Application of Multivariate Statistical Techniques for the Characterization of Ground Water Quality of Lahore, Gujranwala and Sialkot (Pakistan)

Asif Mahmood*, Waqas Muqbool, Muhammad Waseem Mumtaz
and Farooq Ahmad

Department of Chemistry, University of Gujrat, Gujrat, Pakistan

----

Abstract
Multivariate statistical techniques such as factor analysis (FA), cluster analysis (CA) and discriminant analysis (DA), were applied for the evaluation of spatial variations and the interpretation of a large complex water quality data set of three cities (Lahore, Gujranwala and Sialkot) in Punjab, Pakistan. 16 parameters of water samples collected from nine different sampling stations of each city were determined. Factor analysis indicates five factors, which explained 74% of the total variance in water quality data set. Five factors are salinization, alkalinity, temperature, domestic waste and chloride, which explained 31.1%, 14.3%, 10.6%, 10.0% and 8.0% of the total variance respectively. Hierarchical cluster analysis grouped nine sampling stations of each city into three clusters, i.e., relatively less polluted (LP), and moderately polluted (MP) and highly polluted (HP) sites, based on the similarity of water quality characteristics. Discriminant analysis (DA) identified ten significant parameters (Calcium (Ca), Ammonia, Sulphate, Sodium (Na), electrical conductivity (EC), chloride, temperature (Temp), total hardness(TH), Turbidity), which discriminate the groundwater quality of three cities, with close to 100.0% correct assignment for spatial variations. This study illustrates the benefit of multivariate statistical techniques for interpreting complex data sets in the analysis of spatial variations in water quality, and to plan for future studies.

Keywords: Factor analysis; cluster analysis; discriminant analysis; ground water; Lahore; Gujranwala; Sialkot.

Introduction

Water is called matrix of life because it is an essential part of all living systems and is the medium from which life evolved and in which life exists [1]. The quality as well as the quantity of clean water supply is of vital significance for the welfare of humanity [2]. Polluted water is a source of many diseases for human beings [3].

Groundwater is the major source of drinking water in both urban and rural areas. Ground water is also frequently used as the alternative source for agricultural and industrial sector [4].

Distribution of groundwater quality parameters is controlled by complex processes. Ground water typically has a large range of chemical composition [5]. The ground water quality depends not only on natural factors such as the lithology of the aquifer, the quality of recharge water and the type of interaction between water and aquifer, but also on human activities, which can alter these ground water systems either by polluting them or by changing the hydrological cycle [6].

----

*Corresponding Author Email: asifmahmood023@gmail.com

Water quality monitoring has one of the highest priorities in environmental protection policy [7]. A monitoring program that provides a representative and reliable estimation of the quality of ground waters has become an important necessity. Consequently, comprehensive monitoring programs that include frequent water sampling at numerous sites and that consists a full analysis of a large number of physicochemical parameters are to be designed for proper management of water quality in ground waters. Real hydrological data are mostly noisy, it means that they are not normally distributed, often co-linear or autocorrelated, containing outliers or errors etc.

In of order to avoid this problem multivariate methods such as the factor analysis, cluster analysis and discriminant analysis were used. The application multivariable statistical methods offer a better understanding of water quality for interpreting the complicated data sets.

The specific objectives of present study are to: (1) extract latent information about the quality of groundwater (2) classified the sampling stations of each city (3) extract the parameters that are most important in assessing variations in groundwater quality of three cities.

Materials and Methods
Study area

Lahore, Gujranwala and Sialkot are three big cities of northern Punjab. Lahore is the capital of the Punjab province of Pakistan. Lahore is located at 34.94°N, 75.42° E and is 217 metre (711 feet) above sea-level. Gujranwala is located at 32.16° N, 74.18° E and is 226 metre (744 feet) above sea-level. Sialkot is a city situated in the north-east of the Punjab province in Pakistan. Sialkot is located at 32.30°N, 74.32°E and is 256 metre (840 ft) above sea-level. The ground water quality of the study area is mainly affected by the industrialization. Increased population and improper drainage system have potential to influence the ground water quality. Geographical location of study area in Pakistan is shown in (Fig. 1).
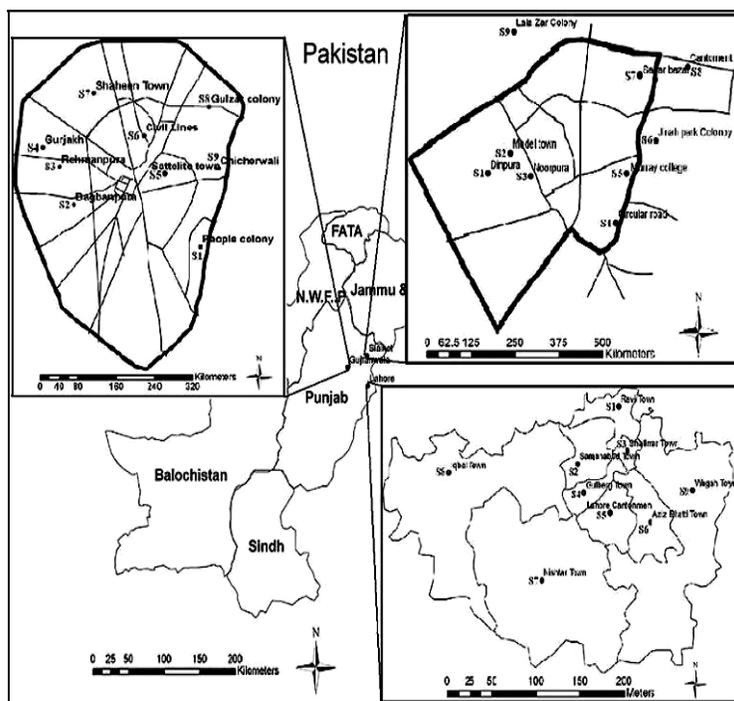
Figure 1. Location of study area in Pakistan

Sample collection

Study area consists of three cities of Punjab (Lahore, Gujranwala and Sialkot). Nine sampling stations were selected from each city. Three water samples were collected from each station. Total 81 samples (27 from each city) were collected. Water samples were collected from turbine pump, municipal supply and hand pump.

Physicochemical analysis of drinking water

The collected samples were analyzed for different physicochemical parameters such as pH, electrical conductivity (EC), temperature (Temp), turbidity, Total dissolved solids (TDS), total hardness (TH), ammonia, nitrate, sulfate, fluoride, chloride, sodium, calcium, magnesium, iron and zinc according to the standard methods (Table 1) [8].

Table 1. **Methods for the determination of physicochemical parameters.**

Parameter	Method of Determination
pH	pH meter
EC	Conductometer
Temp	Thermometer
Turbidity	Nephelometric method
TDS	Gravimetric method
TH	EDTA Titration method
Nitrate	UV spectrophotometric method
Sulphate	Turbidimetric method
Chloride	Argentometric method
Fluoride	SPADN method
Ammonia	Titration method
Na, Ca	Flame photometer
Mg, Fe Zn	Atomic absorption spectrophotometer

Electrical conductance (EC), Temperature (Temp), Total dissolved solid (TDS), Total hardness (TH), Sodium (Na), Calcium (Ca), Magnesium (Mg), Iron Fe, Zinc (Zn)

Brief review of three multivariate statistical techniques used in this study

Multivariate statistical techniques can help to simplify and organize large data sets to provide meaningful insight [9]. In the present study, three multivariate statistical techniques were used to evaluate physicochemical parameters of

groundwater samples. The statistical software package SPSS 16 and Statgraphic were used for the multivariate statistical analysis.

Factor analysis

Factor analysis is a very powerful technique applied to reduce the dimensionality of a dataset consisting of a large number of interrelated variables, while retaining as much as possible the variability presented in dataset and with a minimum loss of information [10]. This reduction is achieved by transforming the dataset into a new set of variables—factors, which are orthogonal (non-correlated) and are arranged in decreasing order of importance. FA can also be used to generate hypotheses regarding causal mechanisms or to screen variables for subsequent analysis.

FA can be expressed as:
$$F_i = a_1 x_1 j + a_2 x_2 j + ... + a_m x_m$$
Where F_i = factor
a = loading
x = measured value of variable
i = factor number
j = sample number
m = total number of variables

There are three basic steps to factor analysis:

1. Computation of the correlation matrix for all variables.
2. Extraction of initial factors.
3. Rotation of the extracted factors to a terminal solution [11].

Cluster analysis

Cluster analysis is a major technique for classifying a 'mountain' of information into manageable meaningful piles. It is a data reduction tool that creates subgroups that are more manageable than individual datum. In cluster analysis there is no prior knowledge about which elements belong to which clusters. The grouping or clusters are defined through an analysis of the data.

Hierarchical CA, the most common approach, starts with each case in a separate cluster and joins the clusters together step by step until only one cluster remains [12,13]. The Euclidean

distance usually gives the similarity between two samples, and a distance can be represented by the difference between transformed values of the samples [14].

There are four basic cluster analysis steps:

1. Data collection and selection of the variables for analysis
2. Generation of a similarity matrix
3. Decision about number of clusters and interpretation
4. Validation of cluster solution

Discriminant analysis

Discriminant analysis is a technique for classifying a set of observations into predefined classes. It operates on raw data and the technique constructs a discriminant function for each group [12, 15]. A simple linear discriminant function transforms an original set of measurements on a sample into a single discriminant score [16]. DA involves the determination of a linear equation that will predict which group the case belongs to. The form of the equation or function is:

$$D = v_1 X_1 + v_2 X_2 + v_3 X_3 ....v_i X_i + a$$

Where D = discriminate function
v = the discriminant coeffi cient or weight for that variable
X = respondent's score for that variable
a = constant
i = the number of predictor variables

Data processing

Factor analysis is applied on experimental data standardized through *z*-scale transformation in order to avoid misclassification due to wide differences in data dimensionality [17]. Furthermore, the standardization procedure eliminates the influence of different units of measurements and renders the data dimensionless.

In the standardization, the raw data were converted to unitless form of zero mean and a variance of one, by subtracting from each variable the mean of data set and dividing by standard deviation. This type of ordination reduces the dimensionality of the data set and minimizes the loss of information caused by reduction.

Results and Discussion

Sixteen physicochemical parameters were determined during this study. Descriptive statistics of all the parameters is given in (Table 2). Large standard deviations of most of the parameters revealed their randomly fluctuating concentration levels in the groundwater.

Table 2. **Descriptive Statistics of water quality Data of Lahore, Gujranwala and Sialkot.**

Parameter	Unit	Minimum	Maximum	Mean	Std. Deviation
pH		6.87	9.42	8.193	0.585
EC	uS/cm	660.00	2785.0	1256.6	501.69
Temp	^{0}C	23.00	31.70	27.37	2.23
Turbidity	NTU	0.00	3.00	1.65	1.05
TDS	mg/l	367.00	1368.0	837.11	263.42
TH	mg/l	164.00	1152.0	339.96	216.14
Nitrate	mg/l	6.20	18.90	11.59	3.21
Sulphate	mg/l	54.00	437.00	121.62	81.22
Chloride	mg/l	115.00	375.00	256.11	61.44
Fluoride	mg/l	0.001	0.590	0.169	0.16
Ammonia	mg/l	0.00	0.39	0.079	0.074
Na	mg/l	35.00	300.00	114.43	56.20
Ca	mg/l	85.00	853.00	218.78	159.70
Mg	mg/l	25.00	286.00	91.09	58.30
Fe	mg/l	0.000	0.097	0.028	0.024
Zn	mg/l	0.002	1.470	0.323	0.452

Electrical conductance (EC), Temperature (Temp), Total dissolved solid (TDS), Total hardness (TH), Sodium (Na), Calcium (Ca), Magnesium (Mg), Iron Fe, Zinc (Zn).

Statistical analysis
Correlation between variables

First step in factor analysis is the determination of the parameter correlation matrix. It is used to account for the degree of mutually shared variability between individual pairs of water quality variables. The correlation matrix with which we can observe the relationship between parameters was obtained and tabulated in (Table 3).

Table 3. Correlation coefficients for Sixteen Physicochemical parameters.

	pH	EC	Temp	Turbidity	TDS	TH	Nitrate	Sulphate	Chloride	Fluoride	Ammonia	Na	Ca	Mg	Fe	Zn
pH	1.000															
EC	-.093	1.000														
Temp	-.075	-.206	1.000													
Turbidity	.207	-.028	-.074	1.000												
TDS	-.135	.583**	-.010	-.297**	1.000											
TH	-.040	.628**	.154	-.154	.501**	1.000										
Nitrate	-.172	.151	.366**	-.145	.071	.141	1.000									
Sulphate	-.167	-.021	.051	-.347**	.028	-.032	.020	1.000								
Chloride	.143	.252*	-.040	-.182	.220*	.010	.226*	-.240*	1.000							
Flouride	.037	.036	-.100	.367**	-.079	-.244*	.112	-.176	.099	1.000						
Ammonia	.100	.466**	-.059	.248*	.231*	.524**	-.273**	-.183	-.037	.181	1.000					
Na	.242*	-.087	-.189	.146	-.105	-.149	-.296**	-.298**	.406**	-.071	.054	1.000				
Ca	-.029	.583**	.174	-.128	.472**	.978**	.099	-.017	-.082	-.243*	.543**	-.183	1.000			
Mg	-.067	.531**	.014	-.183	.545**	.889**	-.012	-.101	-.105	-.331**	.413**	-.165	.898**	1.000		
Fe	-.015	.373**	.329**	-.103	.200	.594**	.252*	.104	.071	.092	.527**	-.249*	.591**	.373**	1.000	
Zn	-.188	.166	.452**	-.282*	.299**	.622**	.134	.099	-.289**	-.141	.323**	-.311**	.645**	.568**	.597**	1

*Correlation is significant at the 0.05 level (2-tailed)
**Correlation is significant at the 0.01 level (2-tailed)
Electrical conductance (EC), Temperature (Temp), Total dissolved solid (TDS), Total hardness (TH), Sodium (Na), Calcium (Ca), Magnesium (Mg), Iron Fe, Zinc (Zn)

Correlation studies between different variables are very helpful tools in promoting research and opening new frontiers of knowledge. The study of correlation reduces the range of uncertainty associated with decision making [18].

pH shows inverse relationships with most of the anions and cations, as pH decrease more rock dissolution occurs. EC shows highly significant (p<0.01) positive correlation with six water quality parameters namely TDS, TH, Ammonia, Ca, Mg and Fe. This indicated that these parameters have similar hydrochemical characteristics in the study area. Nitrate does not significantly contribute to conductivity because of its low concentrations.

Temperature is positively correlated with Fe and Zn at highly significant level (p<0.01). Turbidity is correlated positively with Fluoride and negatively with TDS and sulphate at highly significant level (p<0.01). Chloride shows positive correlation with Na and negative correlation with Zn at highly significant level (p<0.01). Ca, Mg, Fe and Zn are positively correlated with each other at highly significant level (p<0.01).

It clear from Table 3 that the relationship between the parameters having high ion character was observed to be stronger than that of the parameters having less ion character.

Factor analysis

81 water samples were collected from three cities and 16 physicochemical parameters were determined. This water quality data was analyzed by using factor analysis. Before conducting FA, the Kaiser–Meyer–Olkin (KMO) [19] and Bartlett's sphericity [20] tests were performed on the parameter correlation matrix to examine the validity of FA. FA was conducted for all samples' physicochemical data set, and the results were 0.705 for the KMO and 891.9 ($p < 0.0001$) for Bartlett's sphericity, indicating that FA may be useful in providing significant reductions in dimensionality.

From data, 5 factors, explaining 74 % of the total variance, was estimated on the basis of Kaiser criterion [21] of the eigenvalues greater or equal 1 and from a Cattel scree plot [22]. A scree plot shows the eigenvalues sorted from large to small as a function of the factor number. After the fifth factor (Fig. 2), starting the elbow in the downward curve, other components can be omitted. Factor was extracted by principal component method and rotated by Varimax. The factor loading, their eigenvalues, and variances are summarized in (Table 4).
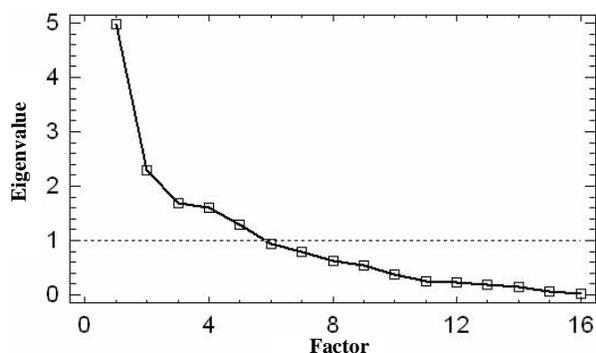
Figure 2. **Scree plot of the eigenvalues.**

Table 4. **Rotated Factor Loading Matrix, eigenvalues, % variance and cumulative variance values.**

Parameter	Factor				
	1	2	3	4	5
pH	-0.034	**0.633**	0.014	0.042	-0.082
EC	**0.722**	- 0.132	- 0.272	0.187	0.415
Temp	0.020	-0.003	**0.873**	-0.110	-0.030
Turbidity	-0.089	0.361	- 0.092	**0.668**	-0.315
TDS	**0.605**	- 0.204	- 0.191	-0.159	0.417
TH	**0.939**	0.002	0.169	-0.130	0.045
Nitrate	-0.025	-0.300	0.571	0.111	0.554
Sulphate	-0.075	-0.625	- 0.010	-0.266	-0.133
Chloride	-0.024	0.372	- 0.022	-0.030	**0.850**
Fluoride	-0.147	-0.060	- 0.001	**0.853**	0.164
Ammonia	**0.696**	0.241	-0.083	0.414	-0.195
Na	-0.143	**0.741**	- 0.254	-0.161	0.173
Ca	**0.938**	- 0.009	0.181	-0.120	-0.052
Mg	**0.887**	- 0.019	- 0.013	-0.257	-0.063
Fe	**0.613**	- 0.081	0.502	0.231	0.063
Zn	**0.615**	- 0.212	0.526	-0.147	-0.237
Eigenvalue	4.970	2.296	1.691	1.602	1.283
% Variance	31.06	14.35	10.57	10.01	8.02
Cumulative %	31.06	45.41	55.98	65.99	74.01

Electrical conductance (EC), Temperature (Temp), Total dissolved solid (TDS), Total hardness   (TH), Sodium (Na), Calcium (Ca), Magnesium (Mg), Iron Fe, Zinc (Zn).

Parameters were grouped based on the factor loading and following factors were indicated:

Factor 1: TH, Ca, Mg, EC, Ammonia, Zn, Fe, TDS
Factor 2: Na, pH
Factor 3: Temperature
Factor 4: Fluoride, Turbidity
Factor 5: Chloride

TH, Ca, Mg, EC, Ammonia, Zn, Fe and TDS marked factor 1, which explained 30.1% of the variance. Factor 1 had a high positive loading in TH, Ca, Mg, EC, Ammonia, Zn, Fe and TDS which were 0.939, 0.938, 0.887, 0.722, 0.696, 0.615, 0.613 and 0.605 respectively. High positive loadings indicated strong linear correlation between the factor and parameters.

Thus, factor 1 can be termed as salinization factor. The electrical conductivity (EC) is positively correlated with the concentration of ions, which can thus be indirectly calculated from EC. Therefore, EC can be regarded as a water salinization index. Simultaneous drought and over-pumping have led to deterioration of the groundwater. EC can be readily measured and used as a surrogate for the presence of the remaining parameters. Reducing the number of these parameters allows resources to be freed up for additional measurements elsewhere. The additional analytes could still be sampled, especially during periods when elevated EC is observed.

Factor 2, with higher loading of Na and pH explained 14.3% of variance. pH of most of the water samples was greater than 7. Alkalinity of water may be due carbonate and bicarbonate of Na. Second factor can be called as alkalinity factor. Factor 3 explained 10.6% of variance and temperature gave most contribution with a loading of 0.873. Climate effects are playing an active role on the 3rd factor. This factor can be denoted as temperature factor.

Factor 4 responsible for 10.0% of total variance and best represented by fluoride and turbidity with loading of 0.853 and 0.668 respectively. Turbidity is due colloidal particles that come from domestic waste water that drain into nearby rivers, canals and streams, and then migrate to water table. 4th factor can be termed as domestic waste factor. Factor 5 represents chlorine and can be called as the chlorine factor. It is also obvious from the lower loading of Fe in fifth factor that the chlorine content does not depend on the iron concentrations in water. Factor 5 explained the 8.0% variance. In (Fig. 3) plot of loadings of first three factors indicates the contribution of different parameters towards first three factors.
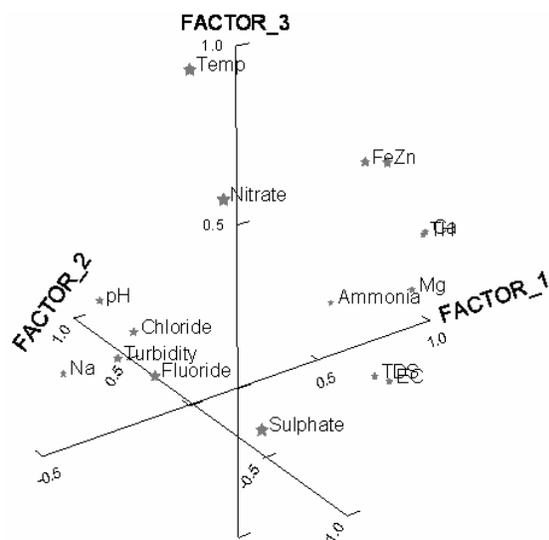
Figure 3. **Plot of loading of first three factors.**

Factor scores for the stations of three cities are calculated to determine the level of pollution, which are given in (Table 5 - 7).

Table 5. **Factor scores of the stations of Lahore.**

Station No.	Factor				
	1	2	3	4	5
1	-1.675	2.097	-1.911	2.337	0.122
2	2.170	2.124	-1.443	0.062	0.118
3	-3.485	1.779	-1.170	2.450	0.715
4	-0.383	-1.391	0.157	3.933	2.756
5	-4.280	1.681	0.283	2.434	-0.431
6	-1.332	2.205	-0.335	2.925	-0.506
7	-4.577	4.384	-3.091	0.970	-1.193
8	-2.733	3.355	-1.844	0.584	-0.600
9	-1.039	1.293	-2.712	1.242	0.372

Table 6. **Factor scores of the stations of Gujranwala.**

Station No.	Factor				
	1	2	3	4	5
1	6.479	-0.226	1.434	-0.839	3.569
2	6.479	-0.226	1.434	-0.839	3.569
3	0.539	-1.339	-1.253	1.051	-1.549
4	5.849	-2.064	3.582	-1.228	2.975
5	0.122	-1.338	0.149	-1.432	-1.609
6	2.067	-0.925	3.059	1.334	-3.478
7	3.144	-0.543	-1.450	-3.011	0.755
8	18.549	-0.545	1.800	-1.036	-0.491
9	3.374	-1.261	3.321	-1.289	-1.257

Table 7. **Factor scores of the stations of Sialkot.**

Station No.	Factor				
	1	2	3	4	5
1	-1.088	-1.219	-1.958	-1.396	1.154
2	-0.617	-0.935	-0.102	0.261	-0.107
3	-4.546	-3.190	-1.820	-1.896	-2.841
4	-4.950	0.455	0.519	-0.128	-1.794
5	-0.319	-1.407	-1.180	-0.007	1.260
6	-3.868	0.562	1.739	-1.944	0.944
7	-4.360	0.000	-1.539	-1.599	-0.125
8	-3.953	1.171	-0.446	-2.146	2.442
9	0.778	-3.972	4.822	-0.827	-0.437

Cluster analysis

In this study, sampling site classification was performed by the use of cluster analysis. Hierarchical CA was performed on the factor scores obtained from factor analysis using Ward's method with squared Euclidean distances as a measure of similarity. Cluster analysis from factor scores of stations reduce the clustering error caused by data error or multicollinearity. Ward's method uses analysis of variance (ANOVA) to calculate the distances between clusters to minimize the sum of squares of any two possible clusters at each step. Results of cluster analysis are represented using dendogram. The distance in dendogram is equal to $(D_{link}/D_{max}) \times 100$, which represents the quotient between the linkage distances for a particular case divided by the maximal linkage distance. The quotient is then multiplied by 100 as a way to standardize the linkage distance [7, 15].

In the dendograms of three cities (Lahore, Gujranwala and Sialkot) water sampling stations are classified into three clusters at $(D_{link}/D_{max}) \times 100 < 60$ as shown in Fig. 4, 5, 6.

On the basis of Cluster analysis stations of Lahore are divided as follows:

Cluster I (Station 1, 9, 6, 3, 5)
Cluster II (Station 7, 8)
Cluster III (Station 2, 4)
Stations of same clusters have the similar pattern of the groundwater quality.
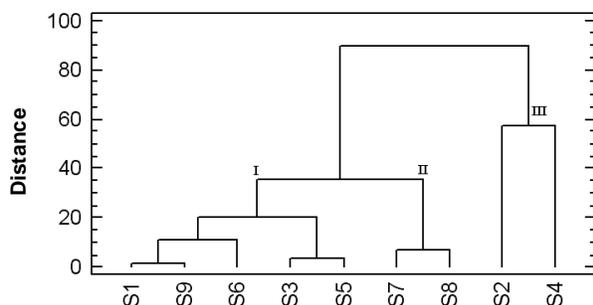
Figure 4. **Dendorgram of the hierarchical cluster analysis of the groundwater quality of sampling stations of Lahore.**

Station 1, 9, 6, 3 and 5 of cluster I are moderately polluted areas of Lahore. Stations have lower loadings in most of the factors but loadings of these stations in factor 1are higher than the loadings of stations of cluster II in factor 1. Factor 1 is major contributor in ground water pollution. Due to this reason stations of cluster I are more polluted than the stations of cluster II. Cluster II is corresponding to station 7and 8 which are less polluted areas of Lahore. These stations have higher loading in factor 2. This indicates the presence of higher concentration of Na and higher values of pH but have high negative loading in factor 1. These stations have negative loading of factor 5, which is chloride factor. This indicates that very less amount of Na is present as NaCl in these stations. Station 2 and 4 of cluster III are highly polluted. Pollution in station 2 is mainly due factor 1 and 2. Pollution in station 4 is mainly due to factor 4 and 5.

On the basis of Cluster analysis stations of Gujranwala are divided as follows:

Cluster I (Station 1, 2, 4)
Cluster II (Station 3, 5, 7, 6, 9)
Cluster III (Station 8)
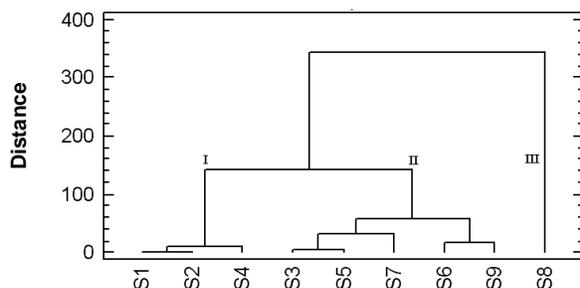
Figure 5. **Dendorgram of the hierarchical cluster analysis of the groundwater quality of sampling stations of Gujranwala.**

Cluster II is corresponding to station 3, 5, 7, 6 and 9 which are less polluted areas of Gujranwala as indicated by factor loadings. Station 1, 2 and 4 of cluster I are moderately polluted. In these stations pollution is mainly due factor 1, 3 and 5. Station 8 of cluster III is highly polluted. In this station pollution is mainly due to factor 1.

On the basis of Cluster analysis stations of Sialkot are divided as follows:

Cluster I (Stations 1, 5, 2)
Cluster II (Stations 3, 4, 7, 6, 8)
Cluster III (Stations 9)
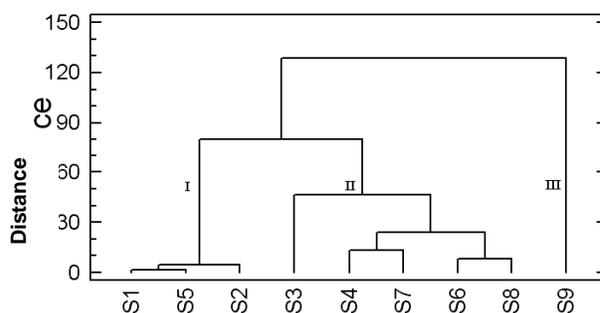
Figure 6. **Dendorgram of the hierarchical cluster analysis of the groundwater quality of sampling stations of Sialkot.**

Cluster 1 contains station 1, 5 and 2 which are moderately polluted areas of Sialkot. In these stations pollution is mainly due factor 1. Factor 1 contains parameters which are very importance in the determination of water quality. Station 3, 4, 7, 6 and 8 belonged to cluster II are less polluted areas because these have high negative loading in most of the factors.

Station 9 corresponding to cluster III is totally different from other stations of city. It is a high polluted area of Sialkot. Station 9 have high positive loading of factor 1 and 3 especially very high loading for factor 3 (4.822) indicates the presence of higher concentration of Fluoride and higher value of turbidity.

It is clear from dendograms of stations of three cities that all the clusters join at the distance of 90 in the case of Lahore, 340 in the case of Gujranwala and 130 in the case of Sialkot. This indicates that variation in the quality of water of Lahore is very less and very large in the stations of Gujranwala.

By comparing the factor loading of stations of three cities given in Table 5, 6 and 7 it is clear that water of Gujranwala is most polluted and Sialkot is less polluted. Overall order of water pollution is: Gujranwala > Lahore > Sialkot.

These cities are totally dependent on ground water for drinking. Sewage system of these cities is defective especially Gujranwala. Most main sewers are 30-50 ft below ground level and are made of 10ft cement sections linked without proper safety seals. Poor connections combined with deteriorating low quality sewer pipes cause a lot of leakage. This outflow from sewer mixes with the water table and the contamination is carried to deeper levels. Industrial wastewater contains toxic chemicals. It is alarming that most industries have been started without proper planning and waste treatment plants. They just dispose of untreated toxic waste into nearby drains, canals or rivers. Lahore, Gujranwala, Sialkot contribute major pollution loads into their water bodies.

Ground water resources in Gujranwala are adequate and due to recharging of the transmissive aquifer are sustainable. However, the shallow water table in the city is being depleted due to the massive use of individual pumps. Also the shallow water is seriously contaminated [23]. Pakistan Council of Research in Water Resources (PCRWR) [24] carried out a survey of major cities of Pakistan among which was Gujranwala. The results of the survey indicated serious contamination problems. Sialkot has a good ground water aquifer, which is recharged by the River Chenab in the northeast and River Ravi tributaries running through the city. The water table is 10-15 meters deep. The upper strata are polluted by industrial waste, however the deeper strata from 90-100 meter are generally considered to be safe. It is estimated that the ground water yield is adequate for Sialkot's future needs as well [25]. According to PCRWR ground water of Sialkot is contaminated. The overall comparison of three cities also made using cluster analysis. Water quality of three cities is different from each other. Lahore and Sialkot combine into one cluster at the distance of 200. This indicates the some extend of similarity between the water quality of Lahore and Sialkot. Cluster of Gujranwala combine to the cluster of Lahore and Sialkot at the distance of

820. This indicates that of quality of water of Gujranwala is totally different from that of Lahore and Sialkot.
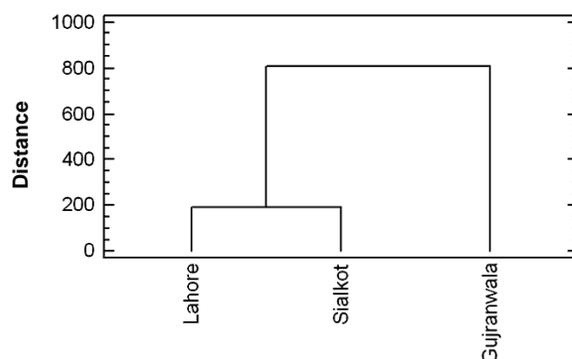
Figure 7. **Dendorgram of the hierarchical cluster analysis of the groundwater quality of three cities.**

Based on the above results, hierarchical CA provided a classification of groundwater quality that aided in designing an optimal spatial monitoring plan with a sharply reduced number of monitoring sites and corresponding costs.

Discriminant analysis

Discriminant analysis was used to find one or two functions (linear combinations) of the observed data (called discriminant functions) that best separate the water quality of three cities and classified the three cities. Standard mode discriminant analysis was applied in present study. DA was applied on raw data. Two discriminate functions (DFs) were found to discriminate the three cities as shown in Table 8. Wilk's Lambda test showed that both functions are statistically significant (Table 9). Furthermore, 100 % of the total variance between the three cities explained by the two DFs. The first DFs explained 66.5% of the total spatial variance, and the second DFs explained 33.5 %. The relative contribution of each parameter to both functions is given in Table 10.

Parameters were grouped based on function coefficients and following functions are indicated:
Function 1:   Ca, Ammonia, Sulphate, Na, EC, Chloride, Temp
Function 2:   TH, Turbidity

In first function Ca, Ammonia, Sulphate, Na, EC, Chloride and Temp exhibited strong

contribution in discriminating the three cities and account for most of the expected spatial variations in the quality of water of three cities, while less contribution exhibited from other parameters. In second function, contribution of TH and turbidity in explaining the spatial variations is major as shown in Table 10. The classification matrix showed that 100.0 % of the cases are correctly classified to their respective groups, as shown in Table 11. The result of classification shows that there are significant differences between three cities, which are expressed by in terms of two discriminate functions.

Table 8. **Eigen-values for two discriminant function for three cities.**

Function	Eigen-value	% Variance	Cumulative %
1	9.305	66.5	66.5
2	4.681	33.5	100

Table 9. **Wilks' Lambda test of DFs for spatial variation of ground water quality of three cities.**

Test of Function(s)	Wilks' Lambda	Chi-square	Sig.
1 through 2	0.017	286.923	0.000
2	0.176	122.473	0.000

Table 10. **Discriminant function coefficients of spatial variation of ground water quality of three cities.**

Parameter	Function	
	1	**2**
pH	-0.090	0.305
EC	**0.828**	-0.175
Temp	**0.562**	-0.844
Turbidity	0.177	**0.925**
TDS	-0.096	-0.046
TH	-3.032	**0.818**
Nitrate	0.442	-0.059
Sulphate	**1.149**	-0.634
Chloride	**0.605**	0.385
Fluoride	-0.318	0.467
Ammonia	**1.159**	0.174
Na	**0.860**	0.092
Ca	**1.282**	-1.130
Mg	-0.314	0.155
Fe	-0.345	0.464
Zn	-0.464	0.301

Electrical conductance (EC), Temperature (Temp), Total dissolved solid (TDS), Total hardness   (TH), Sodium (Na), Calcium (Ca), Magnesium (Mg), Iron Fe, Zinc (Zn).

Table 11. **Classification results for discriminant analysis of Three Cities.**

City	% correct *a*	Predicted Group Membership		
		1	2	3
Lahore	100.0	27	0	0
Gujranwala	100.0	0	27	0
Sialkot	100.0	0	0	27

100.0% of original grouped cases correctly classified.

Scores of two functions were plotted. Plot of scores of two functions clearly classified the three cities, as shown in Fig. 7. This indicates the difference in the quality of groundwater of three cities.
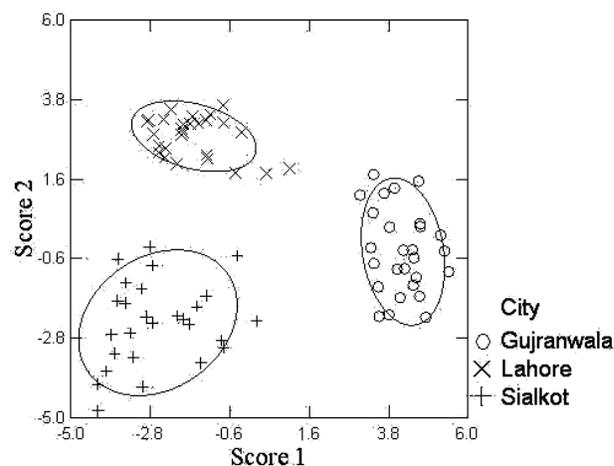
Figure 8. **Canonical Scores Plot.**

Conclusion

The study showed that the analysis of hydrochemical data using the multivariate statistical techniques such as factor analysis, cluster analysis and dicriminant analysis can give some information not available at first glance. Factor analysis is an effective means of manipulating, interpreting, and representing data concerning groundwater pollutants. Factor analysis converted the sixteen parameters in to five factors, which explained the data set with minimum loss of information. The first factor termed as salinization factor, explained 31. 1% of the total variance. The second factor can be called as alkalinity factor, which explained 14.3% of the total variance. Third factor is temperature factor, which explained 10.6% of the total variance. Fourth factor can be

termed as domestic waste factor, which explained 10.0% of the total variance. Remaining 8.0 % of the total variance is explained by fifth factor, which termed as chloride factor. Hierarchical cluster analysis grouped nine sampling stations of each city into three clusters, i.e., relatively less polluted (LP), moderately polluted (MP) and highly polluted (HP) stations, based on the similarity of water quality characteristics. It provides a useful classification of the surface watercourses in the study area that can be applied to the optimization of future spatial monitoring network with lower cost. Discriminant analysis indicated the ten significant parameters (Ca, Ammonia, Sulphate, Na, electrical conductivity, chloride, Temp, TH, Turbidity), which discriminate the groundwater quality of three cities. It is also classified the three cities 100% correctly. Therefore, DA allowed a reduction in the dimensionality of the large data set and indicated a few significant parameters responsible for large variations in water quality that could reduce the number of sampling parameters. Hence, this study illustrates that multivariate statistical methods are an excellent exploratory tool for interpreting complex water quality data sets and for understanding spatial variations, which are useful and effective for water quality management.

Reference

1. F. Franks, Water: a Matrix of Life .Second Edition, RSC Paperbacks. UK, (2000) 225
2. S. S. Dara, Water pollution. In: A textbook of Environmental chemistry and Pollution control, S. Chand & Company Ltd., New Delhi, (2008) 65.
3. A. Nag, Waste water analysis. In: Analytical techniques in agriculture, biotechnology and environmental engineering. Prentice-Hall of India Publishers, New Delhi, (2006) 103.
4. A. Mishra and V. Bhatt, E-*Journal of Chemistry*, 5 (2008) 487.
5. C. A. J. Appelo and D. Postma, Geochemistry, groundwater and pollution. Balkema, Rotterdam. (1998).
6. B. Helena, B. Pardo, M. Vega, Barrado, J. M. Fernandez and L. Fernandez, *Water Ress.,* 32 (2000) 19.
7. V. Simeonov, Jw. Einax, I. Stanimirova and J. Kraft, *Anal. Bional. Chem.* 374(2002) 898.
8. APHA, Standard Methods for the Examination of Water and Wastewater. 20th edn, American Public Health Association/American Water Works Association/Water Environment Federation, Washington DC, USA. (1998).
9. M. Laaksoharju, C. Skarman and E. Skarman, *Applied Geochemistry,* 14 (1999) 861.
10. J. F. Hair, Multivariate data analysis (3rd ed.). New York: Macmillan, (1992).
11. Ho. Robert, Handbook of univariate and multivariate data analysis and interpretation with SPSS. Chapman & Hall/CRC USA, (2006) 391.
12. J. Lattin, D. Carroll and P. Green, Analyzing multivariate data. New York: Duxbury, (2003).
13. J. McKenna, *Environmental Modelling and Software,* 18 (2003) 205.
14. M. Otto, Multivariate methods. In: R. Kellner, J. M. Mermet, M. Otto and H. M. Widmer, (Eds.), Analytical chemistry. Weinheim: Wiley-VCH. (1998).
15. D. A. Wunderlin, M. D. P. Diaz, M. V. Ame, S. F. Pesce, A. C. Hued and M. D. Bistoni, *Water Res.,* 35 (2001) 2881.
16. Sanchez Lopez, F. J.; Gil García, M. D.; Martinez Vidal, J. L.; Aguilera, P. A.; Garrido Frenich, *Environ. Moni. Assess.,* 93 (2004) 17.
17. C. W. Liu, K. H. Lin and Y. M. Kuo, *Science of the Total Environment,* 313 (2003) 77.
18. D. M. Joshi, N. S. Bhandari, A. Kumar, A. and N. Agrawal, *Rasayan J. Chem.,* 2 (2009) 579.
19. H. Kaiser, *Psychometrika,* 35 (1970) 401.
20. M. S. A. Barrlett, *Journal of the Royal Statistical Society,* 16 (Series B), (1954) 296.
21. H. Kaiser, *Psychol. Meas.* 20 (1960) 141.
22. R. D. Cattel, *Multivariate Behav. Res.,* 1 (1966) 245.
23. World Bank, Urban Water Supply and Sewerage Reform Strategy. Status Quo Report Gujranwala. (2006).
24. PCRWR, Pakistan Council of Research in Water Resources. Anaual report 2002-2003. (2003) 75.
25. World Bank, Urban Water Supply and Sewerage Reform Strategy. Status Quo Report Sialkot. (2006).