater salinization and arsenic pollution.

IV. CONCLUSION

The published application of principal component analysis and or factor analysis (PCA/FA) increased tremendously in the recent years. Various researchers revealed the usefulness of the techniques, and most of the studies reviewed were able to successfully use the results of multivariate statistical analysis. These researchers confirmed the usefulness of those techniques in the analysis of large, complex sets of data and in the planning of measuring networks for efficient control of sources of water resources.

REFERENCES

[1]Tebbutt, T. H. Y. (1998*). Principles of water quality control.* Butterworth-Heinemann.

[2] Cieszynska, M., Wesolowski, M., Bartoszewicz, M., Michalska, M. and Nowacki, J. (2012). Application of physicochemical data for water-quality assessment of watercourses in the Gdansk Municipality (South Baltic coast). *Environmental Monitoring and Assessment*, 184, 2017–2029.

[3] Pinto, U. and Maheshwari, B. L. (2011). River health assessment in peri-urban landscapes: An application of multivariate analysis to identify the key variables.*Water research*, 45, 3915–3924.

[4] Silva Alves, R. I., Oliveira Cardoso, O., Abreu Tonani, K. A., Julião, F. C., Trevilato, T. M. B. and Segura-Muñoz, S. I. (2012). Water quality of the Ribeirão Preto Stream, a water course under anthropogenic influence in the southeast of Brazil. *Environmental Monitoring and Assessment*, doi: 10.1007/s10661-012-2622-0.

[5] Adekola, F. A. & Eletta, O. A. A. (2007). A study of heavy metal pollution of Asa River, Ilorin; trace metal monitoring and geochemistry. *Environmental Monitoring and Assessment*, 125, 157-163.

[6] Jindal, R. and Sharma, C. (2011). Studies on water quality of Sutlej River around Ludhiana with reference to physicochemical parameters. *Environmental Monitoring and Assessment*, 174, 417–425.

[7]Tobiszewski, M., Tsakovski, S., Simeonov, V. and Namie´snik, J. (2010). Surface water quality assessment by the use of combination of multivariate statistical classification and expert information. *Chemospher*e, 80, 740–746.

[8] Wundelin, D. A., Diaz, M. P., Ame, M. V., Pesce, S. F., Hued, A. C. and Bistoni, M. (2001). Pattren recognition techniques for the evaluation of spatial and temporal variation in water quality: a case study: Suquia river basin (Cordoba, Argentina). *Water Research*, 35, 2881–2894.

[9] Hair, J. F., Anderson, R. E., Tatham, R. L. and William, C. (1995). *Multivariate data analysis with readings*. Englewood Cliffs, New Jercy.

[10] Panigrahi, S., Acharya, B. C., Panigrahy, R. C., Nayak, B. K., Banarjee, K. and Sarkar, S. K. (2007). Anthropogenic impact on water quality of Chilika lagoon RAMSAR site: a statistical approach. *Wetlands Ecology and Management*, 15, 113–126.

[11]Ledesma, R. D. and Valero-Mora, P. (2007). Determining the Number of Factors to Retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment, Research & Evaluation*, 12, 1–11.

[12] Pejman, A. H., Bidhendi, G. R. N., Karbassi, A. R., Mehrdadi, N., Bidhendi, M. E. et al. (2009). Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques. *International Journal of EnvironmentalScience and Technology*, 6, 467–476.

[13] Mustapha, A. and Aris, A. Z. (2012). Spatial aspects of surface water quality in the Jakara Basin, Nigeria using chemometric analysis. *Journal of Environmental Science and Health, Part A*, 47, 1455–1465.

[14] Harrad, S., Batty, L. and Diamond, M. (2008). *Student projects in environmental science*. John Wiley, New York.

[15] Olsen, R. L., Chappell, R. W. and Loftis, J. C. (2012). Water quality sample collection, data treatment and results presentation for principal components analysis. *Water Research,* 3110–3122.

[16] Sevilla, J. B., Lee, C. H. and Lee, B. Y. (2010). Assessment of spatial variations in surface water quality of Kyeongan stream, South Korea using multi-Variate statistical techniques. *Sustainability in Food and Water*, 18,39–48.

[17] Mustapha, A. 2012. Identification of anthropogenic influence on water quality of Jakara River, northwestern Nigeria. *Journal of Applied Sciences in Environmental Sanitation*, 7(1), 11-20.

[18]Tiri, A. and Boudoukha, A. 2010. Hydrochemical analysis and assessment of surface water quality in Koudiat Medour Reservoir, Algeria. European Journal of Scientific Reseach 41 (2): 273–285.

[19] Chabukdhara, M. and Nema, A. K. (2012). Assessment of heavy metal contamination in Hindon River sediments: A chemometric and geochemical approach. *Chemosphere*, 87, 945–953.
[20] Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, 68, 532–560.
[21] Leach, N. L., Barrett, K. C. and Morgan, G. A. 2005. *SPSS for intermediate statistics: Use and interpretation.* Lawrence Erlbaum, Mahwah, New Jersey.
[22] Liu, W. C., Yu, H. L. and Chung, C. E. (2011). Assessment of water quality in a subtropicalAlpine Lake using multivariate statistical techniques and geostatistical mapping: a case study. *International Journal of Environmental Research and Public Health*, 8, 1126–1140.
[23] Mustapha, A. and Nabegu, A. B. 2011. Surface water pollution source identification using principal component and factor analysis in Getsi River, Kano Nigeria. *Australian Journal of Basic and Applied Sciences*. 5(12): 1507–1512.
[24] Mustapha, A., Aris, A. Z., Juahir, H. and Ramli, M. F. 2012. Surface water quality contamination source apportionment and physicochemical charcterization at the upper section of the Jakara Basin, Nigeria, *Arabian Journal of Gosciences,* doi: 10.1007/s12517-012-0731-2
[25] Mustapha, A. (2008). An assessment of the suitability of water in Jakara-Getsi River system for fadama production at the Kano Region, Kano State. *Techo-Science Africana Journal*, 1(1), 118-124.
[26] Hinton, P. R. and Brownlow, C. (2004). *SPSS explained*. Routledge, New York.
[27] Armah, F. A., Obiri, S., Yawson, D. O., Pappoe, A. N. M. and Akoto, B. (2010). Miningand heavy metal pollution: Assessment of aquatic environments of Tarkwa, Ghana using multivariate statistical analysis. *Journal of Environmental Statistic*, 1, 1–13.
[28] Chau, K. and Muttil, N. (2007). Data mining and multivariate statistical analysis for ecological system in coastal waters. *Journal of Hydroinformatics*, 9, 305–317.
[29] Shrestha, S. and Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software*, 22, 464–475.
[30] Huang, F., Wang, X., Lou, L., Zhou, Z. and Wu, J. (2010). Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques. *Water Research*, 44, 1562–1572.
[31] Juahir, H., Zain, S. M., Aris, A. Z., Yusoff, M. K. and Mokhtar, M. B. (2010). Spatial assessment of Langat river water quality using chemometrics. *Journal of Environmental Monitoring*, 12, 287–295.
[32] Onojake, M. C., Ukerun, S. O. and Iwuoha, G. (2011). A statistical approach for evaluation of the effects of industrial and municipal wastes on Warri Rivers, Niger Delta, Nigeria. *Water Quality, Exposure and Health*, 3, 91–99.
[33] Vidal, M., López, A., Santoalla, M. C. and Valles, V. (2000). Factor analysis for the study of water resources contamination due to the use of livestock slurries as fertilizer. *Agricultural Water Management* , 45, 1–15.
[34] Yidana, S. (2009). The hydrochemical framework of surface water basins in southern Ghana. *Environmental Geology*, 57, 789–796.
[35] Ouyang, Y. (2005). Evaluation of river water quality monitoring stations by principalcomponent analysis. *Water Research*, 39,2621–2635.
[36] Xu, H., Yang, L. Z., Zhao, G. M., Jiao, J. G., Yin, S. X. and Liu, Z. P. (2009).Anthropogenic impact on surface water quality in Taihu Lake region, China. *Pedosphere,* 19 (6): 765–778.
[37] Zhou, F., Guo, H., Liu, Y. and Jiang, Y. (2007). Chemometrics data analysis of marine water quality and source identification in Southern Hong Kong. *Marine Pollution Bulletin*, 54 ,745–756.
[38] Koklu, R., Sengorur, B. and Topal, B. (2010). Water quality assessment using multivariate statistical methodsï£¡a case study: Melen River System (Turkey). *Water Resources Management*, 24, 959–978.
[39] Li, Y., Xu, L. and Li, S. (2009). Water quality analysis of the Songhua River Basin using multivariate techniques. *Journal of Water Resource and Protection*, 1,110–121.
[40] Altın, A., Filiz, Z. and Iscen, C. F. (2009). Assessment of seasonal variations of surface water quality characteristics for Porsuk Stream. *Environmental Monitoring and Assessment* 158, 51–65.
[41] Varol, M. and ¸Sen, B. (2009). Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. *Environmental Monitoring and Assessment,* 159, 543–553.
[42] Calijuri, M. L., Couto, E. A., Santiago, A. F., Camargo, R. A. and Silva, M. D. (2011). Evaluation of the influence of natural and anthropogenic processes on water quality in Karstic Region. *Water, Air and Soil Pollution*, doi: 10.1007/s11270-011-1012-5.
[43] Han, S., Kim, E. and Kim, S. (2009). The water quality management in the Nakdong River watershed using multivariate statistical techniques. *KSCE Journal of Civil Engineering*, 13, 97–105.
[44] Wong, W. S. (2005). Using multivariate statistical techniques to examine the spatial variability of physiochemical elements and water quality in Hong Kong's Rivers. *Asian Geographer,* 24,129–150.
[45] Tanrıverdi, Ç., Alp, A., Demirkıran, A. R. and Üçkarde¸s, F. (2010). Assessment ofsurface water quality of the Ceyhan River basin, Turkey. *Environmental Monitoring and Assessment*, 167, 175–184.
[46] Aris, A. Z., Praveena, S. M., Abdullah, M. H. and Radojevic, M. (2011). Statistical approaches and hydrochemical modelling of groundwater system in smalltropical island. *Journal of Hydroinformatics* doi: 10.2166/hydro.2011.072.
[47] Omo-Irabor, O. O., Olobaniyi, S. B., Oduyemi, K. and Akunna, J. (2008). Surface and groundwater water quality assessment using multivariate analytical methods: A case

study of the Western Niger Delta, Nigeria. *Physics and Chemistry of the Earth, Parts A/B/C*, 33, 666–673.
[48] Cloutier, V., Lefebvre, R., Therrien, R. and 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.
[49] Morell, I., Giménez, E. and Esteller, M. V. (1996). Application of principal components analysis to the study of salinization on the Castellon Plain (Spain). *Science of the total environment*,177, 161–171.
[50] Liu, C. W., Lin, K. H. and Kuo, Y. M. (2003). Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment*, 313, 77–89.