Sādhanā (2022)47:251
https://doi.org/10.1007/s12046-022-02018-7

Application of chemometric modeling for identification of pollution sources from drains of Ghaggar River, Punjab, India

HARNEET KAUR[1], ANITA RAJOR[2] and AMRITPAL SINGH KALEKA[1,*]

[1] Department of Zoology and Environmental Sciences, Punjabi University Patiala, Punjab 147 002, India
[2] School of Energy and Environment, Thapar University Patiala, Punjab 147 004, India
e-mail: harneet.nirwal18@gmail.com; anitarajor@gmail.com; apskaleka@gmail.com

MS received 23 June 2022; revised 24 September 2022; accepted 8 October 2022

Abstract. In the present research work, physicochemical analysis on the drainage system was conducted by including three drains of Ghaggar River (Punjab) India. The study was focused on the drains, namely Sirhind Choe (S1), Patiala River (S2), and Dhakanshu Drain (S3). These drains receive untreated sewage, agricultural runoff, and industrial effluents from several major cities, towns, and villages of Punjab. In this study, fifteen physicochemical parameters were determined from water samples collected from each drain in five seasons during 2017-2018. The analytical results of all three sites showed that total suspended solids (TSS), Chemical oxygen demand (COD), and Biochemical Oxygen Demand (BOD) were found above the permissible limit of the Center Pollution Control Board (CPCB), India. Further, "Chemometric Modeling" was used to interpret complex data. The statistical tools such as Correlation Analysis, Principal Component Analysis (PCA), Factor Analysis (FA), Cluster Analysis (CA), and Discriminant Analysis (DA) were performed on the data set. A strong interrelationship between the parameters of each site was depicted in the correlation analysis. The PCA/FA identified the agricultural and domestic pollution from the study area. The temporal CA reduced the seasonal data into four significant clusters. The spatial DA predicted four parameters viz., total hardness, total dissolved solids, temperature, and sulfate as important discriminant variables with 100% correct assignment. In this study, chemometric modeling contributed to understanding the water quality patterns in wastewater drains scientifically and efficiently. The statistical tools identified major pollution contributors and temporal ways that provide sustainable water resource management guidelines.

Keywords. Water pollution; wastewater drains; Ghaggar River; physicochemical parameters; chemometric modeling; agrarian activities.

1. Introduction

All over the world, water scarcity and its pollution demand monitoring and proper management [1]. Consistent monitoring of all physicochemical, and other parameters can be helpful in preventing water pollution [2].

In India, several reports of government agencies depicted that 80% of India's surface water is polluted, and sewage is a significant source of pollution [3]. In 2009, Center Pollution Control Board (CPCB) revealed that major towns and cities produce more than 38 billion liters of wastewater every day. Out of the total generated sewage, only 30% is collected, and less than 20% of collected sewage is treated. The rest is discharged into ponds, lakes, rivers, and seas [3]. Also, in Punjab, some reports indicated about degraded quality of surface water and groundwater [4, 5].

Punjab is one of the most fertile states with a total area of 50,362 km^2, i.e. 1.54% of India's total geographical area. The state is extremely reliant on the agrarian economy. From the Green revolution era (1969), extensive agricultural activities have been initiated there. The pollution from the agricultural and industrial sectors is adversely affecting the natural water sources of the state. This state is part of the Indo-Gangetic plains and has three major rivers: Beas, Sutlej, and Ravi. Besides this, the Ghaggar River also flows in the southwest part of the state [4].

River water pollution has become a major concern in Punjab because water quality is being deteriorated day by day. In order to manage water resources, Punjab Pollution Control Board (PPCB) monitors thirty-seven sampling sites of the aquatic ecosystem quarterly, and sufficient data is available on the analytical study of all these rivers [5]. There are several other studies that emphasize on water quality monitoring of individual rivers of Punjab such as Sutlej, Ravi, Beas and Ghaggar [6–11]. Although these previous analytical works have been studied in detail, but insufficient attention has been paid to the wastewater drains or small tributaries. Consequently, water quality data on

*For correspondence
Published online: 22 November 2022

local drains of Punjab state is not available. These wastewater drains directly deal with pollution and are responsible for polluting the major rivers of Punjab by discharging untreated water into rivers. Thus, there is a requirement for consistent monitoring plans for small water bodies, tributaries, and drains so that preventive measures can be taken to reduce river water pollution. Furthermore, During 2005-2007 a health survey was conducted by the Punjab Pollution Control Board (PPCB) and Post Graduate Institute of Medical Education & Research, Chandigarh (PGIMER) on the five drains of Punjab. In this study, they revealed that the people who live in the villages along wastewater drains were found to have many reproductive and other serious health issues. This study strictly suggested proper monitoring of wastewater drains [12]. Based on such recommendations and to address the previous overlooked aspect of water quality analysis, current comprehensive research work was designed on the water chemistry of wastewater drains.

The main objectives of the present study were to investigate the physicochemical parameters of three wastewater drains of Ghaggar river and find out the dominant pollution sources in these drains by using statistical approach. Also, temporal variations were studied through statistical model viz., chemometric modeling, which helps to validate the raw data and increases the integrity of the current research work. This method helps in reducing data complexity and is useful for better understanding of water quality of studied sites.

2. Methodology

2.1 *Site description*

Ghaggar River is one of the major rivers of northern India. It originates from the Shivalik foothill ranges of the Himalayas in Himachal Pradesh. It flows between the Yamuna and Satluj in Himachal Pradesh and enters Haryana, Punjab, and Rajasthan. In Rajasthan, it disappears in the world's 17^{th} largest desert, the "Thar Desert." It is a seasonal river, and rainwater is its primary source. However, due to intensive agricultural activities in Punjab and Haryana, the Ghaggar River also receives agricultural runoff along with the industrial and domestic waste. The river water is utilized for irrigation, drinking (cattle), and rarely for washing or bathing [13].

Three monitoring stations: Sirhind choe (S1), Patiala River (S2), and Dhakanshu Drain (S3), were selected in the current research work (table 1; figure 1). These three open drains of Ghaggar River carry wastewater from the agricultural, domestic, and industrial sectors. The Sirhind choe drain is designated by S1 that covers Patiala and Fatehgarh Sahib Region and is 60.97 km long. It carries untreated sewage and agricultural runoff from adjoining cities, towns, and villages. At the same time, the Patiala River (S2) is the most important tributary creek that joins the Ghaggar River and is a 71.08 km long drain. It carries mainly sewage water from Patiala city and adjacent towns and villages. The third site Dhakanshu drain (S3) is 9.90 km long and directly joins Ghaggar. It carries mainly industrial effluents of major cities of Punjab such as Mohali, Chandigarh, Patiala, and Rajpura. It also receives agricultural runoff and untreated sewage.

2.2 *Sampling*

Wastewater sample collection and testing was done according to standard methods. The water samples were collected during 2017-2018 in five different seasons viz., post-monsoon, winter, post-winter transitional, summer, and monsoon [14]. The Grab-sampling method was followed for the wastewater sample collection. Before sampling, containers were adequately washed, then rinsed with distilled water. During sample collection, all containers were rinsed with sewage water from respective sampling sites. Samples were collected underneath the surface of water and sampling containers were opened below the surface with the mouth directed toward the flow of water. Plastic sampling containers of 5L were filled up to the brim. Samples for COD (Chemical Oxygen Demand) were collected in separate plastic containers (1L), and 1 ml/L sulfuric acid (H_2SO_4), with molar concentration of 18 mol/L, was added at the time of sampling in order to preserve samples at Ph<2 that inhibit the bacterial growth in samples. All the samples were transported to the laboratory in the icebox and stored at 4° C [15].

Table 1. Water quality monitoring Sites.

Sites	Coordinates	Length (Km)	Pollution Sources
Sirhind choe (S1)	30° 30' 27.72'' N 76° 14' 22.56'' E	60.97	Agricultural and Domestic
Patiala River (S2)	30° 20' 4.92'' N 76° 25' 33.96'' E	71.08	Agricultural, Domestic and Industrial
Dhakanshu Drain (S3)	30° 30' 39.8412'' N 76° 38' 33.8892'' E	9.90	Agricultural, Domestic and Industrial

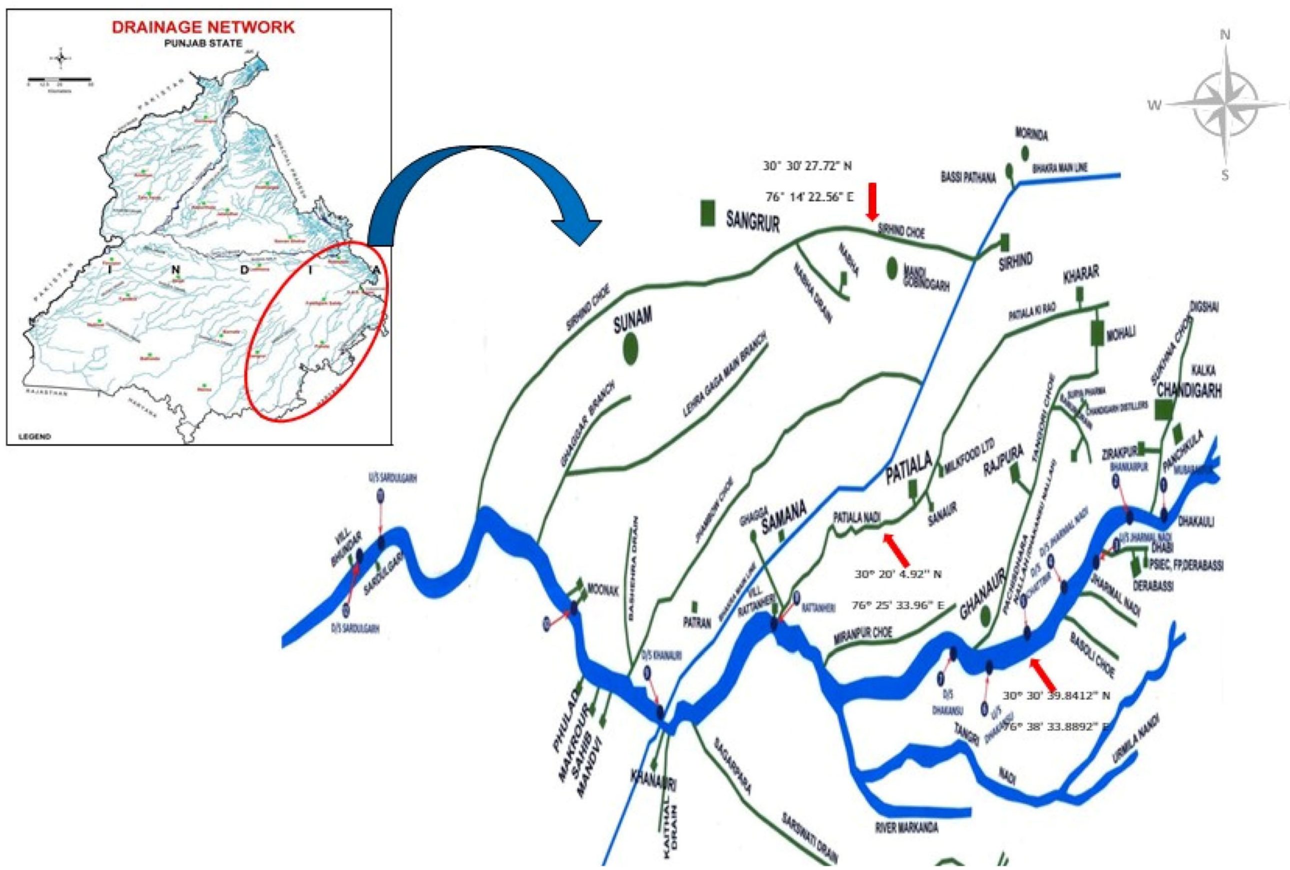

Figure 1. Drainage map of the study area. [Source: PPCB]

2.3 *Analytical methods*

In the present study, 15 physicochemical parameters viz., Temperature (Temp), pH, Chemical oxygen demand (COD), Biochemical oxygen demand (BOD), Total Total Alkalinity (TA), Acidity, Total Hardness (TH), Total dissolved solids (TDS), Total suspended solids (TSS), Sulfate (SO_4-S), Total Phosphorous (TP), Nitrite (NO_2-N), Nitrate (NO_3-N), Ammonical Nitrogen (AN) and Organic Nitrogen (ON) were analyzed by following standard methods given by APHA, 2012 for each parameter [15, 16]. pH of water samples was measured by using pH meter (Hanna, India) at the time of sampling. COD and BOD were measured by using "Open reflux method" and "Winkler's method", respectively. "EDTA Titric method" was used to determine "Total Hardness" of water samples. Alkalinity and acidity were measured by Titritic methods. Sulfate and Phosphate were determined by using ''Gravimetric Method'' and "Persulfate Digestion Method", respectively. Nitrate and Nitrite were assessed by Ultraviolet Spectrophotometric Screening Method. Total Kjeldahl Nitrogen method (TKN) was used to determine ammonical nitrogen and organic nitrogen.

All parametric values were compared with permissible limits for inland surface water prescribed by the Central Pollution Control Board (CPCB) and Bureau of Indian standards (BIS) to observe the general characteristics of the water chemistry of each site.

2.4 *Statistical analysis*

Descriptive data statistics and multivariate chemometric tools such as Pearson's correlation analysis, Cluster analysis (CA), Principal Component Analysis (PCA)/Factor Analysis (FA), and Discriminant Analysis (DA) were employed to analyze complex data set. The software SPSS-20 version was used for statistical computations.

2.4a *Correlation analysis*: To know the relationship between different parameters, Pearson's correlation analysis was performed. Pearson's correlation coefficient (r) values ranging from -1 to 1 were selected to measure the strength of the interrelationship between two parameters. Correlation coefficient values, $r>0.7$ indicate strong correlation, r = 0.5-0.7 (moderate correlation) and $r<0.5$ represent weak interactions between parameters [17].

2.4b *Principal Component Analysis (PCA)*: It is a technique to convert the original variables into new, uncorrelated variables, called the principal components (PCs), linear alignments of the actual variables. This technique converts all variables of the data set to a new coordinate system so that the most remarkable variances by any computation of the variables represented by the first coordinate (called the first principal component), the second most significant variance by the second coordinate, and so on [18]. Prior to perform PCA, data suitability was tested by applying Kaiser Criterion (Kaiser-Meyer-Olkin Test). The data of all three studied sites justified this criterion, therefore it proved that the collected data is acceptable to perform PCA [19]. The principal components (PCs) with eigenvalues of more than one were retained, and scree plots were used to select the significant PCs. The parameters with eigenvalues less than one were eliminated due to their low significance.

2.4c *Factor analysis (FA)*: Factor analysis is a statistical method, which further rejects the less important variables from components extracted by PCA. In FA, varimax rotation dragged a more influential group of variables known as varifactor (VF). The Varimax rotation is an orthogonal rotation of parameters and this rotation maximizes the variance of the loadings within the columns of the matrix. After varimax rotation the columns in the matrix are denoted as varifactors. The factor loadings in each varifactor were classified by following the criteria of Liu *et al* [20], i.e., strong, moderate, and weak loadings were corresponding to absolute values of >0.75, 0.75-0.50, and 0.50-0.30, respectively. Varifactors attained from factor analysis are used to identify the pollution contributors that significantly affect the water quality [21, 22].

2.4d *Cluster analysis*: Cluster analysis (CA) is a method to organize the variables into groups or clusters based on their proximity or resemblance in characteristics [23]. In the present study, Hierarchical agglomerative clustering mode was used, which is a common and widely used method and illustrates the output by dendrogram. This mode of clustering divides each parameter into its own individual cluster in the first step so that the initial number of clusters equals the total number of parameters. At consecutive steps, similar cases or clusters are combined together until every case is assembled into one single cluster. In final stage, significant clusters were selected on the basis of their coefficient values [24].

2.4e *Discriminant analysis (DA)*: Discriminant analysis is a method to predict the variables, differentiating two or more groups [21]. The analysis was carried out by creating discriminant functions (DFs) by following the equation given below:

$$f(G_i) = K_i + \sum_{j=1}^{n} W_{ij} + P_{ij} \quad (1)$$

Where i is the number of groups (G), k_i is the constant inherent for each group, n is the number of parameters used for the classification of a set of data into a specified group, W_{ij} is the weight coefficient allotted by DA to a given selected parameter (P_{ij}) [25].

3. Results and discussion

The physicochemical data of water samples collected from three drains (Sirhind Choe, Patiala River, and Dhakanshu Drain) of Ghaggar River was used to present the current research work. Five sampling campaigns were conducted during 2017-2018 to observe temporal variations. The analytical values were compared with CPCB and BIS standard values for inland surface water.

3.1 *Physicochemical analysis*

Seasonal values for physicochemical parameters of each site are summarized in tables 2, 3, and 4 electronic supplementary material. The annual descriptive summary is given in tables 2, 3, and 4.

The pH range of 6.4-8.6 was recorded with the annual mean of 7.8, 7.6, and 7.9 at S1, S2, and S3, respectively. It has been observed that increased pH values were found in the monsoon season at all sites. A probable reason for this could be the soil erosion and agricultural runoff that proliferate the concentration of carbonates and bicarbonates [26]. A maximum temperature of 36°C in the summer and 14°C in the winter were recorded. The annual mean temperature was 26, 27, and 28°C at S1, S2, and S3, respectively. The COD concentration found in the range 40-400 mg/L, and the annual mean values 81.6, 248.2, and 157.6 mg/L were observed at S1, S2, and S3, respectively. This study found COD concentration above the permissible limit at S2 (winter season, post-winter transitional season, and summer season) and S3 (post-winter transitional season). This increased concentration suggested high organic pollution at studied sites. BOD_5 levels were recorded from 5 to 67.5 mg/L with the annual average concentrations of 22.2, 31.4, and 28.7 mg/L at S1, S2, and S3, respectively. BOD was found above the permissible limit at S2 (summer season and post-winter transitional season) and S3 (winter season). The elevated BOD levels in the summer suggested that organic pollution load increases due to low water levels. Also, bacterial fauna flourishes at high temperatures and consumes more oxygen from the water to degrade organic compounds [27, 28]. The Total Alkalinity range in wastewater was recorded as 126-522 mg $CaCO_3$/L with

Table 2. Statistical summary of water quality in Sirhind Choe (Site- 1) in 2017-2018.

Parameters	Minimum	Maximum	Mean	Median	Std. Deviation	Standards (mg/l)
pH	6.8	8.3	7.8	8	0.6	5.5-9.0(CPCB)
Temperature (°C)	14	33	25.6	25	7.4	NA
COD	40	120	81.6	80	28.5	250 (CPCB)
BOD(5 Days 20°C)	12	48	22.2	16	15.1	30(CPCB)
Alkalinity	222	514	360	308	130.1	600(IS)
Acidity	26	368	131.6	78	138.1	NA
TH	104	360	264	256	105.7	300(BIS)
TDS	349	673	574.4	630	135.4	500(BIS)
TSS	7	153	74.2	50	64.2	100 (CPCB)
SO_4-S	49.4	185.3	89.9	62	55.5	400(CPCB)
TP	1.5	2.4	1.9	1.8	0.3	5 (CPCB)
NO_2-N	0.7	45.7	15.8	14.0	18.4	NA
NO_3-N	0.1	4.1	1.6	0.1	2.1	10 (CPCB)
AN	7.2	15.3	10.7	10.5	3.3	50 (CPCB)
ON	1.8	3.6	2.4	2.2	0.7	NA

All the standards for inland surface water have been adopted from The Environment (Protection Rules), 1986 (Rule number 3). The BIS standards are adopted from IS 10500- 2012.

Table 3. Statistical summary of water quality in Badi Nadi (Site- 2) in 2017-2018.

Parameters	Minimum	Maximum	Mean	Median	Std. Deviation	Standards (mg/l)
pH	6.4	8.5	7.6	7.8	0.8	5.5-9.0(CPCB)
Temperature (°C)	15	33	26.6	27	7.2	NA
COD	41	400	248.2	280	136.4	250 (CPCB)
BOD(5 Days 20°C)	5	67.5	31.4	22	26.4	30(CPCB)
Alkalinity	180	522	322.6	225	170.3	600(IS)
Acidity	14	500	165.2	88	192.8	NA
TH	14	294	190	186	110.6	300(BIS)
TDS	380	924	753.4	900	235.1	500(BIS)
TSS	80	320	158.6	130	93.5	100 (CPCB)
SO_4-S	99	227	163.7	140	59.4	400(CPCB)
TP	0.8	2.78	1.8	2.06	0.8	5 (CPCB)
NO_2-N	1.8	16.9	8.0	3.1	7.8	NA
NO_3-N	0.0	2.0	0.9	1.1	0.9	20 (CPCB)
AN	2.1	27	15.7	22.5	11.6	50 (CPCB)
ON	2.5	3.5	2.8	2.6	0.4	NA

All the standards for inland surface water have been adopted from The Environment (Protection Rules), 1986 (Rule number 3). The BIS standards are adopted from IS 10500- 2012.

annual mean values of 360, 322.6, and 218.2 mg $CaCO_3$/L at S1, S2, and S3, respectively. However, all the values were within the limit, but high Total Alkalinity values were found in the post-winter transitional and summer seasons at all sites. A decreasing trend was also observed in the monsoon season that might be due to the reduction in the ionic concentration of water in heavy rain. The acidity values in water samples ranged from 14 to 500 mg $CaCO_3$/L with the annual mean values of 131.6, 165.2, and 100.8 mg $CaCO_3$/L. Total hardness was recorded from 14 to 360 mg/L with an annual average concentration of 574.4, 190, and 108 mg/L at S1, S2, and S3. A high concentration of total hardness was found in the post-monsoon season and winter season at all sites, and these values were found above the permissible limit at S1. TDS values ranging from 340 to 924 mg/L with the annual mean of 574.4, 753.4, and 413.4 mg/L were observed at S1, S2, and S3, respectively. TDS levels at S1 and S2 were found above the permissible limits throughout the study. Whereas a significant reduction was observed in the monsoon season at all sites. It might be due to the reason that water gets diluted in heavy rain [29, 30]. TSS values were recorded from 7 to 450 mg/L, and the annual mean concentrations were found as 74.2, 158.6, and 232.4 mg/L at S1, S2, and S3, respectively. TSS

Table 4. Statistical summary of water quality in Dhakanshu Nallah (Site-3) in 2017-2018.

Parameters	Minimum	Maximum	Mean	Median	Std. Deviation	Standards (mg/l)
pH	6.9	8.4	7.9	8.1	0.6	5.5-9.0(CPCB)
Temperature ($^{\circ}$C)	17	36	28	28	7.0	NA
COD	40	360	157.6	120	122.1	250 (CPCB)
BOD(5 Days 20^{0}C)	5.4	66	28.7	25	23.2	30(CPCB)
Alkalinity	126	345	218.2	162	101.4	600(IS)
Acidity	20	326	100.8	54	126.8	NA
TH	26	194	108	98	81.7	300(BIS)
TDS	340	464	413.4	410	51.3	500(BIS)
TSS	116	450	232.4	177	142.3	100 (CPCB)
SO_4-S	62	156.4	91.5	70	38.9	400(CPCB)
TP	1.4	2.8	2.0	1.9	0.5	5 (CPCB)
NO_2-N	1.2	45.3	17.5	17.3	18.2	NA
NO_3-N	0.0	4.6	1.1	0.4	2.0	20 (CPCB)
AN	8.2	23	19.4	22.3	6.3	50 (CPCB)
ON	2.7	3	2.9	2.9	0.2	NA

All the standards for inland surface water have been adopted from The Environment (Protection Rules), 1986 (Rule number 3). The BIS standards are adopted from IS 10500- 2012.

concentration was found above the permissible limit (excluding winter, post-winter transitional season, and summer season at S1 and winter at S2). Increased levels of TSS in the monsoon season were recorded, which might be due to the addition of the eroded soil from prepared agricultural fields. Also, falling off debris due to the high-velocity wind and rainfall could have contributed particles to the running water. The increased water level and its rapid flow in the monsoon season may have re-suspended the sediments settled on the bottom [31]. The sulfate concentration was recorded in the range of 49.4-227 mg/L, which was found within the limit. The average values found were 89.9, 163.7, and 91.5 mg/L at S1, S2, and S3, respectively, and a high sulfate concentration was recorded in the monsoon season. Its probable reason might be the application of gypsum in paddy fields that come with surface runoff into the water body. A high sulfate concentration at S2 also suggested the enormous discharge of domestic sewage, municipal waste, and agricultural runoff [32]. Throughout the study, total phosphorus was recorded within the limit, and its concentration ranged from 0.8 to 2.8 mg/L. The annual mean concentrations of total phosphorus were observed as 1.9, 1.8, and 2.0 mg/L at S1, S2, and S3, respectively. These high total phosphorus values revealed the discharge of agricultural runoff and fertilizers [11]. The amount of nitrite ranged from 0.7 to 45.7 mg/L was recorded in this study, and annual mean concentrations were found as 15.8, 8.0, and 17.5 mg/L at S1, S2, and S3, respectively. A sudden increase in nitrite levels in the summer season was observed at S1 and S3. In the summer, these elevated nitrite levels suggested that high organic matter and nitrate have been reduced to nitrite [33]. A very high concentration may be due to high amounts of free ammonia and solid organic waste [34]. The nitrate concentration was observed as 0.0 to 4.6 mg/L, and the annual mean concentrations found 1.6, 0.9, and 1.1 mg/L at S1, S2, and S3, respectively. The nitrate sources in studied sites can be domestic or municipal sewage [35]. Another reason for this could be the uncontrolled use of fertilizers and pesticides for paddy crops [36]. The maximum and minimum concentration of ammonical nitrogen was recorded as 27 mg/L and 2.12 mg/L, respectively, and the annual mean values of ammonical nitrogen were recorded as 10.7, 15.7, and 19.4 mg/L at S1, S2, and S3, respectively. The presence of ammonia in the wastewater suggested the source of domestic pollution [31]. The average concentrations of organic nitrogen were recorded as 2.4, 2.8, and 2.9 mg/L at S1, S2, and S3, respectively, and the maximum and minimum values of organic nitrogen were observed as 3.58 mg/L and 1.84 mg/L.

The results of present study were in accordance with the other studies such as Pareek *et al* and Pareek *et al* [37, 38] observed similar water quality patterns from the Ghaggar River. Ghai *et al* [39] observed high values of BOD in the Ghaggar River. Sharma *et al* [11] found comparable findings and shown the deteriorated water quality at river Sutlej, Punjab (India). Sharma and Walia [40] showed similar water quality patterns in river Sutlej at the confluence point of Budha nallah, the most polluted wastewater drain of Punjab. Soni *et al* [41] revealed that pollution in river Sutlej was due to the discharge of wastewater drain into the river. High concentrations of BOD and phosphate were also recorded from river Beas and Sutlej [10]. Kumar *et al* [9] reported agricultural, industrial pollution in the river Beas. Singh *et al* [42] observed similar findings from water samples of Budha Nallah in Punjab. Chauhan and Sagar [8] analyzed water samples of river Beas and reported a high concentration of BOD, TDS, and phosphate.

Parmar and Bhardwaj [6] reported high pollution loads in Harike Lake at the confluence point of rivers Sutlej and Beas. Kundu *et al* [7] observed significant Spatio-temporal variations in the Ghaggar River. Matta *et al* and Matta *et al* [43, 44] suggested that the water quality of the Ganga River is underprivileged due to agricultural and domestic pollution.

3.2 *Cluster analysis*

Temporal cluster analysis was applied to detect the homogeneity between seasons. The dendrogram obtained by Ward's Method is given in figure 2. Based on agglomeration coefficient values, four significant clusters were selected [45]. The dendrogram shows that cluster 1 and cluster 3 included parametric values of the monsoon season and the post-monsoon season, respectively. Cluster 2 and cluster 4 had the winter and summer season cases, respectively, and both clusters also included the post-winter transitional season. Therefore, findings revealed that temporal patterns to water quality were consistent with four seasons only (post-monsoon season, winter season, summer season, and monsoon season). These results indicated that the CA is a helpful method for the reliable classification of waters. It can be beneficial to make ideal strategies for upcoming seasonal sampling plans by reducing the sampling frequency [46]. It has been suggested that sampling for four seasons in one year is sufficient for rapid water quality assessment. Many researchers have used Hierarchical agglomerative cluster analysis (CA) for spatial variations [47–53]. It was also observed that temporal CA is not a common practice. However, it is a very effective method to reduce sampling frequency and associated expenditures [54–56].

3.3 *Spatial variability*

Spatial variations in water quality parameters were evaluated through correlation matrix, principal component analysis, and factor analysis.

3.3a *Correlation analysis*: In the correlation matrices, several parameters revealed strong interactions with each other. Many of them exhibited positive interactions, and some of the parameters displayed negative interactions. The correlation matrix of each site is given in (tables 8, 9 and 10 of electronic supplementary material). The pH showed a high negative correlation with acidity in all three studied areas. Total hardness exhibited a strong correlation with pH and acidity at S1 and S2. There was a strong correlation of COD with TDS, TSS, sulfate, phosphate, and organic nitrogen that revealed biodegradable and non-biodegradable organic pollution was there at studied sites [57].

On the other hand, BOD and COD have not exhibited a significant strong correlation, which proved the presence of non-biodegradable organic sources [51, 58]. Total phosphorus correlated reasonably well with TSS, COD, and TDS suggesting its anthropogenic and natural origin [51]. Nitrate did not show a significant correlation with other parameters that reflected agricultural and domestic sewage sources [59]. In the present study, a significant strong correlation between several parameters remarked about their common origin, mainly from industrial, agrarian, and municipal sources [60, 61].

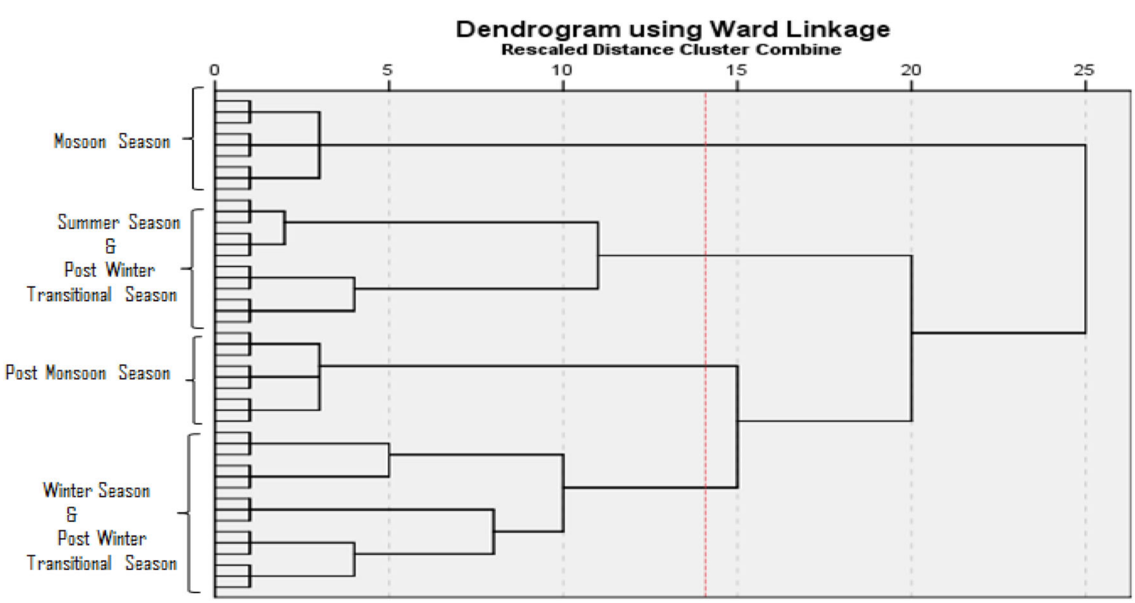

Figure 2. Dendrogram showing temporal analysis of water quality monitoring.

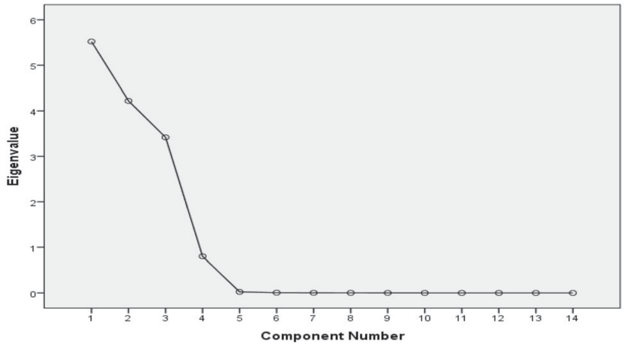

Figure 3. Principal component analysis for Site 1 (Sirhind Choe).

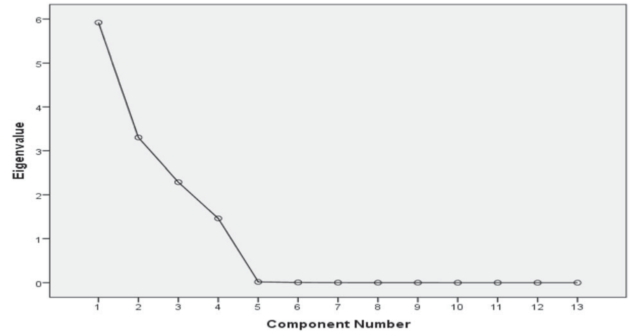

Figure 4. Principal component analysis for Site 2 (Badi Nadi).

3.3b *Identification of potential pollution sources*: The Principal Component Analysis/Factor Analysis was applied to identify pollution sources at studied sites.

Sirhind Choe (S1)

The KMO (Kaiser-Meyer-Olkin Test) value for this site was 0.659, and PCA generated three principal components (PCs) that explained 94% of the total variance. The graphical view of extracted PCs is given in the scree plot (figure 3). As mentioned above, parameters with the factor loadings value> 0.75 were selected as the most influencing variables (table 5).

The first varifactor (VF1) explained 39.4% of the total variance, showed strong positive loading on temperature and strong negative loading on TDS, TP, and NO_3-N. This factor represented nonpoint sources such as surface runoff from agricultural areas with nitrate and phosphate-based fertilizers [51, 62]. The second varifactor (VF2) accounted for 30.15% of total variance has strong positive loading on pH and strong negative loading on COD and acidity. It could be attributed to the point sources: domestic wastewater, industrial effluents, and discharge from wastewater treatment plants [52]. The third varifactor (VF3) explained 24.42% of total variance showed strong positive loading on BOD, Total Alkalinity, NO_2-N, and AN. It represented the point sources such as municipal sewage and industrial wastewater discharge. High loading of NO_2-N and AN could be contributed by surface runoff with nitrogenous-based fertilizers from agricultural lands [63, 64].

Table 5. Loadings of Variables on Varifactors for Site 1 (Sirhind Choe).

	Varifactors		
Parameters	VF1	VF2	VF3
pH	.109	**.963**	.202
Temperature	**.827**	.244	.449
COD	− .419	**− .891**	.159
BOD	.308	− .169	**.811**
TA	− .101	.399	**.837**
Acidity	− .004	**− .989**	− .123
TDS	**− .888**	− .420	.150
TSS	.659	− .692	− .252
SO_4-S	.589	.481	− .649
TP	**− .866**	− .069	− .191
NO_2-N	.204	.388	**.865**
NO_3-N	**− .886**	.335	− .208
AN	− .020	− .041	**.996**
ON	.728	.486	− .452
Eigen values	5.522	4.221	3.420
% Total Variance	39.441	30.150	24.428
% Cumulative Variance	39.441	69.590	94.018

Bold values indicate strong loadings (>0.75).

Patiala River (S2)

The KMO value observed for this site was 0.694. The PCA extracted four significant principal components (PCs) that explained 99.792% of the total variance. The graphical representation of extracted PCs is given in the scree plot (figure 4). The factor-loading matrix for S2 is given in table 6.

The first varifactor (VF1) explained 45.544% of the total variance, showed strong positive loading on COD, TDS, and TP, and strong negative loading on TSS. This factor illustrated the nonpoint sources such as surface runoff from the agricultural area with high solids and phosphate-based fertilizers. The point sources such as domestic and industrial effluents were also recognized by this varifactor [53]. The second varifactor (VF2) accounted for 25.404% of total variance has strong positive loading on pH and NO_3-N and strong negative loading on acidity. This factor was associated with pollution caused by agrochemical application and the decay of natural organic matter [65]. The third varifactor (VF3) explained 17.606% of the total variance and has strong positive loading on BOD and ON. Organic pollution caused by untreated domestic wastewater was identified by this factor [66]. The fourth varifactor (VF4) explained 11.239% of the total variance and has strong positive loading on SO_4-S. This factor is attributed to pollution sources such a soil leaching and surface runoff

Table 6. Loadings of Variables on Varifactors for Site 2 (Badi Nadi).

	Varifactors			
Parameters	VF1	VF2	VF3	VF4
pH	− .185	**.901**	.206	.327
Temperature	− .667	.044	.639	− .378
COD	**.936**	.077	.031	− .342
BOD	.477	.312	**.773**	− .270
Acidity	− .013	**− .912**	− .341	− .225
TH	.684	− .694	− .216	− .053
TDS	**.950**	− .061	.276	− .136
TSS	**− .886**	.459	− .066	.009
SO_4-S	− .246	.247	− .216	**.912**
TP	**.981**	− .128	.121	.081
NO_3-N	.028	**.887**	− .438	− .133
AN	.409	− .736	.136	.520
ON	.117	− .021	**.988**	− .001
Eigen values	5.921	3.303	2.289	1.461
% Total Variance	45.544	25.404	17.606	11.239
% Cumulative Variance	45.544	70.948	88.554	99.792

Bold values indicate strong loadings (>0.75)

Table 7. Loadings of Variables on Varifactors for Site 3 (Dhakansu Nallah).

	Varifactors			
Parameters	VF1	VF2	VF3	VF4
pH	− .252	.099	**.937**	− .188
Temperature	− .239	**.930**	− .278	.009
BOD	.670	.495	.037	.552
TA	.460	**.819**	.026	− .342
Acidity	.120	− .270	**− .937**	.182
TH	.480	− .566	− .356	.566
TDS	**.918**	.066	.263	.287
TSS	**− .994**	.070	.046	.065
SO_4-S	**− .838**	− .121	.506	.163
TP	.221	− .447	**.862**	.082
NO_2-N	− .001	**.958**	.261	.116
NO_3-N	.137	.040	.167	**− .975**
AN	**.939**	− .218	− .222	− .137
Eigen values	4.871	3.488	2.874	1.740
% Total Variance	37.473	26.832	22.104	13.386
% Cumulative Variance	37.473	64.305	86.409	99.795

Bold values indicate strong loadings (>0.75)

from the agricultural area because ammonium sulfate fertilizers are commonly used in crop fields [67].

Dhakanshu Drain (S3)

The KMO value observed for this site was 0.527. The PCA reduced whole data into four significant principal components (PCs) that explained 99.795% of the total variance. The Scree plot (figure 5) is presented for visualization of extracted PCs**.** The factor-loading matrix is given in table 7.

The first varifactor (VF1) explained 37.473% of the total variance, has strong positive loading on TDS and AN, and strong negative loading on TSS and SO_4-S. These loading indicated that pollution sources were due to extensive agricultural activities and domestic discharge in the studied area [68, 69]. The second varifactor (VF2) explained 26.832% of the total variance and showed strong positive loading on temperature, Total Alkalinity, and NO_2-N. This

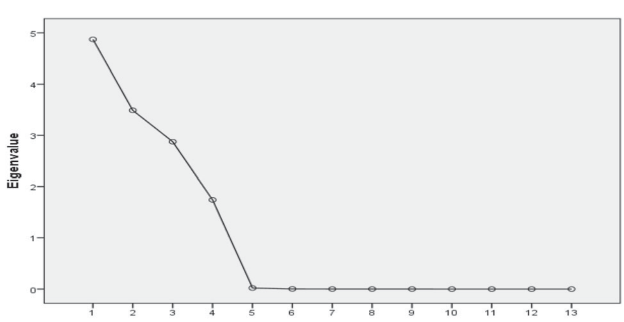

Figure 5. Principal component analysis for Site 3 (Dhakanshu Nallah).

Table 8. Results of discriminant analysis for spatial variations.

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	Df	Sig.
1 through 2	.000	296.979	24	.000
2	.033	124.427	11	.000

Table 9. Canonical Discriminant Function Coefficients.

	Function	
	1	2
pH	2.827	1.584
Temperature	**9.142**	− .400
COD	3.486	1.205
TA	3.887	− .182
Acidity	− 3.685	1.177
TH	**20.547**	.475
TDS	**− 10.109**	**3.167**
SO_4-S	**5.699**	1.719
TP	1.127	− 3.346
NO_2-N	− .724	.405
NO_3-N	3.252	− .607
AN	− 2.382	− .022

factor recognized the pollution sources such as agricultural runoff, poorly degraded domestic wastewater, and physiochemical sources of variability [52]. The third varifactor (VF3) accounted for 22.104% of the total variance. It

Table 10. Classification Function Coefficients.

	Site		
	Site 1	Site 2	Site 3
pH	1367.771	1308.035	1249.253
Temperature	**208.738**	183.497	176.096
COD	5.632	5.104	4.744
TA	1.114	.547	.382
Acidity	− 1.316	− .780	− .729
TH	**28.319**	24.358	22.975
TDS	**− 6.723**	− 5.294	− 5.153
SO_4-S	**12.840**	11.020	9.873
TP	85.681	7.727	62.098
NO_2-N	− 5.576	− 4.419	− 4.468
NO_3-N	258.952	218.637	211.199
AN	− 39.197	− 33.274	− 31.276
(Constant)	− 10740.622	− 9142.497	− 8230.378

exhibited a strong positive loading on pH and TP and strong negative loading on acidity. This factor represented the influence of agricultural pollution caused by the extensive use of phosphate-based fertilizers [70]. The fourth vari-factor (VF4), explained 13.386% of the total variance, has strong negative loading on NO_3-N. This factor identified nonpoint sources such as agricultural runoff with nitrogen-based fertilizer and untreated domestic wastewater [58, 71].

3.3c *Spatial discriminant analysis*: The present study performed the spatial DA using step-wise modes on the raw data set consisting of 15 parameters. The two significant discriminant functions (DFs) were retained based on eigenvalues, and these discriminant functions DF1 and DF2 explained 79.3% and 20.7% of the total variance, respectively. The canonical correlation of both functions revealed that 99% of the total variance was explained in this analysis. The standard mode included all 15 parameters, and the forward step-wise and backward step-wise methods

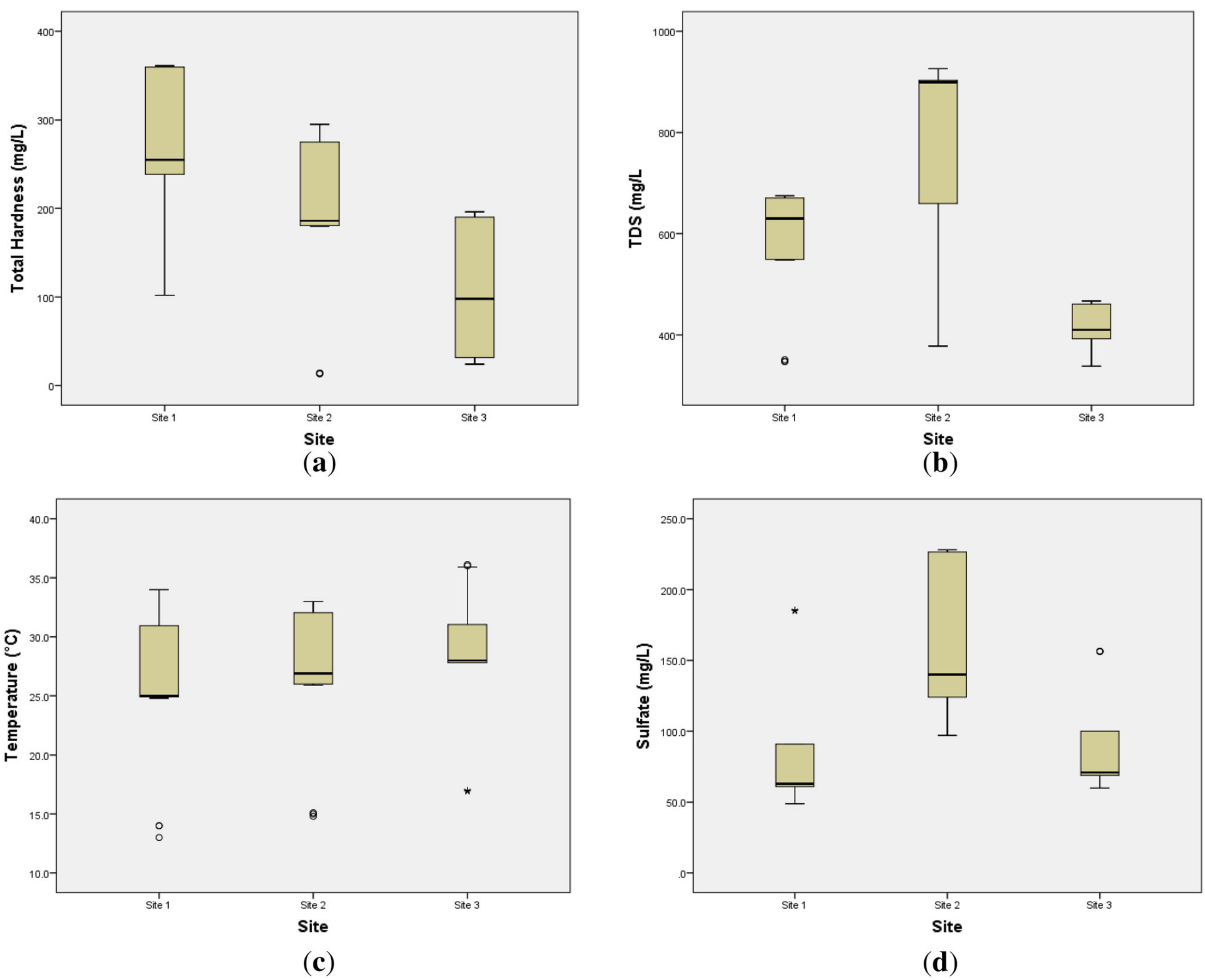

Figure 6. Box and Whisker of major spatial water quality discriminant Parameters.

included only 12 parameters for analysis. The values of Wilk's lambda and the Chi-square for each discriminant function (DF) are shown in table 8. The values of Wilks' lambda is a amount of how well each function splits cases into groups and smaller values indicate greater discriminatory capacity of the function. The associated Chi-square values hypothesis that means of the functions are equal throughout the groups and smaller values of its significance level indicate that discriminant functions does separate the groups in a better way. Therefore, the obtained values of Wilk's lambda and the Chi-square described that the analysis is capable of explaining discrimination between groups.

The results of spatial DA (backward step-wise mode) predicted TDS, TP, SO_4-S, COD, AN, TH, Total Alkalinity, NO_3-N, temperature, Acidity, pH, and NO_2-N as discriminant parameters. The most important discriminant parameters of spatial variations were selected by their standardized canonical discriminant function coefficients (table 9). Discriminant function coefficients revealed that TH was the primary discriminant parameter followed by TDS, temperature, and SO_4-S. Further, classification functions (table 10) showed variations attributed to site-1 (Sirhind Choe).

Therefore, we can conclude that the Sirhind choe drain (S1) showed more discrimination than the other two drains. Box and whisker plots for identified parameters by spatial DA are given in figure 3. The results (figure 6a) showed the highest concentration of TH at S1 that revealed high inorganic pollution at this site [72]. The very high concentration of TDS (figure 6b) was found at S2 revealed inorganic pollution that might be due to gypsum soil from agricultural fields, industrial discharge, and domestic waste that deposit minerals at this site [49, 55]. However, great fluctuations were observed at S1 for TDS. Sulfate concentration was very high at S2, suggesting agricultural and domestic pollution at this site [26]. On the other hand, at S1, again noticeable variations were observed (figure 6d).

4. Conclusion and significance of the study

In the current research work, physicochemical analysis on the drainage system was conducted by including three major drains of the Ghaggar river (seasonal river) Punjab, India. The seasonal data of 2017-2018 revealed that parameters such as COD, BOD, TH, TDS, and TSS were above the permissible limits of CPCB. In correlation analysis, physicochemical parameters at each site were highly correlated, suggesting their common origin. The cluster analysis revealed the similarities of the post-winter transitional season with the summer and winter seasons. Hence sample frequency was reduced to four seasons instead of five seasons. The PCA extracted significant components for S1, S2, and S3 that explained 94%, 99.792%, and 99.795% of the total variance. Further, all the varifactors allotted in FA showed that studied areas were highly affected by agricultural and domestic pollution. The DA displayed the best reduction in the data set as it recognized only four major discriminant parameters (out of a total of fifteen parameters) that were responsible for spatial variations. Overall, this study concluded that water was highly polluted as many parameters were above the permissible limit. The selected sites were affected by agricultural and domestic pollution, and it is necessary to execute the plans to prevent water quality degradation. Multivariate statistical analysis used in the current study is a widely used, reliable, and efficient technique for environmental data. This study will be highly beneficial for local government in making water protection policies and sustainable quality management. Moreover, a validated identification of primary pollution sources from the studied area was achieved. It can be helpful in deciding the upcoming sampling plans for further research. This study suggested that water should be treated before these drains join the Ghaggar River as these drains are a major cause of river pollution. Also, mismanaged farming events should be controlled around the studied area, and awareness programs must be launched to let the farmers learn about the eco-friendly approaches for agriculture.

Author contributions All authors contributed to the study's conception and design. Material preparation, data collection, and analysis were performed by Harneet Kaur. The first draft of the manuscript was written by Harneet Kaur. Supervision, review and editing were made by Anita Rajor and Amritpal Singh Kaleka. All the authors read and approved the final manuscript.

Funding No funding was received for conducting this study.

Declarations

Conflict of interest The authors declare that there are no conflict of interest.

References

[1] Dahunsi S O, Owamah H I, Ayandiran T A and Oranusi U S 2014 Drinking water quality and public health of selected communities in South-Western Nigeria. *Water Qual. Expo. Health* 6: 143–153

[2] Chandra R, Singh S and Raj A 2006 Seasonal bacteriological analysis of Gola River water contaminated with pulp paper mill waste in Uttaranchal India. *Environ. Monit. Assess.* 118: 393–406. https://doi.org/10.1007/s10661-006-1508-4

[3] Narayanan N 2015 *Horrifying fact: Almost all India's water is contaminated by sewage*. Scroll.in. https://scroll.in/article/print/737981

[4] Tiwana N S, Jerath N, Ladhar S S, Singh G, Paul, R, Dua D K and Parwana H K 2007 State of Environment; Punjab-2007. Punjab State Council for Science & Technology, Punjab, India

[5] Central Ground Water Board (CPCB) 2014 Water Quality Issues and Challenges in Punjab. http://cgwb.gov.in/WQ/Punjab%20Book%20Final%20for%20Printing.pdf
[6] Parmar K S and Bhardwaj R 2012 Water quality index and fractal dimension analysis of water parameters. *Int. J. Environ. Sci. Technol.* 10: 151–164
[7] Kundu S, Gill G S, Kishore N, Singh R R and Kumar A 2012 Assessment of Spatio-temporal variations in surface water quality of Ghaggar River (North-Western, India) utilized for drinking and agricultural purposes. *Afr. J. Biotechnol.* 11(79): 14428–14432
[8] Chauhan B S and Sagar S K 2013 Impact of Pollutants on Water Quality of River Sutlej in Nangal Area of Punjab, India. *Biol. Forum.* 5(1): 113–123
[9] Kumar V, Sharma A, Chawla A, Bhardwaj R and Thukral A K 2016 Water quality assessment of river Beas, India, using multivariate and remote sensing techniques. *Environ. Monit. Assess.* 188: 137. https://doi.org/10.1007/s10661-016-5141-6
[10] Najar G N, Bhardwaj A and Pandey P 2017 A spatio-temporal water quality assessment of the Beas and Sutlej Rivers at the Harike Wetland: A Ramsar site in Punjab, India. *Lakes Reserv. Res. Manag.* 22: 364–376. https://doi.org/10.1111/lre.12193
[11] Sharma C, Jindal R, Singh U B and Ahluwalia A S 2018 Assessment of water quality of river Sutlej, Punjab (India). *Sustain. Water Resour. Manag.* 4: 809–822. https://doi.org/10.1007/s40899-017-0173-9
[12] Thakur J S, Prinja S, Singh D, Rajwanshi A, Arora S, Prasad R, Parwana H and Kumar R 2015 Genotoxicity And Adverse Human Health Outcomes Among People Living Near Highly Polluted Waste Water Drains In Punjab, India. *World J. Pharm. Res.* 4(1): 895–908
[13] Kaushik A, Sharma H R, Jain S, Dawra J and Kaushik C P 2010 Pesticide pollution of River Ghaggar in Haryana. *India. Environ. Monit. Assess.* 160: 61–69. https://doi.org/10.1007/s10661-008-0657-z
[14] The Economic and Statistical Organization, Punjab (ESOPB) 2012 Environment Statistics of Punjab. Chandigarh, India. https://www.esopb.gov.in/#gsc.tab=0
[15] American Public Health Association (APHA) 2012 Standard methods for examination of water and wastewater. 22nd edn. DC, United States, Washington
[16] Rajor A 2006 Experimental Methods for General and Environmental chemistry. Standard Publishers Distributors, Delhi, India
[17] Helena B, Pardo R, Vega M, Barrado E, Fernandez J M and Fernandez L 2000 Temporal evolution of groundwater composition in alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.* 34: 807–816. https://doi.org/10.1016/S0043-1354(99)00225-0
[18] Rencher A C 2002 Methods of Multivariate Analysis (Second Edition). John Wiley, Hoboken. https://www.ipen.br/biblioteca/slr/cel/0241
[19] Kaiser H F and Rice J 1974 Little Jiffy, Mark IV. *Educational and Psychological Measurement* 34: 111–117
[20] Liu C W, Lin K H and Kuo Y M 2003 Application of factor analysis in the assessment of groundwater quality in a Blackfoot disease area in Taiwan. *Sci. Total Environ.* 313: 77–89. https://doi.org/10.1016/S0048-9697(02)00683-6
[21] Singh K P, Malik A, Mohan D and Sinha S 2004 Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study. *Water Res.* 38(18): 3980–3992. https://doi.org/10.1016/j.watres.2004.06.011
[22] Sarbu C and Pop H F 2005 Principal component analysis versus fuzzy principal component analysis. A case study: the quality of Danube water (1985–1996). *Talanta* 65: 1215–1220. https://doi.org/10.1016/j.talanta.2004.08.047
[23] Vega M, Pardo R, Barrado E and Deban L 1998 Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res.* 32: 3581–3592. https://doi.org/10.1016/S0043-1354(98)00138-9
[24] McKenna J E 2003 An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environ. Model. Softw.* 18: 205–220. https://doi.org/10.1016/S1364-8152(02)00094-4
[25] Wunderlin D A, Diaz M P, Ame M V, Pesce S F, Hued A C and Bistoni M A 2001 Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: suquia river basin (Cordoba, Argentina). *Water Res.* 35: 2881–2894. https://doi.org/10.1016/S0043-1354(00)00592-3
[26] Jordan T E, Whigham D F, Hofmockel K H and Pittek M A 2003 Nutrient and sediment removal by a restored wetland receiving agricultural runoff. *J. Environ. Qual.* 32(4): 1534–1547. https://doi.org/10.2134/jeq2003.1534
[27] Sanap R R, Mohite A K, Pingle S D and Gunale V R 2006 Evaluation of water qualities of Godawari River with reference to physico-chemical parameters, district Nasik (MS), India. *Pollut. Res.* 25(4): 775–778
[28] Haque M A, Jewel M A S and Sultana M P 2019 Assessment of physicochemical and bacteriological parameters in surface water of Padma River, Bangladesh. *Appl. Water Sci.* 9: 10. https://doi.org/10.1007/s13201-018-0885-5
[29] Jindal R and Sharma C 2011 Studies on water of Sutlej River around Ludhiana with reference to physiochemical parameters. *Environ. Monit. Assess.* 174: 417–425. https://doi.org/10.1007/s10661-010-1466-8
[30] Singh U B, Thakur R K, Verma R and Ahluwalia A S 2013 Nutritional studies of *Chara corallina*. *Recent Res. Sci. Technol.* 5(3): 10–14
[31] Ojok W, Wasswa J and Ntambi E 2017 Assessment of Seasonal Variation in Water Quality in River Rwizi Using Multivariate Statistical Techniques, Mbarara Municipality, Uganda. *Water Resour. Prot.* 9: 83–97. https://doi.org/10.4236/jwarp.2017.91007
[32] Bagga K K 2018 Study and Analysis of Pollution in Khan River. *International Journal of Advance Engineering and Research Development* 5(8): 30–34
[33] Arauzo M, Valladolid M, Valladolid M and Martínez-Bastida J J 2011 Spatio-temporal dynamics of nitrogen in river-alluvial aquifer systems affected by diffuse pollution from agricultural sources: Implications for the implementation of the Nitrates Directive. *J. Hydrol.* 411(1): 155–168. https://doi.org/10.1016/j.jhydrol.2011.10.004
[34] Alleman J E 1984 Elevated Nitrite Occurrence in Biological Wastewater Treatment Systems. *Water Sci. Technol.* 17: 409–419
[35] Matta G, Srivastava S, Pandey R R and Saini K K 2017 Assessment of physicochemical characteristics of Ganga Canal water quality in Uttarakhand. *Environ Dev. Sustain.* 19: 419–431. https://doi.org/10.1007/s10668-015-9735-x

[36] Voudouris K, Panagopoulos A and Koumantakis J 2000 Multivariate Statistical Analysis in the Assessment of Hydrochemistry of the Northern Korinthia Prefecture Alluvial Aquifer System (Peloponnese, Greece). *Nat. Resour. Res.* 9(2): 135–146. https://doi.org/10.1023/A:1010195410646

[37] Pareek R K, Khan A S, Srivastva P and Roy S 2018 Assessment of Physicochemical Properties of Ghaggar River. *J. Biol. Chem. Chron.* 4(1): 19–25

[38] Pareek R K, Khan A S, Srivastva P and Roy S 2019 Water Pollution Assessment of Ghaggar River. *International Journal of Technical Innovation in Modern Engineering & Science* 5(3): 1–8

[39] Ghai D, Jatayan M and Kumar S 2018 Water Quality Study of Ghaggar River In Zirakpur Region, Mohali District, Punjab. *International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)* 4(5): 1315–1324

[40] Sharma N and Walia Y K 2017 A Study to Assess the Water Quality of Satluj River at Wajipur Kalan, Ludhiana, Punjab. *J. Adv. Basic Sci.* 5(1): 08–13

[41] Soni D K, Mishra P K, Mishra R K and Pandey P 2017 Assessment of Water Quality of River Sutlej in Himachal Pradesh and Punjab (India) Using Statistical Tool. *J. Appl. Chem.* 6(5): 792–798

[42] Singh G, Singh, D D and Sharma S K 2013 Effect of Polluted Surface Water on Groundwater: A Case Study of Budha Nullah. *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)* 5(5): 01-08

[43] Matta G, Kumar A, Nayak A, Kumar P, Kumar A and Tiwari A K 2020 Determination of water quality of Ganga River System in Himalayan region, referencing indexing techniques. *Arab. J. Geosci.* 13: 1027. https://doi.org/10.1007/s12517-020-05999-z

[44] Matta G, Naik P K, Machell J, Kumar A, Gjyli L, Tiwari A K and Kumar A 2018 Comparative study on seasonal variation in hydro-chemical parameters of Ganga River water using comprehensive pollution index (CPI) at Rishikesh (Uttarakhand) India. *Desalination Water Treat.* 118: 87–95. https://doi.org/10.5004/dwt.2018.22487

[45] Yim O and Ramdeen K T 2015 Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. *Quant. Meth. Psych.* 11(1): 8–21. https://doi.org/10.20982/tqmp.11.1.p008

[46] Shrestha S and Kazama F 2007 Assessment of Surface Water Quality Using Multivariate Statistical Techniques: A Case Study of the Fuji River Basin, Japan. *Environ. Model. Softw.* 22: 464–475. https://doi.org/10.1016/j.envsoft.2006.02.001

[47] Taoufik G, Khouni I and Ghrabi A 2017 Assessment of physico-chemical and microbiological surface water quality using multivariate statistical techniques: a case study of the Wadi El-Bey River, Tunisia. *Arab. J. Geosci.* 10(7): 181. https://doi.org/10.1007/s12517-017-2898-z

[48] Martinez-Tavera E, Rodriguez-Espinosa P F, Shruti V C, Sujitha S B, Morales-Garcia S S and Munoz-Sevilla N P 2017 Monitoring the seasonal dynamics of physicochemical parameters from Atoyac River basin (Puebla), Central Mexico: multivariate approach. *Environ. Earth Sci.* 76: 95. https://doi.org/10.1007/s12665-017-6406-2

[49] Chow M F, Shiah F K, Lai C C, Kuo H Y, Wang K W, Lin C H, Chen T Y, Kobayashi Y and Ko C Y 2016 Evaluation of surface water quality using multivariate statistical techniques: a case study of Fei-Tsui Reservoir basin, Taiwan. *Environ. Earth Sci.* 75: 6. https://doi.org/10.1007/s12665-015-4922-5

[50] Hamzah F M, Jaafar O, Jani W N F A and Abdullah S M S 2016 Multivariate Analysis of Physical and Chemical Parameters of Marine Water Quality in the Straits of Johor, Malaysia. *J. Environ. Sci. Technol.* 9: 427–436. https://doi.org/10.3923/jest.2016.427.436

[51] Barakat A, El Baghdadi M, Rais J, Aghezzaf B and Slassi M 2016 Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. *Int. Soil. Water Conserv. Res.* 4: 284–292. https://doi.org/10.1016/j.iswcr.2016.11.002

[52] Phung D, Huang C, Rutherford S, Dwirahmadi F, Chu C, Wang X, Nguyen M, Nguyen N H, Do C M, Nguyen T H and Dinh T A D 2015 Temporal and spatial assessment of river surface water quality using multivariate statistical techniques: a study in Can Tho City, a Mekong Delta area, Vietnam. *Environ. Monit. Assess.* 187: 229. https://doi.org/10.1007/s10661-015-4474-x

[53] Zhang Y, Guo F, Meng W and Wang X Q 2009 Water quality assessment and source identification of Dalia river basin using multivariate statistical methods. *Environ. Monit. Assess.* 152: 105–121. https://doi.org/10.1007/s10661-008-0300-z

[54] Zhao Z W and Cui F Y 2009 Multivariate statistical analysis for the surface water quality of the Luan River, China. *J. Zhejiang Univ. Sci.* 10(1): 142–148. https://doi.org/10.1631/jzus.A0850094

[55] Ogwueleka T C 2015 Use of multivariate statistical techniques for the evaluation of temporal and spatial variations in water quality of the Kaduna River, Nigeria. *Environ. Monit. Assess.* 187: 137. https://doi.org/10.1007/s10661-015-4354-4

[56] Anteneh Y, Zeleke G and Gebremariam E 2018 Assessment of surface water quality in Legedadie and Dire catchments, Central Ethiopia, using multivariate statistical analysis. *Acta. Ecologica. Sinica.* 38: 81–95. https://doi.org/10.1016/j.chnaes.2017.05.005

[57] Reynolds D M and Ahmad S R 1997 Rapid and direct determination of waste water BOD values using a fluorescence technique. *Water Res.* 31(8): 2012–2018. https://doi.org/10.1016/S0043-1354(97)00015-8

[58] Vieira J S, Pires J C M, Martins F G, Vilar V J P, Boaventura R A R and Botelho C M S 2012 Surface Water Quality Assessment of Lis River Using Multivariate Statistical Methods. *Wat. Air and Soil Poll.* 223: 5549–5561. https://doi.org/10.1007/s11270-012-1267-5

[59] Helena B A, Vega M, Barrado E, Pardo R and Fernández L 1999 A Case of Hydrochemical Characterization of an Alluvial Aquifer Influenced By Human Activities. *Wat. Air and Soil Poll.* 112: 365–387

[60] Alkarkhi A F M, Ahmad A and Easa A M 2009 Assessment of surface water quality of selected estuaries of Malaysia: multivariate statistical techniques. *Environmentalist* 29: 255–262. https://doi.org/10.1007/s10669-008-9190-4

[61] Bhuiyan M A H, Rakib M A, Dampare S B, Ganyaglo S and Suzuki S 2011 Surface Water Quality Assessment in the Central Part of Bangladesh Using Multivariate Analysis. *KSCE J. Civ. Eng.* 15(6): 995–1003. https://doi.org/10.1007/s12205-011-1079-y

[62] Ibrahim A, Juahir H, Toriman M E, Mustapha A, Azid A and Isiyaka H A 2015 Assessment of Surface Water Quality Using Multivariate Statistical Techniques in the Terengganu River Basin. *Malaysian J. Anal. Sci.* 19(2): 338–348

[63] Wang Y F, Huang K L, Li C T, Mi H H, Luo J H and Tsai P J 2007 Emissions of fuel metals content from a diesel vehicle engine. *Atmos. Environ.* 37(33): 4637–4643. https://doi.org/10.1016/j.atmosenv.2003.07.007

[64] Juahir H, Zain S M, Yusoff M K, Hanidza T I T, Armi A S M, Toriman M E and Mokhtar M 2011 Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environ. Monit. Assess.* 173(1): 625–641. https://doi.org/10.1007/s10661-010-1411-x

[65] Yang Y H, Zhou F, Guo H C, Sheng H, Liu H, Dao X and He C J 2010 Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods. *Environ. Monit. Assess.* 170: 407–416. https://doi.org/10.1007/s10661-009-1242-9

[66] Zhang X, Wang Q, Liu Y, Wu J and Yu M 2011 Application of multivariate statistical techniques in the assessment of water quality in the Southwest New Territories and Kowloon, Hong Kong. *Environ. Monit. Assess.* 173: 17–27. https://doi.org/10.1007/s10661-010-1366-y

[67] Varol M and Sen B 2009 Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. *Environ. Monit. Assess.* 159: 543–553. https://doi.org/10.1007/s10661-008-0650-6

[68] Simeonov V, Stratis J A, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofoniou M and Kouimtzis T 2003 Assessment of the surface water quality in Northern Greece. *Water Res.* 37: 4119–4124. https://doi.org/10.1016/S0043-1354(03)00398-1

[69] Gazzaz N M, Yusoff M K, Ramli M F, Aris A Z and Juahir H 2012 Characterization of spatial patterns in river water quality using Chemometric pattern recognition techniques. *Mar. Pollut. Bull.* 64(4): 688–698. https://doi.org/10.1016/j.marpolbul.2012.01.032

[70] Behrendt H 1993 Point and Diffuse Loads of Selected Pollutants in the River Rhine and its Main Tributaries. Report RR 93-1, International Institute of Applied System Analysis, Laxenburg, Austria, WP-92-015

[71] Absalon D and Matysik M 2007 Changes in water quality and runoff in the Upper Oder River Basin. *Geomorphology* 92(3): 106–118. https://doi.org/10.1016/j.geomorph.2006.07.035

[72] Xu Y, Xie R, Wang Y and Sha J 2015 Spatio-temporal variations of water quality in Yuqiao Reservoir Basin, North China. *Front. Environ. Eng. Sci.* 9(4): 649–664. https://doi.org/10.1007/s11783-014-0702-9