Hydrochemical Assessment of Surfacewater and Groundwater Quality at Bank Infiltration Site

M K N Shamsuddin[1], S Suratman[1], M F Ramli[1], W N A Sulaiman[1], and A Sefie[2]
[1]Hydrogeology Research Centre, National Hydraulic Research Institute of Malaysia, 43300 Seri Kembangan,Selangor, MALAYSIA
[2]Faculty of Environmental Studies, University Putra Malaysia, 43400 Serdang, Selangor, MALAYSIA

E-mail: nizar@nahrim.gov.my

Abstract: Groundwater and surface water quantity and quality are an important factor that contribute for drinking water demand and agriculture use. The water quality analysis was assessed using multivariate statistical analyses based on analytical quantitative data that include Discriminant Analysis (DA) and Principal Component Analysis (PCA), based on 36 water quality parameters from the rivers, lakes, and groundwater sites at Jenderam Hilir, which were collected from 2013 to 2014 (56 observations) . The DA identified six significant parameters (pH, NO_2^-, NO_3^-, F, Fe^{2+}, and Mn^{2+}) from 36 variables to distinguish between the river, lake, and groundwater groups (classification accuracy = 98%). The PCA had confirmed 10 possible causes of variation in the groundwater quality with an eigenvalue greater than 1, which explained 82.931% of the total variance in the water quality data set.

Keywords: Hydrochemical, water quality, multivariate analysis, principal component analysis, discriminant analysis.

1. Introduction

Bank Infiltration (BI) is a surface water (i.e. rivers and lakes) and groundwater process interact which surface water is forced to flow through aquifer into pumping well that are installed on the banks of river and lake [1]. BI which important systems for water supply in managed aquifer recharge (MAR), has been used for water supplies in Eroupe along the Rhine, Elbe, Danube, and Seine Rivers for over a centuryand it has provided the majority of the drinking water for large cities in Russia, USA, China and other countries over the last few decades [2].

The obvious advantage of BI is the conjunctive use of infiltrated surface water and groundwater from the alluvial catchments of intake structures, which ensure long-term productivity and stability of the water supply. Additionally, surface water contaminants can be significantly removed or degraded as the infiltrating water moves from the river/lake to the production wells due to a combination of physicochemical and microbiological processes [3, 4]. However, as noted by Schmidt et al. [5], poor surface water quality, heavy clogging of the riverbed, and accidental pollution have already greatly threatened BI systems. Importantly, the sustainability of BI is affected by particulate organic matter, which intensifies physical or chemical clogging of riverbed on various spatial and temporal scales [3]. The degree of these interactions significantly depends on meteorological, fluvial, anthropogenic, and geological processes [6]. Such interactions significantly affect the quantity and hydrochemical composition of water bodies on the local and regional scales. Therefore, the location of groundwater and surface water interaction should be identified and its mechanisms understood for effective environmental management. The hydrochemical composition of alluvial systems is largely influenced by precipitation, regional geology, aquifer–stream transfer, and anthropogenic contamination. Water quality offers insight into various flow pathways, geological settings, and physical and chemical processes in the water [7, 8, 9]. If water flux data are available, similarities in the water composition of neighboring groundwater and surface water bodies, such as in the iron ratio or the concentration of total dissolved solids (TDS), can be used to qualitatively or statistically infer groundwater–surface water interaction [e.g. 9,10,11], or to quantitatively estimate fractions of water from different sources or end-members with distinct hydrochemistry [5,12,13]. In Malaysia, the impact of pollutants on groundwater is highly influenced by surface water, the main source for treating domestic water supply. However, some aquifers are locally recharged, and a short period takes place between water entry into the aquifer and water distribution for consumption. In this case, the groundwater may retain surface water characteristics and contaminants. The pollution of rivers reduces their suitability and requires costly water treatment. It is very important to investigated the feasibility of BI as an alternative procedure for water demand purposely.BI methods improved the quality of surface water for drinking water use in Malaysia which is based on the percolation of river water through the ground into an aquifer.contaminants from the surface water are removed by filtration, adsorption, reduction, and biodegradation through percolation process [2,3,14].

Multivariate methods such as discriminant analysis (DA) and principal component analysis (PCA) have become increasingly popular in the qualitative and statistical analysis and interpretation of environmental data sets, especially where flux data are unavailable. These techniques simplify large data sets by grouping them into components or clusters based on the relationships between specified variables [9,15]. Both DA and PCA have been extensively used in hydrogeology to identify groundwater facies for the study of natural groundwater evolution [16,17,18,19,20] and anthropogenic contamination [18,19]. Several studies have also utilized multivariate statistics to classify surface water [21,22,23]. Guler and Tyhne [8] has classified both groundwater and surface water bodies on multi- and single-watershed scales respectively to characterize the hydrogeology of an area and identify localized areas of groundwater recharge and discharge. Kumar et al. [24] utilized a similar statistical methodology to identify the source of contaminated groundwater flowing to an urban reach of the Yamuna River in India, and to classify groundwater and surface water bodies according to hydrochemical similarities. This study aims to understand the hydrochemical influence on BI in the study area and to determine whether multivariate analysis can help to delineate the interaction between surface water and groundwater. It also aims to expand knowledge about this relation and convince water quality managers to use BI method to reduce certain parameters.

2. Methodology

2.1 Study Site

Located at the downstream confluence of the Langat and Semenyih rivers, the study area in Jenderam Hilir is a water reservoir surrounded by lakes. The flat Langat riverbank was chosen to test the effectiveness of BI. The study area is located in Selangor within the Langat Basin, extending between 2° 53′ 28.56″ N and 2° 53′ 39.75″ N and between 101° 42′ 03.78″ E and 101° 44′ 14.58″ E, covering an area of 10 km2 (Figure 1).
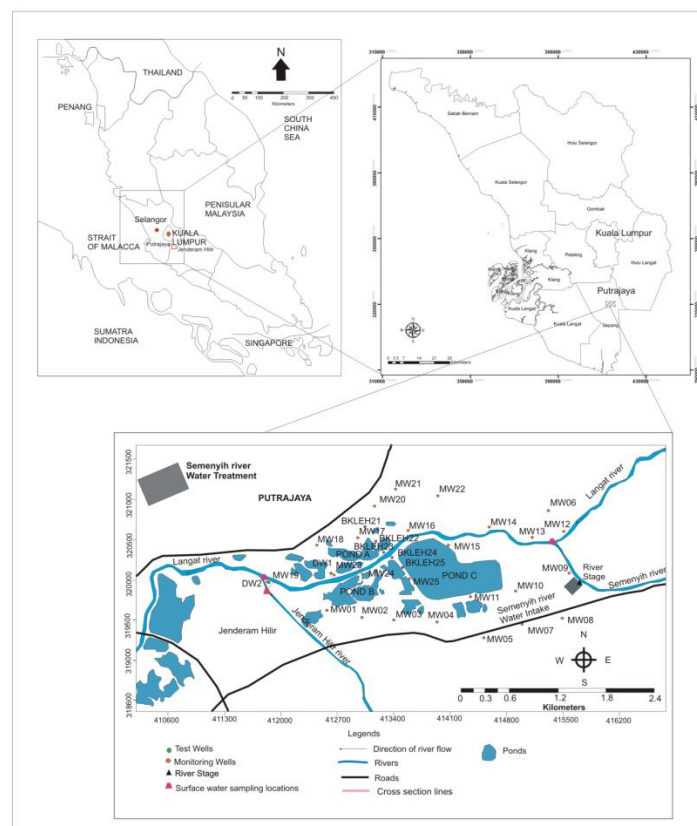
Figure 1. Location of study area, major river systems and location of 25 monitoring wells (MW) and test wells (DW).

2.2 Monitoring and Test Wells

Eight wells were drilled with a rotary drilling machine by wash boring the bedrock. Seventeen holes were drilled by using a motorized auger to a depth of 2 m to 3 m below the water table. Twenty-five monitoring wells (MW) were then constructed at the drilled locations, mostly on the downstream area of the confluence of the Semenyih and Langat rivers (Figure 1).

2.3 Data Collection and treatment

Fifty-six water samples were collected from rivers, lakes, and groundwater from December 2013 to October 2014 during rainy and dry season About 15 water samples collected from the upstream and downstream Langat and Jenderam Hilir Rivers were marked RW, and 13 water samples (LK) were obtained from three lakes (Ponds A, B, and C). Twenty-eight groundwater samples were also taken (from DW1 and DW2, the production well (PW1), and MW2, MW11, MW12, MW13, MW16, MW19, MW23, and MW25) and marked as GW. The sampling locations are shown in Figure 1. For groundwater sampling, the wells were pumped with submersible pumps for more than 30 min to reduce the volume of waste to dispose. Each sample bottle was filled with the water sample without entrapping air bubbles to prevent the biochemical and surface reactions during transportation and storage Each bottle was labeled with its corresponding sampling station and time and were kept at 4 °C to minimize microbial activity in the water [25].

The samples were prevent clogging during analysis with spectrometry instruments and to obtain dissolved ions for metal analysis [25]. The samples were then acidified with HNO_3 until pH<2 to prevent the precipitation of components, such as metal oxides and hydroxides, and retard biological activity. Calibrated Multi-parameter probes (SevenGo pro probe and SevenGo Duo pro probe, Mettler Toledo AG, Switzerland) were used to measure electrical conductivity (EC), TDS, and pH in situ.. Bicarbonate (titration using 0.02 N HCl) and chloride ions (argentometric method using 0.0141 N $AgNO_3$) were analyzed on-site by using unfiltered samples [25]. Meanwhile, the filtered samples were separated into two polyethylene bottles, one for the analysis of sulfate (SurfaVer 4 HACH method) and nitrate (NitraVer 5 HACH method) and the others to determine cations

and metals which were analyzed using flame atomic absorption spectrometry (FAAS, Shimadzu AA6800) for cations, inductively coupled plasma mass spectrometry (ICP-MS, ELAN DRC-e, Perkin Elmer) as for trace metal. The sampling, preservation, and transportation to the laboratory of the water samples were based on the Standard Method for Water and Wastewater Analysis [25]. All of the reagents were from analytical or equivalent grade and contaminant-free. All the laboratory equipments were pre-cleaned with 5% (v/v) concentrated nitric acid (HNO_3) and then rinsed with distilled water to ensure that any contaminants or traces of cleaning reagents were removed before the analysis. Polyethylene bottles which are free from metal-containing material were used to collect the water samples to minimize interference of heavy metal analysis[25]. The accuracy of the result was determined by replicating samples (n^2) with relative standard deviation and assessed by diluting ICP Multi-Element Mixed Standard III (Perkin Elmer) in a series of concentrations with the same acid mixture used for sample dissolution. The water quality parameters that were considered for the analysis are turbidity, color, pH, EC, manganese (Mn), iron (Fe), nitrate nitrogen (NO^3-N), nitrite nitrogen (NO_2-N), TDS, total solids, chloride (Cl), ammonium (NH4), fluoride (F), carbonate (CO_3), bicarbonate (HCO_3), sodium (Na), calcium (Ca), potassium (K), aluminum (Al), magnesium (Mg), mercury (Hg), cadmium (Cd), selenium (Se), arsenic (As), chromium (Cr), silver (Ag), copper (Cu), zinc (Zn), sulfate (SO4), silica (SiO_2), cobalt (Co), barium (Ba), strontium (Sr), nickel (Ni), and phosphorus (P). All of the concentrations were expressed in milligrams per liter (mg/L).

2.4 Data analyses
The data set was screened according to the charge balance error (CBE) proposed by Freeze and Cherry [7] to identify electrically unbalanced samples:

$$CBE = \frac{\sum \upsilon mc - \sum \upsilon ma}{\sum \upsilon mc + \sum \upsilon ma} x100\% \qquad (1)$$

Where;
  υ is the absolute value of the ionic valence,
  mc is the molarity of the cationic species, and
  ma is the molarity of the anionic species.

CBEs were calculated on each site by using the median concentration of the cations and anions. the inclusion of these samples in the calculations affecting the CBEs by less than 0.5%, the CBEs based on major ion concentrations were considered significant for this study. Although 36 water quality parameters were available, only 28 parameters consistently included in 56 sample data sets were chosen for multivariate analysis (data matrix: 28 × 56)The statistical analysis for water quality data sets were subjected to univariate analysis; range, mean, and standard deviation and multivariate analysis; and DA and PCA. DA was first applied to the groups or clusters among the stations, followed by PCA to extract and distinguish potential factors or sources of pollution contributing to variations in water quality measures.

2.5 Analytical Methods
DA is a multi-component statistical technique used to classify objects into mutually exclusive and exhaustive groups according to a set of independent variables.and and builds a discriminant function (DF) that operates on raw data for each group [26,27,28]:

$$f(\text{Gi}) = k_i + \sum_{j=1}^{n} \left(w_{ij} P_{ij} \right) \qquad (2)$$

Where;
  i is the number of groups (G),
  ki is the constant inherent to each group,
  n is the number of parameters used to classify a set of data into groups, and
  wj is the weight coefficient assigned by DF analysis (DFA) to a given parameter (pj).

DA was used to determine whether the groups differed in terms of the mean of a variable and use such variable to predict group membership [26]. All of the parameters (n = 28) were employed in the DA using the

standard, forward-stepwise, and backward-stepwise modes. In the forward-stepwise mode, the variables were included step by step, beginning with the most significant variable until no significant changes were obtained.

PCA is an analytical technique for multivariate data.that reduces many variables to few variables, without sacrificing too much of the information. PCA provides information on the most significant parameters based on spatial and temporal variations that describe the whole data set by excluding the less significant parameters with minimum loss of original information [4,27,29,30]. The principal component (PC) can be expressed as

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + ... + a_{im}x_{mj} \qquad (3)$$

where
    z is the component score,
    a is the component loading,
    x is the measured value of the variable,
    i is the component number,
    j is the sample number, and
    m is the total number of variables

The PCs generated by PCA are sometimes not readily interpreted; therefore, rotating the PCs by varimax rotation is advised. The VF coefficients with correlations greater than 0.75 are considered as "strong" significant factor loadings, 0.75 to 0.50 as "moderate," and 0.50 to 0.30 as "weak" [31]. The PCA explains relationships between measured variables, and thus be used to infer the hydrogeochemical processes that control water chemistry[32,33]. The PCA takes the data organized in a correlation matrix and rearranges them in a manner that better explains the structure of the underlying system that produced the data. Therefore, the matrix of correlation coefficients measures how well the variance of each constituent can be explained by its relationships with each of the variables.

3. Results and Discussion

3.1 Hydrochemical Characteristics
The descriptive statistics of the measured three-year data set are summarized in Table 1, including the univariate (mean, standard deviation, minimum, and maximum) overview of the chemistry of surface water and groundwater in the study area. The pH of the groundwater sample analyzed in this study within the permissible limit of drinking water standard of 5.5 to 9.0 [34]. Samples from surface waters show pH values that the mildly acidic surface water significantly increased the pH values of the groundwater. Meanwhile, EC for groundwater in the study area varies from about 62 µS/cm to 166 µS/cm with the mean value of 118 µS/cm. Low conductivity was observed in the Langat River upstream (SL1) and downstream (SL2) of the study area. High turbidity value for surface water in the study area ranges between 328 and 699 NTU. The groundwater samples in the study area show low turbidity, indicating that BI in Jenderam Hilir can significantly reduce the turbidity value. Sulfate concentration in the groundwater lies between 3.0 and 14.0 mg/l while in the surface water it ranges from 3.0 mg/l to 15.0 mg/l. Nitrate in natural water is due to organic sources or industrial and agricultural chemicals. Although nitrogen is an essential constituent of protein in all living organisms but the presence of , nitrate with a concentrations greater than 50 mg/l can cause cyanosis or blue baby syndrome among infants [35] and cancer in adults (WHO, 2008). The permissible limit for NO^{3-} in raw water is 10 mg/l (DSM, 2010).

The NO^{3-} concentration in groundwater and surface water in the study area varies from 1.80 mg/l to 39.0 mg/l. In Jenderam Hilir, high NO^{3-} concentration of 39.0 mg/l can be observed at MW16 for the groundwater and 16.0 mg/l at SL1 for the surface water. The high nitrate concentration of nitrate could have resulted from the seepage of the liquids due to the occurrence of sewage and septic tanks, industrial effluents, and agricultural chemicals nearby. As MW16 is near the oil palm estate. Meanwhile, all of the groundwater samples from eight wells showed Fe^{2+} concentration higher than the drinking water standard, i.e. 0.3 mg/l. The Fe^{2+} concentration of groundwater was within 14.0 mg/l to 39.0 mg/l. Fe is a common constituent of various primary minerals such as biotite, pyroxenes, and amphiboles. The geochemical conditions of the study area are assumed to result in a predominant distribution of ferrous and ferric irons, though the definite existence of these minerals in the area has not been confirmed. The Ca^{2+} concentration ranges from 1.50 mg/l to 32.0 mg/l. The permissible limit of

Ca^{2+} concentration is 75 mg/l in drinking water standard (WHO, 2008). The Na^+ concentration in the wells and surface water ranges is below the permissible limit of the Malaysia Drinking Water Standard (DSM, 2010). Ionic K^+ occurs at fairly low concentrations in groundwater, and the major sources of K concentration in the study area could be from rainfall, rivers, potash feldspar, and micaceous minerals. The K^+ concentration is in the range of show low K^+ concentrations.

3.2 Surface and Ground Water Classification

Figure 2 illustrates the Piper trilinear diagram for the data obtained from the chemical analysis of surface water and groundwater samples from the study area. The diagram consists of two lower triangles that show the percentage distribution, on the milliequivalent basis, of the major cations (Ca^{2+}, Mg^{2+}, Na^+, and K^+) and major anions (SO_4^{2-}, Cl^-, CO_3^-, HCO_3)..

Three chemical types of surface water and groundwater in the study area were recognized as (i) Calcium–Sodium–Sulfate–Chloride–Bicarbonate facies (Ca^{2+}–Na^+–SO_4—Cl^-–HCO_3^-), (ii) Calcium–Chloride–Bicarbonate facies (Ca^{2+}–Cl^-–HCO_3^-) and, (iii) Calcium–Magnesium–Chloride–Bicarbonate facies. The Ca^{2+}–HCO_3^- water type prevails in all the groundwater and surface water samples. The major cations dominant in the groundwater are Ca^{2+} and Mg^{2+}. In well MW16, the Na^+ and K^+ ions are dominant compared to the other major cations. The water classification shows that the surface water and groundwater in this study area are connected.
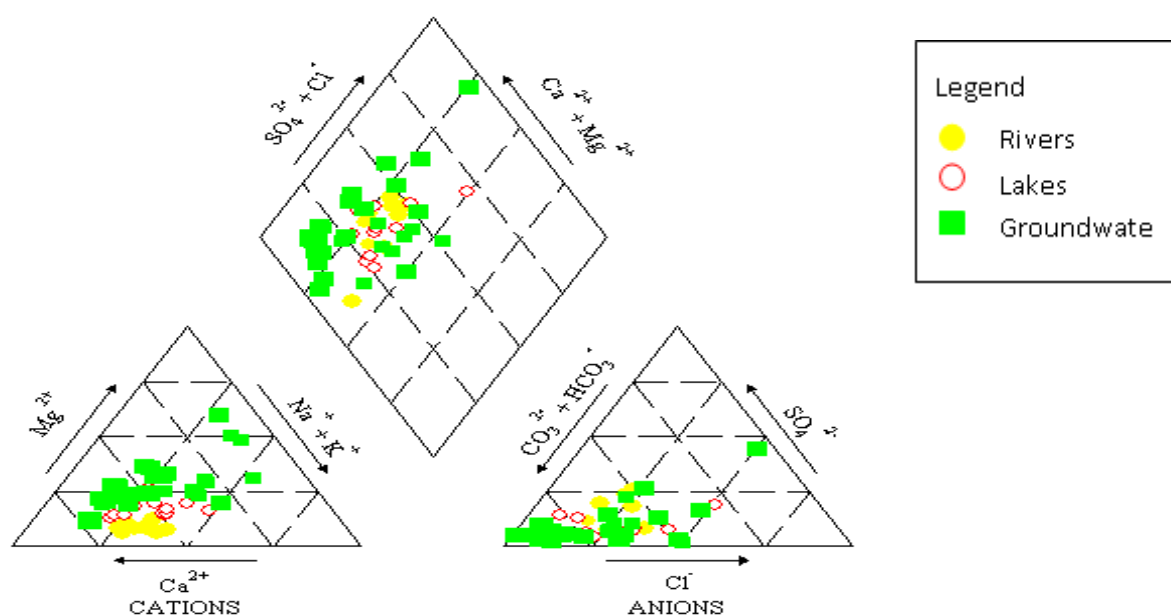

Figure 2. Piper trilinear diagram for the data obtained from chemical analysis of river, lake and groundwater sample from Jenderam Hilir.

Table 1: Descriptive statistics (mean, SD (standard deviation), minimum and maximum) of the surface water quality of Jenderam Hilir.

Parameter	pH	Turbidity	EC	TS	TDS	CO_3	Cl	SO_4	NO_2	NO_3	F	P	NH_4	Ag	Al
Unit		NTU	µS/cm	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l
River (N=9)															
Mean	6.84	199.8	68.2	167	61	0.67	10.3	7.67	0.66	9.05	0.3	0.17	0.57	0.1	1.6
SD	0.39	229.92	60.22	98	23	0.24	6.23	3.69	1.35	4.25	0.1	0.26	0.75	0.1	2.7
Min	6.3	8	2	60	16	0.5	4	1.5	0	3.3	0.3	0.01	0.25	0	0.2
Max	7.7	699	161	370	98	1	24	15	3.9	16	0.5	0.81	2.4	0.3	11
Lakes (N=13)															
Mean	6.94	8.44	60.31	12.2	53	22.3	0.62	29.9	7.69	2.85	0	1.17	0.31	0.2	0.3
SD	0.18	9.3	39.94	12.6	27	11	0.22	12.9	4.23	1.69	0	0.91	0.11	0.3	0.2
Min	6.7	0.5	2	3.1	20	4	0.5	3	2	1.5	0	0.25	0.25	0	0.3
Max	7.3	25	147	38	108	40	1	53	18	7	0.1	2.6	0.5	0.8	00.7
Groudnwater (N=28)															
Mean	6.46	158.92	84.85	20/7	177	55.6	0.65	48.3	9.5	4.15	0.9	2.57	0.32	0.3	2
SD	0.48	223.14	67.14	16.2	195	27.2	0.23	29.8	7.57	3.94	4.6	7.51	0.12	0.3	4.7
Min	5.6	10	2	4.6	44	22	0.5	0.5	0.5	1.5	0	0.25	0.25	0	0.3
Max	8.1	1196	228	47	150	150	1	124	36	14	24	39	0.5	0.9	25

Parameter	As	Ba	Ca	Co	Cr	Cu	Fe	Hg	K	Mg	Mn	Na	Ni	Pb	Sr
Unit	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l	mg/l
River (N=9)															
Mean	0	0.1	11	0	1.3	0.1	2.7	0.6	3.6	0.9	0.32	5.2	0.1	0	0
SD	0	0	5.1	0	4.9	0	1.6	0.3	1	0.3	0.48	2.8	0	0	0
Min	0	0.1	6.1	0	0	0.1	1.4	0.1	2.1	0.5	0.05	2	0.1	0	0
Max	0	0.1	24	0.1	19	0.1	7.9	1	4.8	1.5	1.5	10	0.1	0	0.1
Lakes (N=13)															
Mean	0.1	0.1	0	0.1	7.6	0	0.1	1.5	1.3	6.7	1.45	0.3	3.1	0.1	0
SD	0.1	0.1	0	0	3.4	0	0	1.7	2.5	12	0.46	0.2	0.9	0	0
Min	0	0.1	0	0.1	3.2	0	0.1	0.1	0.1	2.5	0.9	0.1	1.5	0.1	0
Max	0.3	0.3	0	0.1	15	0.1	0.1	4.9	9	46	2.6	0.6	4.9	0.1	0.1
Groudnwater (N=28)															
Mean	0.1	0.6	0.1	0.7	9.5	0	0.1	20	1.1	4.6	3.38	1.4	3.8	0.1	0.3
SD	0.1	1.6	0.1	1.7	7.2	0	0	9.7	1.6	2.3	1.6	1.5	3.6	0	1
Min	0	0.1	0	0.1	0.9	0	0.1	4.9	0.1	1.9	1.3	0.1	1	0.1	0
Max	0.3	7.7	0.2	7.9	32	0.1	0.1	39	8	12	7.8	6	20	0.1	5.2

3.3 Surface and Groundwater Spatial Pattern Discrimination
In this study, the DFs were performed based on the raw data from three types of water samples, and similar classification matrixes (CMs) treated as dependent variables while water quality parameters were treated as independent variables The accuracy of classification using the standard, forward stepwise, and backward stepwise mode DAs were 100% (28 discriminant variables), 92% (6 discriminant variables), and 98% (13 discriminant variables), respectively (Table 1). The Wilk's Lambda test for standard mode gave a Lambda value of 0.027. The null hypothesis states that the means of vectors of the three groups (RW, LK, and GW) are similar to one another. The alternative hypothesis states that at least one of the means of vectors is different from one another. Since the enumerated p-value is lower than the significant alpha level (i.e. 0.005), one should reject the null hypothesis and accept the alternative hypothesis. The risk of rejecting the null hypothesis while it is true is lower than 0.001%. Thus, the three groups of samples are indeed different from one another because of their different characteristics. DFs and CMs obtained from standard, forward stepwise, and backward stepwise modes of DA are shown in Table 2.

Table 2. Classification matrix for DA of spatial variations in surface water and groundwater interaction study at Bank.

Samples	Samples assigned by DA			Total	% correct
	GW	LK	RW		
Sampling DA mode (28 variables)					
GW	27	0	0	27	100.00%
LK	0	13	0	13	100.00%
RW	0	0	10	10	100.00%
Total	27	13	10	50	100.00%
Forward stepwise mode (6 variables)					
GW	26	1	0	27	96.30%
LK	0	13	0	13	100.00%
RW	0	3	7	10	70.00%
Total	26	17	7	50	92.00%
Backward stepwise mode (13 variables)					
GW	27	0	0	27	100.00%
LK	0	13	0	13	100.00%
RW	0	1	9	10	90.00%
Total	27	14	9	50	98.00%

Figure 3 presents the discriminant variables by DA into three groups, RW, LK, and GW, with 100% correct assignments. Using the forward stepwise mode, DA, pH, NO_2^-, NO_3^-, F^-, Fe^{2+}, and Mn^{2+} were found to be the significant variables. This result indicated that these parameters, which possess high variation, constituted the anaerobic redox parameters containing Fe^{2+} and Mn^{2+}. In this study, redox process play an important role in changing the concentration of Mn^{2+} and Fe^{2+}, which represent the minor portion of reduction processes in the aquifer, whereas the presence of NO_3^- is highly correlated with the existing major ion concentrations in the aquifer. In the backward stepwise mode of DA, 13 parameters (pH, TDS, EC, CO_3^-, HCO_3^-, NO_2^-, NO_3^-, P, NH_4^-, Al, Fe^{2+}, Mg^{2+}, and Na^+) were considered as the most significant parameters (p<0.05), as shown in their high spatial variation in the data set.

As identified by DA (backward step and forward step modes), box and whisker plots of selected discriminating parameters were constructed to evaluate different patterns associated with different water sample sources. Figure-3 shows the box and whisker plots of six variables of these water quality parameters during the period December 2013 to October 2014. The trend for NO_3^- and Fe^{2+} suggested that the average concentration in rivers was the highest, followed by the concentration in lakes. In groundwater, the average concentration was the lowest, while Mn^{2+} had the inverse trend. Within groundwater, most of the Mn^{2+} detected was naturally

occurring. Six parameters that produced high variations based on the forward stepwise DA were used for further analysis.
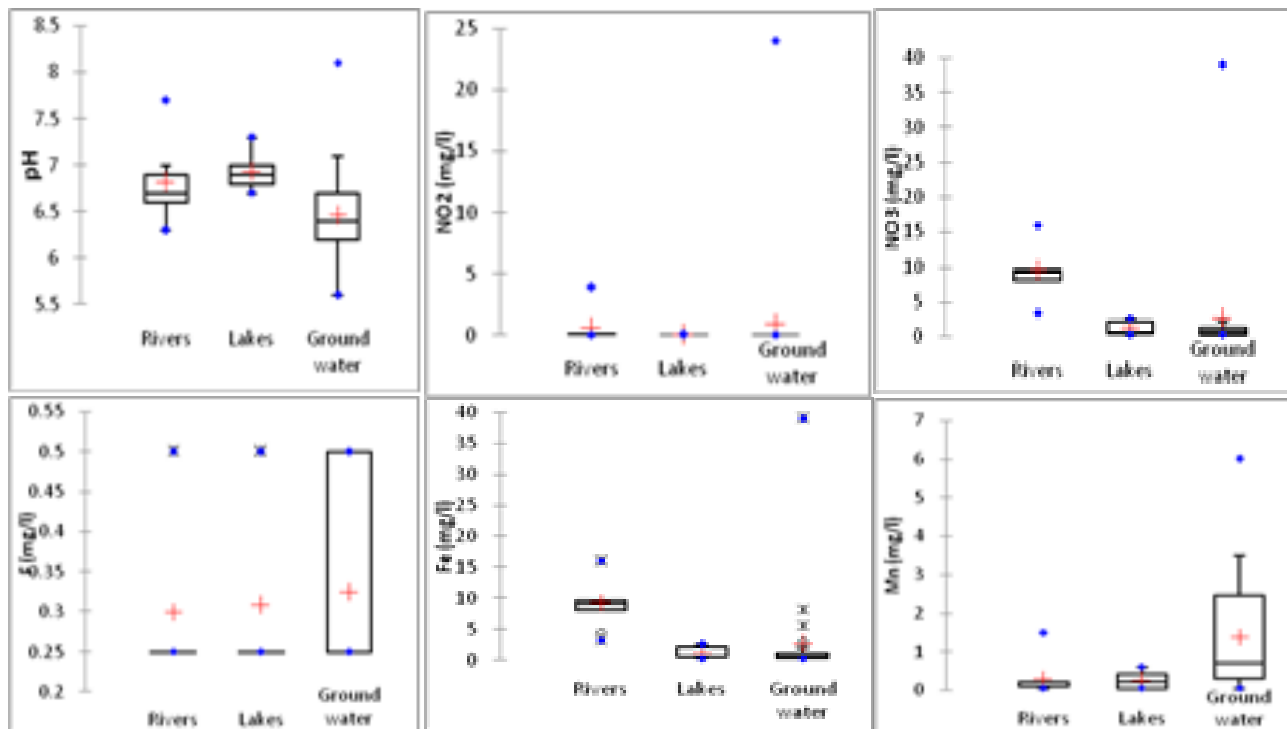

Figure 3. Box and whisker plots of some parameters separated by DA associated with the water quality data of Langat and Jenderam Hilir rivers, lakes and groundwater.

3.4 Correlation between Variables

The first step in factor analysis (FA) is to determine the parameter correlation matrix. The correlation matrix is used to account for the degree of mutually shared variability between individual pairs of water quality variables. PCA only considers the common hydrochemical variables of all samples taken from rivers, lakes, and groundwater. The correlation matrix of the 28 variables (Table-4) allows us to distinguish several relevant hydrochemical relationships (indicated by underlined values): (a) high and positive correlations between turbidity, total solid, NO^{2-}, NO^{3-}, NH_4, Al^{3+}, C^{2+}, and SO_4^{2-} (r=0.504–0.821) are attributed to river water hydrochemistry; (b) strong and positive correlations between HO^{3-}, F, P, As, Fe^{2+}, $Mg^{2+,}$ Na^+, and Cl^- represent groundwater influence toward rivers (r = 0.516–0.905); and (c) high and positive K and Hg are characterized toward the water chemistry of lakes (r = 0.604). However, pH, SiO_2, total solid, HO_3, P, Ag, and Al have weak correlations with all the variables, which indicates that all variables have different origins. Only Ag shows good correlation with conductivity, due to the increase of conductivity along with the dissolution of metals through ion exchange or oxidation–reduction reaction in a groundwater aquifer system. Na shows moderate correlation with Cl (0.68), indicating that a major source of these elements was influenced by marine deposits (Kenny Hills Formation), a process during groundwater flow through rock in the study area. The Na^+ and Cl^- contents detected in certain samples may also suggest the dissolution of chloride salts. The dissolution of halite in water release is equal to the concentrations of Na^+ and Cl^- in the solution.

3.5 Source Identification Using PCA and Factor Analysis (FA)

PCA was employed in this study to investigate the compositional pattern of common hydrochemical variables of all samples taken from rivers, lakes, and groundwater. PCA then strengthens its role in identifying the dominant process that controls the chemical components in both surface water and groundwater. An eigenvalue gives a measure of the significance factor, where the highest eigenvalue resembles the most significant. Eigenvalues greater than 1 are considered significant. Table-3 summarizes the PCA results, including the loadings, eigenvalue, variance contribution rate of each Varimac Factors (VF), and cumulative variance contribution rate. The loadings with significant absolute value in each VF are highlighted.

To reduce the overlap of original variables over each principal component, a varimax rotation was conducted. The 10 factors are extracted to represent the contributions that influence the chemical composition of surface water and groundwater. The varimax rotated Factor Analysis (FA) results are presented in the Table-4 with probable source types. In the Table-4, the eigenvalue is set to 1.0 as a threshold in order to limit the number of extracted factors. Factor loadings equal to or greater than 0.5 are considered significant. Ten factors obtained explain 83% of the total variance.

In Table 3, VF 1 explains 17.171% of the variance, which is highly attributed to As, Fe, and Mg. This factor suggests the significant contribution of groundwater hydrochemistry and of redox conditions. VF2 elucidates 13.836% of the total variance, which has strong positive loading of F, P, and CO_3. This VF represents, during pre-monsoon, the high loading factor of F concentration. The important source of F in groundwater is fluoride-bearing minerals such as fluorite, apatite, amphiboles, and micas, which may be due to the ion exchange of F, leaching of F-containing minerals, higher evapotranspiration, and longer residence time of water in the aquifer. These factors are also significant in groundwater flow. VF3 describes 9.59% of the total variance provided by a strong positive loading of color, turbidity, and total solid, which represents the surface runoff of sources. VF4 (8.715% of total variance) has a strong positive loading of EC, Sr, and SiO_2. This VF can be explained by the groundwater contact with rock or weathered rock (Kenny Hills formations).

The hydrolysis and weathering of ranging silicate minerals to clay minerals such as kaolinite and illite are also explained by the fourth factor. Meanwhile, VF5, which contains 8.257% of the total variance, includes TDS, Ca, NH_4, and SO_4 as the highest mean concentrations found in river water and groundwater interaction, as well as in the discharge of untreated sewage and industrial waste into rivers. The association of TDS with higher Ca concentrations is caused by tanneries that use calcium carbonate to process raw hides and their effluent discharges into the river course. The substances in this VF are characteristic of input from surface water. They represent the substantial influence of surface water of the Langat River infiltration and control the groundwater quality. VF6 with 6.948% of the total variance has strong positive loadings of Na+ and Cl-, which indicate the groundwater interaction with rock (Kenny Hills formation) and are deposited into the marine environment. VF7 explains 5.638% of the strong positive loading of Al, NO_2, and NO_3, which represent agricultural wastes based on nitrate found in the upstream of Jenderam Hilir.

Table 3. Pearson correlation coefficients for 13 hydrogeochemical variables of water samples. Coefficients are significant at the 0.05 level and those higher than 0.70 were shown as bold fonts.

Variables	pH	Colour	Tur	EC	TS	TDS	Na	K	Mg	Ca	Mn	Fe	Al	Hg	Cl	F	P	Ag	As	Cr	Sr	SiO$_2$	CO$_3$	HCO$_3$	NH$_4$	SO$_4$	NO$_2$
pH																											
Colour	-0.01																										
Tur	-0.15	**0.65**																									
EC	0.07	0.21	0.25																								
TS	-0.16	**0.68**	**0.73**	0.3																							
TDS	-0.23	0.03	0.18	0.39	0.29																						
Na	-0.15	-0.14	-0.05	-0.04	0.04	0.15																					
K	0.06	-0.09	-0.11	0.12	-0.08	0.08	-0.04																				
Mg	-0.2	-0.02	0.09	0.39	0.17	0.43	0.23	0.09																			
Ca	0.08	-0.21	-0.09	0.49	0.06	**0.72**	0.08	0.14	0.28																		
Mn	-0.17	-0.12	0.06	-0.03	0.08	0.14	-0.14	-0.02	0.36	0.07																	
Fe	-0.23	0.15	0.25	0.43	0.3	0.33	0.09	-0.03	**0.8**	0.05	0.25																
Al	0.01	-0.06	0.18	0.14	0.23	0.03	0.1	-0.07	-0.05	0.02	-0.16	0.17															
Hg	-0.01	-0.06	-0.13	0.03	-0.09	-0.11	0.2	**0.64**	0.04	-0.18	-0.11	0.01	-0.03														
Cl	-0.19	-0.22	-0.09	-0.18	0.02	0.02	**0.68**	-0.05	0.27	-0.03	-0.11	0.12	-0.01	0.09													
F	0.2	-0.11	-0.16	0.33	-0.03	0.01	0.05	-0.08	0.06	0.2	-0.16	0.27	0.38	-0.04	-0.04												
P	0.19	-0.09	-0.1	0.41	0.01	0.05	0.04	-0.06	0.09	0.22	-0.14	0.31	0.4	-0.02	-0.1	**0.91**											
Ag	-0.02	-0.17	-0.1	**-0.78**	-0.14	-0.02	0.08	-0.14	-0.12	-0.11	0.02	-0.25	-0.13	-0.29	0.25	-0.37	-0.44										
As	-0.1	0.19	0.25	0.15	0.21	0.14	-0.04	-0.07	**0.51**	-0.08	0.2	**0.74**	0.16	-0.11	0.07	0.27	0.29	0.01									
Cr	0.11	-0.03	0.4	0.06	-0.04	-0.17	-0.06	-0.05	-0.16	-0.06	-0.08	-0.1	0.14	-0.08	-0.07	-0.09	-0.02	0	-0.09								
Sr	-0.17	-0.04	0.03	-0.15	-0.04	0.04	0.01	0.02	0.2	0.04	0.24	0.1	-0.05	-0.04	0.06	-0.02	-0.05	0.19	0.26	-0.03							
SiO$_2$	-0.34	0.12	0.1	0.21	0.04	0.21	-0.06	0.26	0.22	-0.07	0.22	0.16	-0.06	0.35	-0.06	-0.38	-0.28	-0.4	-0.12	-0.14	-0.13						
CO$_3$	0.19	-0.12	-0.14	0.27	0.01	0.1	0.14	-0.08	0.02	0.24	-0.18	0.23	0.38	-0.05	0.06	**0.95**	**0.86**	-0.3	0.25	-0.09	-0.03	-0.38					
HCO$_3$	0.23	0.2	0.25	0.33	0.37	0.49	0.13	0.05	0.29	0.34	0.03	0.28	-0.05	-0.06	0.14	-0.09	-0.06	-0.01	0.16	-0.03	-0.15	0.07	0				
NH$_4$	-0.11	-0.01	0.13	0.35	0.08	**0.57**	-0.02	0.17	**0.58**	0.41	0.08	0.32	-0.08	-0.02	-0.01	-0.14	-0.12	-0.1	-0.01	-0.04	-0.04	0.31	-0.15	0.36			
SO$_4$	0.01	-0.06	0.09	0.12	0.08	0.45	0.18	0.03	0.03	**0.52**	-0.06	-0.09	0.19	-0.17	0.08	0.16	0.17	0.15	-0.12	0.02	0.11	-0.12	0.21	0.16	0.23		
NO$_2$	0.02	-0.03	-0.01	0.14	-0.04	-0.01	0.01	-0.05	0.1	-0.17	-0.11	0.32	**0.5**	-0.01	0.01	0.22	0.23	-0.1	0.35	-0.02	-0.03	0	0.22	-0.04	-0.03	-0.02	
NO$_3$	0.08	-0.06	0.04	0.16	0.03	0.12	0.18	-0.09	-0.12	0.04	-0.2	0.1	**0.61**	-0.08	0.08	0.34	0.35	-0.11	0.19	0.1	-0.04	-0.13	0.42	0.06	-0.1	0.31	**0.82**

Table 4. Loading of 28 hydrochemical variables on nine significant varifactors (VF).Coefficients are significant at the 0.05 level and those higher than 0.50 were shown as bold fonts.

Variables	VF1	VF2	VF3	VF4	VF5	VF6	VF7	VF8	VF9	VF10
pH	-0.163	0.225	-0.117	-0.093	-0.046	-0.221	-0.004	**0.78**	0.087	0.136
Colour	0.015	-0.066	**0.877**	0.085	-0.129	-0.16	-0.049	0.09	-0.015	-0.122
Tur	0.145	-0.132	**0.836**	0.036	0.06	-0.039	0.058	-0.075	-0.066	0.397
EC	0.278	0.339	0.225	**0.633**	0.381	-0.137	0.032	0.101	0.088	0.149
TS	0.129	0.022	**0.894**	0.058	0.144	0.072	0.008	-0.03	-0.058	-0.053
TDS	0.249	-0.046	0.158	0.087	**0.842**	0.061	0.058	-0.04	-0.021	-0.168
Na	0.022	0.082	-0.015	-0.02	0.101	**0.889**	0.047	-0.065	0.076	-0.002
K	0.012	-0.075	-0.078	0.011	0.151	-0.092	-0.034	0.075	**0.893**	-0.01
Mg	**0.864**	-0.021	-0.024	0.123	0.293	0.199	-0.088	-0.046	0.057	-0.034
Ca	0.031	0.227	-0.116	0.064	**0.874**	-0.023	-0.132	0.081	-0.018	0.015
Mn	0.46	-0.184	-0.096	-0.1	0.103	-0.273	-0.202	-0.328	-0.08	0.001
Fe	**0.883**	0.178	0.185	0.176	0.047	0.078	0.143	-0.046	-0.019	-0.049
Al	-0.05	0.324	0.15	0.034	0.053	0.058	**0.694**	-0.153	-0.011	0.176
Hg	-0.021	-0.022	-0.048	0.199	-0.191	0.189	-0.022	-0.033	**0.866**	-0.063
Cl	0.155	-0.052	-0.09	-0.138	-0.018	**0.884**	0.01	-0.012	-0.006	-0.024
F	0.089	**0.952**	-0.075	0.063	0.027	-0.01	0.133	0.032	-0.035	-0.045
P	0.114	**0.916**	-0.036	0.141	0.053	-0.051	0.157	0.015	-0.011	0.021
Ag	-0.093	-0.39	-0.123	-0.782	0.018	0.177	0.003	0.11	-0.234	-0.065
As	**0.774**	0.236	0.226	-0.245	-0.143	-0.059	0.212	0.026	-0.044	-0.068
Cr	-0.091	-0.06	0.056	0.017	-0.062	-0.017	0.056	0.06	-0.051	**0.953**
Sr	0.287	0.057	-0.003	**-0.601**	0.083	-0.107	-0.097	-0.407	0.165	0.101
SiO$_2$	0.13	-0.458	0.044	**0.582**	0.079	-0.068	0.046	-0.328	0.3	-0.164
CO$_3$	0.041	**0.917**	-0.04	0.008	0.092	0.097	0.177	0.061	-0.043	-0.076
HCO$_3$	0.272	-0.14	0.311	0.09	0.433	0.163	0.001	**0.59**	-0.005	-0.065
NH$_4$	0.36	-0.279	-0.046	0.284	**0.605**	-0.021	-0.034	0.097	0.05	0.029
SO$_4$	-0.235	0.165	0.072	-0.247	**0.733**	0.105	0.176	-0.1	-0.008	0.066
NO$_2$	0.235	0.075	-0.072	0.052	-0.131	-0.053	**0.905**	0.043	-0.009	-0.056
NO$_3$	-0.054	0.251	0.015	-0.022	0.115	0.089	**0.902**	0.034	-0.04	0.062
Eigenvalue	4.802	3.877	2.687	2.44	2.312	1.945	1.579	1.272	1.195	1.112
Variability (%)	17.149	13.847	9.595	8.715	8.257	6.948	5.638	4.543	4.267	3.971
Cumulative %	17.149	30.996	40.591	49.306	57.563	64.511	70.149	74.692	78.959	82.931

Al possesses higher concentrations derived from mining waste [36]. This VF describes the lakes–groundwater–river water interaction. The presence of Al ions in the stream may result from industrial wastes, but is more likely to come from the wash water of drinking water treatments plants. This VF8 explains 4.543% of the strong positive loading of pH and HCO3. The relationship between HCO3 and pH reveals that the CO3–CO2 equilibrium is an important pH buffer system in the surface water and groundwater of the study area. HCO3 mainly originates from the dissolution of carbonate rocks present in the aquifer (Kumar et al. 2011), and from the reaction of CO2 with silicate minerals. These factors are strongly related to surface runoff. VF9 explains 4.267% of the total variances of high K and Hg loadings. These components reflect the signatures of natural water recharge and water–soil–rock. The existence of K is mainly due to the rock–water interactions on potassium bearing feldspars and clay minerals, such as illlite and biotite-rich minerals. Hg describes the lakes and groundwater interaction from a former mining area. Past mining pollution had strong positive loadings on Hg, representing heavy metal pollution that may have originated from metal smelting. The last principal component, VF10 (3.91% of the total variance), contributes high and positive Cr, a specific pollutant. This finding provides evidence of industrial pollution such as dyeing or paint-processing operations.

4. Conclusion

The result of the statistical analysis offers an overview of the main process responsible for the characteristic of the BI site in Jenderam Hilir; (i) The DA gave the best result of the redox parameters that contain Fe and Mn under anaerobic conditions, (ii) This study also found that the redox process had a significant role in influencing the variation of hydrogeochemistry concentrations in the studied aquifer systems, (iii) Mainly, the quality of groundwater, which was near the river bank of the Langat River and the lakes in Jenderam Hilir was affected by quality changes due to bank filtration, (iv) The study of rivers–lakes–groundwaters interactions at different sampling sites yielded an important data reduction, where six parameters (pH, NO2, NO3, F, Fe, and Mn) were close to 98% correct assignment, (v) On the groundwater flow paths, the effects of surface water fluctuations decreased with increasing distance from the river bank due to dilution, adsorption, reduction, and biodegradation, (vi) The variability of the concentrations in groundwater was also dependent on the changing water levels in the Langat River and the lakes, (vi) High water levels led to high infiltration potentials and less time for purification processes due to the shorter residence time in the aquifer. On contrary, the effects of back-flooding during high water levels were not significant. The temporal (residence time) and spatial dimension (distance of flow path) of the aquifer passage determined the balance in groundwater quality for its use as drinking water, (vii)The quality of Jenderam Hilir groundwater is mainly dependent on the aquifer`s material type, material quantity, and quality of infiltrating surface water from the Langat River.

References

[1]   Ray C, Grischek J, Schubert JZ, Wang and TF, Speth (2002) A perspective of riverbank filtration. J Am Water Works Assoc 94(4); 149-160

[2]   Tufenkij N, Joseph NR and Menachem E (2002) The Promise of Bank Filtration – A simple technology may inexpensively clean up poor-quality raw surface water. Environmental Science and Technology 1: 423-428

[3]   Hiscock, KM and Grischek T (2002) Attenuation of groundwater pollution by bank filtration. J Hydrol 266:139–144

[4]   Singh KP, Malik A, Singh VK et al. (2005) Chemometric analysis of groundwater quality data of alluvial aquifer of Gangetic plain, North India[J]. Anal Chim Acta, 550: 82– 91.

[5]   Schmidt A, Gibson JJ, Santos R, Schubert M Tattrie K and Weiss H (2010) The contribution of grounfdwater discharge to the overallof to typical boreal lakes in alberta /canada estimated from a radon mass balance. Hydrol Earth Syst Sci. 14-7989, doi;10.5194/hess-14-79-2010

[6]   Winter TC, Harvey JW, Franke OL and Alley WM (1998) Groundwater and surface water, in: A single resources, edited by; Survey, U.G. US government printing ofiice, Denver, Colorado.

[7]   Freeze RA and Cherry JA (1979) Groundwater, Prenyice-Hall, New Jersey

[8]   Guler C and Thyne GD (2004) Delineation of hydrochemical facies distribution in a regional groundwater system by means of fuzzy c-means clustering Water Resour Res 40: 1-11

[9]   Kumar PJS, Jegathambal P and James EJ (2011) Multivariate and geostatistical analysis of groundwater quality in Palar River Basin International Journal of Geology 4 (5): 108-119p

[10]   Burden RJ (1982) Hydrochemical variation in a water –table aquifer beneath grazed pastureland, J Hydrol 21: 61-75

[11]   Taylor CB, Wilson DD, Brown LJ, Stewart MK, Buren RJ and Brailsford GW (1989) Sources and flow of north Canterbury Palins groundwater, New Zealand, J Hydrol 106: 311-340

[12]   Vazquez-Sune E, Carrera J, Tubau I, Sanchez-Villa X and Soler A (2010) An approach to indetify urban groundwater recharge, Hydrol Earth Syst Sci. 12: 2085-2097, doi 10.5194/hess-14-2085-2010

[13]   De Louw PGB, Van der Velde Y and Van der Zee SEATM (2011) Quantifying water and salt fluxes in a lowland polder catchment dominated by boil seepage; a probabilistic end-,member mixing approach. Hydrol Earth Syst Sci 15:2101-2117. Doi:10.5194/hess-15-2101-2011

[14]   Irmscher R and Teermann I (2002) Riverbank filtration for drinking water supply- a proven method, perfect to face today's challenges. Water Sci Technol Water Supply 2(5–6):1–8

[15]   Mencio A and Mas-pla J (2008) Assessment by multivarate analysis of groundwater-surfacewater interactions in urbanized Mediteranean streams, J Hydrol 352:355-366

[16] Banoeng-Yakubo B, Yidana SM and Nti E (2009) Hydrochemical analysis of groundwater using multivariate statistical methods-The Volta Region, Ghana. KSCE J Civ Eng 13(1): 55-63

[17] Cloutier V, Lefebvre R, Therrien R and Savard MM (2008) Multivarate statistical analysis of geochemical data as indicative of the hydrochemical evolutionof groundwater in a sedimentary rock aquifer system. J Hydrol 353:294-313.

[18] Demlie M, Wohnlich S, Wisotzky F and Gizaw B (2007) Groundwater recharge, flow and hydrogeochemical evolution in a complex volcanic aquifer system central Ethiopia, Hydrogeol J, 15, 1169-1181

[19] Koh DC, Chae GT, Yoon YY, Kang BR, Koh GW. and Park KH (2009) Baseline geochemical characteristics of groundwater in the mountainous area of Jeju Island, South Korea. Implications for degree of mineralization and nitratae contamination. J Hydrol. 376: 81-93

[20] Delgado-Outeirino I, Araujo-Nespereira P, Cid-Fernandez JA, Mejuto JC, Martinez-Carballo E and Simal-Gandra J (2011) Classification of thermal waters based on their inorganic fingerprint and hydrogeothermal modelling. Hydrol Earth Syst Sci Discuss 8:4559-4581. Doi 10.5194/hessd-8-4559-2011

[21] Alther GA (1989) A simplified statistical sequence applied to routine water quality analysis. A case history, Groundwater 17:556-561

[22] McNeil VH, Cox ME and Preda M (2008) Assesment of chemical water types and their spatial variation using multi-stage cluster analysis, Queensland, Australia, J Hydrol. 310: 181-200

[23] Panda UC, Sundaray SK, Rath P, Nayak BB and Bhatta D (2006) Application of factor and cluster analysis for characterization of river and estuarine water systems-A case study; Mahanadi River (india), J Hydrol 331: 434-445

[24] Kumar M, Ramanathan A and Keshari AK (2009) Understandng the extent of inteactions between groundwater and surface water trough major ion chemistry and multivariate statistical techniques, Hydrol Process 23: 297-310

[25]   APHA (2005) Standard methods for the examination of water and wastewater, 21st ed. American Water Works Association, Water Environment Federation, Washington

[26] Johnson RA and DW Wichern (1992) Applied multivariate statistical analysis. 3rd ed. Prentice- Hall, Englewood Cliffs, New Jersey

[27] Singh KP, Malik A and Mohan D et al. (2004) Multivariate statistical techniques for the evaluation of spatial and temporal variation in water quality of Gomti River (India) a case study. J Water Research 38: 3980–3992.

[28] Wunderlin DA, Diaz MDP and Ame MV et al. (2001) Pattern recognition techniques for the evaluation of spatial and,temporal variation in water quality. A case study: Suquia river basin (Cordoba Argentina). J Water Research, 35: 2881–2894.

[29] Kannel PR, Lee S, Kanel SR and Khan SP (2007) Chemometric application in classification and assessment of monitoring locations of an urban river system. Analytica Chimica Acta 582: 390-399

[30] Juahir H, Zain MS and Yusoff MK (2010b) Spatial water quality assessment of Langat River Basin (Malaysia) using envirometric techniques. Environment Monitoring Assess 173: 625-641

[31] Liu CW, Lin KH and Kuo YM (2003) Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. The Science of the Total Environment 313: 77–89

[32] Suk H and Lee K (1999). Characterization of a ground water hydrochemical system through multivariate analysis: Clustering into groundwater zones. Ground Water, 37: 358-366

[33] Helena B, Pardo R, Vega M, Barrado E, Fernandez JM and Fernandez L (2000) Temporal evaluation of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. Water Research 34: 807–816.

[34] Department of Standards Malaysia (2010) Malaysian Standard: Drinking water-quality requirements (MS 2320:2010). Department of Standard Malaysia, 19p.

[35] Young CP, Oakes DB and Wilkinson WS (1976) Prediction of future nitrate concentration in Groundwater. Groundwater, 4(6). 426-438.

[36] Lim WY, Aris AZ and Praveena SM (2012) Application of the chemometric approach to evaluate the spatial variation of water chemistry and th identification of the sources of pollution in Langat River, Malaysia. Doi 10.1007/s1517-012-0756-6.