Journal of Mines, Metals and Fuels, *71(9): 1155-1173; 2023. DOI: 10.18311/jmmf/2023/35432*

Print ISSN : 0022-2755

Journal of Mines, Metals and Fuels

Contents available at: *www.informaticsjournals.com/index.php/jmmf*

Modelling of Surface Water Quality and Spatial Mapping in Mahanadi Basin, Odisha

Abhijeet Das*

Research Scholar, Department of Civil Engineering, C.V. Raman Global University (C.G.U), Bhubaneswar, Odisha - 752054, India; das.abhijeetlaltu1999@gmail.com, abhijeetlaltu1994@gmail.com

Abstract

This study aimed at evaluating the surface water quality within the urban reach of Mahanadi River basin in Odisha. For this study, yearly water quality data were acquired from the State Pollution Control Board between low flow (between June and August) and high flow (between November and January) periods as part of a 5-year (2016–2021) monitoring programme (measuring 11 variables at 21 primary sites), to conduct an integrative investigation into the hydrogeochemical mechanisms that affect the water quality for drinking. Different methods like Multivariate Statistical Techniques (MST) such Pearson Correlation Matrix (PCM), Hierarchical Cluster Analysis (HCA), and Principal Component Analysis/Factor Analysis (PCA/FA) were applied. However, the relationship between the various factors revealed that anthropogenic activities and physicochemical parameters directly affect the fundamental ionic component of the water samples. Furthermore, with the help of plots, HCA was carried out and divided the 21 sampling sites into three major groups to verify the factor analysis results. Additionally, factor analysis has been conducted and four factors were extracted. Hence, the four components accounts for 87.17% and 67.22% of the total variance of the data composition for both seasons i.e., LFP and HFP, explaining the causes of water quality degradation. The four factors account for 'salinity', 'hardness', 'domestic sewage' and industrial effluents properties and were named accordingly. These factors revealed about point-based contamination, such as wastewater from commercial and residential sources, non-point sources, including such farm crops, and natural phenomena, such as inundation, rock, and soil disintegration. Further, this ongoing investigation also sheds light on core procedures, municipal regulations, and continual pollution monitoring to reduce the amount of pollution entering the river from both diffuse and point sources..

Keywords: *Effluents, Factor Scores, HCA, Mahanadi River, Multivariate, PCA, PCM.*

1.0 Introduction

Rivers are valuable natural resources that are heavily utilized in many different sectors, including manufacturing, agriculture, forestry, aquaculture, cattle, transportation, and recreation, largely as sources of drinking and utility water.[1] Hence, water quality describes the physical, microbial, and chemical properties of the substance. The availability and quality of potable fresh water sources are steadily reducing as a result of factors like climate change, environmental degradation, urbanization, agricultural activities, and industrialization.[2] It is noticed that one of the key drivers behind the reuse of waste water for peri-urban and urban irrigation is the rising demand for water to meet the rising population's food needs. However, surface runoff from agricultural regions, which transports man-made pollutants, pesticides, and fertilizers, has also had an impact on the rivers. Since recent years, it has seen a significant rise in the degradation of surface water quality as a result of inadequate management, lack of control over the outflow of liquid waste, and other human practices that pose serious risks to health of people.[3] So, water sources were also, gets contaminated

*Author for correspondence

by rapid industrialization, population development, and urbanization, and numerous localities experiences frequent water shortages in remote areas. Furthermore, the world's population is rising, and obtaining clean water resources for both urban and rural areas, is now becoming increasingly more difficult, particularly in countries that are experiencing a water shortage.[4] Numerous research has been undertaken in the past regarding the assessment of water quality, with a focus particularly on the hydro-chemical properties of water quality for drinking and irrigation, contamination processes, and ion distribution patterns. As a result, there is now a severe issue with the rising trend of surface water pollution.[5] It is nowadays, a challenging cause, to discern between point and diffuse sources of pollution, in vast industrial zones. Statistical techniques based on covariance structure and Euclidean distance are applicable to parameter concentrations and result into an accurate correlation of elemental connections. Huge water quality data sets can be analysed with little loss of critical information by using both multivariate and univariate statistical methods. Furthermore, it is frequently utilised to comprehend and assess the seasonal and geographic fluctuations of water quality measures. However, numerous studies has been performed on utilising a water quality measure and multivariate statistical methods, to evaluate the Swat River in northern Pakistan.[6] The findings showed that, except for nickel (Ni), lead (Pb), and copper (Cu), which had mean concentrations of 34, 60, and 56 %, respectively, within the World Health Standards' permitted levels. Also, another researcher displays a study that uses Cluster Analysis (CA) and Water Quality Index (WQI) to evaluate the water quality in the vicinity of Al-Zaatari Camp in Jordan using samples taken from the Basalt and B2/A7 aquifers (26 and 4 samples respectively).[7] Additionally, it was determined from the Gibb's diagram that the results in the research region are impacted by the operations of both types of ion exchange and rock weathering. So, through data reduction and classification, the numerical method for determining the quality of surface and ground water uses several multivariate statistical techniques, such as PCA/FA and HCA. This strategy aids in the simplification and organising of data sets to provide helpful generalisations and insights into water systems for the efficient management of water resources and the speedy eradication of pollution problems. A multivariate statistical technique called PCA can be used to reduce the amount of data when there is a lot of information accessible.[8] The factor analysis technique is useful for analysing surface water data with a wide range of variations. Hence, studies using this technique, produce findings, that are easier to understand since the data dimensionality is reduced to a smaller set of variables that can account for all the parameters. As a result, at the final stage, the factor scores provide a quantifiable relationship between the samples and the components, indicating how much each factor affected the samples. It is suggested that one of the multivariate statistical methods used to assess relative similarity in the homogeneity of observed parameters is Cluster Analysis (CA). The goal of this mechanism, is to compare data from related or comparable groups to one another with data from separate or unrelated groups.[9] The stronger the similarity (or homogeneity) inside a group and the greater the dispersion between groupings, the better or more distinct the clustering. The final visual representation of the clustering process is the resulting dendrogram, which dramatically reduces the dimensionality of the original data and shows a diagram of the groups and their proximity. Using a Geographic Information System (GIS) and geostatistical analysis was carried out to assess the spatial extent of the research area's surface water quality using Inverted Distance Weighted (IDW) approaches.[10] The IDW approach was employed to produce geographical distribution maps. It functions as a deterministic tool that estimates each unmeasured site or locator's value using measured values close by the prediction location.[11] However, these GIS approaches have shown to be particularly significant for data analysis and integration to produce relevant outputs for discovering, evaluating, and managing key resources. In view of this, there is no single study that has offered a multivariate statistical methodology to describe the surface water quality in this area. So, there is a study gap in this area, thus more discussion is required to better understand the level and sources of contamination. In conjunction with the above context, the primary goal of the current research paper aimed to investigate and evaluate/identify potential sources that might contaminate surface water by using Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA) and spatial distribution mapping. Overall, this proposed research is a way to assess the effects of different water quality management strategies implemented in the Mahanadi River, Odisha.

2.0 Study Area

The study area lies within geographic coordinates 19°21'N and 23°35'N and 80°30'E and 86°50'E, which represents a gateway to the districts of Odisha, Chhattisgarh and less portions of Maharashtra, Jharkhand, and Madhya Pradesh. It referred as a major east-flowing river system in Peninsular India. The total area is around 141600 Km^2. An average of 1360 mm of rain falls annually, primarily during the months of June and September, with the occasional strong cyclonic storm. In the river basin, a tropical monsoon climate predominates with average annual temperatures ranging from 15.8 °C to 28.7 °C. Our study area is limited to 851 km in total. In this reach, 357 km of the river runs in Chhattisgarh State, and the rest is in Orissa. After travelling 28 Km from Chhattisgarh, it enters Odisha belt, where slum residents, workshops, small-scale industry, and rice mills heavily invade. Additionally, the river drains into the stony beds of the Hirakud dam, which is located 10 Km from Sambalpur city, across the Mahanadi stream.[12] Further, the river is used as a dump yard for all the solid garbage from these places. It is observed that significant land uses exist in the basin namely forestry, agriculture, quarrying, and urban areas and, the most common varieties of soil found are mixed red, black, red, and yellow soils (laterite soils). Figure 1 shows the location of the Mahanadi Watershed.

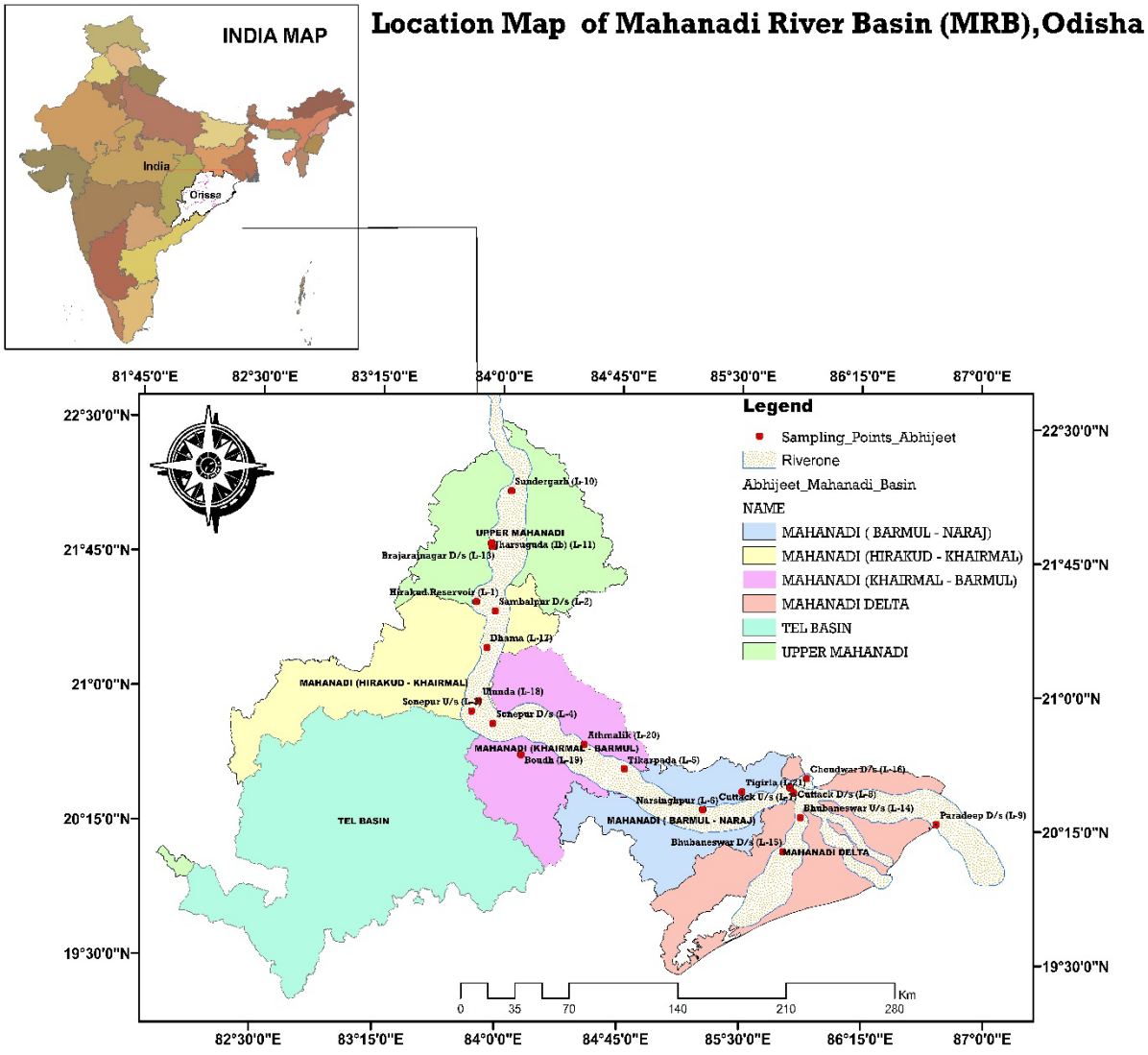


Figure 1. Map of study area with twenty-one sampling site location.

3.0 Sampling Collection and Analysis

The study attempts to locate a prospective surface water accessibility zone in the Indian Mahanadi Watershed. For the study region, 11 samples were taken over a 5-year period from 21 sampling stations during two flow periods: low flow (between June and August) and high flow (between November and January) (2016-2021), to describe the surface water's qualitative conditions in the research area. A Global Positioning System (GPS) was used to locate the sampling site. All evaluations were completed using the established standard operating procedures, given by the American Public Health Association.[13] 11 physiochemical parameters namely: Sulphate (SO_4^{2-}), Biochemical Oxygen Demand (BOD), Sodium (Na^+), Magnesium (Mg^{2+}), Nitrate-Nitrogen (NO_3-N), Potassium (K^+), Total Kjeldahl Nitrogen (TKN), Calcium (Ca^{2+}), Electrical Conductivity (EC) and Chemical Oxygen Demand (COD) and Total Dissolved Solids (TDS), were used to calculate and delineate the surface water vulnerable zones for drinking purpose, using PCA, GIS and CA. The Dichromate and Winkler's Azide method was used in the laboratory to test BOD and COD. By electrometric, EC and TDS. Titrimetric was used to measure Ca^{2+} and Mg^{2+}, whereas Na^+ and K^+ assessed using flame photometry, and SO_4^{2-}, NO_3-N and TKN, using a spectrophotometer respectively. In order to ensure quality, distilled water (blank samples) was analysed in parallel twice. Hence, Standard Reference Materials (SRM) is used as an accuracy for checked method its quality control. In addition, the Charge Balance Error (CBE) of the principal ions in all the samples, served as a measure of the data accuracy, which is computed as 100 * (Cations-Anions) / (Cations + Anions). The results obtained, were used to obtain the analytical precisions for ions, which usually falling within 5%.

4.0 Methodology

The main goal of water quality assessments was to ascertain the makeup of water. The efficacy of surface water as drinking water was assessed in the current study. The degree of similarity between the selected independent and dependent variables is measured via correlation analysis. Thus, Pearson Correlation Matrix (PCM) was investigated to discover any correlations between the observed data in order to decipher factor loadings during PCA. To determine the nature and strength of the link between the various water quality metrics, PCM, also known as r, was estimated. At a significant threshold of $p < 0.05$, a correlation coefficient value around zero indicates there is no association between two variables, whereas a correlation coefficient value of -1 or 1 indicates a strong relationship between two variables. Strongly correlated parameters are those with $r > 0.6$, whereas moderately correlated parameters have r between 0.4 and 0.6. A multivariate statistical technique called Principal Component Analysis (PCA) is used to identify the main sources of variance in a set of data that includes a lot of interconnected variables. Its goal is to create new variables, known as Principal Components (PC), from a group of original, pre-existing variables. It was performed in three steps: computing the correlation matrix by calculating the correlation coefficient, estimating the factor loadings, and interpreting the results.[14] The eigen values and eigen vectors are derived from the covariance matrix of the input variables. The loadings represent the contribution of the specified variable whereas the eigen values of the PC represent their related variance. A PC that has an eigen value greater than unity preserves most of the variability of the data set and is therefore, taken into consideration, for evaluation. The most important parameters, as indicated by the principal components, characterize the data with little information loss. To determine whether data is appropriate for factor analysis, the Kaiser Meyer Olkin (KMO) test is typically utilized. The statistic identifies the percentage of variance in variables with equal variance. For factor analysis, the data must have a KMO value larger than 0.5. KMO values of 0.5 to 0.7 are deemed adequate, whereas 0.7 and higher are deemed good. Factor Analysis (FA), a group of multivariate statistical procedures, is primarily concerned with data reduction and the generation of a set of component loadings with the highest variance of the squares of the loadings in order to improve the interpretability of the factor solutions without changing the underlying mathematical structure.[15] However, in this study, Clustering has been chosen as a categorization approach, that divides observations into groups, each of which is homogeneous, and unique from others technique. Nevertheless, it is a well-known classification strategy that seeks to ascertain the proximity or similarity between the variables to be grouped, and is known as 'Hierarchical Cluster Analysis' (HCA). It is typically demonstrated using a phylogenetic

tree, a two-dimensional network that offers a clear pictorial view of the project. Here, the linkage between groups and Euclidean distances as a degree of similarity, were generally taken in HCA, on the standardization data. In this ongoing area, the Ward's method, which takes use of squared Euclidean distances, is applied to the entire data frame. Finally, several comparisons were performed using PCA and CA to detect the possible risk, linked to urban outflows and anthropogenic activities, in order to demonstrate the existence of significant differences in water quality from one monitoring site to another and between seasons.[16]

5.0 Results and Discussion

In Table 1, the statistical summaries of the physicochemical properties of the surface water samples from the 21

Table 1. Statistics of surface water chemistry (all values in mg/L except EC in µS/cm)

	Low Flow Period (LFP)		High Flow Period (HFP)	
Parameters	**Minimum**	**Maximum**	**Minimum**	**Maximum**
EC	144.86	967	147	880
TDS	86.54	1000	89	986
Na^+	3.7	38.7	11.18	33.1
K^+	0.9	3.75	1.5	3
Ca^{2+}	37.3	75.9	36	76.2
Mg^{2+}	10.9	37	13.5	41.3
SO_4^{2-}	4.83	364.7	0.251	326
NO_3-N	1.85	6.4	0.235	12.64
TKN	3.44	11.069	1.3	21.84
BOD	1.273	3.71	0.6	3.1
COD	6.63	25.22	7.4	17.7

Table 2. Correlation matrix among water quality parameters during low flow period (LFP)

Parameters	**EC**	**TDS**	**Na^+**	**K^+**	**Ca^{2+}**	**Mg^{2+}**	**$S0_4^{2-}$**	**NO_3-N**	**TKN**	**BOD**	**COD**
EC	1.00	1.00	-0.08	0.06	-0.03	-0.02	1.00	0.21	0.68	0.09	0.17
TDS	1.00	1.00	-0.07	0.06	-0.03	-0.01	1.00	0.20	0.68	0.09	0.17
Na^+	-0.08	-0.07	1.00	0.00	0.03	0.61	-0.08	-0.01	-0.08	-0.07	-0.10
K^+	0.06	0.06	0.00	1.00	-0.44	0.21	0.06	0.06	0.06	0.45	0.37
Ca^{2+}	-0.03	-0.03	0.03	-0.44	1.00	0.18	-0.03	0.15	-0.06	-0.13	-0.06
Mg^{2+}	-0.02	-0.01	0.61	0.21	0.18	1.00	-0.01	-0.17	-0.24	-0.07	-0.10
$S0_4^{2-}$	1.00	1.00	-0.08	0.06	-0.03	-0.01	1.00	0.21	0.68	0.10	0.18
NO_3-N	0.21	0.20	-0.01	0.06	0.15	-0.17	0.21	1.00	0.47	0.81	0.77
TKN	0.68	0.68	-0.08	0.06	-0.06	-0.24	0.68	0.47	1.00	0.36	0.33
BOD	0.09	0.09	-0.07	0.45	-0.13	-0.07	0.10	0.81	0.36	1.00	0.93
COD	0.17	0.17	-0.10	0.37	-0.06	-0.10	0.18	0.77	0.33	0.93	1.00

Table 3. Correlation matrix among water quality parameters during High Flow Period (HFP)

Variables	EC	TDS	Na^+	K^+	Ca^{2+}	Mg^{2+}	$S0_4^{2-}$	NO_3- N	TKN	BOD	COD
EC	1.00	0.46	0.12	0.05	0.26	0.07	-0.13	0.15	0.23	0.31	0.03
TDS	0.46	1.00	-0.02	-0.18	0.08	-0.05	-0.22	0.12	-0.05	-0.06	0.00
Na^+	0.12	-0.02	1.00	0.30	-0.30	0.66	-0.22	0.14	0.05	0.34	0.02
K^+	0.05	-0.18	0.30	1.00	0.07	0.39	0.06	-0.23	0.15	-0.03	-0.03
Ca^{2+}	0.26	0.08	-0.30	0.07	1.00	0.09	-0.01	0.06	0.14	-0.10	-0.04
Mg^{2+}	0.07	-0.05	0.66	0.39	0.09	1.00	0.08	0.17	0.12	0.37	0.13
$S0_4^{2-}$	-0.13	-0.22	-0.22	0.06	-0.01	0.08	1.00	0.34	-0.07	0.39	-0.39
NO_3-N	0.15	0.12	0.14	-0.23	0.06	0.17	0.34	1.00	0.13	0.52	-0.29
TKN	0.23	-0.05	0.05	0.15	0.14	0.12	-0.07	0.13	1.00	0.42	-0.16
BOD	0.31	-0.06	0.34	-0.03	-0.10	0.37	0.39	0.52	0.42	1.00	-0.31
COD	0.03	0.00	0.02	-0.03	-0.04	0.13	-0.39	-0.29	-0.16	-0.31	1.00

sites are shown. The results are described in detail in the paragraphs that follow. The degree of affiliation between all the surface water characteristics, which has a significant impact on the water quality, was revealed using the Pearson correlation, a univariate statistical method. It is noticed that, higher correlation coefficients between various factors, can be seen in the correlation matrix (Table 2, 3).

As a result of the LFP, EC and TDS associated with TKN and SO_4^{2-} levels, indicating that the water is highly mineralized and possibly indicating cation exchange, surface infiltration, and water circulation. The EC serves as a "method of detecting TDS and also, it makes suggestions for a range of geochemical processes, including rock-water interaction, major evaporation, direct and reverse ion exchange, silicate degradation, and significant evaporation. Additionally, the presence of TKN in water typically denotes an inadequate process for the breakdown of organic matter and is a reliable sign that urban effluents have contaminated a river. The

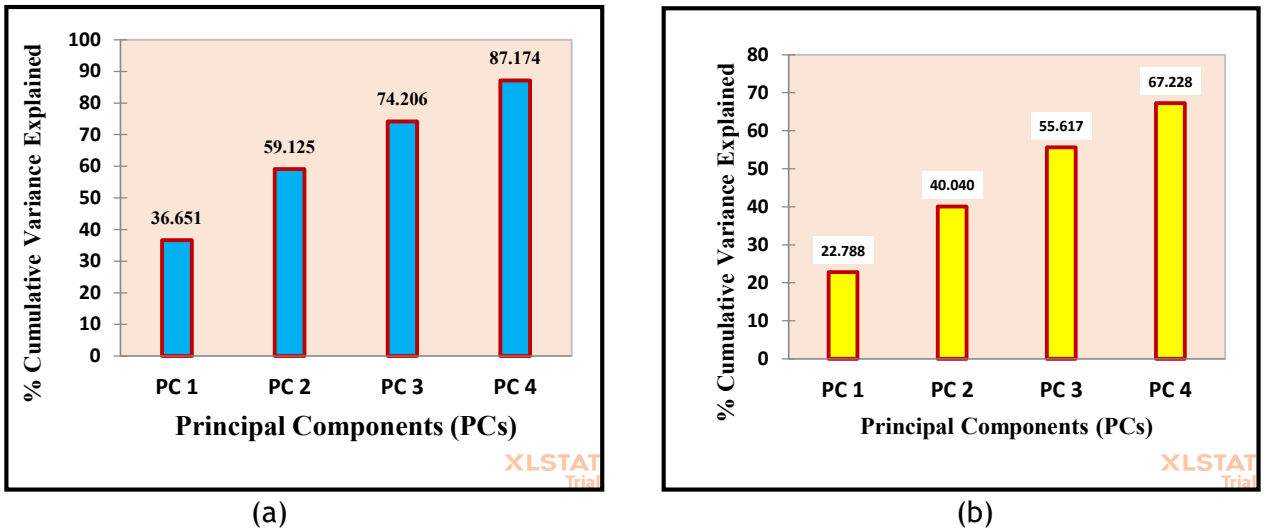


Figure 2. % Cumulative variance using principal component analysis during (a) LFP and (b) HFP.

positive association between Na^+ and Mg^{2+} indicates human activity and weathering of carbonate rocks. Mg^{2+} would probably come from trash discharges, residential wastewater, and soil leaching.[17] Consequently, these metals could endanger surface water. However, the strong association between SO_4^{2-} and TKN indicates that fertilizer dissolves after irrigation, together with human and animal wastes, and that BOD and COD are ultimately correlated with NO_3-N. The levels of NH_4^+ and NO_3^-, the most oxidised forms of nitrogen, are probably correlated with both intensive agricultural activity and direct industrial and urban river discharges. It demonstrates that NO_3-N came from nitrate fertilizers and plant nutrients that leached out during rigorous farming techniques. In later stage, Na^+ and Mg^{2+} were found to have a strong positive correlation with HFP. It comes from agricultural, anthropogenic, and carbonate weathering processes.[18] Further, the dominance is caused by the degradation of Na^+-Mg^{2+} containing minerals, the cation-exchange process, and industrial and/or agricultural activity. Also, Mg^{2+} is one of the major dissolved polyvalent metallic ions that contribute to hardness in water. For instance, to identify the factors affecting river water quality, principal component analysis-factor analysis was performed to normalize water quality data (11 variables) from low and high flow seasons. The KMO and Bartlett's test of sphericity results were used to determine whether the data was appropriate for PCA/FA. These tests were run to make sure the sample size was sufficient and that each variable was independent. Further, the sphericity values from Bartlett's test were $8.66*10^3$ and $4.2*10^3$, respectively ($p < 0.05$), whereas the KMO values for low flow and high flow seasons were 0.89 and 0.83, respectively. Strong, moderate, and light loadings are indicated by factor loadings > 0.75, 0.50-0.75, and 0.3-0.5, respectively. Thus, four factors in all were included in the current investigation, which accounts for about 87.174% and 67.228% of the total variation for both seasons. Figure 2 reveals the proportion of each component's variability in the data set. In order to identify the most crucial variables, that could adequately highlight the significant variations along the Mahanadi River, PC-1, 2, 3, and 4 were used in this study. The results show that many factor loadings are above 0.75 in low flow seasons, indicating strong factor loadings, whereas many are below 0.75 in high flow seasons, indicating moderate factor loadings. This may be, as a result of, dilution impact on water quality indicators, during the high flow season.

To characterize the number of PCs necessary to comprehend the fundamental database schema, the scree plot is employed.[19] It was proposed that at least 87.174 % and 67.228 % of the variation in both periods could be explained by the cumulative proportion of variance, which in PCA reflects the basic rule for dimension reduction. The scree plot shows a decline in slope preceding the

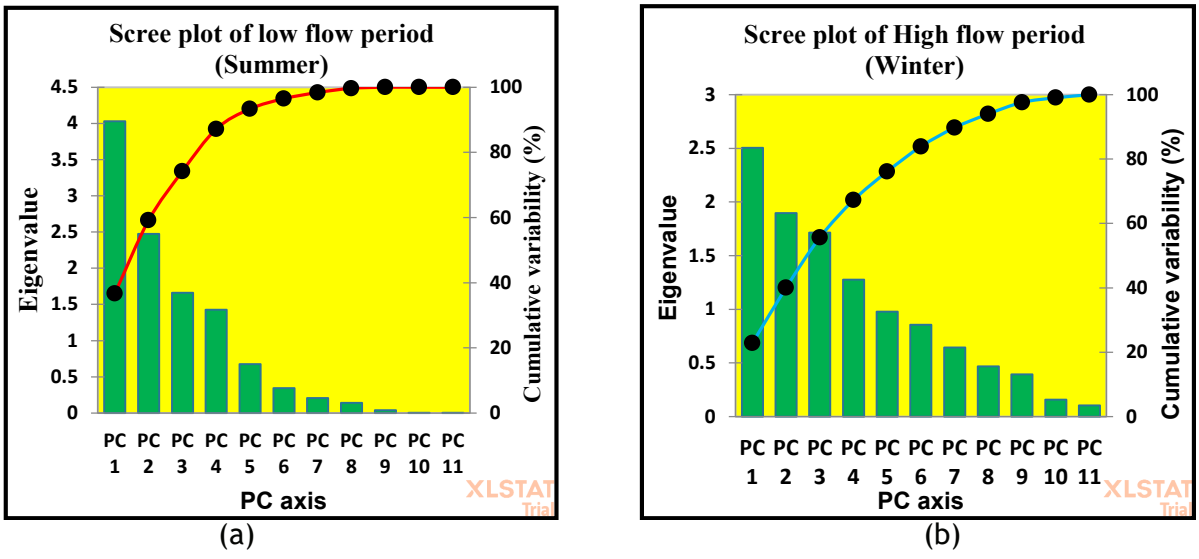


Figure 3. The Scree plot during (a) LFP and (b) HFP.

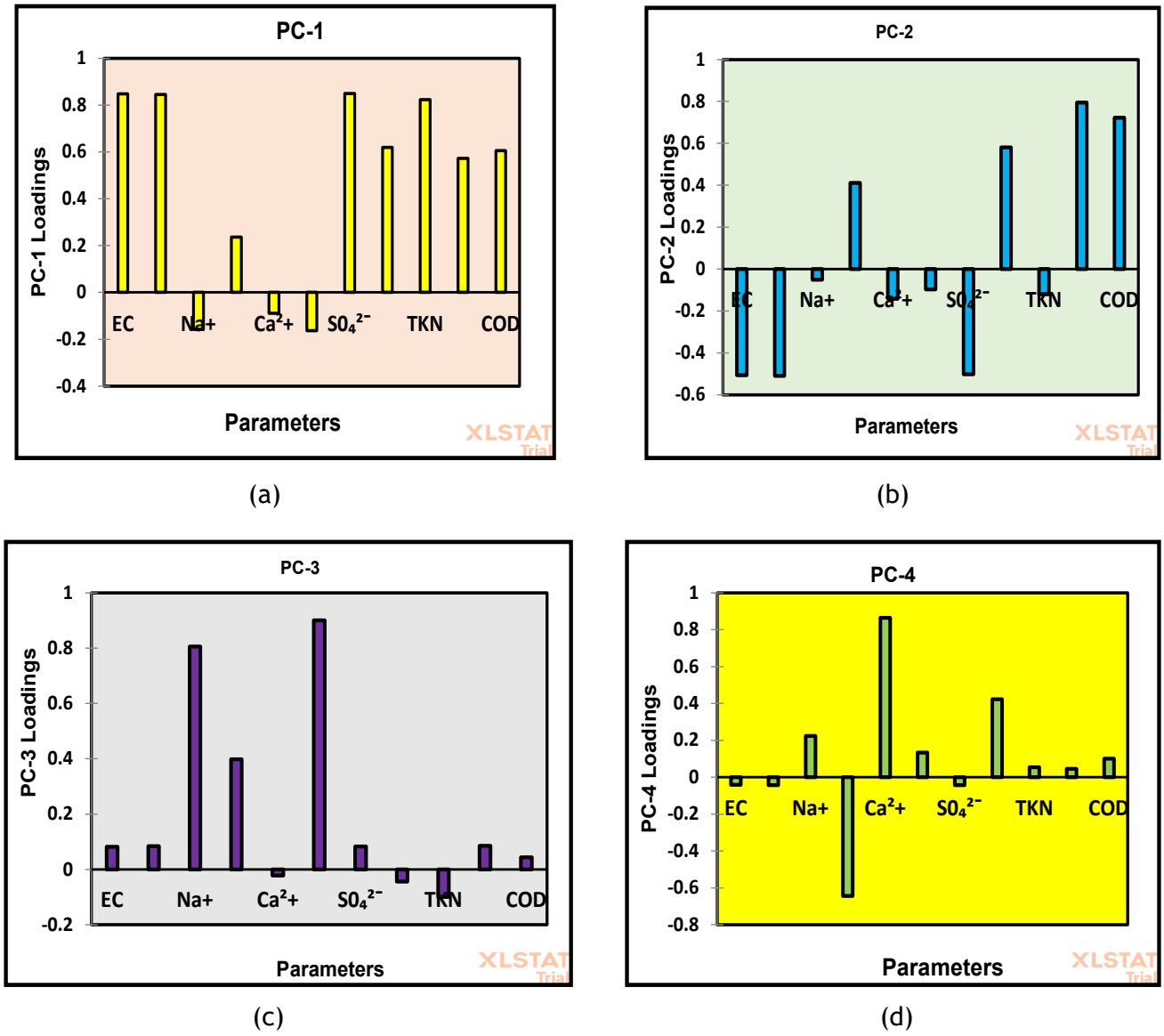


Figure 4. PCA loadings in LFP for various factors namely, (a) PC-1, (b) PC-2, (c) PC-3, and (d) PC-4.

fourth value, showing that four components dominate in both flow seasons (Figure 3).

Four factors were adequate to account for 87.17% of the variance for the Low Flow Period (LFP) (Figure 4). In evaluating the components, the most important variables in the ones with large loadings have been considered and expressed in Table 4. The four major components that make up the overall variance are as follows: PC-1: 36.65%; PC-2: 22.47%; PC3: 15.08%; and PC4: 12.97%. PC 1 showed a significant positive load on EC, TDS, SO_4^{2-}, and TKN and a moderate positive load on NO_3-N, BOD, and COD, accounting for 36.651% of the total variation. The concentration of dissolved solids is indicated by EC. Organic, inorganic, and other microbiological contaminants that are suspended in water samples have an impact on TDS. Also, SO_4^{2-} can be found in soil minerals and rock. The quality of the surface water is eventually impacted by ionic or reverse ion exchange, which was revealed by TKN to represent geogenic activity. Urban domestic and industrial wastewater that has been partially or not at all treated is one form of organic pollution that can be measured by COD. Additionally, PC-1 implies a

process of decomposition, which results in negative Ca^{2+}, Mg^{2+}, and Na^{+} loading and positive BOD. Soil leaching and industrial pollution are to be considered as of combined pollution, according to PC-1. Ca^{2+} mostly as a result of its entrance through rock formations including those of limestone, dolomite, and gypsum. Mg^{2+} is called to be 2nd most important factor in water hardness after calcium that occurs naturally, as it entails about zero negative effects on human and animal health. However, at high levels, it is toxic to living organisms. Continuous use of water with Na^{+}, may cause the soil's physical structure to collapse. Additionally, it decreases water entry into the soil, which may eventually result in the growth of excessive weeds, the rotting of seeds, and an obstruction of water flow to the roots. Further, 2nd Principal Component (PC-2) were kept to account for 22.474% of the dataset variability. COD

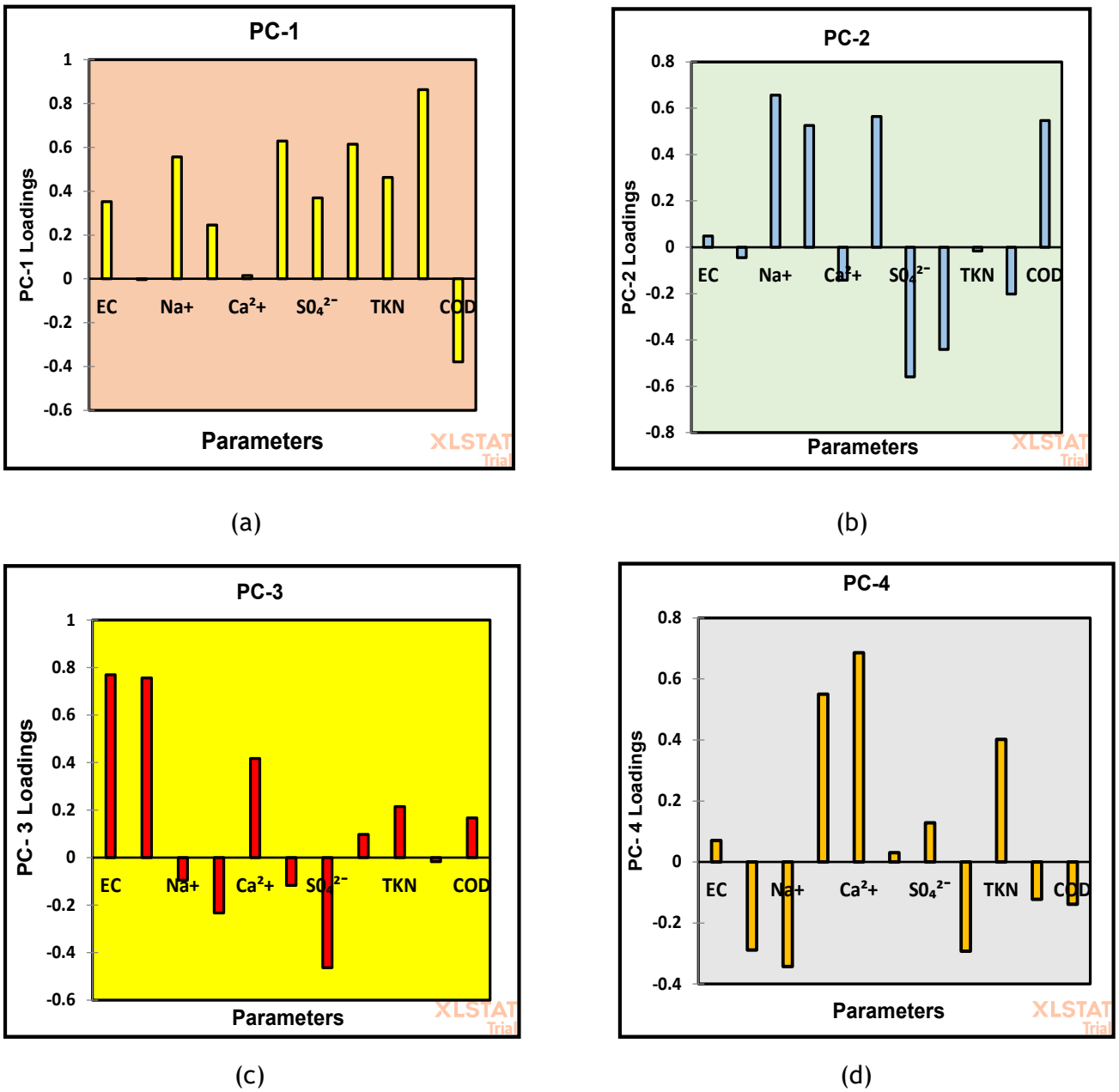


Figure 5. PCA loadings in HFP for various factors namely, (a) PC-1, (b) PC-2, (c) PC-3, and (d) PC-4

and BOD were the major sources of saturation in the 2nd component. Since organic matter contains nitrogenous materials, this is clear that aquatic ecosystem resources are evaluated using this key criterion because, it shows the level of organic contaminants, that deplete oxygen, in the water body. Additionally, these variables are well known to be important predictors of river contamination. PC-2 is an indicator of organic contamination, that suggests the input of wastewater gets discharged into the river. Mg^{2+} and Na^{+} were very Positively Charged in factor 3 (PC-3), which accounted for 15.081% of the overall variance. The breakdown of magnesium calcite and gypsum could produce magnesium. Na^{+} showing origins from surface water overuse and feldspar weathering (plagioclase bearing). In determining whether the water is suitable for irrigation, Na^{+} is a key cation to consider since its higher score causes the interchange of Mg^{2+} and Ca^{2+} ions, which lowers soil permeability and harms the drainage system.[20] PC-3 is comparable to soil leaching. Finally, Factor 4 (PC-4), which makes up 12.968% of the overall factor, has a substantial positive load for Ca^{2+} and a moderate positive load for NO_3-N. Ca^{2+} produced by the weathering of carbonates. Leaching of NO_3^- fertilizers and plant nutrients from intense agricultural activities results in the formation of NO_3-N. Consequently, PC-4 can be linked to soil leaching and pollution. In fine, PCA is effective in describing 67.23% of the total cumulative variation in the water quality parameters for the study for the High Flow Period (HFP) shown in (Table 5) and its loadings are shown in (Figure 5). According to the results, PC-1 showed high positive loadings on BOD and moderate positive loadings on Mg^{2+}, Na^{+}, and NO_3-N. It also explained 22.788% of the total variations. Fertilizers and clay minerals play a significant role in regulating Mg^{2+} in surface water. On the other side, sewage systems could contaminate surface water and cause an increase in Na^{+} levels. In case of NO_3-N, the level of its content was impacted by the influence of waste water effluents on the surface water in the area, which can be attributed to the use of fertilisers (both chemical and natural) for agricultural activities. The outcomes of PC-2 (17.251%) showed that Mg^{2+}, Na^{+}, and COD may have different anthropogenic origins. However, the concentration of Mg^{2+} and COD is lower at all monitoring locations, showing that soil

Table 4. Loadings of the significant principal components in Low Flow Period (LFP)

Parameters	PC-1	PC-2	PC-3	PC-4
EC	0.85	-0.51	0.08	-0.04
TDS	0.85	-0.51	0.08	-0.04
Na^{+}	-0.16	-0.05	0.81	0.22
K^{+}	0.24	0.41	0.40	-0.64
Ca^{2+}	-0.09	-0.14	-0.02	0.86
Mg^{2+}	-0.16	-0.10	0.90	0.13
SO_4^{2-}	0.85	-0.50	0.08	-0.04
NO_3-N	0.62	0.58	-0.04	0.42
TKN	0.82	-0.12	-0.10	0.05
BOD	0.57	0.80	0.09	0.05
COD	0.61	0.72	0.04	0.10
Eigenvalue	4.03	2.47	1.66	1.43
Variance (%)	36.65	22.47	15.08	12.97
Cumulative (%)	36.65	59.13	74.21	87.17

weathering and erosion are the main sources of rock mineral contributions. The extracted factors have a variance of 15.578% in PC-3, which means 15.58% of the measured parameter variability is reported from these factors. Hence, it is accounted for substantial positive loadings on EC and TDS, which are mostly caused by industrial effluents and residential sewage, respectively. Ca^{2+} and K^{+} had moderately positive loadings on the final factor PC-4, which explained 11.611% of the overall variance. Through mass bathing, the use of detergents, the discharge of industrial effluent, and domestic sewage, the Ca^{2+} component of TH gets diluted in the water. Leaching from fertilizer and salt patches in the geology could be the causes of K^{+}. It shows that minerals including calcium and magnesium have particularly contributed to mineral dissolution in the water. Hence, PC-4 denotes organic contamination and, is explained by the same anthropogenic practices. In conclusion, the four extracted principal components for both periods represent four distinct processes, namely agricultural sites (fertilizers and micronutrients), geological impact (due to weathering), and other processes (dissolution of halite's), that are accountable for various component loadings. Domestic and industrial effluent wastewater contamination, as well as regional monsoon conditions, induce wastewater environmental damage.[21]

Factor scores are used to define the relationship between variables known as factors as well as to reflect the overall variance of the data set and the similarities between the observations. The two approaches, that are most frequently used to obtain them are regression and weighted least squares. The regression technique is used to calculate the factor scores in the research work (Figure 6-9). The positive zones show that factor's dominance. The regions where factor I predominates during LFP season are those where factor I predominates during HFP season. The spatial characterization of factors II, III, and IV shows that they are in the direction of the river's flow, indicating that secondary salts are being leached and that agricultural activities have an impact on anthropogenic activity. In the end, factor scores show that lithology, anthropogenic activities, and agricultural practices are the

Table 5. Loadings of the significant principal components in High Flow Period (HFP)

Parameters	PC-1	PC-2	PC-3	PC-4
EC	0.35	0.05	0.77	0.07
TDS	0.00	-0.05	0.76	-0.29
Na^{+}	0.56	0.66	-0.10	-0.34
K^{+}	0.25	0.53	-0.23	0.55
Ca^{2+}	0.02	-0.14	0.42	0.69
Mg^{2+}	0.63	0.57	-0.12	0.03
SO_4^{2-}	0.37	-0.56	-0.46	0.13
NO_3-N	0.61	-0.44	0.10	-0.29
TKN	0.46	-0.02	0.21	0.40
BOD	0.86	-0.20	-0.02	-0.12
COD	-0.38	0.55	0.17	-0.14
Eigen value	2.51	1.90	1.71	1.28
Variance (%)	22.79	17.25	15.58	11.61
Cumulative %	22.79	40.04	55.62	67.23

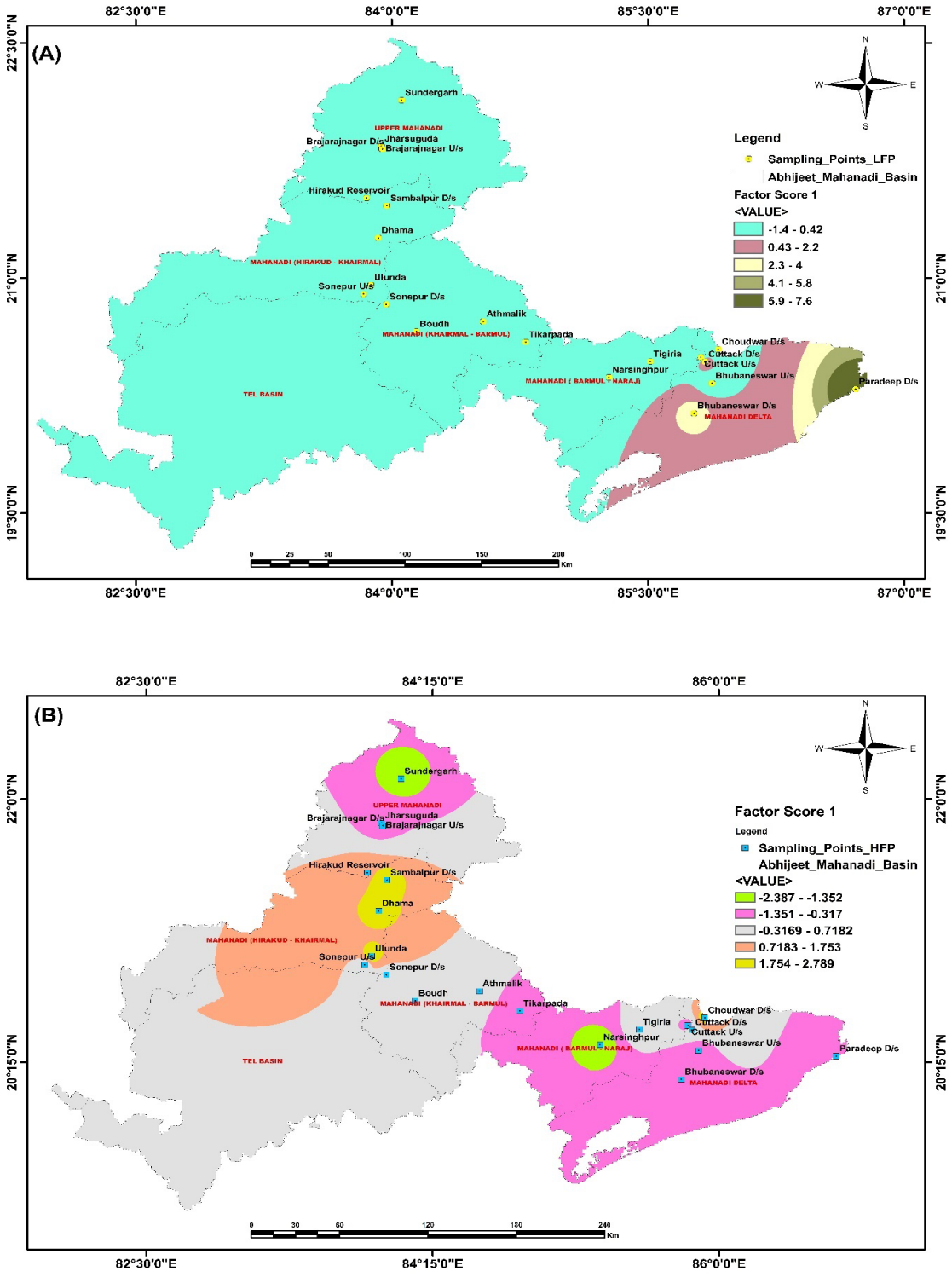


Figure 6. Spatial maps of Factor 1 during (a) LFP and (b) HFP in the research area.

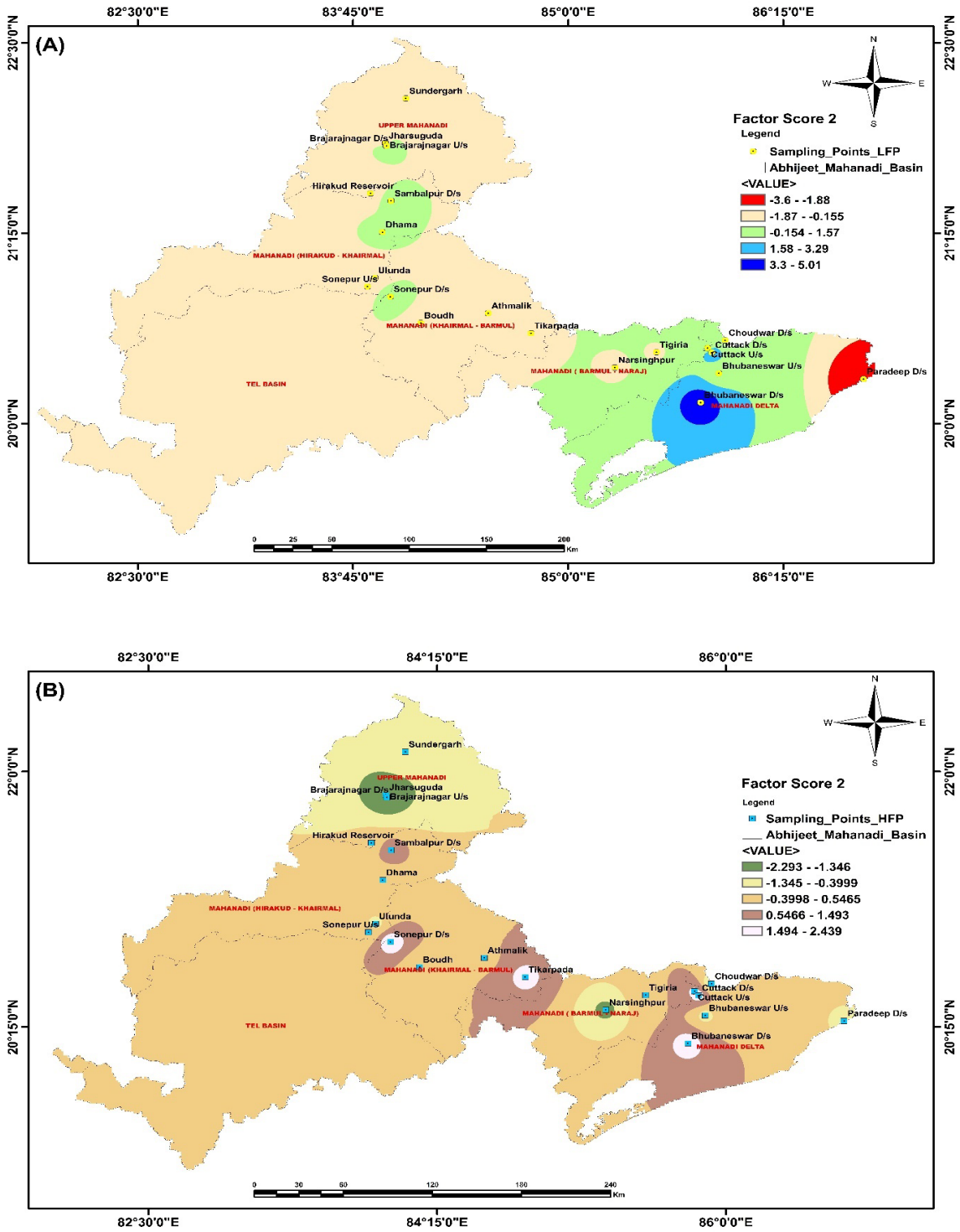


Figure 7. Spatial maps of Factor 2 during (a) LFP and (b) HFP in the research area.

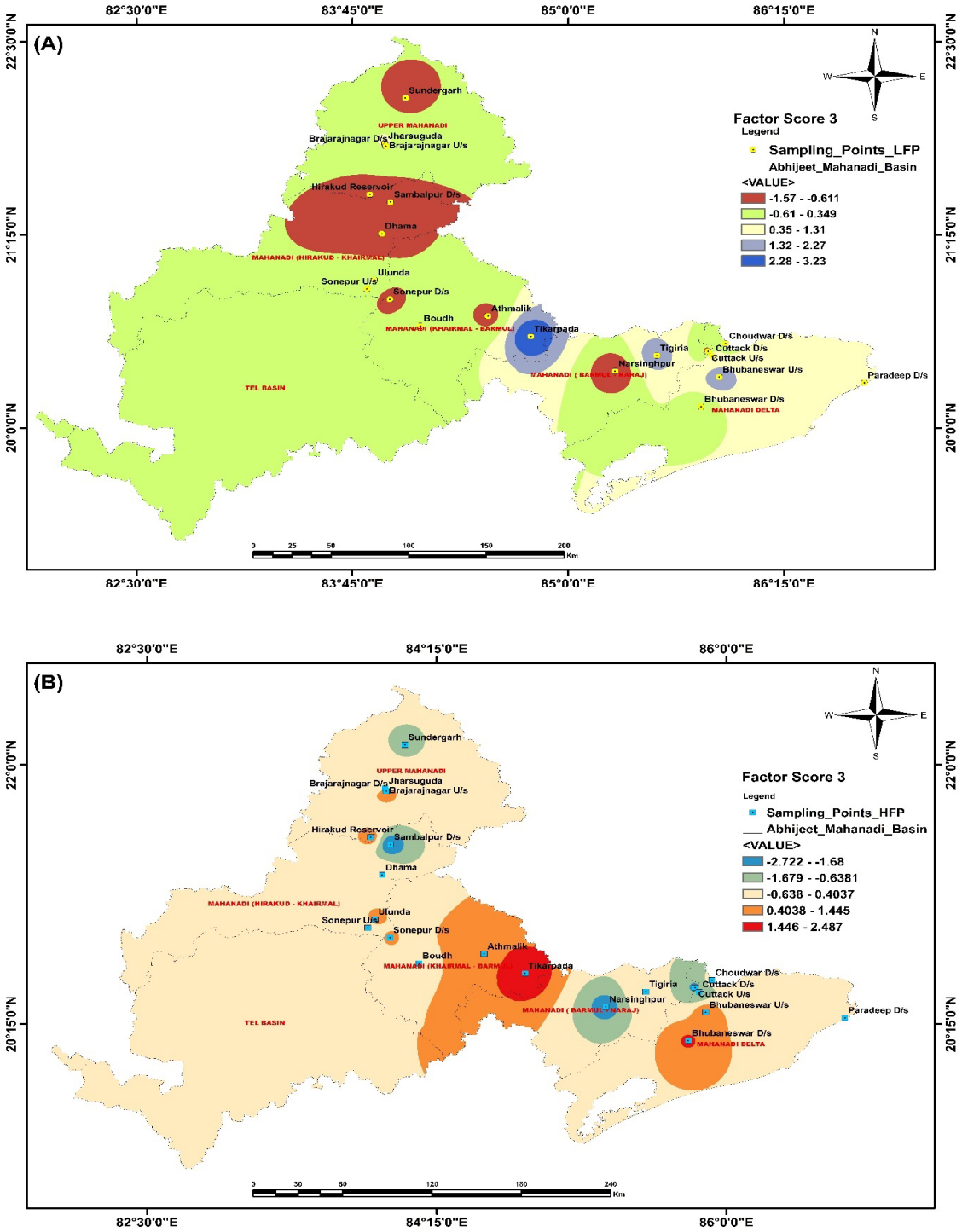


Figure 8. Spatial maps of Factor 3 during (a) LFP and (b) HFP in the research area.

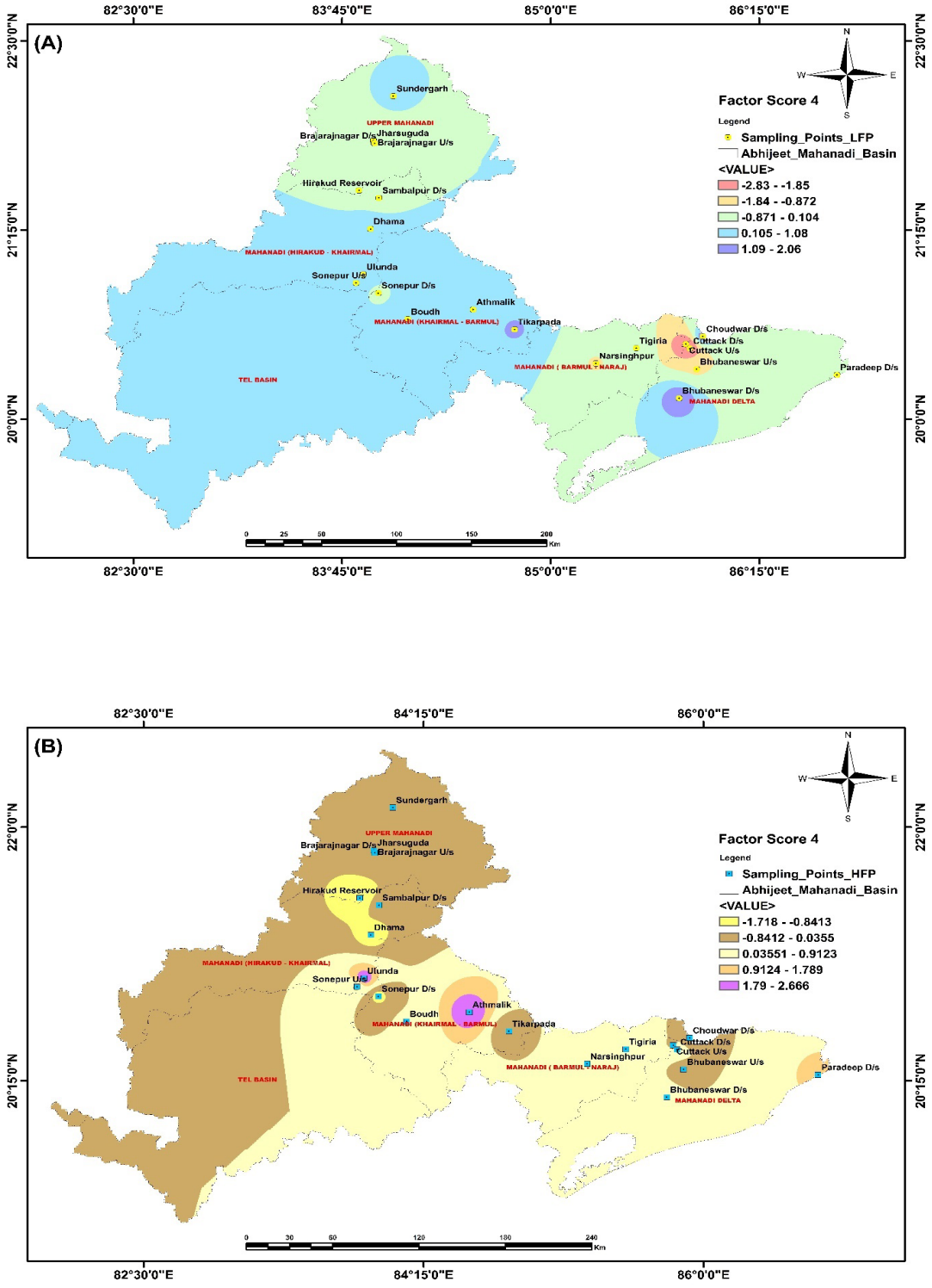


Figure 9. Spatial maps of Factor 4 during (a) LFP and (b) HFP in the research region.

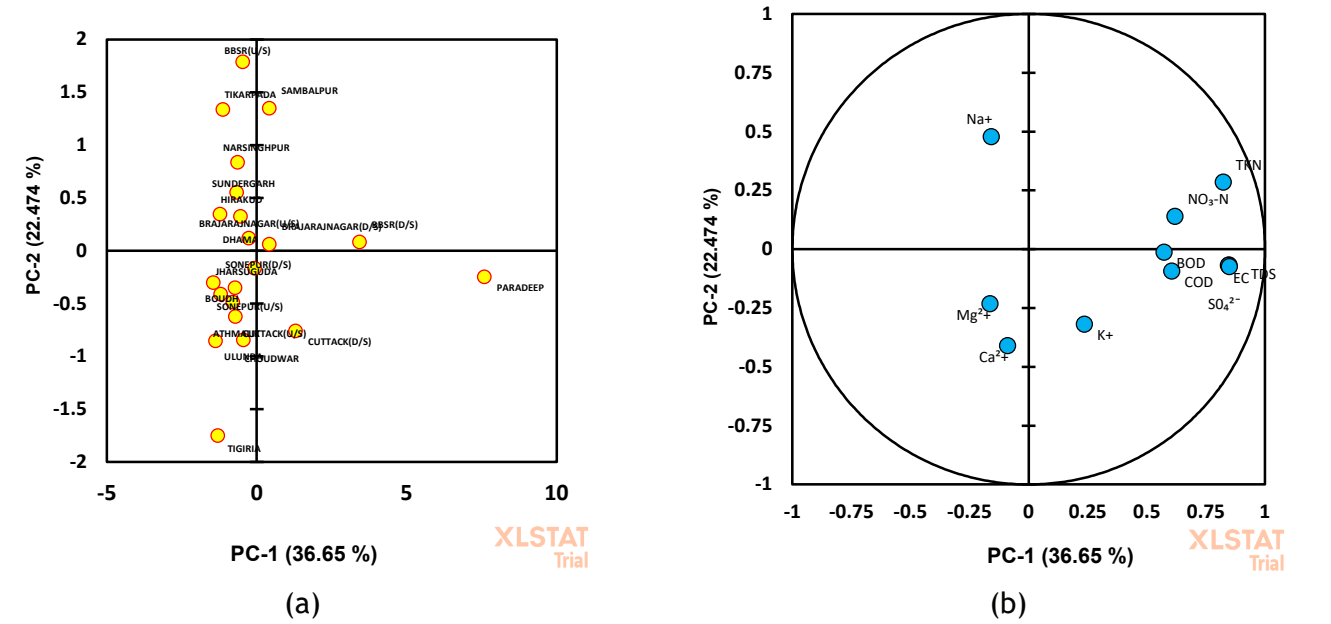


Figure 10. Cluster analysis of (a) sampling sites and (b) parameters for LFP.

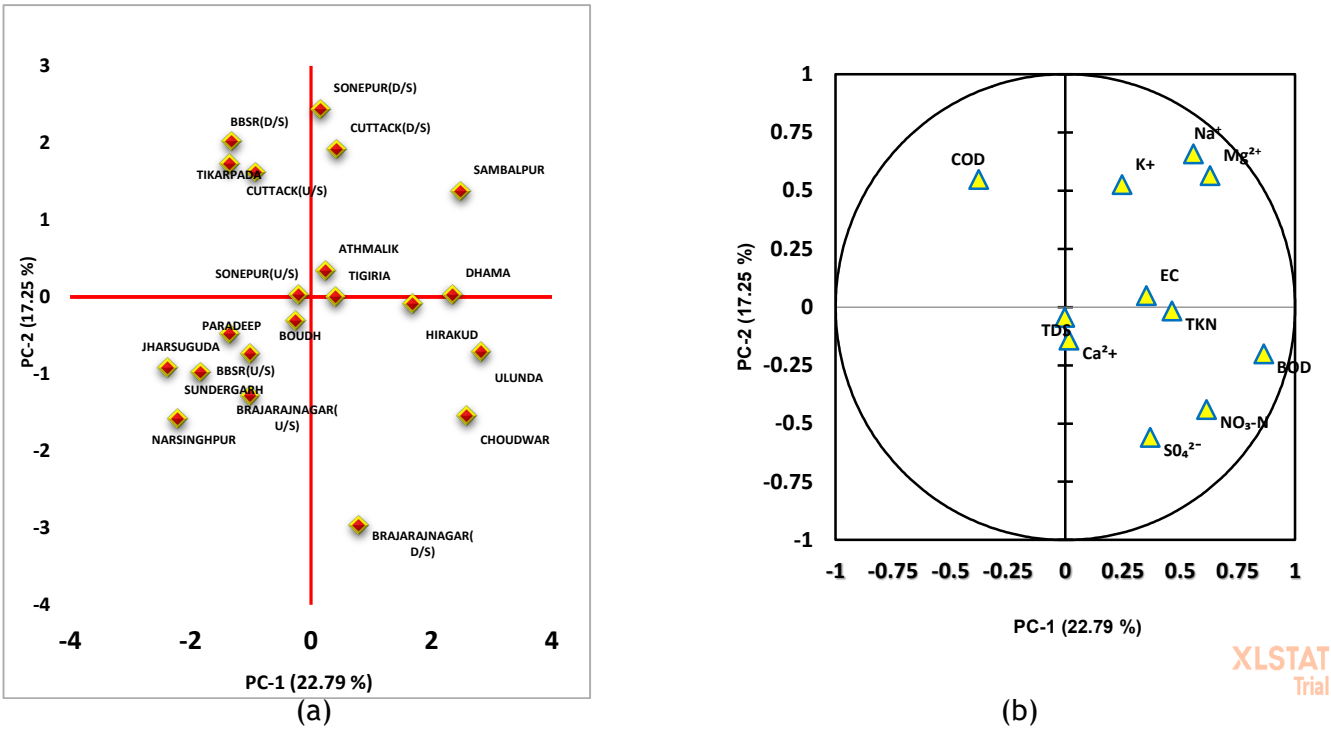


Figure 11. Cluster analysis of (a) sampling sites and (b) parameters for HFP.

primary controlling factors in the research area. For the purpose of grouping river network monitoring stations, HCA finds commonality. In a cluster analysis (Figure 10), three groups were found that shared common traits and contributions from background natural sources for both seasons. During LFP, Group I, consisting of (L-14,

5, 6, 10, 1, 12, 17, 4, 3, 11, 19, 20, 7, 18, 16, and 21), was diffused on the negative side of PC-1. This bunch having the parameters like Na^+, Mg^{2+} and Ca^{2+}. Longer residence times in water, precipitation, and irrigation return flow are indicative of higher Na^+. It implies the effects of infiltration water's water dissolution. Group II had the locations of (L-2, 13, and 15), which were located on the top section of PC-2. These locations had TKN and NO_3-N as parameters. This component denotes the combined impact of anthropogenic activities, leaching, and dissolution. Group III containing L-8 and 9, was situated at the right PC-1 direction. It was accompanied by BOD, COD, K^+, SO_4^{2-}, EC and TDS. BOD and COD indicating the involvement of an organic load from urban areas' drains. This cluster identifies the areas where farmers are applying fertilisers, especially NPK and organic chicken, sheep, and goat manure. Additionally, the surface water characteristics in these places are significantly impacted by the penetrated water.

During HFP (Figure 11), Group I, consisted of (L-15, 5, 7, 3, 9, 19, 11, 14, 10, 12, and 6). TDS and COD were its defining characteristics. This relationship implies that surface water is dominated by infiltration water, which interacts with ambient CO_2 and CO_2 generated from the organic matter breakdown of manure. All of these parameters considerably affect the COD and TDS. This cluster represents 52.38% of the total samples. Group II had the sites of (L-4, 8, 2, 20, 21, and 17) and composed of K^+, EC, Na^+ and Mg^{2+}. The two groups mentioned above represent urban waste discharge as a point source of pollution. Besides, Group III comprised of (L-1, 16, 13, and L-18) and it is composed of Ca^{2+}, TKN, SO_4^{2-}, NO_3-N and BOD. It is dominated by 28.57% of the total sites. This group displays as a result of anthropogenic contamination, agricultural operations such the usage of fertilizers, and the discharge of evaporated elements from rock layers. The factors, which include human mixed sources including human waste, manure waste, and agricultural runoff in addition to surface water leaching to surface water, have a significant impact on this association. In the current study, data from all stations for both times may be used to quickly analyse the water quality, which reduces monitoring expenses and the chance of data loss when trying to derive meaningful results.

6.0 Conclusion

The study's main objective was to evaluate the effects of anthropogenic activities and urban discharges on water quality and nutrient fluxes exported into the Mahanadi River Basin. To determine the probable polluted agents and their impact on water quality metrics, multivariate data analysis approaches like CA and PCA/FA are adopted. The correlation between EC and TDS in LFP indicates a strong association because of the linear relationship between the dissolved ions and conductivity. It is obvious that minerals have a greater relative impact on salinity than other ions. According to the results of the CA, the monitoring sites can be divided into three clusters with equivalent patterns in water quality and background contributions. However, PCA/FA is an effective method to find the factors accountable for alteration/pollution of water quality. It is noticed that the first four components, together accounts, for 87.17% of the total variance in the data set, in which the first principal component explains 36.65% of the total variance, the 2nd PC exhibits 22.47%, the third PC exhibit 15.08% and fourth PC is 12.97% of the total variance respectively, in LFP. So, in HFP, the PC result explains 67.22% of the total variance in the data set. The PC1 was dominated about 22.79%, while the 2nd, 3rd and 4th PC explains about 17.29%, 15.58% and 11.61% respectively. The first element seems to be connected to the existence of solid elements, whereas the second element seems to be connected to the effect of water cloudiness brought on by TDS. Once more, the third element seems to be connected to the modifications in water quality brought on by the presence of soluble salts. Human actions are likely responsible for the subsequent 4th PC. During periods of low flow, higher levels of EC, TDS, TKN, COD, and BOD were noticed and it reveals that domestic water and agricultural runoff were the major factors influencing the water quality. However, during the low season, the volume of the river decreases, increasing the concentration of nutrients in the water. In the later stage, in case of high flow period, higher levels of BOD, COD, NO_3-N, and Mg^{2+} were observed and it shows anthropogenic activities linked to the often-used manure for fertilising have an impact on rivers. Industrial and domestic wastewaters, soil weathering and erosion, soil leaching, organic pollution, and natural pollution are all significant polluting factors in the river system, with

seasonal changes in their quantifiable inputs. The findings of the correlation matrix were supported by the PCA and CA. The distribution of chief scores indicates the amount of pollutants, that enter metropolitan areas during the rainy season is more than it is during the dry one because of the significantly higher river discharges that occur when it rains, which result in much more waste being dumped into the rivers. Therefore, this study suggests methods like correlation analysis, PCA, and CA, which are effective, for pinpointing pollution sources in the area under inquiry and evaluating geographic variations.

7.0 Acknowledgements

The author is thankful to C.V. Raman Global University (CGU), Bhubaneswar, Odisha, India, for providing the necessary facilities to accomplish this work. Conceptualization, chemical analysis, writing original draft, data curation and validation, review and editing has been solely done by the author itself.

8.0 References

1. Das A. Multivariate statistical approach for the assessment of water quality of Mahanadi basin, Odisha. Materials Today: Proceedings. 2022 Jan 1; 65:A1-1.
2. Egbueri JC. Prediction modeling of potentially toxic elements' hydrogeopollution using an integrated Q-mode HCs and ANNs machine learning approach in SE Nigeria. Environmental Science and Pollution Research. 2021 Aug; 28(30):40938-56.
3. Han Y, Lam JC, Li VO, Guo P, Zhang Q, Wang A, Crowcroft J, Wang S, Fu J, Gilani Z, Downey J. The effects of outdoor air pollution concentrations and lockdowns on Covid-19 infections in Wuhan and other provincial capitals in China.
4. Das A. Assessment of potability of surface water and its health implication in Mahanadi Basin, Odisha. Materials Today: Proceedings. 2023 May 26.
5. Karuppannan S, Panneerselvam B, Haji M, Murugesan B, Shube H, Serre Kawo N. Evaluation of groundwater quality and its suitability for drinking and cultivation practices in and around around the deltaic regions of South India using DWQI, IWQI, and GIS. In: Climate Change Impact on Groundwater Resources: Human Health Risk Assessment in Arid and Semi-arid Regions 2022 Aug 25 (pp. 181-200). Cham: Springer International Publishing.
6. Jehan S, Ullah I, Khan S, Muhammad S, Khattak SA, Khan T. Evaluation of the Swat River, Northern Pakistan, water quality using multivariate statistical techniques and water quality index (WQI) model. Environmental Science and Pollution Research. 2020 Nov; 27:38545-58.
7. Obeidat M, Awawdeh M. Assessment of groundwater quality in the area surrounding Al-Zaatari Camp, Jordan, using cluster analysis and water quality index (WQI). Jordan Journal of Earth & Environmental Sciences. 2021 Sep 1;12(3).
8. Omar PJ, Gaur S, Dwivedi SB, Dikshit PK. Development and application of the integrated GIS-MODFLOW model. Fate and transport of subsurface pollutants. 2021:305-14.
9. Ravi R, Aravindan S, Shankar K, Balamurugan P. Suitability of groundwater quality for irrigation in and around the main Gadilam river basin on the east coast of southern India. Archives of Agriculture and Environmental Science. 2020 Dec 25; 5(4):554-62.
10. Serele C, Pérez-Hoyos A, Kayitakire F. Mapping of groundwater potential zones in the drought-prone areas of south Madagascar using geospatial techniques. Geoscience Frontiers. 2020 Jul 1; 11(4):1403-13.
11. Das A. Water Criteria Evaluation for Drinking Purposes in Mahanadi River Basin, Odisha. InInternational Conference on Trends and Recent Advances in Civil Engineering 2022 Aug 20 (pp. 237-260). Singapore: Springer Nature Singapore.
12. Das A, Goya A, Soni A. Use of water quality indices and its evaluation to verify the impact of Mahanadi River basin, Odisha. In: AIP Conference Proceedings 2023 Jul 27 (Vol. 2721, No. 1). AIP Publishing.
13. Eatson DL, Clesceri S, Rice EW, Greenberg AE. Standard methods for the examination of water and wastewater. 21st edn edn. USA: Centennial Edition. 2005:4-138.
14. Chen X, Lu J, Zhu J, Liu C. Characteristics of denitrifying bacteria in different habitats of the Yongding River wetland, China. Journal of Environmental Management. 2020 Dec 1; 275:111273.
15. Shekar PR, Mathew A. Evaluation of morphometric and hypsometric analysis of the Bagh River basin using remote sensing and geographic information system techniques. Energy Nexus. 2022 Sep 1; 7:100104.
16. Singh KR, Goswami AP, Kalamdhad AS, Kumar B. Development of irrigation water quality index incorpo-

rating information entropy. Environment, Development and Sustainability. 2020 Apr; 22:3119-32.
17. Thacker H, Shah Y, Borah AJ, Jadeja Y, Thakkar M, Bhimani S, Chauhan G. Assessment of groundwater potential zones across Katrol hill fault, Kachchh, Western India: A remote sensing and GIS approach. Open Journal of Geology. 2022 Feb 18; 12(2):111-35.
18. Das A, Goyal A, Soni A. Deciphering surface water quality for irrigation and domestic purposes: A case study in Baitarani Basin, Odisha. In: AIP Conference Proceedings 2023 Jul 27 (Vol. 2721, No. 1). AIP Publishing.
19. Ustaoğlu F, Islam MS. Potential toxic elements in sediment of some rivers at Giresun, Northeast Turkey: A preliminary assessment for ecotoxicological status and health risk. Ecological indicators. 2020 Jun 1; 113:106237.
20. Xiao J, Wang L, Deng L, Jin Z. Characteristics, sources, water quality and health risk assessment of trace elements in river water and well water in the Chinese Loess Plateau. Science of the Total Environment. 2019 Feb 10; 650:2004-12.
21. Wang Z, Wang Y, Chen L, Yan C, Yan Y, Chi Q. Assessment of metal contamination in coastal sediments of the Maluan Bay (China) using geochemical indices and multivariate statistical approaches. Marine Pollution Bulletin. 2015 Oct 15; 99(1-2):43-53.