

## Research Article

# Investigating the Impact of Anthropogenic and Natural Sources of Pollution on Quality of Water in Upper Indus Basin (UIB) by Using Multivariate Statistical Analysis

Mansoor A. Baluch  and Hashim Nisar Hashmi

Faculty of Civil and Environmental Engineering, University of Engineering & Technology, Taxila, Pakistan

Correspondence should be addressed to Mansoor A. Baluch; mansoorbaluch@gmail.com

Received 16 June 2019; Accepted 12 September 2019; Published 16 November 2019

Guest Editor: Lisa Yu

Copyright © 2019 Mansoor A. Baluch and Hashim Nisar Hashmi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Water quality of the Indus River around the upper basin and the main river was evaluated with the help of statistical analysis. In order to analyze the similarities and dissimilarities for identifying the spatial variations in water quality of the Indus River and sources of contamination, multivariate statistical analysis, i.e., principle component analysis (PCA), cluster analysis, and descriptive analysis, was done. Data of 8 physicochemical quality parameters from 64 sampling stations belonging to 6 regions (labeled as  $M_1$ ,  $M_2$ ,  $M_3$ ,  $M_4$ ,  $M_5$ , and  $M_6$ ) were used for analysis. The parameters used for assessing the water quality were pH, dissolved oxygen (DO), oxygen reducing potential (ORP), electrical conductivity (EC), total dissolved solids (TDS), salinity (%), and concentration of arsenic (As) and lead (Pb), respectively. PCA assisted in extracting and recognizing the responsible variation factors of water quality over the region, and the results showed three underlying factors including anthropogenic source pollution along with runoff due to rain and soil erosion were responsible for explaining the 93.87% of total variance. The parameters which were significantly influenced by anthropogenic impact are DO, EC, TDS (negative), and concentration of Pb (positive), while the concentration of As, % salinity, and ORP are affected by erosion and runoff due to rain. The worst pollution situation for regions  $M_1$  and  $M_6$  was due to the concentration of As which was approximately 400  $\mu\text{g/l}$  (i.e., 40 times higher than minimum WHO recommendation). Furthermore, the results also indicated that, in the Indus River, three monitoring stations and five quality parameters are sufficient to have a reasonable confidence about the quality of water in this most important reserve of Pakistan.

## 1. Introduction

One of the most influential elements regarding environment concerns all over the world is the anthropogenic distribution, e.g., sewage discharge into the river, reclamation process of the land, and climate change due to atmospheric deposition effects [1]. In the recent years, much attention is paid by various researchers in assessing the quality of surface water because of its direct connection not only with human beings but also with other species [2]. The major factors affecting the quality of river water usually are morphology of the basin and regional atmosphere along with climate change, and both natural and anthropogenic effects are responsible for governing these effects [3]. Wastewater from human activities like from industries, agriculture, and

natural degradation, e.g., weathering, affects water quality impede not only for drinking purpose but also not fit for agriculture and other utilization. Clean water is an imperative part for the affluence of human society, but the major serious environmental problem of the last century was the vandalization of the land aquatic system [4]. The most harmful pollution to the surface water bodies is due to the role of increasing contamination from industrial and urban wastewater along with the runoff from the agricultural land. The characteristics of the surface contaminations are considered as a nonpoint source, while industrial and municipal discharges are considered as a main point source due to the influenced responsibilities of water pollution [5–7]. Precipitations due to the climate change, surface flow, ground water flow, and the flow due to the pump are the major ones

responsible for the discharge of the river along with the pollutant concentration on the surface water [5]. It is therefore an effective, long lasting monitoring protocol because the surface water needs an essential knowledge of hydromorphological, hydrochemical, and hydrobiological characteristics [8]. There are other important variations, e.g., spatial and temporal variations. For such kinds of variation monitoring, there should be some conceptual and essential estimations of the surface water quality [9], and some standards, e.g., here in this study, such as FAO-29 guidelines and WHO recommendations (given in Table 1) were utilized for assessing the water quality.

Researchers have utilized number of techniques and methods to unveil the possible hidden sources of pollutions for various rivers around the world. For example, in another study, Aalami et al. [10] implemented the structured best management practices (BMPs) to define quantity-based sustainability index (SI). They had done this to link the water quality with the outflow of reservoir. Couto et al. [11] figured out the trace element distribution and their possible sources while defining the pollution indices, i.e., geoaccumulation (Igeo) and enrichment factor (EF) during the investigation for anthropogenic pressure of Ave River Portugal. Rahman et al. [12] investigated the drinking water sources in the southwest coastal area of Bangladesh and found that the availability of trace metals originated both from natural and anthropogenic sources. Salam et al. [13] utilized inductively coupled plasma-optical emission spectrometry (ICP-OES) for estimating the trace elements and found that the concentrations of heavy metals at downstream of Perak River water were higher than the concentrations of upstream at the Perak River Basin, Malaysia. Barceló et al. [14] investigated the waters from Uzunçayır Dam (Tunceli), Turkey, and concluded that the water of this dam is safe for agriculture and household use. Das Sharma et al. [15] assessed the heavy metal concentration at Kolleru lake in Andhra Pradesh, India, with multiple index values and revealed the fact that the lake was influenced anthropogenic input. Bhuyan et al. [16] assessed the water quality of the old Brahmaputra River and figured that hidden sources including industrial effluents, municipal wastes, and agricultural activities are responsible for degradation of water quality of the Old Brahmaputra River. The Gorganrood River water quality was also evaluated with new WQI, and it was found that the parameters that were directly affected due to anthropogenic impact were given more weights to assess the quality of water [17]. Sediment and suspended particulate matters of the Tadjan River (southern part of the Caspian Sea) were analyzed, and the major sources of pollution were pulp and paper mill, dairy factory, and municipal sewage [18].

A few notable studies are also present in the literature about the water quality and trace elements present in the Indus River. These studies disclosed the fact that the concentrations of toxic heavy metals are higher in the Indus River [19–21]. The Pakistani environment is badly affected by geochemical pollution through natural processes, i.e., deposition of alluvial material through floods in the Indus Delta, volcanic eruptions, as well as anthropogenic activities such as sewage irrigation, animal manure, along with

fertilizers and pesticides. This contaminated soil can lead to high levels of various trace elements, including As in the Indus River [22]. Although aforementioned studies pointed about the higher concentration of As and other toxic heavy metals in Indus River water, unfortunately no real effort has been made for assessing the actual hidden sources of pollution. It is therefore, during this multivariate statistical technique, utilized to divulge the actual sources of pollution of the Indus River.

*1.1. Background Literature.* Multiple multivariate statistical techniques are used to estimate the different parameters of surface water quality index with morphological status of land. The techniques used are the hierarchy cluster analysis, principal component analysis, and discriminate analysis, respectively. These statistical techniques allow estimating multifaceted data matrices for understanding the quality of surface water quality in a better way and also permit for distinguishing the possible factors having significant influence on the water system. Statistical technique also provides a tool for water flow management system along with the solution of pollution issues. It has also been used for characterizing and evaluating the parameters between surface and the fresh quality of water and helpful for exploring the effect of anthropogenic sources on the spatial and temporal variations river quality [23–25].

Pejman et al. [26] have used cluster analysis (CA), principal component analysis (PCA), and factor analysis (FA) for calculating the seasonal and spatial variations effects on the water quality of Haraz Basin. During the four seasons, i.e., summer and autumn of 2007 and winter and spring of 2008, sets of eight water quality parameters were collected and analyzed by CA and PCA. According to results, the parameters, which were significant on one season, may not be significant on the other season. Mustapha and Abdu [27] investigated the water quality parameters for the Jakara River. They have used multiple linear regression and principal component analysis (PCA) to identify the most influential parameter responsible for water pollution. Furthermore, they used PCA to figure out the origin of quality parameters which serve in identifying the hidden pollution sources. Bhattacharyya et al. [28] have used principal component analysis (PCA)/factor analysis (FA) and cluster analysis (CA) along with correlation analysis multivariate statistical technique for calculating the water quality index of the Damodar River in India. The study revealed the existence of some harmful chemicals within the river. PCA/FA interpreted the major responsible factors for affecting the water quality were geogenic and anthropogenic factors. Shrestha and Kazama [29] used cluster analysis (CA), principal component analysis (PCA), factor analysis (FA), and discriminant analysis (DA) for the evaluation of spatial/temporal variations and huge complex water quality index for the Fuji River Basin. Around twelve parameters were evaluated at thirteen different sites. FA analysis indicated that the major responsible parameters for degrading the quality of the water system were discharge and organic pollutions. More recently, Khan et al. and Zafar et al. [30, 31]

TABLE 1: Irrigation water quality framework guidelines along with WHO recommended concentration of As in  $\mu\text{g/l}$ .

Potential water quality problem	Units	None	Degree of restriction on using slight to moderate	Severe
Salinity (affects $\alpha$ on water availability)				
$E_{c_w}$	dS/m	<0.7	0.7–3.0	>3.0
Or				
TDS	mg/l	<450	450–2000	>2000
Infiltration (affects infiltration rate of water into the soil and evaluates using $E_{c_w}$ and SAR together)				
SAR				
= 0–3 and $E_{c_w}$ =		>0.7	0.7–0.2	<0.2
= 3–6		>1.2	1.2–0.3	<0.3
= 6–12		>1.9	1.9–0.5	<0.5
= 12–20		>2.9	2.9–1.3	<1.3
= 20–40		>5.0	5.0–2.9	<2.9
Specific ion toxicity (affects sensitive crops)				
Sodium ( $\text{Na}^+$ )				
Surface irrigation	SA	<3	3–9	>9
Sprinkler irrigation	me/l	<3	>3	
Chloride ( $\text{alpha}^-$ )				
Surface irrigation	Me/l	<4	4–10	>10
Sprinkler irrigation	Me/l	<3	>3	
Boron (B)	Mg/l	<0.7	0.7–3.0	>3.0
Trace elements				
Miscellaneous effects (affects susceptible crops)				
Nitrogen ( $\text{NO}_3^{2-N}$ )	mg/l	<5	5.0–30	>30
(Overhead sprinkling only) pH	me/l	<1.5	1.5–8.5	>8.5
		Normal range		
		6.5–8.4		
As concentration	$\mu\text{g/l}$	<10	>50	

have used the multivariate statistics for assessing the effluent discharge on the irrigation water and Indus River water quality in KPK, Pakistan, and found that effluent discharge results in the sever degradation of water quality for irrigation purpose. Moreover, using PCA and CA, they conclude that the Indus River is found strongly affecting city effluent in KPK Province of Pakistan. In short, multivariate statistics has been used by number of researchers for analysis of water quality over the years in order to obtain useful information about required number of monitoring stations, pollution sources, and spatial and temporal variations assessment [32–34].

## 2. Materials and Methods

**2.1. Information about Monitoring Parameters and Stations.** Researchers [35] have discussed the frequently used parameters for characterizing the quality of water which includes the important physical and chemical parameters. Physical parameters are temperature, pH, electrical conductivity (EC), total dissolved solvents (TDSs), and salinity, while the chemical parameters are dissolved oxygen (DO), oxygen reducing potential (ORP), COD, BOD, anions ( $\text{Cl}^-$ ,  $\text{Cl}^-$ ,  $\text{NO}_3^-$ ,  $\text{SO}_4^{2-}$ , and  $\text{PO}_4^{3-}$ ), and toxic heavy metals (Cd, Pb, and As).

The 8 physicochemical parameters which include pH, TDS, EC, DO, ORP, salinity, and concentration of As and Pb were chosen based on the available literature and data [36] for applying multivariate statistics. The data of 64 samples

collected from 6 regions of upper Indus Basin and main Indus River were originally taken by Ahmad et al. The data [36] were utilized for the analysis of water quality over the region and obtaining useful information about the required number of monitoring stations, pollution sources, and spatial variations assessment. The six regions were labeled as  $M_1$  to  $M_6$  monitoring stations with the following details:

- (i)  $M_1$ : Gilgit to Khunjerab pass
- (ii)  $M_2$ : Gilgit to Chitral
- (iii)  $M_3$ : Gilgit to Nalter
- (iv)  $M_4$ : Gilgit to Deosai plain
- (v)  $M_5$ : Skardu to Gilgit
- (vi)  $M_6$ : main Indus River

All regions with marked sample collection points are shown in Figure 1.

## 3. Statistical Analysis

All the analyses are made while using XSTAT software (trial version) and Microsoft Excel version 2007.

**3.1. Hierarchy Cluster Analysis.** Cluster analysis is basically a cluster of multivariate database whose basic operation is to provide tools for assembling objects based on individual parametric characteristics. The primary method for classifying objects in cluster analysis is that every object is similar

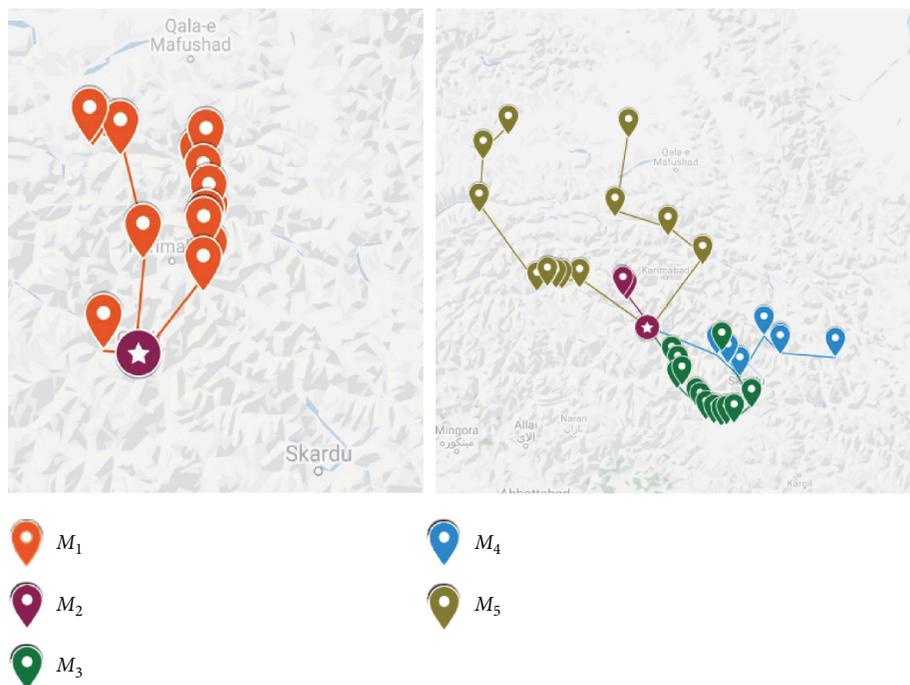


FIGURE 1: Map showing the six regions of the Indus River with marked samples collection point.

comparatively to the other object with respect to predefined descriptive criteria. The resulting outcomes of cluster object correspond to high-internal homogeneity and also very high-external heterogeneity in order to differentiate between relevant and irrelevant variables. Therefore, variables in a cluster analysis must be supported through conceptual considerations. The common approach for analyzing similarities relationship between the samples through whole data set is hierarchical agglomerative clustering and is normally illustrated by a dendrogram. The main aim of this study is to calculate a relationship between individual cluster to identify the similarity and dissimilarity between each samples of monitoring sites. Dendrogram tells the parametric mean value for every sampling site. There are three statistical groups comprising all the calculating values:

- (1) Group A: this cluster could be regarded as generally less polluted areas. There should be low or moderate industrialization and urbanization. Consequently, the human effect also should be very low.
- (2) Group B: generally, this type of cluster could be regarded on highly polluted areas such as industrial sites. Pollutants in the sample from industrial sites generally are involved from industrial wastewater treatment plants along with sewage wastewater from agriculture activities.
- (3) Group C: this cluster could be viewed as generally moderate-contaminated (MC) sites. These locales may get contamination from nonpoint sources

**3.2. Principal Component Analysis (PCA).** PCA is a linear combination of factors, which intended to change the first factors into new, uncorrelated factors (axes), called as

principal components. The new axis lies along the direction of the maximum variance. PCA is not only used to normalize factors for the sack of comparison between the samples but also to find the factors of pollutants that influence each sample. In PCA, there is a new cluster of factors called a rotation of axis involved that is used to divide the original variables into groups. It is applied to find a compositional influential factor involved in each sample.

PCA provides certain explanation for the most valuable factor which depicts the full information set interpretation. PCA summaries the statistical correlation between the compositions among the water samples with least reduction of information. The following equation is used to express the principal component analysis:

$$Z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \dots + a_{im}x_{mj}, \quad (1)$$

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

## 4. Results and Discussion

**4.1. Descriptive Measures of River Water Quality Data.** The XSTAT software (trial version) was employed in this study for PCA and CA of water quality data after normalizing the data to unit variance and zero mean following the approach reported in [37]. Due to high-dispersion-measured quality parameter's data for all regions (i.e.,  $M_1-M_6$ ), as shown in Figure 2, it is extremely difficult to rely on either mean value or variance for assessing the water quality. Therefore, coefficient of variation (CV) of quality parameters focused in this study is calculated from the ratio

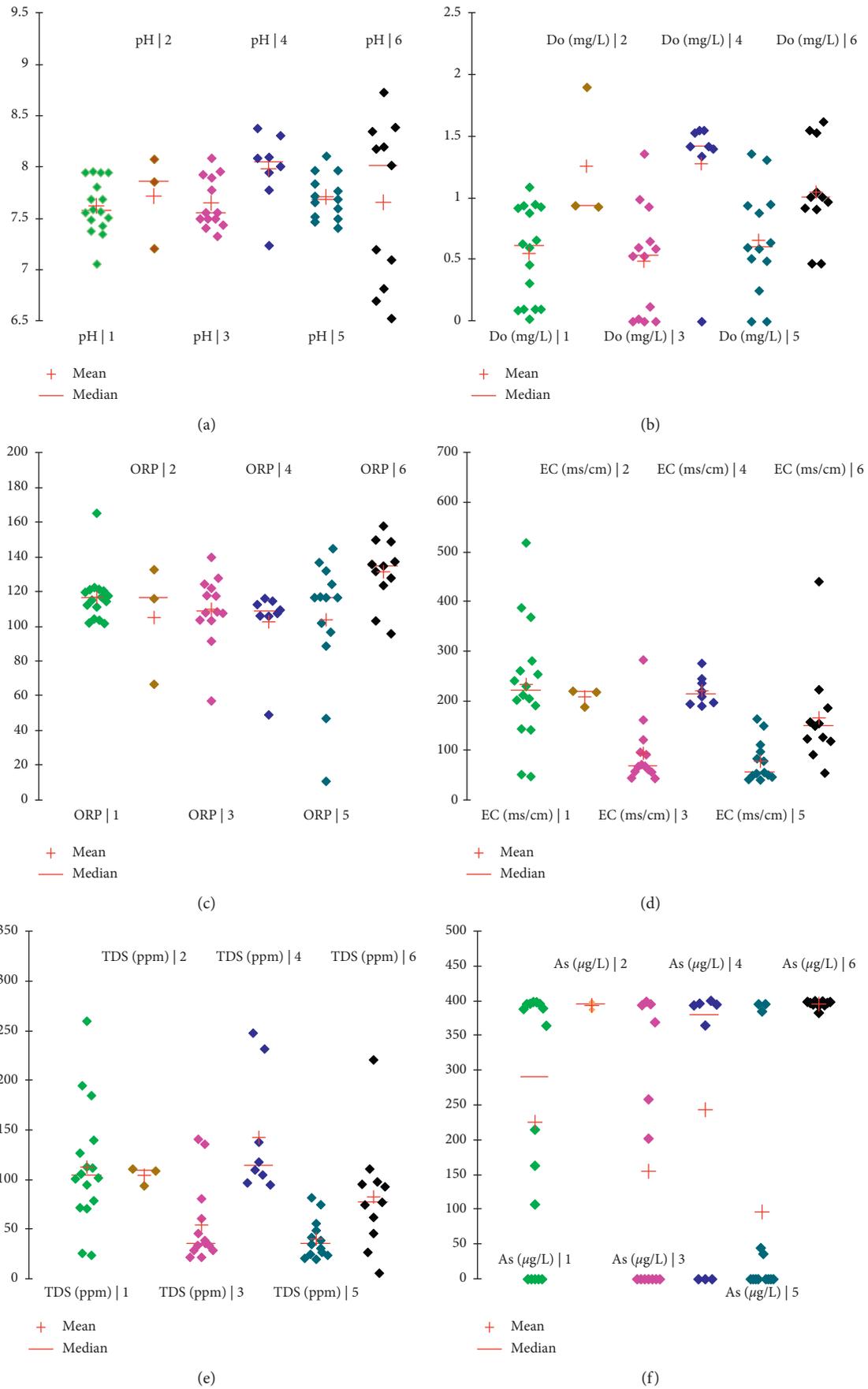


FIGURE 2: Descriptive analysis of water quality data.

of standard deviation to mean calculated. The range, mean, median, dispersion, and standard deviation of each quality parameter is shown in Figure 2 in the form of descriptive data representation, while CV is given in Table 2.

The mean and median values of pH at observation points ranging from  $M_1$  to  $M_5$  overlap each other, while there is a slight difference with median having a higher value than for the main Indus River (i.e.,  $M_6$  observation station). This suggests that pH do not vary much, and smaller CV values point to fact that spatial effects on pH are minor on relative scale. These findings are consistent with the findings reported in literature [33, 38]. Heavy rainfall might be responsible for this consistency because it is the “only” common factor in abovementioned studies. DO, ORP, EC, TDS, % salinity, and As and Pb concentration have higher dispersion of data at all active observations sites/stations/point as evident from Figure 2(a)–2(f) thus giving high values of CVs (see Table 2). The highest CVs of almost all the aforementioned physicochemical parameters are observed at  $M_3$ , while the highest mean and median values of DO and TDS are observed at  $M_4$ , and ORP, As, and Pb are observed at  $M_6$ . The mean and median values of As and Pb are also significant higher at all stations along with the higher values of CVs. Largest variations in the mean values of “DO” and “ORP” point to the high level of organic pollutants from various sources and decomposition [33, 37, 39]. Highest mean and median values of As from the start of the Indus River suggest that it is polluted with the traces of As from the start and increases with flow when it remains in the upper Indus Basin. However, at the main Indus River, its concentration is maximum, but the smallest value of CV is suggesting that As is coming in the Indus River from effluents of the upper basin. The TDS mean and median values at  $M_3$  is highest with larger CVs, as evident from Figure 2 and Table 2, which is a confirmation of high inorganic loading at this station because of soil erosions and agriculture activities [39, 40].

**4.2. Correlation Matrix.** The results of correlation analysis are given in Table 3 as a correlation matrix, which is calculated by combining the mean values of quality parameters of six regions belonging to the upper Indus Basin, i.e., from  $M_1$ – $M_6$ . It can be seen from the data that most of measured parameters in the upper Indus region have a positive correlation. The strong positive correlation of pH, EC, TDS, and As and Pb concentration with DO is quite evident from the correlation matrix, and the reason for this correlation is explained in the section below where PCA of current data set is performed. In addition, EC is strongly correlated with TDS and the concentration of Pb. TDS and other inorganic matters/metals like Pb and As correlates well with EC, as reported by [35]. It is worth mentioning here that the concentration of As is positively correlated with all measured parameters with relatively strong dependence on ORP, EC, TDS, and salinity.

**4.3. Cluster Analysis (CA).** On the basis of similarities in water quality features, various sampling stations of UIB are

grouped into clusters. However, employing this analysis technique planning for future events with an optimum number of sampling stations is possible. In addition, this strategy is also helpful in reducing the monitoring cost with better understanding of prevailing factors involved in a system under investigation. Figure 3 shows the dendrogram obtained from hierarchy cluster analysis which generated three groups on the basis of similar characteristic features. Groups 1, 2, and 3 correspond to relatively bad, fair, and good quality regions of UIB, respectively. In aforementioned dendrogram shown in Figure 3,  $M_3$  and  $M_5$  formed one group, i.e., group 3 consists of comparatively less polluted site, which is attributed to the fact of almost negligible human activities at  $M_3$  and  $M_5$ . The water at  $M_3$  and  $M_5$  comes from the snow-covered peaks and glaciers, which falls into main subriver through side nallas, and most of side nallas are clear and transparent with a little or no domestic effluent. Group 2 consists of sample points belonging to  $M_2$  and  $M_4$ , and analysis shows that this group has slightly less quality water as compared to group 3, i.e., it is a moderately polluted group. The human excreta effluent discharge points along with dumping of domestic and commercial waste, which belongs to the population in these regions, are mainly responsible for degradation of water quality.

Group 1 which consists of sampling points belonging to  $M_1$ , i.e., from Gilgit to Khunjerab pass, and  $M_6$  (main Indus River) outlines the most polluted group of the upper Indus Basin. There are the following reasons for these sampling sites to be at highest polluted side:

- (i) Higher concentration of arsenic and lead
- (ii) Higher water turbidity due inclusion of silts and rocks
- (iii) Picking of herbs and shrubs for medical use and illegal cutting of forest and trees of environmental importance in an uncontrolled manner
- (iv) Higher amount of dissolved oxygen (DO)
- (v) Higher amount of oxidation reducing potential (ORP)
- (vi) Higher percentage of salinity
- (vii) Human excreta effluent discharge
- (viii) Industrial waster along the main Indus River

It is observed from analysis, however, that sampling can be done for the points/stations with similar quality parameters and human activities. Also, on the basis of cluster analysis, the upper Indus Basin can be classified as either less, moderately, or highly polluted regions. The regions  $M_1$  (from Gilgit to Khunjerab pass which also include the sampling points of Attabad lake) and  $M_6$  (main Indus River) are the most polluted area because of higher concentration of Arsenic which is 40 times higher than the WHO recommended limits, i.e.,  $10 \mu\text{g/l}$ . Furthermore, a rapid and more efficient assessment of upper Indus Basin water quality is possible while using the spatial cluster analysis of various regions of the upper Indus Basin. Moreover, this analysis is also helpful for designing an optimal, more feasible, and cost-effective monitoring strategy. The sample collections

TABLE 2: Coefficient of variation (CV) in water quality parameters of the upper Indus Basin.

Quality parameters	$M_1$	$M_2$	$M_3$	$M_4$	$M_5$	$M_6$
pH	0.05862	0.03395	0.03235	0.04454	0.02798	0.10346
Do (mg/L)	0.44337	0.69481	0.90625	0.40862	0.65668	0.37309
Do (%)	0.48706	0.93223	0.93134	0.45269	0.63853	1.53101
ORP	0.32578	0.12623	0.18369	0.2137	0.35976	0.14286
Cond (ms/cm)	0.08591	0.51238	0.6919	0.13514	0.52085	0.6101
TDS (ppm)	0.08877	0.5367	0.74332	0.43104	0.49662	0.67449
Salinity (%)	0.05973	0.6935	0.66504	0.41818	0.61572	1.58771
As ( $\mu\text{g/l}$ )	0.01382	0.80201	1.17808	0.82926	1.74784	0.01202
Pb ( $\mu\text{g/l}$ )	0.08533	0.95421	0.52915	1.92162	0.64507	1.29853

TABLE 3: Correlation matrix of water quality parameters.

	pH	Do	ORP	EC	TDS	Salinity	As	Pb
pH	1							
Do	<b>0.6642</b>	1						
ORP	-0.5311	-0.1018	1					
EC	0.3056	<b>0.5279</b>	0.1066	1				
TDS	<b>0.5924</b>	<b>0.6226</b>	-0.0676	<b>0.9394</b>	1			
Salinity	-0.2175	0.2530	<b>0.8862</b>	0.04051	-0.0411	1		
As	<b>0.0023</b>	<b>0.7085</b>	<b>0.4812</b>	<b>0.5967</b>	<b>0.4762</b>	<b>0.5991</b>	1	
Pb	-0.5468	-0.9222	0.08630	-0.7916	-0.8084	-0.1233	-0.7643	1

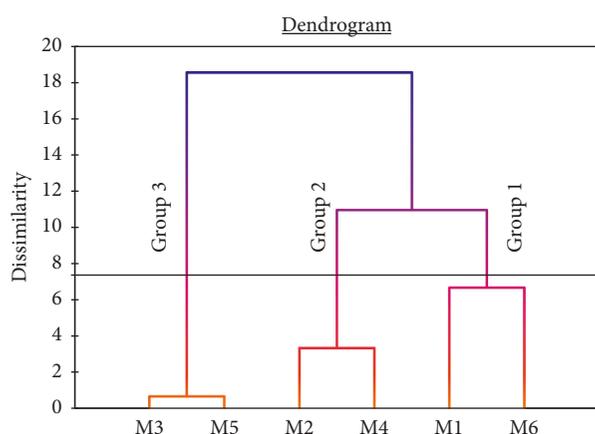


FIGURE 3: Dendrogram representing the cluster of monitoring stations.

from the upper Indus Basin is extremely difficult due to geography of this region [36, 38, 41], and this study suggests that number of monitoring stations can be reduced to only three stations instead of collecting the samples from whole upper Indus Basin.

**4.4. Box Plots of Water Quality Parameters.** In order to have in-depth examination of cluster analysis shown in Figure 3, a set of box plots of eight physicochemical quality parameters are shown in Figure 4. The spread of boxes for all parameters belonging to group 3 is smaller. This indicates the fact that quality parameters are less affected by environment and human activities, and water quality in this group is good which is also evident from cluster analysis shown above. The

concentration of As in samples belonging to this group is smaller and less dependent on temperature/environmental/seasonal alterations. Now coming to group 2, pH, TDS, and As concentrations have box plots with a higher spread (see Figures 4(a)–4(h)) which suggests the significant influence of environmental changes in this region. Among the abovementioned quality parameters, box plots show that pH of water in this region is more sensitive to seasonal/environmental alterations and thus results in degradation of water quality from group 2.

Group 1 box plots for EC, DO, ORP, salinity, and As concentration show largest spread of data along with bigger top and bottom whiskers for abovementioned quality parameters which is another confirmation of the fact that water quality parameters for sampling stations belonging to group 1 are more affected due to seasonal/environmental alterations. Furthermore, larger spread of data as evident from larger top and bottom whiskers also suggests that concentrations for aforementioned parameters are highest at group 1, thus confirming the role of soil erosion in degrading the water quality. The dissolved oxygen (DO) suffer more on moving from group 3 to group 1 (see Figure 4(b)) which is the confirmation of the fact that, among all three groups, group 1 is the most polluted as far as organic contents are concerned. This finding is also consistent with the box spread for salinity of samples belonging to group 1 (see Figure 4(f)) because DO and salinity are negatively correlated [38]. The soil erosion due to illegal uprooting of herbs and shrubs of medical importance, land sliding, and domestic/industrial waste are the main reasons for water pollution (as evident from cluster analysis) in this group. The results shown in Figures 3 and 4 are quite similar with the river statistical characteristics reported by number of researchers [32, 33, 38].

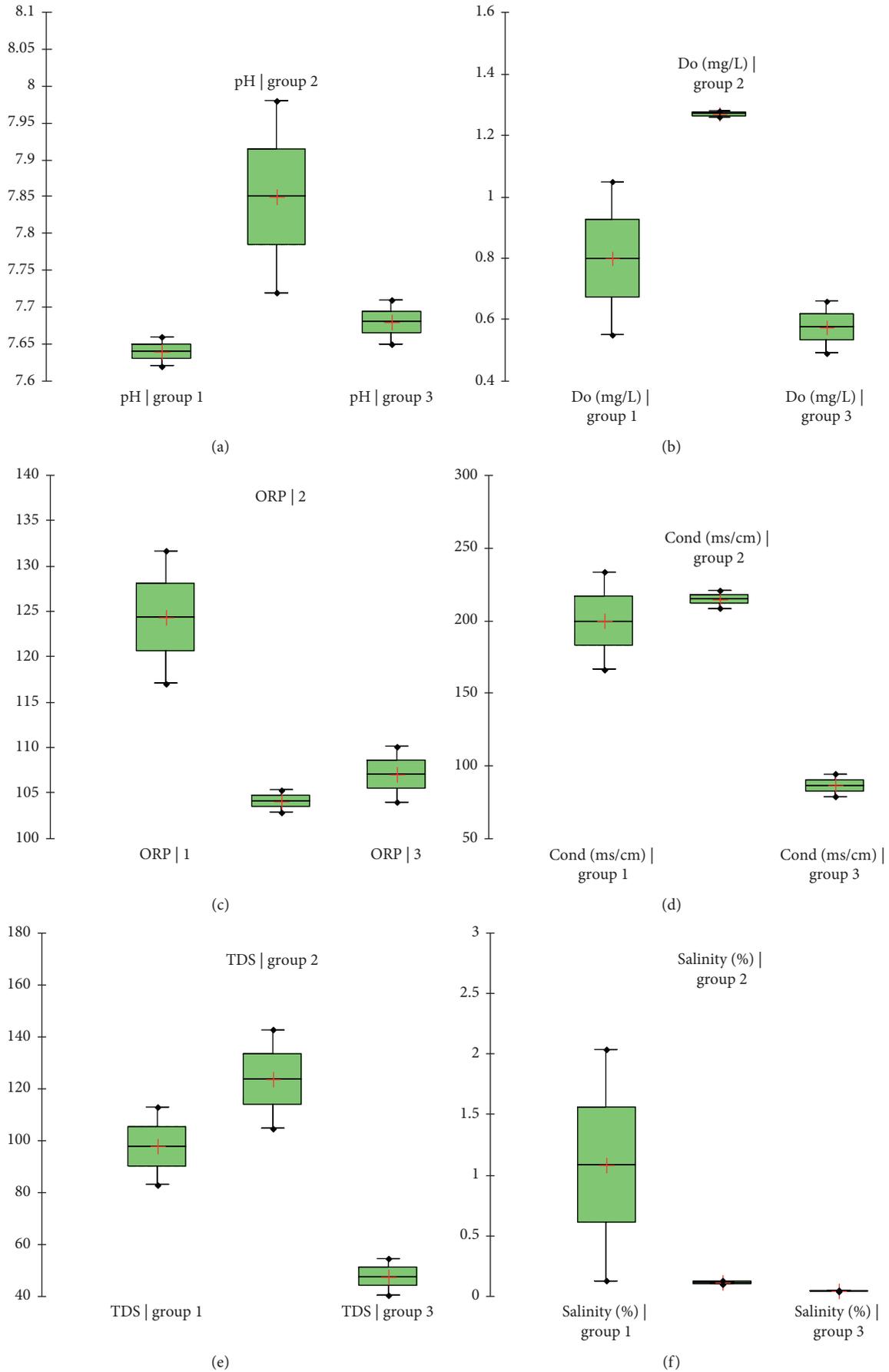


FIGURE 4: Continued.

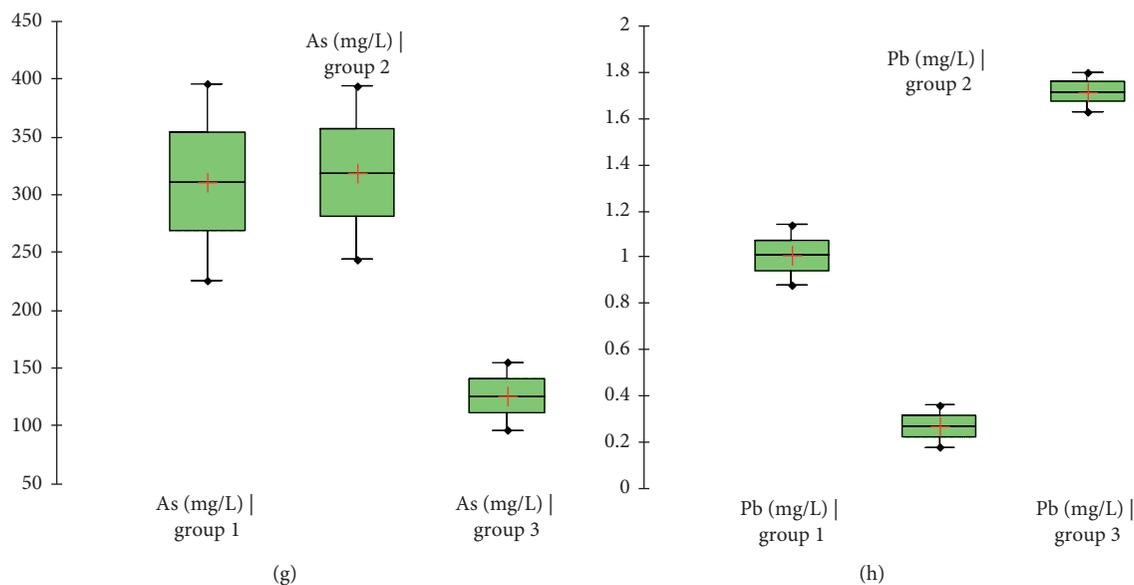


FIGURE 4: Box plots of (a) pH, (b) DO, (c) ORP, (d) EC, (e) TDS, (f) salinity, (g) As, (h) and Pb for different spatial clusters.

**4.5. Principal Component Analysis (PCA) of Upper Indus River Data.** The eigenvalues obtained after performing the PCA of the quality parameters belonging to six regions, i.e.,  $M_1$  to  $M_6$  of the upper Indus Basin, are shown in Table 4. By retaining the eigenvalues  $> 1$ , it is evident from the results of PCA of current data that over 83% of the information of original data is explained by first two sets of eigenvalues. The variance percentage for first two eigenvalues is 83% and for the first three Eigen values is 94%. These percentages are far better than previously reported variance percentages of the first two and three eigenvalues with a focus on river quality data [32, 33, 38], thus confirming the application of PCA to the current set of data with a reasonable confidence. The eigenvalues for six regions, i.e.,  $M_1$ – $M_6$ , are 4.210, 2.462, 0.839, 0.464, 0.024, and  $< 0.024$ , respectively. The higher eigenvalue for  $M_1$  suggests the large dispersion of data for these regions. The higher concentration of As for all sampling points and significant higher dispersion in DO, ORP, salinity, and the concentration of As (as evident from the descriptive analysis and box plots) are reasons for obtaining higher eigenvalue during the PCA for this particular region. For  $M_2$ , large dispersion of pH, TDS, and As concentration data of samples belonging to this region is responsible for obtaining the eigenvalue 2.462 (see Table 4). For region  $M_6$ , smaller dispersion of data for all quality parameters give the lowest eigenvalue, i.e.,  $< 0.024$ .

According to PCA theory as explained by [37], parameter loadings (which are the projections of water quality parameters on PCs axes) are the correlation coefficients among variables and coefficients. The factor loading retained PCs, as shown in Table 4, and are classified as follows [38]:

- (i) Factor loading  $> 0.75$  is classified as “strong”
- (ii) Factor loading value between 0.75 and 0.50 is classified as “moderate”

TABLE 4: Factor loadings of quality parameters.

	PCA-1	PCA-2	PCA-3
pH	-0.574	<b>-0.617</b>	-0.380
Do	<b>-0.894</b>	-0.057	-0.426
ORP	-0.036	<b>0.965</b>	0.118
Cond	<b>-0.846</b>	0.001	<b>0.526</b>
TDS	<b>-0.882</b>	-0.203	<b>0.347</b>
Salinity	-0.221	<b>0.893</b>	-0.312
As	<b>-0.765</b>	<b>0.552</b>	-0.051
Pb	<b>0.975</b>	0.072	0.049
Eigenvalue	4.210	2.462	0.839
Variability	52.630	30.772	10.490
Cumulative	52.630	83.402	93.892

- (iii) Factor loading value between 0.50 and 0.30 is classified as “weak”

The variance corresponding to PC-1 is 52.63% with a strong negative factor loading ( $> 0.7$ ) by DO, EC, TDS, As, and Pb (positive). Parameters indicating the organic pollution are related with anthropogenic pollution sources are represented by this PC, as reported in the literature [38]. The consumption of oxygen at a larger extent during the fermentation (anaerobic) process by a higher concentration of dissolved organic matters results in formation of ammonia and organic acid which then causes a decrease of water pH because of hydrolysis. The negative loading of pH and strong positive loading of Pb shown in Table 4 are also in favor of our argument. In addition to this DO and TDS, As with strong negative loading factors confirms their negative correlation with anthropogenic pollution sources and agrees with already reported studies [24, 38]. The quality parameters strongly contributing are % salinity and ORP, while pH and the concentration of As are the moderate contributor to PC-2. This reveals the fact that this PC is the representative of seasonal variations (flow of solids from elevated

mountains and other sources during rain) and soil erosions because of illegal cutting of forests and uprooting the herbs of medical importance [36] at massive scale and agriculture activities. The negative loading of pH and strong positive loading of salinity (see Table 4) are also in favor of our argument here.

The plot of quality parameters in the first two PCs space is shown in Figure 5. From the figure, it is evident that PC-1 has strong negative loading on pH, DO, % salinity, Pb concentration, strong positive loading on the concentration of As, and moderate loading on EC and ORP. This shows that PC-1 is affected by organic and inorganic pollution due to soil erosion, interaction of Indus River water with arsenic rich bedrocks, and flowing of domestic wastewater into the Indus River through a number of creeks. The strong negative loading of PC-1 on DO and pH is quite similar to the observation reported in the literature. Zafar et al. [31, 38] confirms the presence of organic acids and DO negative correlation with organic pollutants [38]. PC-2 which is only 4.39% spread (as compared PC-1 of quality parameters has strong positive loading) on ORP, strong negative loading on As concentration, and weak to moderate negative loadings on TDS, pH, DO, and % salinity, respectively. This is the further confirmation of the fact that major source of pollution in the upper Indus region is either because of toxic heavy metal like As and Pb or anthropogenic pollution sources. In addition to these, there is some contribution of agriculture activities which have positive correlation with rain and flooding.

The negative correlations of aforementioned parameters (i.e., pH, DO, Pb, and % salinity) with inorganic/organic pollution, as shown in Figure 5, is also evident from strong negative PC-1 loading (see Figure 5) of these parameters. Both inorganic and organic pollutants mainly affect the TDS, whereas inorganic pollutants affect EC only. The moderate positive PC-1 loading of EC points to the fact that EC is mainly dependent on the concentration of As and Pb. The concentrations of As and Pb have strong positive and negative loading on PC-1, respectively (see Figure 5), thus making the loading of EC moderate. TDS has weak positive loading on PC-1 and weak positive loading on PC-2, suggesting the influence of other parameter loadings with more dominant PC-1 contributions. Figure 6 represents the projection of quality parameter on PC space with their confidence intervals and is the conclusive statement of PCA of quality parameters because it shows the true pictures about monitoring the exact number of quality parameters for water quality assessment in the upper Basin of Indus River or upper Indus Basin. According to the PCA presented in Figure 6, four parameters ORP, DO, and the concentrations of As and Pb are sufficient to have a reasonable confidence on the quality of water in the upper Indus Basin.

Figure 7 shows a bipolar plot of the quality parameters at different stations in the first two PCs space. It is evident from figure that the six regions of the upper Indus Basin, which are under investigation, form three groups: (1)  $M_1$  and  $M_6$ , (2)  $M_1$  and  $M_6$ , and (3)  $M_3$  and  $M_5$ . The group 3 here has strong positive loadings on PC-1, which means that 52.63% of total variance belongs to group 3. With respect to active

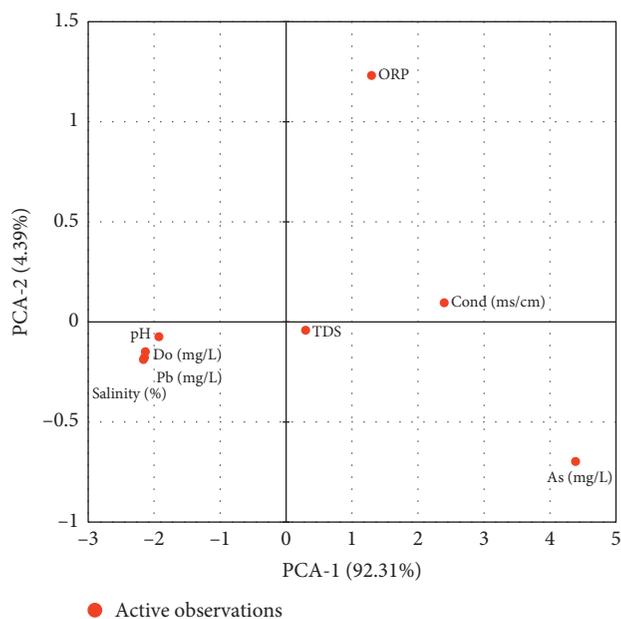


FIGURE 5: PCA-1 and PCA-2 loadings of water quality parameters.

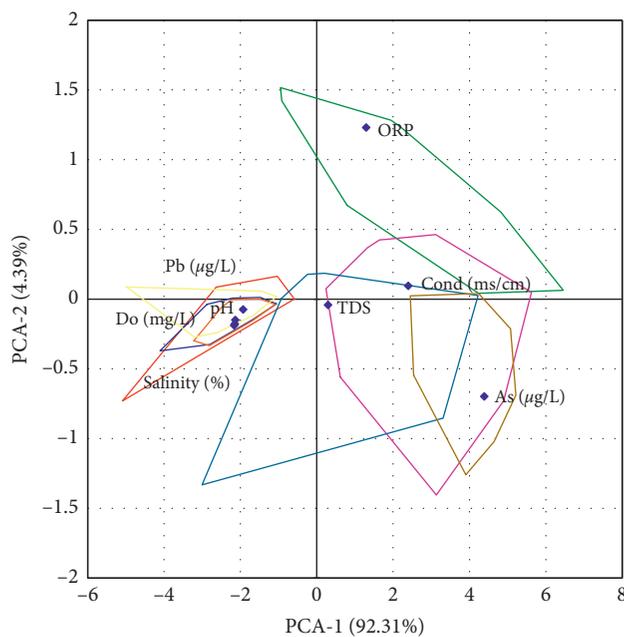


FIGURE 6: PCA-1 and PCA-2 loadings of water quality parameters with confidence intervals.

parameters,  $M_3$  and  $M_5$  form a least polluted group with only Pb as the notable pollutant (see Figure 7). On the contrary, group 2, which consists of  $M_2$  and  $M_4$ , has strong negative loadings of active variables on PC-1 and PC-2. However, out of eight physicochemical parameters EC, DO, TDS, and pH are the most prominent active variables (see Figure 7) for this group which make group 2 a moderately polluted group.  $M_1$  active observation site/station (as shown in Figure 6) has weak positive and negative loadings on PC-1 and PC-2, but almost all the active parameters are contributing to the net

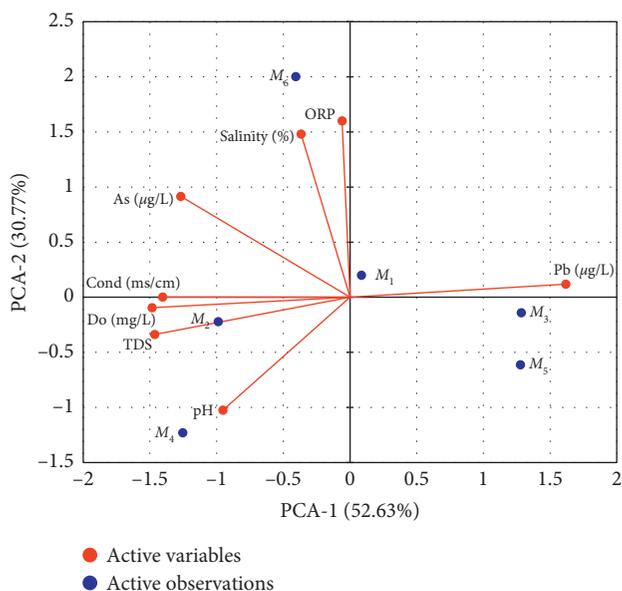


FIGURE 7: Bipolar PCA plot.

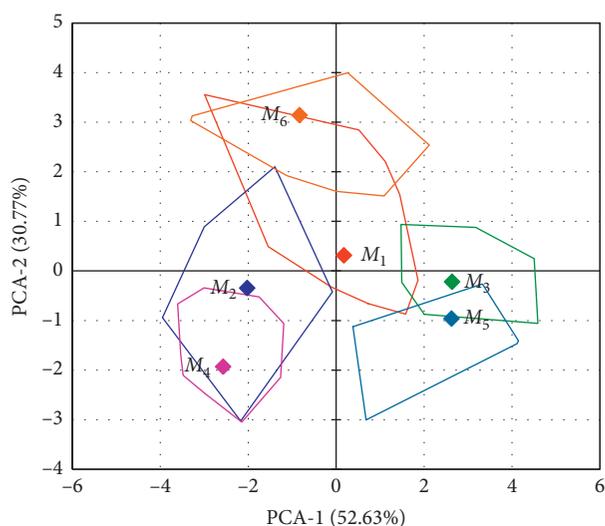


FIGURE 8: Representation of spatial behavior of the sampling network in the PC space.

pollution of  $M_1$ , as evident from Figure 7, thus making  $M_1$  as a member of the worst polluted group, i.e., group 1. The concentration of As and Pb, % salinity, and ORP are major contributors at  $M_6$  (see Figure 7), thus making the  $M_6$  active observation/site/station a very strong negative loading on 30.77% of total variance, i.e., PC-2. The higher concentration of As (i.e.,  $399.32 \mu\text{g/l}$ , which is eight times higher than national limit and 40 times higher than the World Health Organization (WHO) permissible limits), significantly increases in % salinity due to agriculture activities along the passage of the main Indus River, and inclusion of industrial/domestic wastewater in the main Indus River is responsible for making  $M_6$  as a member of the worst polluted group, i.e., group 1. The results of PCA shown in the bipolar plot of Figure 7 are in good agreements with the results of CA (see Figure 3) and box plots (see Figures 4(a)–4(h)).

In addition to this, Figure 8 which shows the projection of active observation sites on PC space with their confidence intervals suggests that, instead of sampling whole upper Indus Basin and Indus River, only three monitoring stations are sufficient for stating a conclusive statement about the quality of water of the upper Indus Basin and main Indus River.

## 5. Conclusion

CA reveals that upper Indus Basin monitoring can be done while dividing the region into three clusters, representing good quality water ( $M_3$  and  $M_5$ ), moderate quality water ( $M_2$  and  $M_4$ ), and worst quality water ( $M_1$  and  $M_6$ ) cluster, respectively. Three latent factors are responsible for explaining the 93.87% of total variance as confirmed by PCA. The parameters, which are significantly influenced by anthropogenic impact, seasonal variations, and soil erosion, are DO, EC, TDS, and concentration of As ( $\mu\text{g/l}$ ). The concentration of Pb ( $\mu\text{g/l}$ ) is influenced by anthropogenic sources only, while the (%) salinity and ORP are influenced due to agriculture activities and runoff, as confirmed by PCA. The box plot results are also supporting the PCA outcomes. The highest pollution is observed at  $M_1$  and  $M_6$  with the concentration of As being the dominant factor in making this group as the worst polluted group. The results of current study could be utilized for designing a comprehensive monitoring protocol for monitoring the Indus River, thus saving this important resource of Pakistan.

## Data Availability

Data will be made available when required.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

Engr. Mansoor Ahmad Baluch acknowledges the technical and moral support provided by the University of Engineering and Technology Taxila, 47050, Pakistan.

## References

- [1] R. Shabbir and S. S. Ahmad, "Use of geographic information system and water quality index to assess groundwater quality in rawalpindi and islamabad," *Arabian Journal for Science and Engineering*, vol. 40, no. 7, pp. 2033–2047, 2015.
- [2] M. Kavurmaci and A. K. Üstün, "Assessment of groundwater quality using DEA and AHP: a case study in the Serflikochisar region in," *Turkey Environmental Monitoring and Assessment*, vol. 188, no. 4, p. 258, 2016.
- [3] A. Alamgir, M. A. Khan, J. Schilling, S. S. Shaukat, and S. Shahab, "Assessment of groundwater quality in the coastal area of Sindh province," *Pakistan Environmental Monitoring and Assessment*, vol. 188, no. 2, p. 78, 2016.
- [4] M. Terrado, D. Barceló, R. Tauler, E. Borrell, S. D. Campos, and D. Barceló, "Surface-water-quality indices for the analysis

- of data generated by automated sampling networks,” *TrAC Trends in Analytical Chemistry*, vol. 29, no. 1, pp. 40–52, 2010.
- [5] S. Das, M. Majumder, D. Roy, and A. Mazumdar, “Determination of urbanization impact on rain water quality with the help of water quality index and urbanization index,” in *Impact of Climate Change on Natural Resource Management*, pp. 131–142, Springer, Berlin, Germany, 2010.
  - [6] E. Fathi, R. Zamani-Ahmadmohmoodi, and R. Zare-Bidaki, “Water quality evaluation using water quality index and multivariate methods, Beheshtabad River,” *Iran Applied Water Science*, vol. 8, no. 7, p. 210, 2018.
  - [7] I. Mladenović-Ranisavljević and S. Žerajić, “Comparison of different models of water quality index in the assessment of surface water quality,” *International Journal of Environmental Science and Technology*, vol. 15, no. 3, pp. 665–674, 2018.
  - [8] F. Misaghi, F. Delgosha, M. Razzaghmanesh, and B. Myers, “Introducing a water quality index for assessing water for irrigation purposes: a case study of the Ghezel Ozan River,” *Science of the Total Environment*, vol. 589, no. 4, pp. 107–116, 2017.
  - [9] S. H. Ewald, “Water quality evaluation of Al-Gharraf River by two water quality indices,” *Applied Water Science*, vol. 7, no. 7, pp. 3759–3765, 2017.
  - [10] M. T. Aalami, H. Abbasi, and V. Nourani, “Sustainable management of reservoir water quality and quantity through reservoir operational strategy and watershed control strategies,” *International Journal of Environmental Research*, vol. 12, no. 6, pp. 773–788, 2018.
  - [11] C. M. C. M. Couto, R. Ribeiro, A. R. Ribeiro et al., “Spatiotemporal distribution and sources of trace elements in Ave River (Portugal) lower basin: estuarine water, sediments and indigenous flora,” *International Journal of Environmental Research*, vol. 13, no. 2, pp. 303–318, 2019.
  - [12] M. A. Rahman, S. Kumar, A. A. Mohana, R. Islam, M. A. Hashem, and L. Chuanxiu, “Coliform bacteria and trace metals in drinking water, southwest Bangladesh: multivariate and human health risk assessment,” *International Journal of Environmental Research*, vol. 13, no. 2, pp. 395–408, 2019.
  - [13] M. Salam, M. Kabir, L. Yee, and M. Khan, “Water quality assessment of Perak River, Malaysia,” *Pollution*, vol. 5, no. 3, pp. 637–648, 2019.
  - [14] M. Barceló and E. Arslan Topal, “Monitoring of Zn and Cr in Downstream Water from Uzunçayır Dam in Turkey,” *Pollution*, vol. 5, pp. 649–655, 2019.
  - [15] S. Das Sharma, “Risk assessment and mitigation measures on the heavy metal polluted water and sediment of the Kolleru Lake in Andhra Pradesh,” *India Pollution*, vol. 5, pp. 161–178, 2019.
  - [16] M. Bhuyan, M. Bakar, A. S. M. Sharif, M. Hasan, and M. Islam, “Water quality assessment using water quality indicators and multivariate analyses of the old Brahmaputra River,” *Pollution*, vol. 4, no. 3, pp. 481–493, 2018.
  - [17] A. Karbassi, F. Mir Mohammad Hosseini, A. Baghvand, and M. Nazariha, “Development of water quality index (WQI) for Gorganrood River,” *International Journal of Environmental Research*, vol. 5, no. 4, pp. 1041–1046, 2011.
  - [18] M. Saeedi, S. Daneshvar, and A. R. Karbassi, “Role of riverine sediment and particulate matter in adsorption of heavy metals,” *International Journal of Environmental Science & Technology*, vol. 1, no. 2, pp. 135–140, 2004.
  - [19] T. Ahmad, M. A. Kahlown, A. Tahir, and H. Rashid, “Arsenic an emerging issue: experiences from Pakistan,” in *Proceedings of the 30th WEDC International Conference*, pp. 459–466, Vientiane, Laos, October 2004.
  - [20] M. M. Rahman, G. Owens, and R. Naidu, “Arsenic levels in rice grain and assessment of daily dietary intake of arsenic from rice in arsenic-contaminated regions of Bangladesh-implications to groundwater irrigation,” *Environmental Geochemistry and Health*, vol. 31, no. S1, pp. 179–187, 2009.
  - [21] T. Roychowdhury, T. Uchino, H. Tokunaga, and M. Ando, “Survey of arsenic in food composites from an arsenic-affected area of West Bengal, India,” *Food and Chemical Toxicology*, vol. 40, no. 11, pp. 1611–1621, 2002.
  - [22] A. Alamdar, S. Ali Musstjab Akber Shah Eqani, S. Waqar Ali et al., “Human Arsenic exposure via dust across the different ecological zones of Pakistan,” *Ecotoxicology and Environmental Safety*, vol. 126, pp. 219–227, 2016.
  - [23] M. H. Gholizadeh, A. M. Melesse, and L. Reddi, “Analysis of spatiotemporal trends of water quality parameters using cluster analysis in South Florida,” in *Proceedings of the World Environmental and Water Resources Congress 2016*, pp. 519–528, West Palm Beach, FL, USA, May 2016.
  - [24] X. Wang, Q. Cai, L. Ye, and X. Qu, “Evaluation of spatial and temporal variation in stream water quality by multivariate statistical techniques: a case study of the Xiangxi river basin, China,” *Quaternary International*, vol. 282, pp. 137–144, 2012.
  - [25] M. Xiao, F. Bao, S. Wang, and F. Cui, “Water quality assessment of the Huaihe River segment of Bengbu (China) using multivariate statistical techniques,” *Water Resources*, vol. 43, no. 1, pp. 166–176, 2016.
  - [26] A. H. Pejman, G. R. N. Bidhendi, A. R. Karbassi, N. Mehrdadi, and M. E. Bidhendi, “Evaluation of spatial and seasonal variations in surface water quality using multivariate statistical techniques,” *International Journal of Environmental Science & Technology*, vol. 6, no. 3, pp. 467–476, 2009.
  - [27] A. Mustapha and A. Abdu, “Application of principal component analysis & multiple regression models in surface water quality assessment,” *Journal of Environment and Earth Science*, vol. 2, no. 2, pp. 16–23, 2012.
  - [28] R. Bhattacharyya, K. Manoj, and P. Padhy, “Index analysis, graphical and multivariate statistical approaches for hydrochemical characterization of Damodar River and its canal system Durgapur, West Bengal, India,” *International Research Journal of Environment Sciences*, vol. 2, pp. 53–62, 2013.
  - [29] S. Shrestha and F. Kazama, “Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji River basin, Japan,” *Environmental Modelling & Software*, vol. 22, no. 4, pp. 464–475, 2007.
  - [30] I. Khan, A. Khan, M. S. Khan et al., “Impact of city effluents on water quality of Indus River: assessment of temporal and spatial variations in the southern region of Khyber Pakhtunkhwa,” *Pakistan Environmental Monitoring and Assessment*, vol. 190, no. 5, p. 267, 2018.
  - [31] S. Zafar, A. Khan, S. Ullah et al., “Assessing impact of effluent discharge on irrigation water quality in southern region of Khyber Pakhtunkhwa,” *Pakistan Environmental Monitoring and Assessment*, vol. 189, no. 4, p. 156, 2017.
  - [32] K. P. Kazama, A. Malik, D. Mohan, and S. Sinha, “Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)-a case study,” *Water Research*, vol. 38, no. 18, pp. 3980–3992, 2004.
  - [33] K. R. Singh, A. P. Goswami, A. S. Kalamdhad, and B. Kumar, “Assessment of surface water quality of Pagladia, Beki and Kolong rivers (Assam, India) using multivariate statistical techniques,” *International Journal of River Basin Management*, pp. 1–10, 2019.
  - [34] Y. Zhang, Y. Li, Y. Li, W. Da, M. Yu, and Q. Quan, “Application of multivariate statistical methods in the assessment

- of water quality in selected locations in Jialing river basin in Guangyuan, China,” *Water Science and Technology: Water Supply*, vol. 19, no. 1, pp. 147–155, 2019.
- [35] D. A. Cornwell and M. L. Davis, *Introduction to environmental engineering*, Mcgraw-hill Education, Europe, 2012.
- [36] M. Ahmed, K. Nazim, A. Khan, M. F. Siddiqui, and M. U. Khan, “Arsenic and lead level from upper to lower Indus basin of Pakistan,” *FUUAST Journal of Biology*, vol. 7, pp. 183–191, 2017.
- [37] P. J. Shaw, *Multivariate Statistics for the Environmental Sciences*, Wiley, Hoboken, NJ, USA, 2009.
- [38] V. Chounlamany, M. A. Tanchuling, and T. Inoue, “Spatial and temporal variation of water quality of a segment of Marikina river using multivariate statistical methods,” *Water Science and Technology*, vol. 76, no. 6, pp. 1510–1522, 2017.
- [39] N. Yusof, A. Haraguchi, M. A. Hassan, M. R. Othman, M. Wakisaka, and Y. Shirai, “Measuring organic carbon, nutrients and heavy metals in rivers receiving leachate from controlled and uncontrolled municipal solid waste (MSW) landfills,” *Waste Management*, vol. 29, no. 10, pp. 2666–2680, 2009.
- [40] P. Kjeldsen, M. A. Barlaz, A. P. Rooker, A. Baun, A. Ledin, and T. H. Christensen, “Present and long-term composition of MSW landfill leachate: a review,” *Critical Reviews in Environmental Science and technology*, vol. 32, no. 4, pp. 297–336, 2002.
- [41] S. Akhtar, “The south Asiatic monsoon and flood hazards in the Indus river basin, Pakistan,” *Journal of Basic & Applied Sciences*, vol. 7, 2011.



Hindawi

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

