Scientific Research Publishing

Spatial Variability of Ground Water Quality Using HCA, PCA and MANOVA at Lawspet, Puducherry in India

N. Suresh Nathan, R. Saravanane, T. Sundararajan

Department of Civil Engineering, Pondicherry Engineering College, Puducherry, India
Email: suresh_eepdy@yahoo.com

Abstract

In ground water quality studies multivariate statistical techniques like Hierarchical Cluster Analysis (HCA), Principal Component Analysis (PCA), Factor Analysis (FA) and Multivariate Analysis of Variance (MANOVA) were employed to evaluate the principal factors and mechanisms governing the spatial variations and to assess source apportionment at Lawspet area in Puducherry, India. PCA/FA has made the first known factor which showed the anthropogenic impact on ground water quality and this dominant factor explained 82.79% of the total variance. The other four factors identified geogenic and hardness components. The distribution of first factor scores portray high loading for EC, TDS, Na^+ and Cl^- (anthropogenic) in south east and south west parts of the study area, whereas other factor scores depict high loading for HCO_3^-, Mg^{2+}, Ca^{2+} and TH (hardness and geogenic) in the north west and south west parts of the study area. K^+ and SO_4^{2-} (geogenic) are dominant in south eastern direction. Further MANOVA showed that there are significant differences between ground water quality parameters. The spatial distribution maps of water quality parameters have rendered a powerful and practical visual tool for defining, interpreting, and distinguishing the anthropogenic, hardness and geogenic factors in the study area. Further the study indicated that multivariate statistical methods have successfully assessed the ground water qualitatively and spatially with a more effective step towards ground water quality management.

Keywords

1. Introduction

Water covers 78% of the earth's surface, yet its availability for human use is limited.

Ground water is the primary source of drinking water, and it plays a fundamental role in human life and development. Safe potable water is absolutely essential for healthy living. It is ultimate and most suitable fresh water resource for human consumption in urban as well as rural areas. In many areas, ground water is the only available source for drinking purposes. Further it is a finite resource, essential for agriculture, industry and human existence and it plays a key role in meeting the water needs of various user-sectors in India. To sustain and maximize the benefit of this resource, knowledge about the natural hydro-geological and geo-chemical processes, as well as associated human effects on the ground water resource is a must for a comprehensive and complete scientific understanding of the vulnerability of the aquifers to pollution. In this context, rapid increase in human population coupled with expanding urbanization and industrialization has led to a greater imbalance between water availability and demand [1]. As worldwide extraction of ground water is accelerated to meet increasing demand, the significance of the chemical quality of ground water also increases relative to its economic value and usefulness. In recent years, ground water contamination has become an important environmental issue especially in urban areas. Ground water contamination is always the result of human activity in areas where population density is high and human land use is intensive [2]. Virtually any activity whereby chemicals or wastes released to the environment either intentionally or accidentally, has the potential to pollute ground water. When once aquifer becomes contaminated, it is very difficult, expensive and time consuming affair to clean up and may remain unusable for decades. Deterioration of ground water quality due to different geogenic and anthropogenic activities is of great concern, especially in an alluvial aquifer in a coastal area like Puducherry, India [3] [4]. Against this background at Lawspet area in Puducherry, India, the following two human induced activities play a critical role in the ground water contamination scenario [5].

1) Unscientific and indiscriminate Municipal Solid Waste (MSW) dumping and

2) Partially treated or Secondary Wastewater (SWW) land application.

The urban and peri urban areas of Puducherry have been divided into various zones for comprehensive water supply schemes. Being a coastal town, Puducherry entirely depends on ground water for its water supply, as there are no surface water sources. The Puducherry District population was around 9.50 lakhs as per 2011 census and nearly 70% of this population lives in the urban areas and the population growth was close to 28.08% between 2001 and 2011. Presently the population explosion and urbanisation are at peak in Puducherry.

Over drawl of ground water is the only result consequent to these two social events. Due to over exploitation of ground water, all the coastal borewells had been affected and abandoned. Of late Lawspet area in Puducherry which is at a higher elevation, has been identified as a potential source to meet the future water supply requirements. But on the contrary, it is feared that the above said two anthropogenic activities viz., in discriminant MSW dumping and SWW land

application may make the ground water unsuitable for domestic purposes in Lawspet area.

Hence it is pertinent to examine the spatial and temporal variations of ground water attributes and to interpret the results of the same to determine the factors affecting the hydro-geochemistry of ground water, so that suitable remedial measures could be suggested to conserve and sustain ground water resources.

2. Study Area and Current Status

The study area is bounded by latitude 11˚58'16"N and longitude 79˚48'11"E and is located at Karuvadikuppam in Lawspet on the northern part of Puducherry, India. Location map of study area is shown in Figure 1. The ground falls from 53 m to 6 m (Figure 1) within a radial distance of 2.5 km. Alluvial aquifer is probably, the dominant type of aquifer in the study area. Lawspet area receives its major rainfall from the north east monsoon (65%), and also gets some rainfall from south west monsoon (35%). The rainy season is from October to December. The study area receives an annual rainfall of about 1200 mm [5].

The existing Sewage Treatment Plant (STP) at Karuvadikuppam, consists of four facultative oxidation ponds connected in series and the treatment efficiency is about 65%. After treatment the partially treated 12.5 MLD, SWW is discharged directly into a recharge pond located inside STP campus for the past 35 years. Besides, a portion of STP is used as a MSW dump site for the past 10 years. Here MSW has been dumped indiscriminately and unscientifically in an irregular fashion. So it can be said that in the study area, co-disposal of MSW and partially treated SWW are taking place simultaneously within the same campus [5]. The present work therefore focuses

1) To understand the process of controlling components which govern the chemical constitution of ground water;

2) To distinguish the ground water quality evolution process;

3) To determine the spatial variability of ground water quality using Multivariate Statistical Analysis like Hierarchical Cluster Analysis (HCA), Principal Component Analysis (PCA), Factor Analysis (FA) and MANOVA.

3. Methodology

3.1. Sampling and Testing

Due to spatial and temporal variations in ground water chemistry, a monitoring programme that will provide a representative and reliable estimation of the quality of ground water is necessary. So to accurately represent the groundwater quality, a sampling strategy was designed to cover a wide range of borewells at the key locations. Nearly 125 water supply and agrarian borewells are located within a radial distance of 2.5 Km from STP and MSW landfill. Out of which, 20 Public Works Department (PWD) borewells supply drinking water to Muthialpet and Lawspet areas. The remaining are private domestic borewells or agricultural borewells.
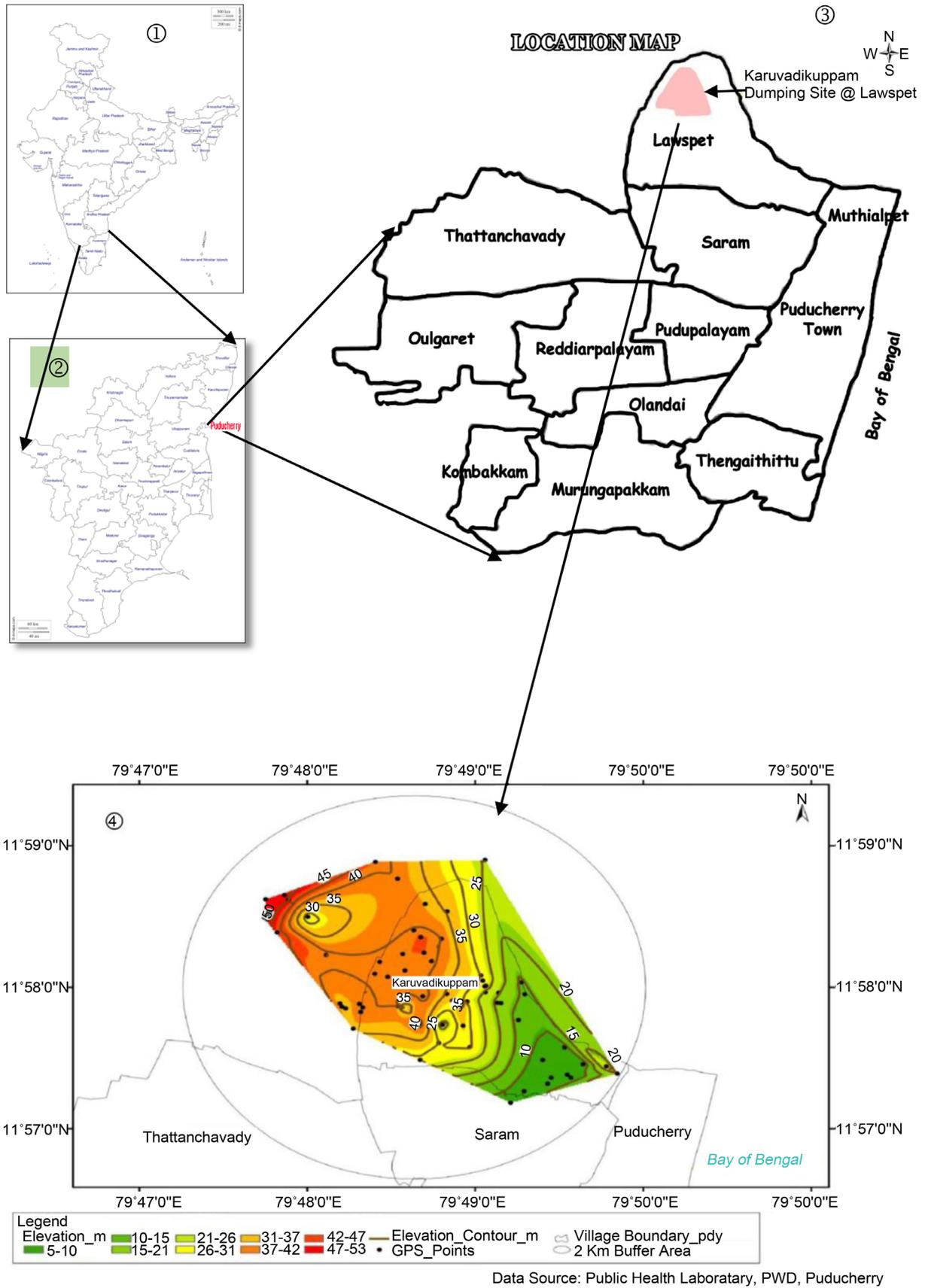
Figure 1. Elevation and location of borewells in study area.

Totally 68 borewells (GPS points) were identified in solid waste dump area, recharge pond area, sewage farm area (existing) and peripheral area (private & Govt.) in order to study the seasonal and spatial variations, as detailed below and depicted in Figure 1.

1) Solid Waste Dump area   2 borewells (newly sunk);
2) Recharge Pond area   5 borewells (newly sunk);
3) Sewage Farm area   3 borewells (existing);
4) Peripheral area   58 borewells (private & Govt.).

All the 68 borewells had been considered for investigation and water samples were collected from the borewells after pumping for 15 minutes. The samples were analysed in the Public Health Laboratory, PWD, Puducherry, India. In Toto 1065 water samples were collected and tested for 17 physio-chemical and 4 bacteriological parameters viz. EC, pH, TDS, Alkalinity, Bicarbonate, Total Hardness, Calcium, Magnesium, Iron, Chloride, Sulphate, Nitrate, Sodium, Fluoride, Potassium, Phosphate, Silica, B.O.D, C.O.D, total coliforms and faecal coliforms. Eventhough the water samples were tested for all physio-chemical and bacteriological parameters, the study was confined to the pollution aspects of ten significant physio-chemical parameters only viz., EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+, and K^+ [5].

3.2. Multivariate Statistical Analyses

The sampling and testing strategyso designed involved frequent water samplings and examination of large number of physicochemical parameters, thereby producing a large data matrix, which needs a complex data interpretation. The application of different multivariatestatistical approaches for the interpretation of these complex data matrices offers a better comprehension of water quality and ecological status of the studied systems, and it allows the detection of the possible factors/sources that control the ground water systems and suggests a useful tool for dependable supervision of water resources as well as quick solutions to pollution related problems. The basic aim of such an analysis is to study the hydro-geochemistry of an aquifer using various statistical methods and to assess and ascertain the deterioration of ground water quality [6] [7]. Further these multivariate statistical techniques also verify the spatial and temporal variations which are brought out by natural and anthropogenic factors.

In the present study an effort has been made to carry out detailed and systematic investigation of hydro geochemical parameters and spatial variability of the ground water quality using multivariate statistical methods like HCA, PCA/FA, and MANOVA without losing important information [8]-[18].

3.2.1. Hierarchical Cluster Analysis (HCA)

The object of HCA technique is to group ground water samples into clusters based on their squared Euclidean distance which is used most commonly as the adopted measure of distance, samples of the same cluster are characterized by high homogeneity whilst samples belonging to different clusters are characterized

by high heterogeneity between them [19] [20] [21] [22]. The levels of the similarity at which observations are merged, are used to construct dendrogram. The dendrogram provides a visual summary of clustering process, presenting a picture of groups and their proximity, with a dramatic reduction in dimensionality of the original data. The most important ten hydro-chemical parameters viz., EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+, and K^+, for 68 borewells, were chosen for HCA in the study area.

3.2.2. Principal Component Analysis (PCA) and Factor Analysis (FA)

PCA and FA have been widely used in environmental sciences and hydro geochemical research and it is a multivariate statistical technique that embodies linear combinations of variables through a correlation-centred approach. PCA and FA are both variable reduction techniques. The PCA is used when variables are highly correlated and reduces the number of observed variables to a smaller number of principal components which account for most of the variance of the observed variables whereas, the FA is a variable reduction technique which identifies the number of latent factors and the underlying factor structure of a set of variables. It also, estimates factors which influence responses on observed variables. The PCA extracts the eigenvalues and eigenvectors from the covariance matrix of original variables. The eigenvalues of the PCs are the measures of their associated variance, the participation of the original variables in the PCs is given by the loadings, and the coordinates of the objects are called scores. In other words, PCA includes correlated variables with the purpose of reducing the numbers of variables and at the same time it explains the same amount of variance with fewer variables (principal components), while FA estimates factors, *i.e.* underlying structure that cannot be quantified directly.

The purpose of FA is to ascertain the minimum number of new variables necessary to replicate various attributes of the data by cutting down the original data matrix from one having (n) variables necessary to describe the (N) samples into a matrix with (m) factors (m < n). It also aims at converting the variables so that the axes become orthogonal, which then admits new independent variables. By this way, the first factor is selected to account for the total variance of the observations to the maximum extent, the second factor to describe the maximum possible residual variance, so on and so forth. In other words, the first factor is evaluated such that the sum of squares of the projections of the points on the factor is highest (factor loadings). Next, to specify the second factor, the points are projected on a plane orthogonal to the first factor and so on for the other factors, each exhibiting less and less of the total variance. On the other hand, the sum of squares of the factor loadings for each variable is the communality and it deliberates the proportion of the total variability of each variable accounted for by the factoring. FA follows three main measures 1) extraction of initial factors 2) rotation of factors and 3) calculation of each factor scores [23] [24] [25]. In this research work PCA/FA are implemented to the hydro-chemical data in the study area to extract principal factors analogous to different sources of variation

in the data and to detect the possible source of contamination spatially.

3.2.3. Multivariate Analysis of Variance (MANOVA)

As the ground water samples were collected at various locations and at different points of time, as the borewells are wide spread radially in all directions, as the borewells are positioned within a radial distance of 2.5 kms from the anthropogenic sources and as the ground level drastically falls for more than 47 m, the analysis on the spatial variability was considered very important. Hence, MANOVA was carried out to evaluate the significant effects of spatial differences (variability) on mean concentration of selected physio-chemical variables of ground water [26] [27]. The intention of performing MANOVA is to address the following queries:

1) What are the influences of independent variables (observation spots/clusters) on the dependent variables (mean concentrations of physio-chemical parameters including EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+, and K^+)?

2) What are the interactions among the independent variables?

To execute MANOVA the following procedure is to be adopted:

1) Descriptive analysis covering average, standard deviation, minimum and maximum values for EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+, and K^+, is to be carried out.

2) MANOVA test using Pillai's Trace, Wilks' Lambda, Hotelling's Trace and Roy's Largest Root are to be performed. The aim is to test whether there are differences in the average levels of physio-chemical compounds in ground water samples between the clusters at $\alpha = 5\%$.

The model for MANOVA with one factor is: $X_{ij} = \mu + \tau_I + \epsilon_{ij}$

where

X = vector of dependent variable (physio-chemical parameters);

μ = overall mean vector;

τ = factor effect/spatial variability vector;

ϵ = error vector; i = level of effect, $i = 1, 2, \cdots, I$; j = the replication, $j = 1, 2, \cdots, J$.

The F and p values from the results of the MANOVA are used for testing of significance and H_o is rejected if $F > F_{(df1, df2, \alpha)}$ or p value $< \alpha$. In this research the dependent variables are EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+, and K^+. The factor is cluster *i.e.* the independent variables are Clusters 1, 2 and 3.

4. Results and Discussion

4.1. Descriptive Statistics

As the borewells are wide spread in all directions and as the ground level difference is more than 47 m in the study area, the data about hydro-geochemistry of the parameters, is very important. As such during this study selected physiochemical properties of the borewells were acquired and considered. Eventhough, the main aim of the study was to statistically establish the spatial variability of ground water quality, it is significant to detail the current status of ground water quality so that the study will be worthwhile to the authorities who are in charge of ground water management and control. Vital physio-chemical attributes viz., EC, TDS,

HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+, and K^+ affected by MSW and SWW exercises, taking into consideration the geology and environmental conditions, were statistically evaluated and documented in Table 1. Correlation Co-efficient matrix has also been generated in order to identify the inter-parameter relationships and presented in Table 2.

4.2. Spatial Similitude and Clustering

HCA was performed on borewells as well as on the selected ten physio-chemical parameters, to determine the spatial similarities and qualitative affinity among the parameters. Elements belonging to the same cluster are likely to have originated

Table 1. Descriptive statistics of chemical parameters.

Statistic	EC	TDS	TH	HCO_3^-	Cl^-	SO_4^{2-}	Na^+	Ca^{2+}	Mg^{2+}	K^+
Borewells	68	68	68	68	68	68	68	68	68	68
Minimum	155	97	60	45	17	7	13	16	6	1
Maximum	2176	1356	464	459	517	126	318	109	68	11
Range	2021	1259	404	414	500	119	305	93	62	10
1st Quartile	432.8	276	155.3	114.3	42.8	25.8	35.3	42.5	15.8	2
Median	1016.0	640	244.5	208.5	202	61	123	59.5	25	3
3rd Quartile	1320.3	830.5	319.8	289	274.8	69	186.3	71	36.3	4
Mean	969.6	610.9	243.3	217.3	189.9	53.2	119.8	57.1	26.5	3.7
Variance (n)	312,728	123,456	10,724	12,283	22,275	721	7539	384	202	4
SD (N)	559.2	351.4	103.6	110.8	149.2	26.9	86.8	19.6	14.2	2
Skewness	0.4	0.4	0.2	0.3	0.5	0.1	0.5	−0.1	0.5	1.9
Kurtosis	−0.8	−0.8	−0.8	−0.9	−0.9	−0.5	−0.9	−0.4	−0.2	3.6
SE of variance	54,838	21,648	1881	2154	3906	126	1322	67	35	1

Note: EC, μS/cm, all other parameters, mg/L.

Table 2. Correlation matrix between the variables.

Variables	EC	TDS	TH	HCO_3^-	Cl^-	SO_4^{2-}	Na^+	Ca^{2+}	Mg^{2+}	K^+
EC	1									
TDS	0.999	1								
TH	0.893	0.891	1							
HCO_3^-	0.729	0.728	0.820	1						
Cl^-	0.979	0.979	0.816	0.616	1					
SO_4^{2-}	0.830	0.831	0.721	0.445	0.803	1				
Na^+	0.979	0.979	0.812	0.666	0.982	0.819	1			
Ca^{2+}	0.814	0.811	0.947	0.793	0.732	0.629	0.724	1		
Mg^{2+}	0.916	0.915	0.955	0.822	0.851	0.734	0.855	0.855	1	
K^+	0.806	0.806	0.715	0.600	0.784	0.665	0.789	0.616	0.715	1

from a common source. The R-mode HCA (parameter clustering) retains four main clusters for analysed parameters. Cluster 1 includes EC and TDS which may be explained by a combination of point sources like MSW dumping and SWW land application. Cluster 2 includes parameters like TH and HCO_3^- and this cluster reflects hardness component indicating the influence of SWW land application. Cluster 3 consists of Na^+ and Cl^- which represents the anthropogenic activities viz., MSW dumping and SWW land application. Cluster 4 involves parameters like Ca^{2+}, Mg^{2+}, K^+ and SO_4^{2-} indicating the geogenic nature of ground water and this may be interpreted as due to weathering of rocks and dissolution of minerals. The resulted parameter dendrogram is shown in Figure 2.

The Q-mode HCA (borewell clustering) has been employed to discover the spatial hydro-chemicalresemblance among the borewells. The borewells which are grouped in a particular cluster share similar characteristics in relation to the investigated parameters. The resulted borewell dendrogram (Figure 3) grouped all the 68 borewells into three statistically significant clusters. The cluster wise designated borewells, the regional distribution of borewells and the profile plot of clusters are presented in Table 3, Figure 4 and Figure 5. Cluster 1 consists of 28 borewells and falls in the "polluted" category. Cluster 2 includes 8 borewells and can be termed as "highly polluted" and the balance 32 borewells are incorporated in Cluster 3 and this cluster is "non polluted".

In Clusters 1 & 2, on the basis of overall chemical composition, characterised by the ion abundances the order of anions and cations is $Cl^- > HCO_3^- > SO_4^{2-} = Na^+ > Ca^{2+} > Mg^{2+} > K^+$. Cl^- and Na^+ dominate these two clusters and the distribution of borewells lies in South-East and South-West directions of the study area. In Cluster 3, the order of affinity in ground water is $HCO_3^- > Cl^- > SO_4^{2-} = Ca^{2+} > Na^+ > Mg^{2+} > K^+$. HCO_3^- and Ca^{2+} dominate this cluster
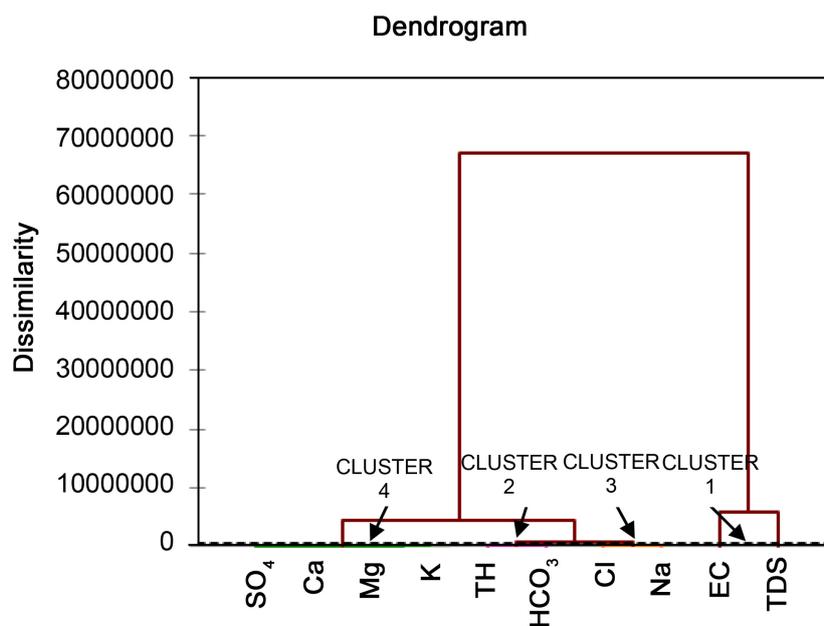
Figure 2. Parameter dendrogram.
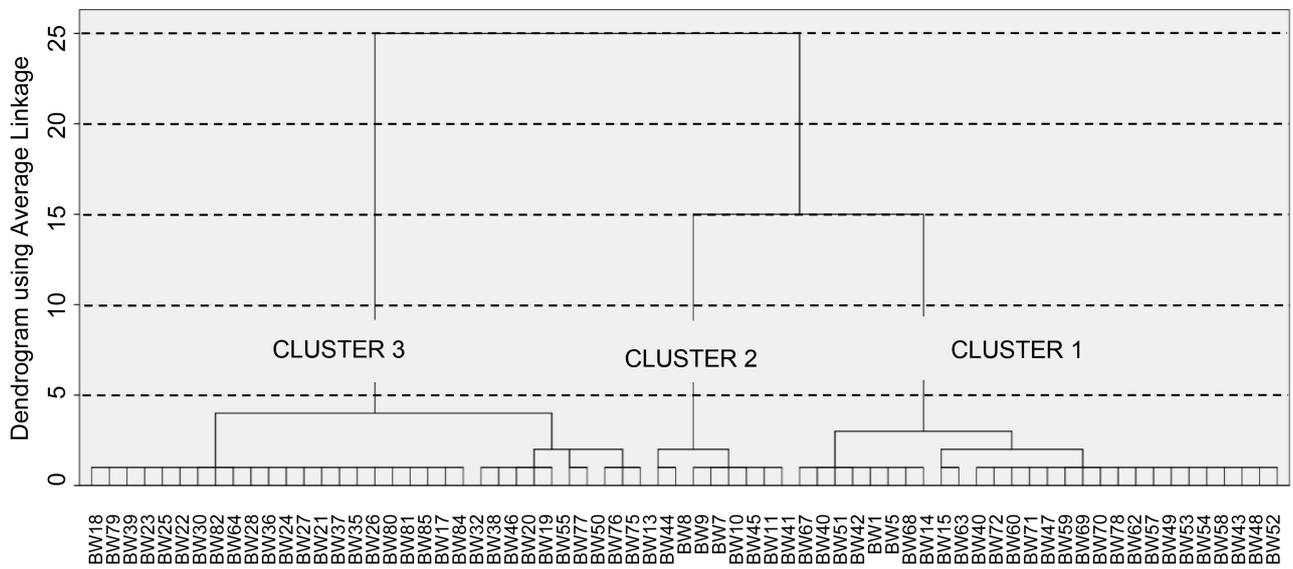
Figure 3. Borewell dendrogram.
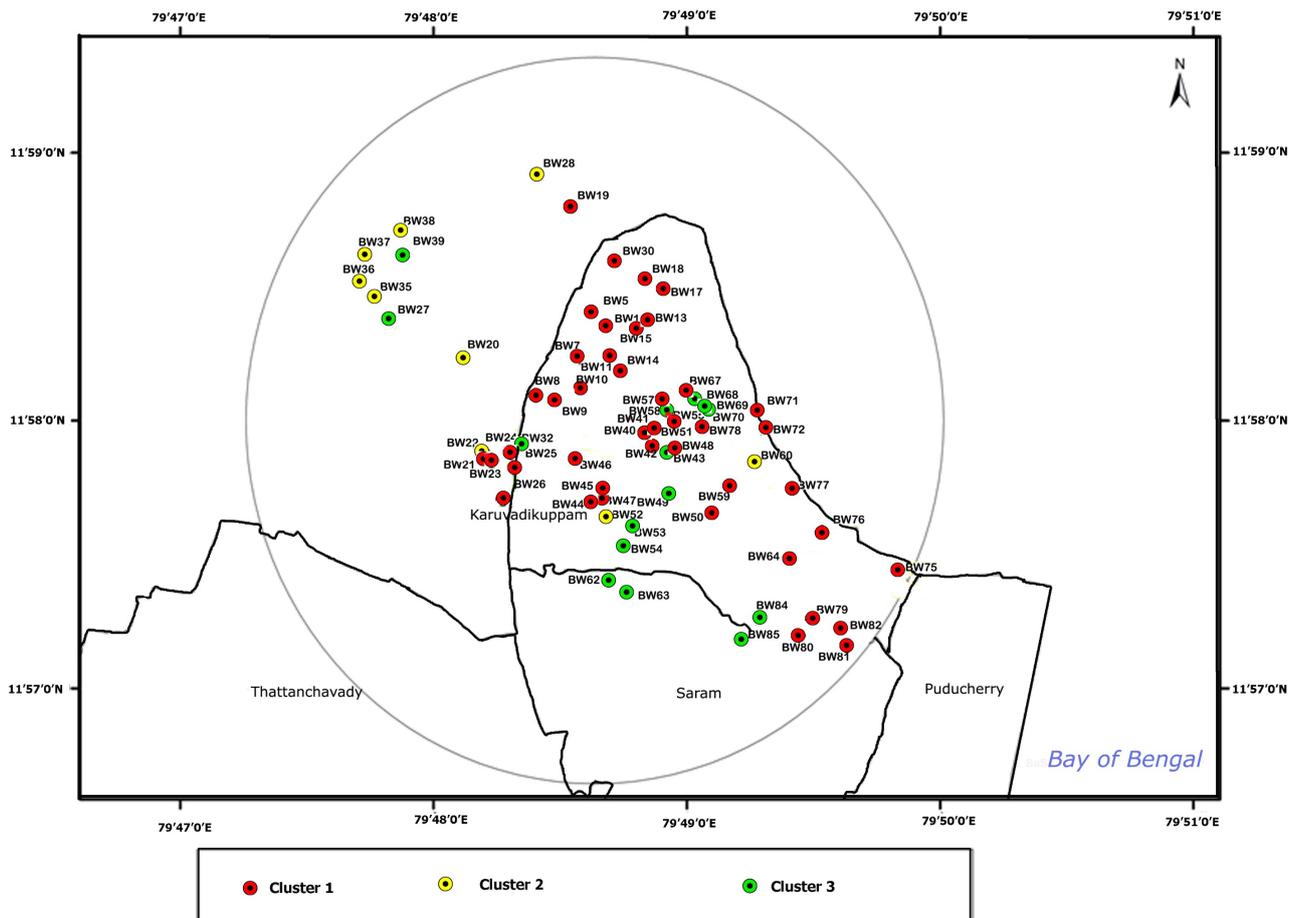
Figure 4. Borewell locations in different clusters.

and the source of HCO_3^- and Ca^{2+} is geogenic and attributed to natural processes suchas dissolution of carbonate minerals in the presence of soil CO_2. This cluster shows spatial variation in North-East, North-West, South-East,
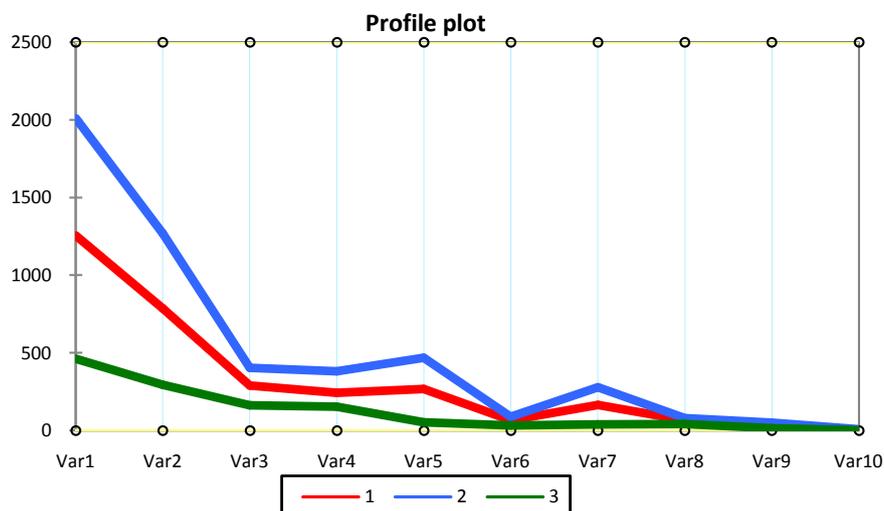
Figure 5. Parameter profile plot.

Table 3. Cluster classification.

Cluster No.	Borewell Designation	Borewell Nos.	Location	Remarks
C1	1, 5, 14, 15, 40, 41, 42, 43, 47, 48, 49, 51, 52, 53, 54, 57, 58, 59, 60, 62, 63, 67, 68, 69, 70, 71, 72 and 78	28 (41%)	Solid waste dump area, Recharge pond area and south-east and south-western parts of study area	Polluted
C2	7, 8, 9, 10, 11, 13, 44 and 45	8 (12%)	Solid waste dump area, Recharge pond area and south-western parts of study area	Highly Polluted
C3	17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 30, 32, 35, 36, 37, 38, 39, 46, 50, 55, 64, 75, 76, 77, 79, 80, 81, 82, 84 and 85	32 (47%)	North-east, north-west and south-east and south-western parts of study area	Non polluted

South-West parts of study area.

4.3. Sampling Adequacy and Bartlett's Sphericity Test

Before FA, the Kaiser Meyer Olkin (KMO) and Bartlett tests were carried out to determine the applicability of the data for FA. KMO is a measure of sampling adequacy and it shows the proportion of variance reflected by the underlying factors. The size of KMO value is not statistically critical, however larger the KMO value more factors are suitable for FA. KMO value > 0.8 is very good and the value < 0.5 is not suitable for FA. Similarly Bartlett's test denotes whether correlation matrix is an identity matrix which would imply that the parameters are unrelated. In a nutshell, it provides the presence of a common factor between relevant matrixes of the parent population, and its statistical significance tests

the suitability of the data for FA. The KMO and Bartlett's test results are presented in Table 4 and interpreted as follows:

Test interpretation:

H_0: There is no correlation significantly different from 0 between the variables.

H_1: At least one of the correlations between the variables is significantly different from 0.

As the KMO value is >0.8 and the computed p-value (Bartlett's test) is lower than the significance level $\alpha = 0.05$, the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted. So FA is effective in reducing the dimensionality.

4.4. Source Identification of Ground Water Contamination

PCA/FA has been used to determine the interdependence among different sets of ground water physio-chemical data and for identifying different sources which are responsible for ground water contamination and to condense the data with a minimum loss of information. PCA estimates the amount of variation in each parameter explained by the factors. PCA/FA was performed using XLSTAT version 2014 software.

Eigenvalues are the amount of variance explained by each factor, each parameter had a variance of 1 with a total variance of 10 (for the selected ten variables) for the entire data set. Factor with eigenvalue > 1 explains more total variation in the data than individual parameter, and factor with eigenvalue < 1 explains less total variation. Therefore only factors with eigenvalue >1 are retained for the interpretation and the retained factors are subjected to varimax rotation. Varimax rotation is an orthogonal rotation method that minimizes the number of variables that have high loading on each factor. The VariFactor (VF) coefficient greater than 0.75 is considered to be strong and indicates high proportion of variance explained by the factor, between 0.50 and 0.75. It is considered as moderate loading while 0.30 - 0.50 as weak significant factor loading, indicating much of that attribute's variance remains unexplained and it is less important.

Primarily R-mode PCA/FA was applied for all the borewells as a whole in the study area. The scree plot (Figure 6) has been utilized to distinguish the number of PCs to be employed to comprehend the rudimentary parameters' structure. The computed percentage of variance with cumulative percentage explained by each factor together with factor loadings after varimax rotation are listed in Table 5 and Table 6. The positive scores demonstrate that all the water samples are

Table 4. KMO measures and Bartlett's test of Sphericity.

Kaiser Meyer Olkin measure of sampling adequacy, 0.862		
	Chi-square distribution	−1628.583
Bartlett's Sphericity Test	DF	−45
	p-value	−<0.0001
	α	−0.05
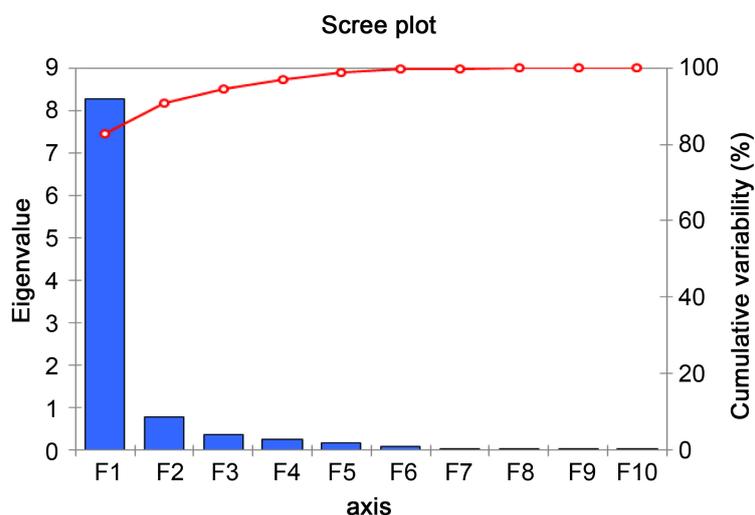
Figure 6. Scree plot.

substantially influenced by the presence of extracted loads on a specific component. Only one factor with eigenvalue > 1 has been extracted from the ground water data matrix, which represents 82.79% of the total variance (Table 5) *i.e.* the first principal component (PC1) explains more than 82.79% of the total variance. VF1 is loaded with EC, TDS, Na^+ and Cl^- (Table 6). The parameters Na^+ (0.797) and Cl^- (0.822) are heavily loaded and this factor represents the human induced activities, such as MSW dumping and SWW application on land. In all other components the eigenvalues are <1 indicating that these components are less significant. Consequently all the parameters except Na^+ and Cl^- do not contribute much to the hydro-chemistry of the study area. However, it may be seen from Table 5 and Table 6 that PC2 explains 7.94% of the total variance with VF2 showing strong loading for HCO_3^- (0.848) and moderate loading for Mg^{2+} (0.524) indicating the geogenic nature of the groundwater and PC2 can be designated as geogenic component. PC3, PC4 and PC5 together contribute only 7.97% of the total variance and VF3, VF4 and VF5 are associated heavily with K^+ (0.811), SO_4^{2-} (0.787) and Ca^{2+} (0.800), showing that these factors are geogenic in nature. Further VF5 exhibits moderate loading with TH (0.672) reflecting the geogenic and hardness components.

The scatter plot between VF1 and VF2 of the variables and the score plot between VF1 and VF2 of the borewells is presented in Figure 7.

As the borewells are located within a radial distance of 2.5 kms. from the polluting sources and as ground elevation falls from 53 m to 6 m, it was decided to perform cluster wise PCA/FA in order to study the inherent characteristic structure of the polluting parameters. R-mode PCA/FA was employed to Clusters 1, 2 and 3 and PCAs, whose eigenvalues > 1 were chosen for study purposes. The variance and cumulative variance of the principal components in all three clusters are presented in Table 7. Similarly after varimaxrotation, the factor loadings of the significant factors of all the three clusters are given in Table 8. In Cluster 1, the first PC explained 51.46% of the total variance. From Table 8 for

Table 5. Principal component analysis for 68 borewells (R-mode).

	PC1	PC2	PC3	PC4	PC5
Eigenvalue	8.280	0.794	0.375	0.252	0.170
Variability (%)	82.798	7.937	3.749	2.516	1.704
Cumulative %	82.798	90.735	94.485	97.000	98.705

Table 6. Factor loadings after varimax rotation for 68 borewells (R-mode).

	VF1	VF2	VF3	VF4	VF5
EC	0.723	0.343	0.312	0.334	0.386
TDS	0.724	0.343	0.313	0.336	0.381
TH	0.415	0.450	0.258	0.302	0.672
HCO_3^-	0.274	0.848	0.215	0.078	0.378
Cl^-	0.822	0.214	0.304	0.280	0.321
SO_4^{2-}	0.489	0.110	0.239	0.787	0.260
Na^+	0.797	0.301	0.303	0.321	0.260
Ca^{2+}	0.338	0.388	0.190	0.203	0.800
Mg^{2+}	0.511	0.524	0.223	0.327	0.498
K^+	0.425	0.246	0.811	0.227	0.222

Table 7. Principal component analysis (cluster wise R-mode).

	Cluster 1			Cluster 2			Cluster 3	
	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2
Eigenvalue	5.146	2.405	1.168	4.776	2.463	1.339	6.204	2.317
Variability (%)	51.464	24.053	11.683	47.759	24.634	13.394	62.041	23.172
Cumulative %	51.464	75.518	87.200	47.759	72.394	85.787	62.041	85.214
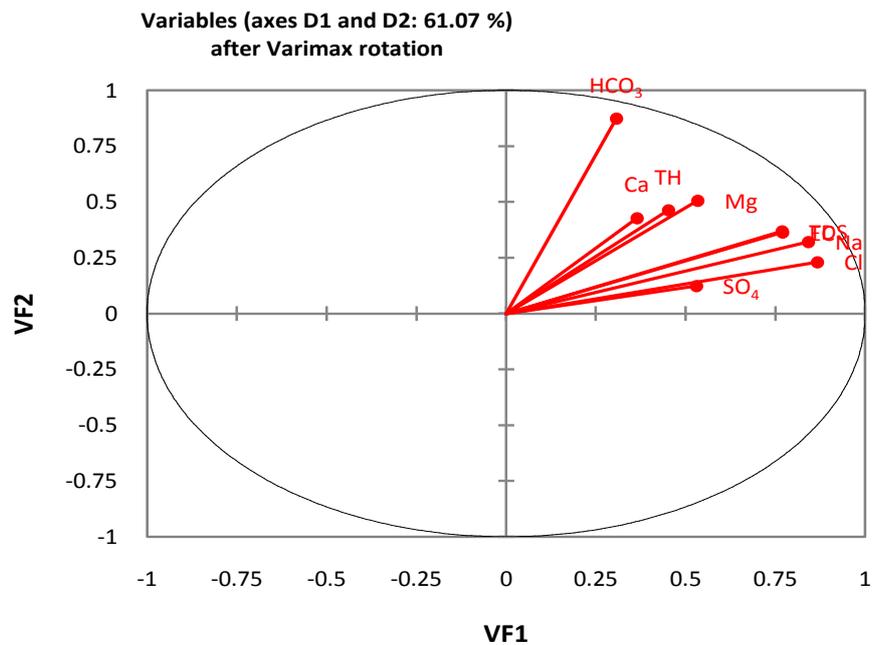
Figure 7. Scatter plot of VF1 vs. VF2.

Cluster 1, it can be seen that the Varifactor 1 (VF1) is heavily loaded with EC(0.937), TDS (0.932), Na^+ (0.892) and Cl^- (0.934) indicating the anthropogenic nature of component/factor. VF2 is heavily loaded with TH (0.954) and Ca^{2+} (0.972) and moderately loaded with HCO_3^- (0.703) and Mg^{2+} (0.734). So VF2 can be designated as geogenic component. VF3 is represented with K^+ (0.987) and can be regarded as geogenic in nature. In Cluster 2, the first 3 components explained 85.787% of the total variance (Table 7) and from Table 8, it can be seen that VF1 is loaded heavily with EC (0.803), TDS (0.783), TH (0.932) and Mg^{2+} (0.816) and moderately loaded with Cl^- (0.703). In VF1 the human induced activities and hardness play a crucial role and VF1 can be thought of a combination of anthropogenic and hardness components. VF2 and VF3 are loaded with HCO_3^- (0.822) and K^+ (0.975) which are geogenic in nature.

In Cluster 3, first 2 components accounted for 85.214% of the total variance (Table 7). VF1 (Table 8) is described heavily with TH (0.945), HCO_3^- (0.815), Ca^{2+} (0.967) and Mg^{2+} (0.894), here geogenic and hardness components play a significant role. However, VF2 is loaded with Na^+ (0.929) and Cl^- (0.944) which reflects human induced activities in ground water contamination.

4.5. Spatial Change Detection

Q-mode PCA/FA was implemented for the selected ten variables as a whole in the study area to investigate the spatial variability of ground water quality particularly to examine and locate the borewells which contribute to the ground water contamination and to find out the contaminant movement direction [28] [29] [30] [31]. The factor scores after varimax rotation are furnished in Table 9. As discussed earlier, VF1 (Table 6) corresponds to anthropogenic component in which Na^+ and Cl^- are the contributing parameters to the ground water contamination and when correlated to factor scores of VF1 (Table 9) it may be

Table 8. Factor loadings after varimax rotation (cluster wise R-mode).

	Cluster 1			Cluster 2			Cluster 3	
	VF1	VF2	VF3	VF1	VF2	VF3	VF1	VF2
EC	0.937	0.307	0.028	0.803	0.465	0.108	0.673	0.644
TDS	0.932	0.317	−0.013	0.783	0.515	0.147	0.645	0.662
TH	0.211	0.954	0.133	0.932	−0.176	0.019	0.945	0.096
HCO_3^-	0.287	0.703	−0.123	0.179	0.822	0.306	0.815	−0.027
Cl^-	0.934	0.111	0.099	0.703	0.021	−0.476	−0.033	0.944
SO_4^{2-}	0.408	−0.060	0.122	0.052	0.387	0.037	0.151	0.352
Na^+	0.892	−0.058	0.030	0.343	0.188	0.109	0.146	0.929
Ca^{2+}	0.015	0.972	0.008	−0.018	−0.830	0.077	0.967	0.173
Mg^{2+}	0.503	0.734	0.169	0.816	0.117	−0.313	0.894	0.121
K^+	0.050	0.092	0.987	−0.035	0.122	0.975	0.316	0.498

Table 9. Factor scores after varimax rotation (Q-mode).

Bore-well	VF1	VF2	VF3	VF4	VF5	Bore-well	VF1	VF2	VF3	VF4	VF5
BW 1	0.201	−1.264	0.395	0.763	1.939	BW44	2.158	−0.196	−1.165	−0.555	1.533
BW 5	0.539	−1.012	1.176	0.069	1.014	BW45	2.351	2.335	−2.145	1.884	−0.554
BW14	−0.256	1.776	−0.570	0.484	0.787	BW17	−0.511	−0.562	0.338	−1.238	0.015
BW15	−1.730	0.384	−0.001	−0.469	4.109	BW18	−0.443	−0.651	−0.082	−0.365	−1.594
BW40	0.844	0.647	−0.595	0.266	0.367	BW19	−2.204	2.474	0.438	−0.903	1.501
BW41	1.669	0.433	−1.570	0.038	0.476	BW20	−1.287	1.730	−0.788	−1.011	0.612
BW42	1.146	0.587	−1.334	−0.047	0.706	BW21	−0.913	0.507	−0.305	−0.361	−1.074
BW43	1.109	0.154	−1.038	0.028	0.118	BW22	−0.623	−0.453	−0.109	−1.212	−0.484
BW47	0.218	−0.861	0.200	0.269	0.354	BW23	−0.349	−0.403	−0.902	−0.962	−0.628
BW48	0.841	0.687	−1.143	0.397	−0.141	BW24	−0.629	−0.874	−0.270	−1.235	0.572
BW49	0.804	0.192	−1.006	0.150	0.292	BW25	−0.268	−0.649	−0.122	−1.539	−0.520
BW51	0.903	0.769	−0.560	0.032	0.215	BW26	−0.336	−0.942	−0.192	−1.202	−0.125
BW52	1.024	0.849	−0.376	−0.308	−0.483	BW27	−1.042	0.511	−0.412	−0.607	−0.225
BW53	0.909	0.134	−0.952	0.126	0.000	BW28	−1.202	0.585	−0.399	−0.999	0.498
BW54	0.438	0.046	−0.159	0.093	0.374	BW30	−0.346	−0.753	−0.138	−0.806	−0.963
BW57	0.811	−1.098	−0.024	0.357	0.147	BW32	−1.953	2.428	−0.338	1.613	−0.806
BW58	0.649	0.316	−0.189	0.453	−0.604	BW35	−0.943	−0.001	−0.263	−1.255	0.195
BW59	0.233	−0.555	0.078	0.662	0.005	BW36	−0.888	1.245	−0.443	−1.201	−0.586
BW60	0.203	−0.222	−0.575	0.339	0.105	BW37	−0.585	0.003	−0.188	−1.407	−0.480
BW62	0.307	−0.560	0.090	0.143	0.378	BW38	−1.157	1.678	−0.913	−0.491	0.436
BW63	−0.165	−0.731	0.184	0.856	0.330	BW39	−0.074	−0.475	0.028	−1.393	−1.854
BW67	1.317	−0.002	−0.023	0.303	0.277	BW46	−1.676	1.242	0.731	−0.104	0.044
BW68	0.984	−0.465	−0.499	0.336	0.830	BW50	−0.052	−1.352	0.144	0.364	−0.892
BW69	−0.108	−0.467	0.939	−0.017	0.427	BW55	0.001	−1.287	0.673	0.173	−0.428
BW70	0.081	−0.431	0.782	0.276	0.193	BW64	−0.445	−1.037	−0.158	0.218	−1.448
BW71	0.188	−0.034	0.157	0.062	−0.116	BW75	−2.484	−1.891	−0.149	3.558	2.205
BW72	0.204	−0.909	−0.557	1.083	0.118	BW76	−0.893	−0.720	0.068	0.761	−0.017
BW78	0.779	−0.367	−0.059	0.047	−0.200	BW77	0.148	−1.276	−0.300	0.816	−0.078
BW7	1.512	1.478	1.032	0.279	−0.774	BW79	−0.349	−0.534	−0.049	−0.753	−1.622
BW8	1.249	−0.410	2.537	−0.045	1.012	BW80	−0.343	−0.799	−0.304	−0.138	−1.020
BW9	1.008	0.554	1.668	0.523	0.683	BW81	−1.466	−0.842	−0.578	2.834	−0.836
BW10	1.393	0.796	3.257	−1.817	0.864	BW82	−0.811	−0.693	−0.263	0.718	−1.271
BW11	−0.185	2.534	3.713	2.891	−2.285	BW84	0.016	−0.406	0.261	−0.504	−1.597
BW13	0.821	0.066	2.840	−0.684	0.883	BW85	−0.344	−0.954	0.476	−0.631	−0.910

seen that many borewells in Clusters 1 and 2 are affected by excess Na^+ and Cl^-.

Even though all the five vari factors in Table 9 were considered for study

purposes, only VF1 is reckoned to be important from the point of view of % variance explained and eigenvalue > 1. In Cluster 1, out of 28 borewells, 23 borewells (82%) in VF1 show positive scores and are affected by the anthropogenic activities viz., MSW dumping and SWW land application. Similarly in Cluster 2, 7 borewells (88%) in VF1 are affected out of 8 borewells. Again in Cluster 1 the worst affected borewells whose positive scores > 0.75 are BW40, BW41, BW42, BW43, BW48, BW49, BW51, BW52, BW53, BW57, BW67, BW68 and BW78 (13 borewells). But in Cluster 2 all the 7 affected borewells show positive scores > 0.75, indicating that they are highly polluted.

Next Q-mode VF2 (Table 9) is related to R-mode VF2 (Table 6) which demonstrate HCO_3^- and Mg^{2+} as geogenic component. The worst affected borewells are BW14 in Cluster 1, BW11 in Cluster 2, BW19, BW20, BW32, BW36, BW38 and BW46 in Cluster 3. Similarly VF3 (Q-mode) relates to K^+ which is geogenic in nature. The borewells which are badly affected by this, are BW69 and BW70 in Cluster 1. VF4 (Q-mode) corresponds to SO_4^{2-} and the borewells which are affected badly are BW75, BW76 and BW81 in Cluster 3. Lastly VF5 (Q-mode) reflects TH and Ca^{2+} and borewells like BW24 and BW35 in Cluster 3 are affected to some extent.

The results of PCA/FA for Q-mode (borewells) in comparison to R-mode (parameters) are summarized in Table 10, which exhibits the overall picture of the borewells that are affected by excess chemical components. The very purpose of using PCA/FA is to address the following questions:

Table 10. Spatial variability of borewells with positive factor scores.

Cluster No.	Borewell Designation (affected)	Borewell Nos.	Vari Factor	Parameters	Location/Direction	Remarks
1	1, 5, 40, 41, 42, 43, 47, 48, 49, 51, 52, 53, 54, 57, 58, 59, 60, 62, 67, 68, 71, 72 and 78	23	VF1	Na^+ and Cl^-	Solid waste dump area, Recharge pond area and south-east and south-west parts of study area	Anthropogenic
2	7, 8, 9, 10, 13, 44 and 45	7	VF1	Na^+ and Cl^-	Solid waste dump area, Recharge pond area and south-west parts of study area	Anthropogenic
1	14 and 15	2	VF2	HCO_3^- and Mg^{2+}	Solid waste dump area	Anthropogenic and Geogenic
2	11	1	VF2	HCO_3^- and Mg^{2+}	Recharge pond area	Anthropogenic and Geogenic
3	19, 20, 21, 27, 28, 32, 36, 38 and 46	9	VF2	HCO_3^- and Mg^{2+}	North-west and south-west parts of study area	Geogenic
1	63, 69 and 70	3	VF3	K^+	South-east parts of study area	Geogenic
3	17, 50, 55 and 85	4	VF3	K^+	North-east and south-east parts of study area	Geogenic
3	64, 75, 76, 81 and 82	5	VF4	SO_4^{2-}	South-East parts of study area	Geogenic
3	24 and 35	2	VF5	TH and Ca^{2+}	North-west and south-west parts of study area	Hardness and Geogenic

1) Which chemical component is responsible for contamination?

2) What are the borewells that are badly affected?

3) How the contaminated borewells are distributed? and

d) What is the direction of the contaminant movement?

Table 10 answers all the above questions.

4.6. Sources of Contamination

Conclusively PCA/FA had been applied to the parameters as a whole in the study area and also cluster wise, further different PCs/VFs were investigated and many factors viz., 1) Anthropogenic 2) Geogenic and 3) Hardness responsible for ground water contamination were identified. Next thing is to find out the source of the contamination based on these PCs and VFs. Consequently the following reasons may be attributed to the sources of contamination.

4.6.1. Anthropogenic Component

The excess Cl^- in ground water is generally considered as an index of ground water contamination and is mainly due to anthropogenic activities in the study area viz.

1) Non-engineered and unplanned Municipal Solid Waste (MSW) dumping and

2) Partially treated or Secondary Wastewater (SWW) application on land.

From the field observations it was ascertained that mean Cl^- in MSW leachate and SWW are 1350 mg/L and 639 mg/L. Similarly the mean Na^+ is 838 mg/L in MSW leachate and 232 mg/L in SWW.

Also from the correlation analysis it can be seen that a very good correlation exists (r = 0.982) between Na^+ and Cl^-. From this, it is observed that MSW dumping and SWW land application are the main causes for excess Cl^- and Na^+ in ground water in the study area.

4.6.2. Geogenic Component

- The primary source of HCO_3^- is the dissolution of minerals like calcite ($CaCO_3$) and dolomite ($CaMg(CO_3)_2$). From Table 2, it can be seen that there is very good relationship between HCO_3^- and Ca^{2+} (r = 0.793), HCO_3^- and Mg^{2+} (r = 0.822). So it can be concluded that excess HCO_3^- is due to calcite and dolomite dissolution, which is geogenic in nature.

- A good correlation between SO_4^{2-} and Ca^{2+} (r = 0.629) indicates that gypsum ($CaSO_4 \cdot 2H_2O$) and anhydrite ($CaSO_4$) are major sources of excess Ca^{2+} which is geogenic in nature.

- Also the "r" value of 0.734 between SO_4^{2-} and Mg^{2+} suggests weathering of Mg-sulphate minerals.

- Generally K^+ is derived from K-feldspar.

- The correlation between Na^+ and SO_4^{2-} (r = 0.819) indicates the dissolution of Na-sulphate minerals.

4.6.3. Hardness Component

Generally cations like Ca^{2+} and Mg^{2+} and anions such as HCO_3^- and SO_4^{2-}

are mainly responsible for the hardness of water. From the field study the following observations were reported.

As discussed earlier there is very good correlation among HCO_3^-, Ca^{2+} and Mg^{2+}. Also there is good relationship among the variables SO_4^{2-}, Ca^{2+} and Mg^{2+}. Further from Table 11 and correlation analysis it can be concluded that hardness of water in the study area is mainly due to human induced activities such as MSW dumping and SWW recharge in addition to leaching from minerals like calcite, gypsum, anhydrite and dolomite.

Primarily it can be said that anthropogenic activities played a very critical role in deteriorating the ground water quality in the study area. Consequent to human induced activities, there is excess Cl^- in many borewells and the contaminant movement is in the south east and south west directions following the ground profile. Secondly the variables like TH, HCO_3^-, Ca^{2+} and Mg^{2+} played a significant role in the ground water chemistry in the north west and south west parts of the study area. Parameters like K^+ and SO_4^{2-} are predominant in the south eastern part of the study area.

Conclusively it is observed that the anthropogenic component is significant in south east and south west directions following the ground elevation. The geogenic and hardness components play a major role in northwest and south west directions. Also one very important observation is that the ground water in north eastern part of the study area is generally not affected either due to anthropogenic activities or due to geogenic/hardness factors.

4.7. MANOVA

Firstly MANOVA was measured to establish the effects of three independent variables *i.e.* Clusters 1, 2 and 3 on the dependent variables (physio-chemical parameters). Relevant tests were done through tests like Pillai's Trace, Wilks' Lambda, Hotellings Trace and Roy's Largest Root.

The hypothesis used is:

H_o: There are no significant differences in the mean values of physio-chemical parameters like EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+ and K^+ with respect to Clusters 1, 2 & 3.

H_1: There are significant differences in the mean values of physio-chemical parameters in relation to Clusters 1, 2 & 3.

Decision making is the rejection of H_o if p value < 0.05 (*a*). It can be concluded, if H_o is rejected, then H_1 is accepted indicating that there are differences in the mean values of EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+ and K^+ between the clusters. The MANOVA test statistics are presented in Table 12

Table 11. Characteristics of MSW and SWW.

	Ca^{2+}	Mg^{2+}	HCO_3^-	SO_4^{2-}
MSW leachate	214	104	522	465
SWW	101	63	383	216

Note: All values are in mg/L.

and Table 13. To evaluate whether one way MANOVA was statistically significant, Wilks' Lambda row needs to be examined along with significance column. Wilks' Lambda test is an appraisal to show how well each function separates cases into groups. It is equal to the proportion of the total variance in the discriminate scores not explained by difference among the groups.

Smaller values of Wilks' Lambda tests indicate greater discriminatory ability of the function. From Table 12, it can be seen that Wilks' Lambda value is 0.042

Table 12. MANOVA tests.

	Effect	Value	F	Hypothesis df	Error df	Sig.
INTERCEPT	Pillai's Trace	0.978	254.704	10.000	56.000	0.000
	Wilks' Lambda	0.022	254.704	10.000	56.000	0.000
	Hotelling's Trace	45.483	254.704	10.000	56.000	0.000
	Roy's Largest Root	45.483	254.704	10.000	56.000	0.000
CLUSTER	Pillai's Trace	1.310	10.830	20.000	114.000	0.000
	Wilks' Lambda	0.042	21.709	20.000	112.000	0.000
	Hotelling's Trace	14.402	39.605	20.000	110.000	0.000
	Roy's Largest Root	13.794	78.628	10.000	57.000	0.000

Table 13. Tests of between-subjects effects.

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	EC	72276261.755	1	72276261.755	2275.637	.000
	TDS	28681433.224	1	28681433.224	2200.223	.000
	TH	3820268.960	1	3820268.960	997.717	.000
	HCO_3^-	3155532.280	1	3155532.280	432.768	.000
	Cl^-	3234281.873	1	3234281.873	1708.628	.000
	SO_4^{2-}	184782.739	1	184782.739	602.628	.000
	Na^+	1220693.053	1	1220693.053	1544.947	.000
	Ca^{2+}	188489.086	1	188489.086	1048.396	.000
	Mg^{2+}	50642.794	1	50642.794	804.893	.000
	K^+	1033.433	1	1033.433	849.477	.000
CLUSTER	EC	19201062.627	2	9600531.313	302.275	.000
	TDS	7547687.783	2	3773843.891	289.501	.000
	TH	480365.235	2	240,182.618	62.727	.000
	HCO_3^-	361264.573	2	180,632.287	24.773	.000
	Cl^-	1391657.257	2	695828.629	367.597	.000
	SO_4^{2-}	29107.043	2	14553.522	47.463	.000
	Na^+	461304.978	2	230652.489	291.921	.000
	Ca^{2+}	14445.546	2	7222.773	40.174	.000
	Mg^{2+}	9625.227	2	4812.613	76.489	.000
	K^+	186.145	2	93.072	76.505	.000

with a significance of 0.000 at p < 0.005 and F (20, 112) = 21.71. It implies that there is a statistically significant difference in the mean values of physio chemical parameters based on the formation of clusters. This shows that these physio chemical parameters have high variations in terms of spatial distributions in the study area. In other words spatial variability is significantly dependent on parameters like EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+ and K^+, across all the three clusters and shows good discriminatory ability with the ground water quality parameters.

To establish how the independent and dependant variables interact, the tests between the subjects effects, is given in Table 13. It can be seen from Table 13, that clusters (spatial distribution) has a statistically significant effect on all the physio-chemical parameters, for example for EC, F (2, 65) = 302.28, p < 0.005. Similarly for all other parameters it can be interpreted in the same manner based on the values in Table 13. Table 13 clearly indicates that there is significant interaction effect of EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl^-, SO_4^{2-}, Na^+ and K^+ in relation to various clusters (spatial distribution) as p values are <0.005. Further as the p values of all the parameters are <0.005, there is no need to perform post-hoc tests.

Precisely the results of MANOVA indicate that there are significant differences exist in the mean values of physio-chemical parameters (dependant variables) in three location variables (clusters) spatially. This phenomenon can be explained in two ways. Firstly anthropogenic activities viz., 1) Indiscriminate MSW dumping and 2) Land application of SWW in the name of ground water recharge which are mainly responsible for the increase in the concentration of EC, TDS, Na^+ and Cl^-. Secondlygeogenic and hardness factors *i.e.* weathering of rocks and dissolution of rock minerals like calcite, dolomite, gypsum, anhydrite, feldspar etc., which are chiefly responsible for increase in the concentrations of parameters like Ca^{2+}, Mg^{2+}, SO_4^{2-}, K^+, HCO_3^- and TH.

4.8. Geo-Statistical Mapping

Ordinary Kriging (OK) is an effective tool for initial decision making of ground water quality management. OK interpolation technique has been employed using ArcGIS (version 10.2) for developing spatial distribution maps of ground water data set (n = 68 borewells) based on PC/VF scores (Table 9). Spatial distribution of VF1 scores in Figure 8 reveals high scores (EC, TDS, Na^+ and Cl^-) in south east and south west parts of the study area. The high scores correspond to anthropogenic activities in and around STP. Spatial distribution of VF2 and VF5 in Figure 9 and Figure 10 shows high scores (HCO_3^-, Mg^{2+}, TH and Ca^{2+}) in northwest and south west parts of study area, indicating hardness and geogenic origin. Figure 11 and Figure 12 exhibit scores (K^+ and SO_4^{2-}) for components VF3 and VF4 in the south eastern part of the study area and the origin is geogenic. Thus the spatial distribution maps of water quality parameters (VF1 to VF5) have contributed a functional and robust visual tool for Environmental Engineers and hydro-geologists towards specifying and adoptive steps.

SPATIAL VARIABILITY-PCA-VF1
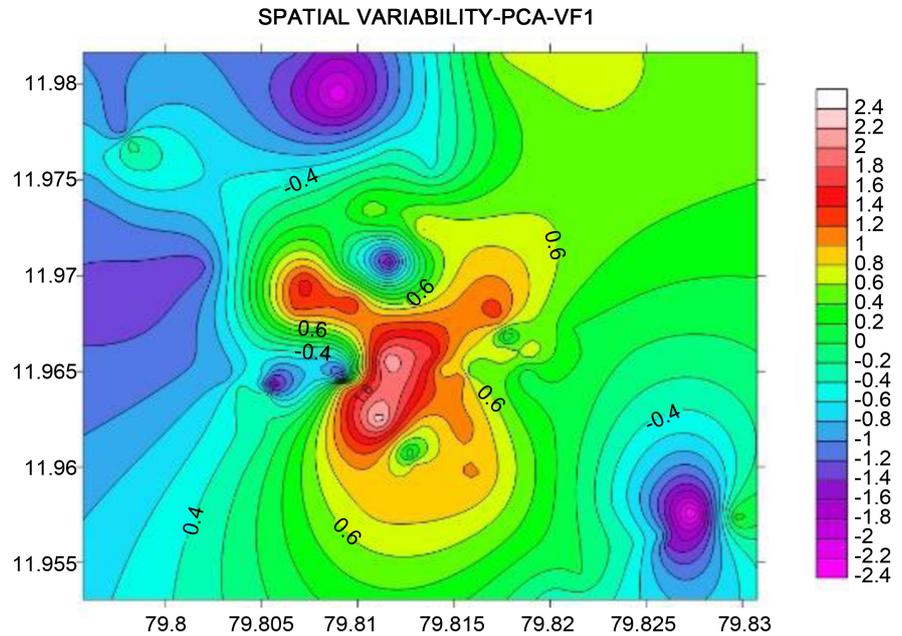
Figure 8. Spatial variability of Na$^+$ and Cl$^-$.

SPATIAL VARIABILITY-PCA-VF2
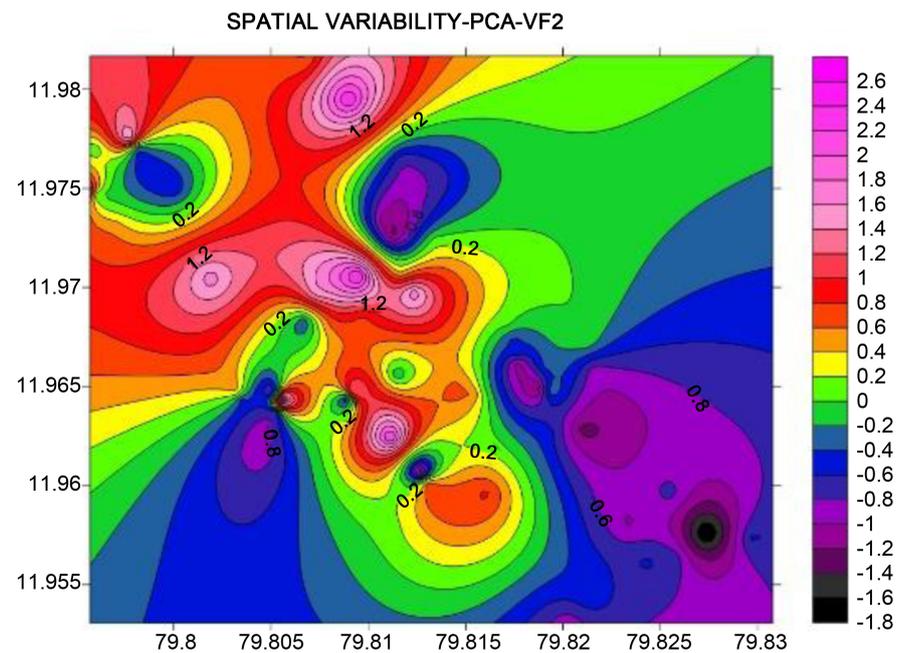
Figure 9. Spatial variability of HCO$_3^-$ and Mg^{2+}.

5. Conclusions

Different multivariate statistical techniques like HCA, PCA/FA and MANOVA were applied in this research work to investigate the spatial variability of ground water quality and to detect the main factors and sources of contamination for effective ground water management at Lawspet area, Puducherry, India. HCA identifies 68 borewells into three well defined clusters reflecting different physio-chemical processes. Based on this information available, it is easy to
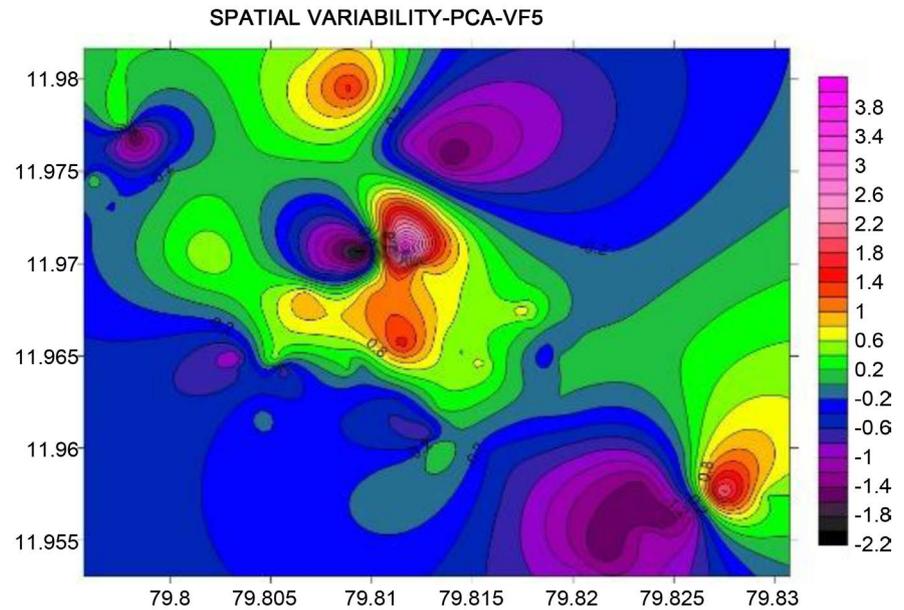
Figure 10. Spatial variability of TH and Ca^{2+}.
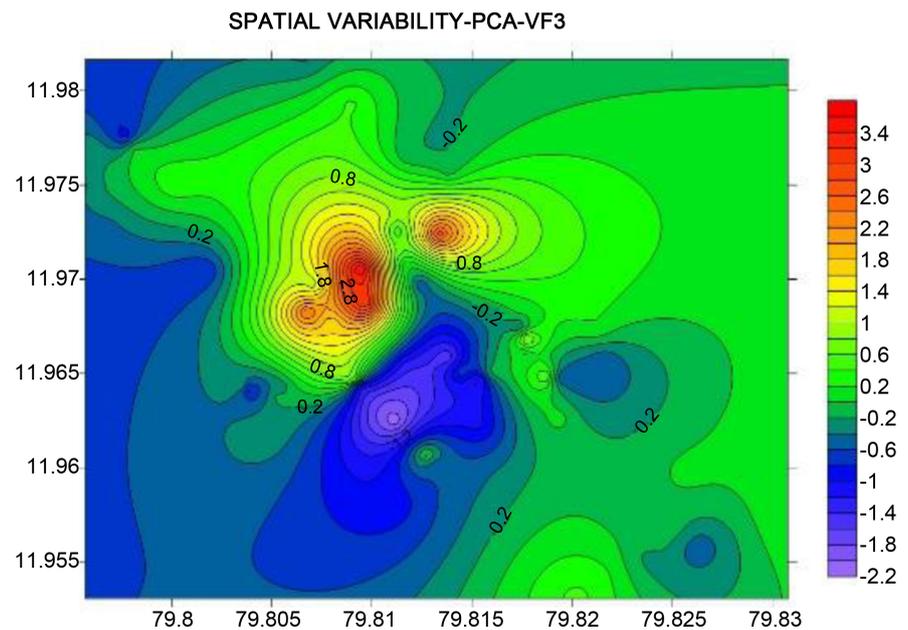
Figure 11. Spatial variability of K$^+$.

work out an optimal strategy to reduce the number of borewells (sampling points) and recurring costs. PCA/FA was used to examine the interdependence of the physio-chemical data and for distinguishing various sources which are accountable to ground water contamination and to summarize the data with least information loss. PCA/FA discloses that anthropogenic, geogenic and hardness factors are responsible for ground water pollution and these factors explained more than 82.79% of the total variance. The Q-mode PCA/FA identified the borewells which are badly affected by the pollution. In other words, the spatial variability of ground water quality has been established. Anthropogenic
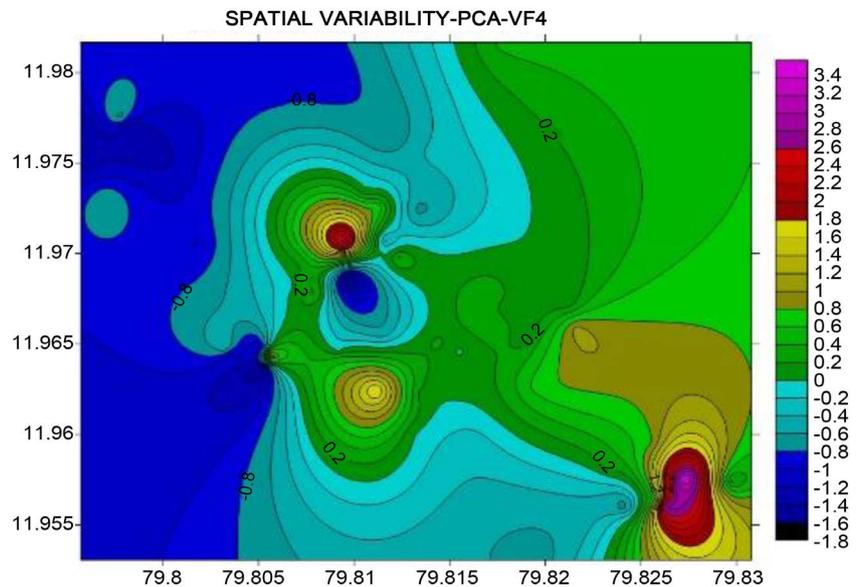
Figure 12. Spatial variability of SO_4^{2-}.

contamination due to increase in the concentration of EC, TDS, Na$^+$ and Cl$^-$ is mainly due to MSW dumping and land application of SWW, which was differentiated by high factor loading of EC, TDS, Na$^+$ and Cl$^-$ in PCA/FA. Natural processes like weathering of rocks and dissolution of rock minerals (geogenic and hardness factors) such as calcite, dolomite, anhydrite, gypsum, feldspar etc., are accountable for higher concentrations in Ca^{2+}, Mg^{2+}, K$^+$, SO_4^{2-}, HCO_3^- and TH.

Convincingly, it is established that anthropogenic operations affected the ground water sources in south east and south west parts of the study area, following the ground elevation. The geogenic and hardness factors affected the ground water in northwest and south west directions. By and large, the north eastern portion of the study area is unaffected. The one-way MANOVA test revealed that there are significant mean differences between EC, TDS, HCO_3^-, TH, Ca^{2+}, Mg^{2+}, Cl$^-$, SO_4^{2-}, Na$^+$ and K$^+$, at $p < 0.005$. The mean differences between Clusters 1, 2 and 3 also show that they are significantly different for all physio-chemical parameters.

The resulting spatial distribution maps based on Q-mode factor scores provide a beneficial and powerful visual tool for researchers and decision makers towards specifying adaptive procedures. This study contributes background information on physio-chemical parameters, polluting chemicals, contaminating factors, potential sources and spatial variation in ground water quality at Lawspet, Puducherry, India.

References

[1] Danquah, L., Abass, K. and Nikoi, A.A. (2011) Antropogenic Pollution of Inland Waters: The Case of the Aboabo River in Kumasi, Ghana. *Journal of Sustainable Development*, **4**, 103. https://doi.org/10.5539/jsd.v4n6p103

[2] Sakizadehand, M. and Ahmadpour, E. (2016) Geological Impacts on Ground Water

Pollution: A Case Study in Khuzestan Province. *Environmental Earth Sciences*, **75**, 1-12.

[3]   Fernandes, P.G., Carreira, P. and Da Silva, M.O. (2008) Anthropogenic Sources of Contamination Recognition-Sines Coastal Aquifer (SW Portugal). *Journal of Geochemical Exploration*, **98**, 1-14. https://doi.org/10.1016/j.gexplo.2007.10.003

[4]   Guler, C., *et al.* (2012) Assessment of the Impact of Anthropogenic Activities on the Ground Water Hydrology and Chemistry in Tarsus Coastal Plain (Mersin, SE Turkey) Using Fuzzy Clustering, Multivariate Statistics and GIS Techniques. *Journal of Hydrology*, **414-415**, 435-451. https://doi.org/10.1016/j.jhydrol.2011.11.021

[5]   Suresh Nathan, N., Saravanane, R. and Sundararajan, T. (2016) Assessment of Ground Water Contamination Due to Co-Disposal of Municipal Solid Waste and Secondary Wastewater on Land at Lawspet in Puducherry, India. *International Journal of Environmental Engineering and Management*, **7**, 35-67.

[6]   Zhang, B., *et al.* (2012) Hydrochemical Characteristics and Water Quality Assessment of Surface Water and Ground Water in Songnen Plain, Notheast China. *Water Research*, **46**, 2737-2748. https://doi.org/10.1016/j.watres.2012.02.033

[7]   Nematollahi, M.J., *et al.* (2016) Hydrogeochemical Investigations and Ground Water Quality Assessment of Torbat-Zaveh Plain, KhorasanRazavi, Iran. *Environmental Monitoring and Assessment*, **188**, 2. https://doi.org/10.1007/s10661-015-4968-6

[8]   Batayneh, A. and Zumlot, T. (2012) Multivariate Statistical Approach to Geochemical Methods in Water Quality Factor Identification; Application to the Shallow Aquifer System of the Yarmouk Basin of North Jordan. *Research Journal of Environmental and Earth Sciences*, **4**, 756-768.

[9]   Cobbina, S.J., *et al.* (2012) Multivariate Statistical and Spatial Assessment of Ground water Quality in the Tolon-Kumbungu District, Ghana. *Research Journal of Environmental and Earth Sciences*, **4**, 88-98.

[10] Nosrati, K. and Van Den Eeckhaut, M. (2012) Assessment of Ground Water Quality Using Multivariate Statistical Techniques in Hashtgerd Plain, Iran. *Journal of Environmental Earth Science*, **65**, 331-344. https://doi.org/10.1007/s12665-011-1092-y

[11] Machiwal, D. and Jha, M.K. (2015) Identifying Sources of Ground Water Contamination in a Hard-Rock Aquifer System Using Multivariate Statistical Analyses and GIS-Based Geostatistical Modeling Techniques. *Journal of Hydrology: Regional Studies*, **4**, 80-110.

[12] Masoud, A.A. (2014) Ground Water Quality Assessment of the Shallow Aquifers West of the Nile Delta (Egypt) Using Multivariate Statistical and Geostatistical Techniques. *Journal of African Earth Sciences*, **95**, 23-137. https://doi.org/10.1016/j.jafrearsci.2014.03.006

[13] Omo-Irabor, O.O., *et al.* (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*, **33**, 666-673. https://doi.org/10.1016/j.pce.2008.06.019

[14] Khan, T.A. (2011) Multivariate Analysis of Hydrochemical Data of the Ground Water in Parts of Karwan-Sengar Sub-Basin, Centralganga Basin, India. *Global NE-ST Journal*, **13**, 229-236.

[15] Spanos, T., *et al.* (2015) Assessment of Ground Water Quality and Hydrogeological Profile of Kavala Area, Northern Greece. *Environmental Physics*, **60**, 1139-1150.

[16] Bhuiyan, M.A.H., *et al.* (2016) Assessment of Ground Water Quality of Lakshimpur District of Bangladesh Using Water Quality Indices, Geostatistical Methods, and Multivariate Analysis. *Environmental Earth Sciences*, **75**, 23. https://doi.org/10.1007/s12665-016-5823-y

[17] El Sharabi, E.S.A. (2015) A Statistical Evaluation of Some Heavy Metals in Ground Water, Taiz, Yemen. *International Journal of Engineering Research and Science & Technology*, **4**, 160-167.

[18] Belkhiri, L. and Narany, T.S. (2015) Using Multivariate Statistical Analysis, Geostatistical Techniques and Structural Equation Modeling to Identify Spatial Variability of Ground Water Quality. *Water Resources Management*, **29**, 2073-2089. https://doi.org/10.1007/s11269-015-0929-7

[19] Kamble, S.R. and Vijay, R. (2011) Assessment of Water Quality Using Cluster Analysis in Coastal Region of Mumbai, India. *Environmental Monitoring and Assessment*, **178**, 321-332. https://doi.org/10.1007/s10661-010-1692-0

[20] Shihab, A.S. and Hashim, A. (2006) Cluster Analysis Classification of Ground Water Quality in Wells within and Around Mosul City, Iraq. *Journal of Environmental Hydrology*, **14**, 1-11.

[21] Lin, C., *et al.* (2010) Multivariate Statistical Factor and Cluster Analyses for Selecting Food Waste Optimal Recycling Methods. *Environmental Engineering Science*, **28**, 349-356. https://doi.org/10.1089/ees.2010.0158

[22] Hosseinimarandi, H., *et al.* (2014) Assessment of Ground Water Quality Monitoring Network Using Cluster Analysis, Shib-Kuh Plain, Shur Watershed, Iran. *Journal of Water Resource and Protection*, **6**, 618-624. https://doi.org/10.4236/jwarp.2014.66060

[23] Heberger, K., Milczewska, K. and Voelkel, A. (2005) Principal Component Analysis of Polymer-Solvent and Filler-Solvent Interactions by Inverse Gas Chromatography. *Colloids and Surfaces A*: *Physicochemical and Engineering Aspects*, **260**, 29-37. https://doi.org/10.1016/j.colsurfa.2005.02.029

[24] Liu, C.W., Lin, K.H. and Kuo, Y.M. (2003) Application of Factor Analysis in the Assessment of Ground Water Quality in a Black Foot Disease Area in Taiwan. *Science of the Total Environment*, **313**, 77-89. https://doi.org/10.1016/S0048-9697(02)00683-6

[25] Lalitha, A., *et al.* (2012) The Evaluation of Ground Water Pollution in Alluvial and Crystalline Aquifer by Principal Component Analysis. *International Journal of Geomatics and Geosciences*, **3**, 285-298.

[26] Tantyet, H., *et al.* (2015) MANOVA Statistical Analysis of Inorganic Compounds in Groundwater Indonesia. *AIP Conference Proceedings*, **1621**, 492-497.

[27] Basu, S. and Lokesh, K.S. (2014) Application of Multiple Linear Regression and Manova to Evaluate Health Impacts Due to Changing River Water Quality. *Applied Mathematics*, **5**, 799-807. https://doi.org/10.4236/am.2014.55076

[28] Hu, K., Li, B., Lu, Y. and Zhang, F. (2004) Comparison of Various Spatialinterpolation Methods for Non-Stationary Regional Soil Mercury Content. *Environmental Science & Technology*, **25**, 132-137.

[29] Wunderlin, D.A., *et al.* (2001) Pattern Recognition Techniques for the Evaluation of Spatial and Temporal Variations in Water Quality. A Case Study: Suquia River Basin (Cordoba-Artgentina). *Water Research*, **35**, 2881-2894. https://doi.org/10.1016/S0043-1354(00)00592-3

[30] Kundu, S. (2012) Application of Statistical Analysis in Assessment of Seasonal and Temporal Variations in Ground Water Quality. *Bulletin of Environment, Pharmacology and Life Sciences*, **1**, 7-11.

[31] Saatsaz, M., *et al.* (2013) Multivariate Statistical Techniques for the Evaluation of Spatial and Temporal Variations in Ground Water Quality of Astaneh-Kouchesfan Plain, Sefid-Rūd Basin, North of Iran. *9th International River Engineering Conference*, Ahwaz, 22-24 January 2013.