Research Article

Investigation of the Soan River Water Quality Using Multivariate Statistical Approach

Zakaullah and Naeem Ejaz

Department of Civil Engineering, University of Engineering and Technology, Taxila, Pakistan

Correspondence should be addressed to Zakaullah; engrzaka@gmail.com

Received 7 November 2020; Revised 23 November 2020; Accepted 3 December 2020; Published 21 December 2020

Academic Editor: Hafiz Muhammad Ali

Copyright © 2020 Zakaullah and Naeem Ejaz. 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.

Evaluating the quality of river water is a critical process due to pollution and variations of natural or anthropogenic origin. For the Soan River (Pakistan), seven sampling sites were selected in the urban area of Rawalpindi/Islamabad, and 18 major chemical parameters were examined over two seasons, i.e., premonsoon and postmonsoon 2019. Multivariate statistical approaches such as the Spearman correlation coefficient, cluster analysis (CA), and principal component analysis (PCA) were used to evaluate the water quality of the Soan River based on temporal and spatial patterns. Analytical results obtained by PCA show that 92.46% of the total variation in the premonsoon season and 93.11% in the postmonsoon season were observed by only two loading factors in both seasons. The PCA and CA made it possible to extract and recognize the origins of the factors responsible for water quality variations during the year 2019. The sampling stations were grouped into specific clusters on the basis of the spatiotemporal pattern of water quality data. The parameters dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), turbidity, and total suspended solids (TSS) are among the prominent contributing variations in water quality, indicating that the water quality of the Soan River deteriorates gradually as it passes through the urban areas, receiving domestic and industrial wastewater from the outfalls. This study indicates that the adopted methodology can be utilized effectively for effective river water quality management.

1. Introduction

Rivers convey water and minerals to territories all around the earth. They have a significant influence on the hydrological cycle and serve as drainage outlets for runoff. Unsafe drinking water and inadequate sanitary conditions are associated with infectious diseases such as cholera, diarrhea, dysentery, and polio which significantly affect human health. According to the World Health Organization (WHO), at least two billion people worldwide use drinking water from a contaminated source containing human excreta [1]. In the last decades, research has revealed that developing countries have been able to control waterborne diseases to some extent through the development of sustainable natural resources, especially water resources, out of concern for the improvement of public health. However, challenges to achieving this workable goal remain, due to the increase in demand for water and the reduction of water availability

resulting from escalating population growth and increased financial stability [2].

Many developing and developed countries are facing degradation of freshwater resources over the last decades up to an alarming situation mainly due to rapid regional development especially in underdeveloped countries and agricultural runoff from fields containing fertilizers and pesticides [3, 4]. The use of agricultural chemical manures and municipal water influence the water quality significantly [5, 6].

As the quality of water is important for human health and ecological systems, its deteriorating quality in developing countries has caused detrimental health impacts such as diarrhea, poor oral hygiene, hepatitis, anemia, and dental caries [7]. Research has revealed that the quality of existing water resources has been deteriorating rapidly with ever increasing urbanization, massive industrialization, agricultural expansion, inadequate management of domestic, industrial and

urban wastes, eroded soil, and transportation infrastructures in recent decades [8–10].

Giri and Singh [11] reported that the escalation of urbanization and industrial development near water tributaries has led to river degradation, water effluence, and disintegration of natural ecological integrity. In most developing countries, the massive contamination of tributaries due to the point source pollution containing harmful chemicals is a major issue and has been widely observed [12, 13]. The utilization of water has been escalating at more than double the increase in population in the recent era. Globally, approximately six billion people are impounding 54% of all the open freshwater contained in streams, lakes, and natural springs [14].

Waterways play a significant role in incorporating and arranging the landscape and in influencing the environmental condition. They are the prime variables in controlling the global water cycle and behave as active elements which carry nutrients and contaminants. In Pakistan, the increasing pace of urbanization and industrialization has caused the severe decline of flora and fauna in the river basin due to the mismanagement of water reservoirs. In particular, untreated sewage, agricultural runoff, and industrial effluents have resulted in a considerable deterioration in water quality. The population residing near rivers has to use this water for many purposes including domestic uses and drinking. Similarly, the quality of surface water is declining due to human development, industrialization, agricultural inputs, transportation, urbanization, animal and human excretions, and domestic waste. Variation in the quality and amount of stream water as a result of characteristic and anthropogenic impacts is generally observed in different waterways throughout the world [15].

Water quality assessment is the most widely accepted way forward for identification of biological issues pertaining to river environment, and fluctuations in the environmental key indicators due to seasonal and time variation are quite helpful for analysis of the contamination level [16].

According to a report published by the Asia Foundation [17], almost 95% of the water in Pakistan is utilized for agro-nomic practices, which compensate 60% of its population engaged in agriculture and livestock, while 80% of the exports depend upon agribusiness. Despite the fact that Pakistan has the world's largest glaciers, it is nevertheless counted among the world's 36 most water-stressed countries. Due to an escalation in population growth as well as to strained relations with the neighboring country concerning transboundary issues of water management, the water situation in the country has become critical, both with respect to its impending waterfront and sustainability, as the demand for water is expected to exceed supply.

The release of sewage water into the water bodies is the primary source of water pollution whereas the secondary source of contamination is the discharge of toxic chemicals from industrial effluents and agricultural nonpoint sources into the water bodies [18]. Low dissolved oxygen problem in the River Ravi (Pakistan) was identified by [19, 20] due to discharge of very high biochemical oxygen demand (BOD) loadings from several wastewater outfalls and surface drains. They also identified concentrations of unionized

ammonia (NH_3) higher than its permissible toxicity levels. Fecal coliforms were also found higher than the recreational and irrigation water quality standards [21]. In particular, in low flow conditions, the water quality of the river does not meet the desired standards for any of its beneficial uses, including fishery, irrigation, recreation, and drinking.

The people frequently suffer from waterborne diseases because of using contaminated drinking water. About 30% of all reported diseases and 40% of deaths in the Pakistan are attributed to fecal contamination of drinking waters [22]. In Pakistan, the main causes of waterborne diseases are contaminated drinking water, lack of sanitary facilities at treatment plants, and release of industrial effluents and domestic sewage into the water bodies [23].

The PCA approach was successfully utilized to interpret a large and complex data matrix for the investigation of water quality of the Açude da Macela reservoir for consumption of the local community and agriculture purposes [24]. Different multivariate integrated techniques were applied to classify the various sampling sites according to spatial and temporal characteristics for the investigation of water quality of Qian-dao Lake in China. The results obtained from this study reveal that CA, discriminant analysis (DA), and PCA are valuable tools for water resource management [25].

The Soan River is among those waterways where housing developments have been established along the banks. According to [26] about 5.5 million residents of Rawalpindi/Islamabad discharge municipal sewage into the Soan River. Due to encroachment along the riverbanks, the ecosystem and surface runoff have drastically affected the pollution levels of the river. Due to the inadequate drainage of wastewater into the river, an excessive increase of heavy metals and nutrients in river water above the national standards has been observed [26].

The aim of this study was to investigate the possible major pollution sources affecting water quality in the urban area of the Soan River by using multivariate statistical approaches through Pearson's correlation, CA, and PCA with the objective of providing support for environmental executives in order to enable them to make better decisions regarding action plans.

2. Materials and Methods

2.1. Description of Study Area. The Soan River is a minor tributary of the Indus River. The origin of the Soan River is the southwestern range of the Murree hills, which are located at longitude $71^{\circ}45'$ to $73^{\circ}35'E$ and latitude $32^{\circ}45'$ to $33^{\circ}55'N$ and passes through the Rawalpindi, Attock, and Jehlum districts of Punjab (Pakistan). The total length of the Soan River is 272 km, and it drains a total catchment area of 11,085 km^2 . Nallah Lai, a drainage channel which originates from the area of the Margalla hills, ultimately falls into this river by draining wastewater into the river near the Soan bus bay. This river provides its water to the Simly Dam which is located in Islamabad and is the largest drinking water reservoir for the Islamabad Capital Territory and its adjacent areas. Agricultural lands are also located along the river, and the use of herbicides and insecticides is common in these farming areas

[27]. The main sources of water in the Soan River basin are precipitation, streamflow, and snow melting found in the Murree region. However, this basin is water deficient, as most of the area is covered and underlain by the rocks of tertiary age. The Ling stream from Lithar and Kahuta joins the Soan River near Sihala before Kaak bridge. The Kaak bridge was constructed over this river at the Islamabad Highway. Furthermore, two more streams near this bridge join the Soan River. One stream is the Korang River, and the other is the Lei stream. After following about 16 km distance upstream of Kalabagh in the southwest direction, the Soan River joins the Indus River [27, 28].

The mean temperature of the river ranges from a minimum of 3°C in winter to a maximum of 45°C in summer. In summer, the river flow decreases, whereas in the monsoon season (July–September), heavy rain results in flash flooding, which causes increased runoff. The river passing through residential societies is contaminated by the dumping sites of the surrounding area. Furthermore, construction waste is also poured into the river and has been since the formation of private housing societies nearby. There are also some industries constructed along the left bank of the river, and these directly pour their effluents into the river without any treatment [29]. This surface water also recharges the ground water aquifer of the region. The average annual rainfall in the study area varies from 250 to 1800 mm from north to southwest. The maximum rainfall occurs during July–September, whereas the minimum takes place in the months of November–December [30].

2.2. Sample Collection and Chemical Analysis. The most critical and polluted urban reach of the Soan River from downstream of Simly Dam (Chakian Bridge) Islamabad to the Bahria Town Phase-VIII Rawalpindi was identified by reconnaissance survey, and seven sampling sites were selected according to natural conditions and industrial zone as well as human activities along the river banks. The water samples were collected for two seasons (premonsoon in the month of May 2019 and postmonsoon in the month of November 2019) from a depth of 0.3 m in clean and sterilized plastic sample bottles of 1.5 L. The water samples collected from the field were placed in an insulated cool box and transported as soon as possible to the PCRWR Water Quality Lab Islamabad and stored at 4°C until they were analyzed [31]. The water quality parameters, their units, and methods of analysis are described in Table 1. The locations of selected sites as shown in Figure 1 were recorded (latitudinal and longitudinal position) using a handheld GPS (Global Positioning System) unit.

2.3. Pearson Correlation Matrix. Pearson correlation matrix is widely employed to evaluate the relationship between the variables of any data set. If the correlation coefficient exists between 1 and -1; it indicates that the strongest relationship may be positive or negative between two parameters, and if it is zero, there will be no correlation between them at a significant level of $p < 0.05$ (Kumar et al., 2006). Pearson's correlation matrix measures the strength, direction, and probability of the linear association between two interval or ratio variables.

2.4. Cluster Analysis (CA). CA enables the researcher to arrange the water samples into groups resulting in high interval homogeneity (within clusters) and high external heterogeneity (between clusters). In CA, a tree-like diagram known as a dendrogram is created, and it summarizes the clusters by significantly decreasing the dimensions of the primary dataset. Euclidean distance is widely used to evaluate the sampling stations which can be characterized by the discrepancy present in the statistical analysis of the data.

The CA approach through the Ward linkage method was used collectively for clustering to classify the samples for investigation of water quality characteristics of the Mahitsy Commune, Central Highland of Madagascar [32]

2.5. Principal Component Analysis (PCA). PCA is a dimension reduction analytical method which reduces a large number of variables into a lower number of factors. It is normally used for data formation and for attaining qualitative inductive data about possible contamination causes. An eigenvalue in the PCA approach shows the significance of any factor, and the factors having higher eigenvalues greater than 1 or equal to 1 are considered the most significant factors as endorsed by [33]. A scree plot is a diagram in the PCA multivariate technique whose Y-axis represents the eigenvalues and whose X-axis shows the number of factors. The point where the gradient of the curve clearly levels off (the elbow) shows the number of factors to be created by the data analysis.

According to the literature review, the PCA has been widely used to determine latent contamination causes, without changing the original characteristics of the variables and by reducing the dimensions of data with negligible loss of primary data and by grouping multiple variables with respect to their geography [8, 34–36].

3. Results and Discussion

3.1. Data Processing and Statistical Analyses of Water Quality Parameters. To evaluate the collected river water quality dataset consistently at different sampling stations, multivariate statistical techniques were employed through Pearson's correlation, CA, and PCA. All statistical computations in the present study were rendered using Microsoft Excel 2016 and IBM SPSS 22.

The statistical description of mean value, range, and standard deviation of selected water quality parameters at seven stations of the Soan River collected during premonsoon and postmonsoon seasons (2019) are provided in Table 2.

The values of measured parameters EC, pH, alkalinity as CaCO_3 , HCO_3^- , Ca, Cl^- , hardness, Mg, Na, K, SO_4 , NO_3 , and TDS are generally within the recommended range at all stations during both seasons, but there is significant change of measured values of DO, BOD, COD, and turbidity.

Lower concentration of DO was observed during both seasons at stations S-4 to S-7 at downstream as compared to upstream stations S-1 and S-2 due to large amounts of organic waste originating from domestic waste and industrial waste water entering into the Soan River containing plenty of biodegradable material causing increase in oxygen demand, and DO is depleted steadily. Lower concentration of DO

TABLE 1: Water quality parameters, abbreviations, units, and analytical methods used for analysis.

Sr. no.	Parameters	Units	Analysis method
1	pH	pH unit	pH meter, Hanna Instruments, model 8519, Italy
2	Electrical conductivity (EC)	$\mu\text{S}/\text{cm}$	EC meter, Hach-44600-00, USA
3	Turbidity (Turb.)	NTU	Turbidity meter, LaMotte, model 2008, USA
4	Alkalinity as CaCO_3	ppm	2320 standard method (1992)
5	Bicarbonates (HCO_3^-)	ppm	2320 standard method (1992)
6	Calcium (Ca)	ppm	3500 Ca-D standard method (1992)
7	Chloride (Cl^-)	ppm	Titration assembly argentometric method APHA (1998)
8	Hardness (hard.)	ppm	
9	Magnesium (Mg)	ppm	Flame photometer
10	Potassium (K)	ppm	Flame photometer
11	Sodium (Na)	ppm	Flame photometer
12	Sulphate (SO_4)	ppm	Spectrophotometric
13	Nitrate (NO_3)	ppm	Spectrophotometric
14	Total dissolved solids (TDS)	ppm	2540C, standard method (2012)
15	Biological oxygen demand (BOD)	ppm	Manometric pressure measurement method by Lovibond BOD system
16	Chemical oxygen demand (COD)	ppm	
17	Dissolved oxygen (DO)	ppm	DO meter method, model Jenway 970
18	Total suspended solids (TSS)	ppm	Oven dry method

FIGURE 1: Map of the study area showing water quality monitoring stations.

TABLE 2: Statistical analysis of water quality parameters of the Soan River (premonsoon and postmonsoon seasons 2019).

Stations	EC	pH	Turbidity	Alkalinity	HCO ₃	Ca	Cl	Hard.	Mg	K	Na	SO ₄	NO ₃	TDS	DO	BOD	COD	TSS
S-1	426	8.3	2.7	1.80	180	40	24	170	17	2.1	36	25	0.4	234	5.7	10	16	10
S-2	541	7.55	6.8	2.00	200	44	36	170	15	6.3	50	30	1.7	325	5.5	15	25	20
S-3	576	7.52	2.9	2.20	220	32	36	180	24	7.0	55	35	4.7	317	5.3	20	32	10
S-4	754	6.65	63	3.00	300	40	18	200	24	8.2	90	40	12	415	2.1	87	420	142
S-5	799	6.97	6.7	3.30	330	40	48	240	34	8.1	88	45	1.7	439	3.4	55	185	20
S-6	1109	7.52	18	4.50	450	60	60	370	53	17.0	78	65	2.8	665	3.5	50	118	60
S-7	1329	7.64	15.1	4.70	550	55	72	470	78	21.0	90	90	3.2	797	3.9	58	133	50
Mini-	426.00	6.65	2.70	1.80	180.00	32.00	18.00	170.00	15.00	2.10	36.00	25.00	0.40	234.00	2.10	10.00	16.00	10.00
Maxi-	1329.00	8.30	63.00	4.70	550.00	60.00	72.00	470.00	78.00	21.00	90.00	90.00	4.70	797.00	5.70	87.00	420.00	142.00
Mean	790.57	7.45	16.46	3.07	318.57	44.43	42.00	257.14	35.00	9.96	69.57	47.14	3.79	456.00	4.20	42.14	132.71	44.57
Std. deviation	325.05	0.52	21.34	1.17	138.01	9.73	19.29	117.43	22.88	6.61	22.24	22.89	3.87	203.18	1.34	28.14	141.92	47.18
PSQCA/NEQS standards	NGVS	6.5-8.5	<5	NGVS	NGVS	NGVS	250	500	NGVS	NGVS	NGVS	NGVS	10	1000	5	80	150	200
S-1	371	7.6	105	1.30	130	28	14	132	15	3.2	23	26	2.7	204	4.5	80	142	270
S-2	403	7.3	82	1.22	122	28	22	140	17	4.8	25	34	3.7	222	4.2	102	190	210
S-3	415	7.18	75	1.30	130	40	19	140	10	4.6	32	32	4.5	228	3.7	90	163	225
S-4	445	7.01	67	1.40	140	40	26	150	12	5.3	35	34	6.1	245	3.5	95	170	190
S-5	872	6.9	58	2.80	280	60	77	270	29	10.0	63	37	2.1	480	2.3	110	300	170
S-6	1342	6.77	25	4.30	430	116	89	400	27	26.0	100	72	17	738	1.7	123	381	82
S-7	550	6.85	45	1.80	180	44	29	190	19	6.8	37	32	8.2	303	3.3	97	177	130
Mini-	371	6.77	25	1.20	122	28	14	132	10	3.2	23	26	2.1	204	1.7	80	142	82
Maxi-	1342	7.6	105	4.30	430	116	89	400	29	26	100	72	17	738	4.5	123	381	270
Mean	628.29	7.09	65.29	2.02	201.71	50.86	39.43	203.14	18.43	8.67	45.00	38.14	6.33	345.71	3.31	99.57	217.57	182.43
Std. deviation	358.46	0.29	25.95	1.15	114.86	30.70	30.35	99.38	7.21	7.94	27.59	15.30	5.14	197.12	1.00	13.94	88.3	62.30
PSQCA/NEQS standards	NGVS	6.5-8.5	<5	NGVS	NGVS	NGVS	250	500	NGVS	NGVS	NGVS	NGVS	10	1000	5.00	80	150	200

NGVS: no guideline value set; NEQS: National Environmental Quality Standards; PSQCA: Pakistan Standards and Quality Control Authority.

FIGURE 2: (a-e) Comparison between premonsoon and postmonsoon water quality variations in the Soan River: (a) dissolved oxygen (DO), (b) biochemical oxygen demand (BOD), (c) chemical oxygen demand (COD), (d) total suspended solids (TSS), and (e) turbidity.

shows river pollution. Figure 2(a) shows that the impact of the wastewater discharge on DO sag is extended to the subsequent stations, i.e., S-6 and S-7. Also, from Figure 2(a), it is evident that DO concentration is higher in the month of

May (i.e. premonsoon) due to the impact of early rainfall events during these months, while the DO levels are low, with an exception of S-4, in postmonsoon samples due to dry weather flow conditions that prevail right after the monsoon

FIGURE 3: (a) Dendrogram of cluster analysis for sampling stations according to surface water quality parameters of the Soan River (premonsoon). (b) Dendrogram of cluster analysis for sampling stations according to surface water quality parameters of the Soan River (postmonsoon).

in the study area. Figure 2(b) manifest higher BOD values observed against the lower values of DO at downstream stations due to industrial and populated accommodations. Low concentration level of BOD indicates that the river water is free from biological pollution [37] whereas high BOD is harmful for aquatic life as it deteriorates the DO level [38]. Observed COD concentrations show a similar pattern. High COD value (i.e., 2 times of BOD) at S-5 directs towards the possibility of industrial wastewater discharge that contains nonbiodegradable organics [39]. TSS and turbidity levels are fairly correlated in Figures 2(d) and 2(e), and their higher

values in postmonsoon samples manifest the presence of urban and agricultural runoffs with higher sediment loads.

3.2. Spatiotemporal Variance of Water Quality Parameters by Cluster Analysis (CA). According to Ashfaq et al. [30], there are high spatial and temporal variations of rainfall in the Soan Basin containing about 70% rainfall during the monsoon season (July to September) and the remaining 30% falls in spring and winter seasons. The hierarchical CA technique was used for the assessment of similarity among the sampling sites. A dendrogram was plotted based upon the chemical

TABLE 3: Pearson's correlation matrix (premonsoon).

	EC	pH	Turbidity	Alkalinity	HCO ₃	Ca	Cl ⁻	Hard.	Mg	K	Na	SO ₄	NO ₃	TDS	DO	BOD	COD	TSS
EC	1																	
pH	-0.196	1																
Turbidity	0.200	-0.69	1															
Alkalinity	0.984**	-0.241	0.227	1														
HCO ₃	0.998**	-0.17	0.187	0.985**	1													
Ca	0.752	0.163	-0.02	0.637	0.761*	1												
Cl ⁻	0.855*	0.115	-0.32	0.826*	0.853*	0.706	1											
Hard.	0.972**	0.038	0.032	0.942**	0.977**	0.821*	0.901**	1										
Mg	0.968**	0.009	0.019	0.932**	0.974**	0.846*	0.900**	0.992**	1									
K	0.982**	-0.114	0.137	0.950**	0.970**	0.760*	0.877**	0.969**	0.959**	1								
Na	0.769*	-0.747	0.564	0.794*	0.764*	0.415	0.464	0.611	0.636	0.673	1							
SO ₄	0.985**	-0.073	0.100	0.945**	0.985**	0.844*	0.883**	0.989**	0.993**	0.979**	0.686	1						
NO ₃	0.081	-0.721	0.921**	0.085	0.058	-0.07	-0.395	-0.097	-0.073	0.049	0.484	0.005	1					
TDS	0.997**	-0.143	0.166	0.976**	0.992**	0.764*	0.871*	0.980**	0.970**	0.990**	0.723	0.987**	0.04	1				
DO	-0.544	0.816*	-0.808*	-0.612	-0.542	-0.11	-0.11	-0.36	-0.363	-0.43	-0.903**	-0.421	-0.69	-0.493	1			
BOD	0.597	-0.792*	0.821*	0.636	0.595	0.257	0.151	0.421	0.435	0.489	0.926**	0.496	0.719	0.547	-0.984**	1		
COD	0.276	-0.835*	0.927**	0.322	0.275	0.001	-0.21	0.084	0.097	0.161	0.750	0.166	0.848*	0.222	-0.929**	0.933**	1	
TSS	0.328	-0.662	0.989**	0.357	0.315	0.06	-0.20	0.168	0.148	0.271	0.623	0.227	0.886**	0.299	-0.841*	0.854*	0.912**	1

*Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed).

TABLE 4: Pearson's correlation matrix (postmonsoon).

	EC	pH	Turbidity	Alkalinity	HCO ₃	Ca	Cl	Hard.	Mg	K	Na	SO ₄	NO ₃	TDS	DO	BOD	COD	TSS
EC	1																	
pH	-0.697	1																
Turbidity	-0.812*	0.957**	1															
Alkalinity	0.999**	-0.681	-0.800*	1														
HCO ₃	0.999**	-0.681	-0.800*	1.000**	1													
Ca	0.978**	-0.695	-0.830*	0.976**	0.976**	1												
Cl	0.960**	-0.710	-0.758*	0.956**	0.956**	0.893**	1											
Hard.	0.999**	-0.707	-0.823*	0.999**	0.999**	0.976**	0.957**	1										
Mg	0.828*	-0.567	-0.612	0.832*	0.832*	0.696	0.904**	0.836*	1									
K	0.972**	-0.648	-0.806*	0.969**	0.969**	0.990**	0.874*	0.970**	0.704	1								
Na	0.995**	-0.720	-0.823*	0.992**	0.992**	0.986**	0.954**	0.992**	0.784*	0.971**	1							
SO ₄	0.927**	-0.611	-0.776*	0.917**	0.917**	0.965**	0.812*	0.921**	0.605	0.985**	0.933**	1						
NO ₃	0.759*	-0.621	-0.803*	0.756*	0.756*	0.854*	0.565	0.765*	0.364	0.873*	0.769*	0.893**	1					
TDS	1.00**	-0.697	-0.812*	0.999**	0.999**	0.978**	0.960**	0.999**	0.828*	0.972**	0.995**	0.926**	0.759*	1				
DO	-0.938**	0.866*	0.890**	-0.931**	-0.931**	-0.911**	-0.955**	-0.938**	-0.794*	-0.868*	-0.951**	-0.812*	-0.659	-0.938**	1			
BOD	0.895**	-0.769*	-0.832*	0.874*	0.874*	0.851*	0.903*	0.894**	0.802*	0.868*	0.888*	0.861*	0.661	0.895**	-0.884**	1		
COD	0.978**	-0.667	-0.758*	0.970**	0.970**	0.930**	0.980**	0.974**	0.867*	0.931**	0.970**	0.894**	0.652	0.978**	-0.925**	0.940**	1	
TSS	-0.820*	0.920**	0.986**	-0.810*	-0.810*	-0.825*	-0.756*	-0.834*	-0.657	-0.823*	-0.818*	-0.793*	-0.825*	-0.820*	-0.852*	-0.768*	0.860*	1

*Correlation is significant at the 0.05 level (2-tailed); **Correlation is significant at the 0.01 level (2-tailed).

analysis of all selected sites that grouped these sites into three and two clusters during premonsoon and postmonsoon seasons, respectively, as shown in Figures 3(a) and 3(b).

In the premonsoon season, Cluster 1 contains the stations S-1, S-2, and S-3. The primary land uses around these stations include agriculture and newly developed neighborhoods with less number of residents. Therefore, the pollution levels were found to be lower than in the other stations; it is also evident from high DO levels for these stations in Figure 3(a). Cluster 2 comprises S-4 and S-5 that passes through the developed part of the city receiving large volumes of domestic and industrial wastewaters. High pollution levels during the premonsoon, with high BOD, COD, and low DO levels, can also be seen in Figure 2. The land use is the same in the catchment area of these two stations. Other sources of contamination to this part of the river (S-4 and S-5) are open dumping of solid waste from slaughterhouses and some low-income residential neighborhoods without having a proper solid waste collection and disposal mechanism. Cluster 3 consists of two highly polluted stations at the downstream of the river, i.e., S-6 and S-7. There are two main reasons of such as high pollution (also see Figure 3) at these stations in the premonsoon monitoring: (i) at S-6, the river water is completely mixed with the industrial wastewater discharge from S-5—the meandering river morphology between S-5 and S-6 justifies this mixing length during relatively higher velocities in premonsoon period; (ii) S-6 is the confluence point of Nallah Lai that discharges large volumes of untreated municipal wastewater from Rawalpindi city into the Soan River; and (iii) continuing meandered river morphology covers a mixing length and shows high pollution levels at S-7.

In the postmonsoon period, the river primarily behaves like a sewage drain flowing at very low velocity. Therefore, only two clusters are formed for this monitoring survey, and the cluster configuration differs due to the low flow regime in the river. Cluster 1 contains five stations, i.e., S-1, S-2, S-3, S-4, and S-7. The primary reason for having these stations in a single cluster is the similar DO and COD values obtained at these stations (see Figures 2(a) and 2(c)). Relatively higher DO levels at S-4 and S-7 (i.e., >3 ppm) can be attributed to the sampling errors as the samples were obtained from the river bank, where complete mixing of the wastewater could not have been achieved yet. Cluster 2 contains S-5 and S-6 that shares the similar characteristics as already discussed above. The primary reason for having different clusters is the changing flow pattern in both the seasons that resulted in velocity variations, changing mixing lengths, and absence of nonpoint sources during the postmonsoon period.

The results reveal that the CA technique is a useful approach for the classification of river water at temporal and spatial scales. Therefore, the selection of critical sampling points and the respective costs incurred on future monitoring strategies should be minimized.

3.3. Pearson's Correlation Matrix. Pearson's correlation matrix was prepared, and the correlation coefficient (r) was applied to check the possible relationship among the chemi-

cal parameters of the water quality for the Soan River. According to guidelines established by Cohen [40], there is a strong correlation if the value of the Pearson coefficient (r) is $0.75 \leq r < 1$. Several significant positive strong correlations and few negative correlations were found among some of the parameters which are presented in Tables 3 and 4.

In the premonsoon season, results revealed that most of the parameters show a significant positive correlation, as given below in Table 3:

- (i) EC with alkalinity, HCO_3^- , hardness, Mg, K, SO_4 , and TDS ($r > 0.9$)
- (ii) Turbidity with NO_3^- , COD, and TSS ($r > 0.9$)
- (iii) Alkalinity with HCO_3^- , hardness, Mg, K, SO_4 , and TDS ($r > 0.9$)
- (iv) HCO_3^- with hardness, Mg, K, SO_4 , and TDS ($r > 0.9$)
- (v) Cl^- with hardness, Mg, K, and SO_4 ($r > 0.9$)
- (vi) K and Mg with SO_4 and TDS ($r > 0.9$)
- (vii) TSS with NO_3^- ($r > 0.9$)
- (viii) SO_4 with TDS ($r > 0.9$)
- (ix) BOD with COD ($r > 0.9$)

Alkalinity in water exists due to the presence of bicarbonate (HCO_3^-), carbonate (CO_3^{2-}), and hydroxide (OH^-) of elements such as Ca, Mg, Na, K, and NH_3 while hardness is primarily associated with the presence of carbonates and bicarbonates of Ca and Mg. TDS commonly contains inorganic salts having cations (Ca, Mg, K, and Na) and anions (CO_3 , HCO_3 , Cl, and SO_4). Therefore, TDS are highly correlated with alkalinity, hardness, cations, and anions. EC is essentially a measure of TDS and so was found highly correlated with all these parameters with an R value higher than 0.9. Similarly, turbidity is a measure of TSS. In raw wastewater, COD contains both the soluble and particulate forms of organic matter. Therefore, turbidity was found to be highly correlated with TSS and COD. NO_3^- values are very small (i.e., 3.8 ppm in Table 1) in untreated wastewater due to the absence of nitrification process. So, the main reason of this correlation is the similar trend of the NO_3^- values as of TSS, turbidity, and COD. COD was found highly correlated with the BOD and can be used as a monitoring parameter to avoid the long 5-day BOD test. All the other high correlations found reflect the relationships between alkalinity, hardness, TDS, cations, and anions that have already been described above. TDS and EC are quick tests and can be used to monitor the rest of the parameters using regression equations; the correlation relationship was detected as a strong negative between Na and DO, and DO with COD and BOD ($r > -0.9$). The former makes clear sense as there is essentially no relationship between Na and DO in the river water. DO should have a strong correlation with BOD; the reason for not obtaining a strong correlation between these parameters can be attributed to limited data and sampling errors. No significant correlation of pH was observed with any element in the river water, while there is a slightly negative correlation with other

FIGURE 4: (a, b) Scree plot of the eigenvalues for principal components.

parameters like BOD, COD, TSS, NO_3 , and turbidity. The analysis of the results shows that these water quality parameters cannot be controlled by a single specific factor because they have different pollution sources. Overall, divergent results revealed that common sources are not always linked with significant correlations among each other.

In Table 4, the postmonsoon results revealed that most of the parameters show a significant positive correlation as follows:

(i) EC with alkalinity, HCO_3 , hardness, K, Cl^- , Ca, Na, SO_4 , COD, and BOD ($r > 0.9$)

(ii) EC with TDS which has strong correlation ($r = 1$)

(iii) TSS with turbidity with NO_3 ($r > 0.9$)

(iv) Alkalinity with HCO_3 , hardness, K, Na, SO_4 , COD, and TDS ($r > 0.9$)

(v) HCO_3 with Ca, Cl^- , K, Na, hardness, K, SO_4 , COD, and TDS ($r > 0.9$)

(vi) Cl^- with hardness, Mg, Na, TDS, Ca, COD, and BOD ($r > 0.9$)

(vii) K and Mg with Na, SO_4 , and TDS ($r > 0.9$)

TABLE 5: Principal components, loading factors, eigenvalues, and variance.

Premonsoon season			Postmonsoon season		
Parameters	Components		Parameters	Components	
	PC-1	PC-2		PC-1	PC-2
EC	0.97	0.24	EC	0.85	0.53
pH	0.01	-0.86	pH	-0.34	-0.83
Turbidity	-0.03	0.95	Turbidity	-0.42	-0.88
Alkalinity as CaCO_3	0.94	0.29	Alkalinity as CaCO_3	0.85	0.51
Bicarbonate	0.97	0.23	Bicarbonate	0.85	0.51
Calcium	0.82	-0.07	Calcium	0.75	0.63
Chloride	0.94	-0.25	Chloride	0.90	0.39
Hardness	0.99	0.04	Hardness	0.84	0.54
Mg	0.99	0.04	Mg	0.89	0.19
K	0.97	0.15	K	0.75	0.61
Na	0.62	0.71	Na	0.82	0.56
SO_4	0.99	0.12	SO_4	0.68	0.63
Nitrate (NO_3)	-0.14	0.90	Nitrate (NO_3)	0.35	0.84
TDS	0.98	0.19	TDS	0.85	0.53
DO	-0.33	-0.91	DO	-0.74	-0.61
BOD	0.40	0.91	BOD	0.73	0.57
COD	0.05	0.99	COD	0.89	0.43
TSS	0.10	0.93	TSS	-0.45	-0.86
Eigenvalues	10.86	5.77	Eigenvalues	15.67	1.08
% explained variance	60.35	32.11	% explained variance	87.1	6.01
% cumulative variance	60.35	92.46	% cumulative variance	87.1	93.11

(viii) NO_3 with SO_4 ($r > 0.9$)

(ix) SO_4 with TDS ($r > 0.9$)

It is pertinent to mention here that a significant negative correlation of DO and TSS with all other parameters was observed except for turbidity and pH due to the low flow regime in the river. Discussion on the premonsoon period stated above explains the reasons of the strong correlations between hardness, alkalinity, TDS, EC, cations, and anions for postmonsoon sampling as well. Similar to the premonsoon period, the negative correlation between DO and BOD (and COD) during postmonsoon is due to data limitations and sampling errors. For remaining parameters, this finding validates that there is practically no correlation between DO and other parameters, except TSS that sometimes are organic in nature and contribute to sediment oxygen demand. However, the samples in the present study were not obtained from the bottom, so no such correlation can be established.

3.4. Principal Component Analysis (PCA). The guidelines developed by Kaiser [41] for the PCA technique were followed to determine the significant relationships among the selected sites in order to identify the physiochemical properties of the water quality parameters; the loading factors were calculated and classified as “strong,” “moderate,” and “weak” corresponding to absolute loading values of

>0.75 , 0.50 - 0.75 , and 0.30 - 0.50 , respectively. In order to understand the underlying data structure, scree plot graphs are used for configuring the principal components to be retained based on the eigenvalues. The scree plot also provides the percentage variances explained for each component and provides an indication on how the principal components were obtained [42].

As shown in Figures 4(a) and 4(b), there is a definite change of slope after the 3rd eigenvalue; only two components were retained. Loading values and the explained variance of two retained PCs are shown in Table 5. Only two PCs have eigenvalues > 1 (Kaiser Normalization) which explain almost 92.46% of the total variance in the water quality dataset during the premonsoon season in the water quality dataset whereas 93.11% of the total variance was described during the postmonsoon season, which is prescribed by only two factors in both seasons due to eigenvalues greater than 1. The variables having eigenvalues lower than 1 were ignored due to their low significance [43].

In the premonsoon season, PC-1 recorded 60.35% of the total variance, indicating that high positive loadings of EC, alkalinity, HCO_3 , Ca, Cl^- , hardness, Mg, K, SO_4 , and TDS were observed. In this factor, there are biogenic and anthropogenic (urban wastewater) pollution sources. PC-2 contributes 32.11% of the total variance and comprises strong positive loading on turbidity, NO_3 , BOD, COD, and TSS whereas there is a negative loading on pH and DO

FIGURE 5: (a) PCA plotting for premonsoon season; (b) PCA plotting for postmonsoon season.

parameters. This PC-2 is associated with the leaching of sewage from urban areas and industrial point-source discharge, agricultural nonpoint sources, and livestock operations [44]. Similarly, during the postmonsoon season, PC-1 contributes 87.10% of the total variance, showing high positive loadings of all parameters except pH, turbidity, DO, and TSS which show negative loadings. From Table 5, it is clear that there is negative loading of DO during both seasons which may be indicative of too much bacteria and an excess amount of BOD (untreated sewage, organic discharges).

The results obtained from PCA identified the major pollution sources of the study area which is mainly due to some natural processes and anthropogenic activities. The discharging untreated urban wastewater into the river constitutes the major point anthropogenic contamination source. The nonpoint source which also contributes immensely in Soan River water contamination is from agricultural activities and livestock farming. The PCA plots of premonsoon and postmonsoon are shown in Figures 5(a) and 5(b).

4. Conclusion

In the present study, multivariate statistical techniques were used for the evaluation of pollution variations and source apportionment due to water contamination in the Soan River. The hierarchical CA technique was used for the assessment of similarity among the sampling sites. A dendrogram was plotted based upon the chemical analysis of all selected sites that grouped these sites into three and two statistical significant clusters during premonsoon and postmonsoon seasons, respectively, as shown in Figures 3(a) and 3(b). The PCA and FA made it possible to extract and recognize the origins of or the factors responsible for water quality variations during the year 2019. The parameters such as DO, BOD, COD, turbidity, and TSS were identified as the most significant parameters contributing to water quality, indicating that the water quality of the Soan River deteriorates gradually from upstream to downstream as it passes through the urban areas, receiving domestic and industrial wastewater from the outfalls.

The results of DO, COD, TSS, and BOD prove that this wastewater requires proper treatment to overcome these indicators of pollution before disposal. Moreover, mismanagement causes solid waste to move down to the drainage system, which results in sewer line blockage and consequently in the spread of various types of diseases-causing organisms.

The sustainable integrated water resource management of rivers should be planned through various approaches such as the treatment of wastewater released from industrial zone and municipal locations. It is recommended that a long-term and continued monitoring program should be established to predict the water quality of the Soan River. This can be achieved by the selection of landfill sites to dump the municipal solid waste, adequate treatment of urban and industrial wastewater before releasing in the river, and improvement of agronomic practices in this area. Furthermore, installing an automatic wireless-sensor networking system for river water quality assessment and forecasting may support the environmental protection authority in the implementation of appropriate action plans.

Data Availability

The data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflict of interest regarding the publication of this paper.

Acknowledgments

The authors acknowledge the PCRWR, Ministry of Science and Technology Islamabad, Pakistan, for technical support.

References

- [1] WHO, *Drinking water fact sheets*, World Health Organization (WHO), 2018.
- [2] P. Li and H. Qian, "Water resource development and protection in loess areas of the world: a summary to the thematic issue of water in loess," *Environmental Earth Sciences*, vol. 77, no. 24, p. 796, 2018.
- [3] O. V. Omonona, J. O. Amah, S. B. Olorunju et al., "Hydrochemical characteristics and quality assessment of groundwater from fractured Albian carbonaceous shale aquifers around Enyigba-Ameri, southeastern Nigeria," *Environmental Monitoring and Assessment*, vol. 191, no. 3, p. 125, 2019.
- [4] M. Tripathi and S. K. Singal, "Use of principal component analysis for parameter selection for development of a novel water quality index: a case study of river Ganga India," *Ecological Indicators*, vol. 96, pp. 430–436, 2019.
- [5] L. Ali, A. Rashid, S. A. Khattak, M. Zeb, and S. Jehan, "Geochemical control of potential toxic elements (PTEs), associated risk exposure and source apportionment of agricultural soil in southern Chitral, Pakistan," *Microchemical Journal*, vol. 147, pp. 516–523, 2019.
- [6] A. Rashid, S. A. Khattak, L. Ali et al., "Geochemical profile and source identification of surface and groundwater pollution of district Chitral, Northern Pakistan," *Microchemical Journal*, vol. 145, pp. 1058–1065, 2019.
- [7] M. Memon, M. S. Soomro, M. S. Akhtar, and K. S. Memon, "Drinking water quality assessment in southern Sindh (Pakistan)," *Environmental Monitoring and Assessment*, vol. 177, no. 1–4, pp. 39–50, 2011.
- [8] J. Zhao, G. Fu, K. Lei, and Y. Li, "Multivariate analysis of surface water quality in the Three Gorges area of China and implications for water management," *Journal of Environmental Sciences*, vol. 23, no. 9, pp. 1460–1471, 2011.
- [9] Y. Ouyang, P. Nkedi-Kizza, Q. T. Wu, D. Shinde, and C. H. Huang, "Assessment of seasonal variations in surface water quality," *Water Research*, vol. 40, no. 20, pp. 3800–3810, 2006.
- [10] A. A. Masoud, "Groundwater quality assessment of the shallow aquifers west of the Nile Delta (Egypt) using multivariate statistical and geostatistical techniques," *Journal of African Earth Sciences*, vol. 95, pp. 123–137, 2014.
- [11] S. Giri and A. K. Singh, "Risk assessment, statistical source identification and seasonal fluctuation of dissolved metals in the Subarnarekha River, India," *Journal of Hazardous Materials*, vol. 265, pp. 305–314, 2014.
- [12] Y. Yi, Z. Yang, and S. Zhang, "Ecological risk assessment of heavy metals in sediment and human health risk assessment of heavy metals in fishes in the middle and lower reaches of the Yangtze River basin," *Environmental Pollution*, vol. 159, no. 10, pp. 2575–2585, 2011.
- [13] S. Bhargava, U. Saxena, and A. K. Bhargava, "Some microbiological studies of polluted Kali River water at Meerut," *Biochemical and Cellular Archives*, vol. 9, pp. 155–156, 2009.
- [14] WWAP, *Statics on water use. World Water assessment programme*, WWAP, 2012.
- [15] N. Gupta, S. M. Nafees, M. K. Jain, and S. Kalpana, "Physicochemical assessment of water quality of river Chambal in Kota city area of Rajasthan state (India)," *Rasayan Journal of Chemistry*, vol. 4, no. 2, pp. 686–692, 2011.
- [16] M.-L. Wu, Y. S. Wang, Y. T. Wang et al., "Seasonal and spatial variations of water quality and trophic status in Daya Bay, South China Sea," *Marine Pollution Bulletin*, vol. 112, no. 1–2, pp. 341–348, 2016.
- [17] B. Kundi, *Pakistan's water crisis: why a national water policy is needed*, The Asia Foundation, California St., 2017, 9th floor San Francisco, CA 94104.
- [18] M. Daud, M. Nafees, S. Ali et al., "Drinking water quality status and contamination in Pakistan," *BioMed Research International*, vol. 2017, 18 pages, 2017.
- [19] H. Haider and W. Ali, "Development of dissolved oxygen model for a highly variable flow river: a case study of Ravi River in Pakistan," *Environmental modeling & assessment*, vol. 15, no. 6, pp. 583–599, 2010.
- [20] H. Haider and W. Ali, "Calibration and verification of a dissolved oxygen management model for a highly polluted river with extreme flow variations in Pakistan," *Environmental Monitoring and Assessment*, vol. 185, no. 5, pp. 4231–4244, 2013.
- [21] H. Haider and W. Ali, "Fecal coliform management using a coupled hydrodynamics and water quality model for the river Ravi in Pakistan," *Pakistan Journal of Engineering and Applied Sciences*, vol. 9, pp. 48–57, 2011.
- [22] D. S. Asia, *Water vision 2025, country report, Pakistan, Maharashtra, India, Global Water Partnership*, South Asia Technical Advisory Committee Regional Office, 2000.

- [23] I. Hashmi, S. Farooq, and S. Qaiser, "Chlorination and water quality monitoring within a public drinking water supply in Rawalpindi Cantt (Westridge and Tench) area, Pakistan," *Environmental Monitoring and Assessment*, vol. 158, no. 1-4, pp. 393-403, 2009.
- [24] C. A. B. Garcia, H. L. Garcia, M. C. S. Mendonça et al., "Assessment of water quality using principal component analysis: a case study of the Açude da Macela-Sergipe-Brazil," *Water Resources*, vol. 7, p. 8, 2017.
- [25] Q. Gu, Y. Zhang, L. Ma et al., "Assessment of reservoir water quality using multivariate statistical techniques: a case study of Qiandao Lake, China," *Sustainability*, vol. 8, no. 3, p. 243, 2016.
- [26] S. Nazeer, M. U. Khan, and R. N. Malik, "Phytoplankton spatio-temporal dynamics and its relation to nutrients and water retention time in multi-trophic system of Soan River, Pakistan," *Environmental Technology & Innovation*, vol. 9, pp. 38-50, 2018.
- [27] S. Nazeer, M. Z. Hashmi, and R. N. Malik, "Heavy metals distribution, risk assessment and water quality characterization by water quality index of the river Soan, Pakistan," *Ecological Indicators*, vol. 43, pp. 262-270, 2014.
- [28] F. Iqbal, M. Ali, A. Salam et al., "Seasonal variations of physico-chemical characteristics of river Soan water at Dhoak Pathan bridge (Chakwal), Pakistan," *International Journal of Agriculture and Biology*, vol. 6, no. 1, pp. 89-92, 2004.
- [29] M. Aoun Rizvi and A. I. S. Ullah, "A study on the evaluation of water pollution on Soan river Rawalpindi. M.Sc. (Env. Sci.)," in *College of Earth and Environmental Sciences*, University of the Punjab, Lahore, Pakistan: Lahore, 2004.
- [30] A. Ashfaq, M. Ashraf, and A. Bahzad, "Spatial and temporal assessment of groundwater behaviour in the Soan Basin of Pakistan. University of Engineering and Technology Taxila," *Technical Journal*, vol. 19, no. 1, p. 12, 2014.
- [31] L. S. G. Clesceri, A. E. Greenberg, and A. D. Eaton, *Standard methods for the examination of water and wastewater*, American Public Health Association, Washington, 1998.
- [32] M. Rasolofonirina, V. Ramaroson, and S. D. Ravelomanantsoa, "Assessment of nitrate occurrence in the shallow groundwater of Merimandroso area, Analamanga region, Madagascar using multivariate analysis," *American Journal of Water Resources*, vol. 6, no. 1, pp. 39-47, 2018.
- [33] 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.
- [34] W. D. Alberto, D. María del Pilar, A. María Valeria, P. S. Fabiana, H. A. Cecilia, and B. María de los Ángeles, "Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River Basin (Córdoba-Argentina)," *Water Research*, vol. 35, no. 12, pp. 2881-2894, 2001.
- [35] M. Vega, R. Pardo, E. Barrado, and L. Debán, "Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis," *Water Research*, vol. 32, no. 12, pp. 3581-3592, 1998.
- [36] F. Zhou, G. H. Huang, H. Guo, W. Zhang, and Z. Hao, "Spatio-temporal patterns and source apportionment of coastal water pollution in eastern Hong Kong," *Water Research*, vol. 41, no. 15, pp. 3429-3439, 2007.
- [37] D. Saksena, R. Garg, and R. Rao, "Water quality and pollution status of Chambal river in National Chambal sanctuary, Madhya Pradesh," *Journal of Environmental Biology*, vol. 29, no. 5, pp. 701-710, 2008.
- [38] O. R. Awofolu, Z. Mbolekwa, V. Mtshemla, and O. S. Fatoki, "Levels of trace metals in water and sediment from Tyume River and its effects on an irrigated farmland," *Water SA*, vol. 31, no. 1, pp. 87-94, 2005.
- [39] H. A. Raja and C. A. Shearer, "Freshwater ascomycetes: new and noteworthy species from aquatic habitats in Florida," *Mycologia*, vol. 100, no. 3, pp. 467-489, 2017.
- [40] J. Cohen, *Statistical power analysis for the behavioral sciences*, Erlbaum: Hillsdale, Hillsdale, NJ, 2nd edition, 1988.
- [41] H. F. Kaiser, "The application of electronic computers to factor analysis," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 141-151, 2016.
- [42] C.-W. Liu, K.-H. Lin, and Y.-M. Kuo, "Application of factor analysis in the assessment of groundwater quality in a Black-foot disease area in Taiwan," *Science of the Total Environment*, vol. 313, no. 1-3, pp. 77-89, 2003.
- [43] J. O. Kim, O. Ahtola, P. E. Spector, and C. W. Mueller, *Introduction to factor analysis: what it is and how to do it*, Sage, 1978.
- [44] P. Simeonova, V. Simeonov, and G. Andreev, "Water quality study of the Struma river basin, Bulgaria (1989-1998)," *Open Chemistry*, vol. 1, no. 2, pp. 121-136, 2003.