Multivariate Statistical Approach for the Assessment of Water Quality of Mahanadi Basin, Odisha

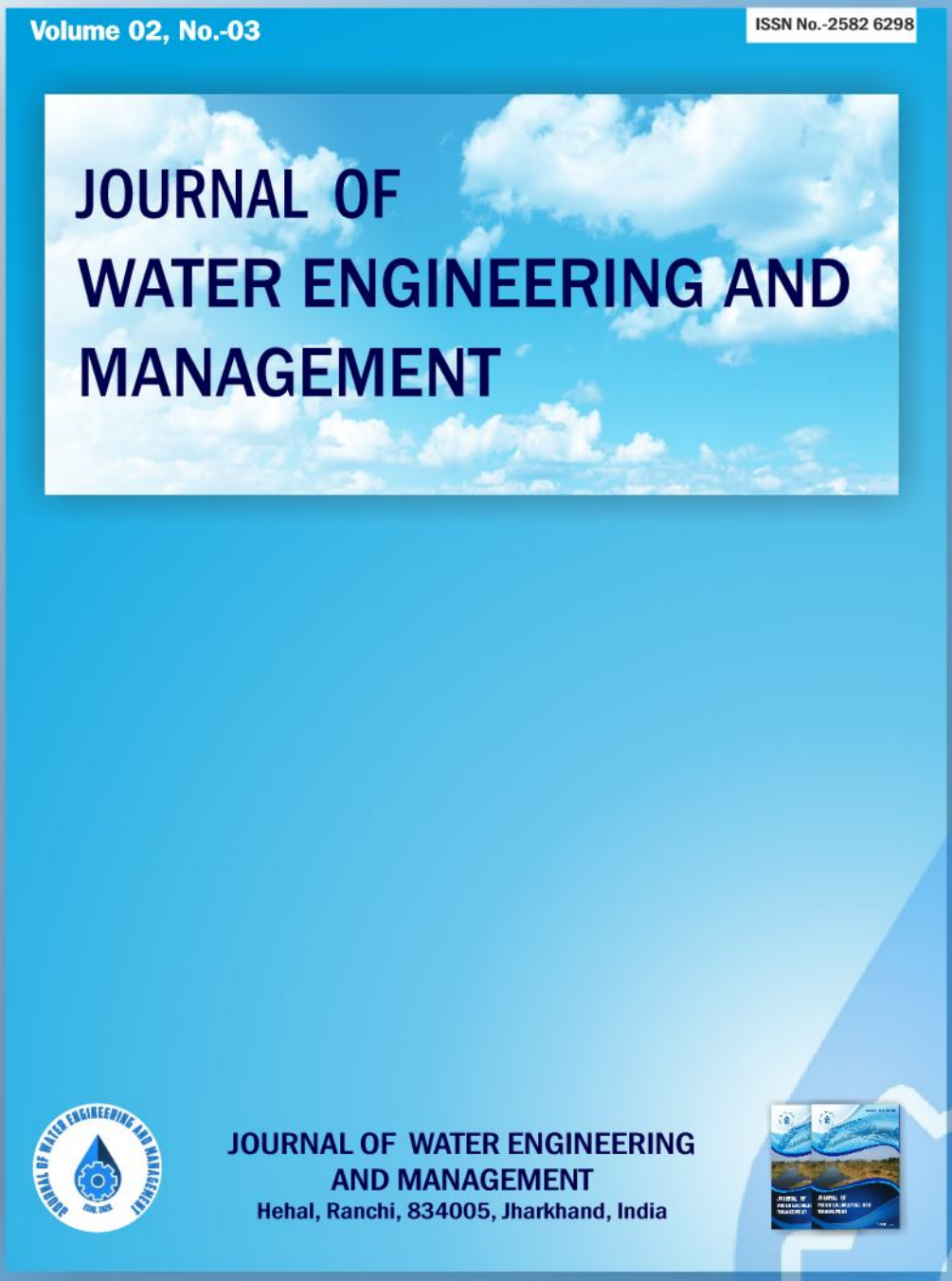
Abhijeet Das

https://doi.org/10.47884/jweam.v2i3pp10-39

Journal of Water Engg. and Management

ISSN 2582 6298
Volume-02
Number- 03

Jr. of Water Engg. and Mgt.
2021, 2(3) : 10-39



Journal of Water Engg. and Management, 2021, 2(3) : 10-39
https://doi.org/10.47884/jweam.v2i3pp10-39

Journal of Water Engg. and Management, 2021, 2(3) : 10-39
https://doi.org/10.47884/jweam.v2i3pp10-39

Research Paper

Multivariate Statistical Approach for the Assessment of Water Quality of Mahanadi Basin, Odisha

Abhijeet Das

Temporary Contractual Faculty, Civil Engineering Department,
College of Engineering and Technology, Ghatika, Bhubaneswar -751003, Odisha, India.
Email- abhijeetlaltu1994@gmail.com

Received on: October 10, 2021, Revised on: Novmber 3, 2021
Accepted on: December10, 2021 Published on: December 31, 2021

ABSTRACT

This research paper explores spatial and temporal water quality fluctuations to examine massive and complex water quality data sets used to quantify the influence of agricultural operations and household pollution sources on the Mahanadi River in Odisha. For a ten-year sampling study, (2008-2018), data sets containing 20 parameters were collected at 19 sampling sites along the river's length. The nineteen sampling locations were also split into three groups i.e. cluster 1 represents low polluted sites, Cluster 2 represents moderately polluted sites and cluster 3 depicts high polluted sites using the hierarchical clustering approach (HCA). It shows spatial and seasonal variations that are frequently symptomatic of contamination from rainfall or other sources. It yields positive results, with three separate groups of similarity across monitoring stations representing the river system's various water quality indicators. The FA/PCA identified the five most important factors, accounting for 93.899 percent of the total variance in the data matrix, allowing the selected parameters to be grouped based on common traits and the frequency of overall changes or variances within each group to be assessed. TSS, TKN, EC, TDS, B, SAR, and Fe have all been correlated with (loading > 0.7) in the 1st PC, which accounted for 43.133 % of the total variance. COD, NH3-N, Free ammonia, and fluoride were all linked to the 2nd PC, which accounted for 23.055 % of the total variation, whereas the 3rd, 4th, and 5th PCs, while accounting for 12.866 %, 8.603 %, and 6.241 % of the total variation, respectively. Five separate VFs with Eigenvalues > 1 were justified by the maximum variance rotation of the PC (original). This illustrates that point source, non-point source and natural occurrences are the primary cause of changes in water chemical concentrations.

Keywords: Mahanadi river, Water quality, Multivariate techniques, Cluster analysis, Factor analysis, Principal component analysis, Maximum variance rotation.

INTRODUCTION :

Water is the most critical requirement for preserving the planet's natural environment's long-term viability, as it is the material backbone for all Earth species' survival (Nowak et al., 2018). As a result of increased water use, the gap between supply and demand for water resources has expanded, demanding more rigorous restrictions for the use and conservation of waterbodies (Jalbaniet al., 2008). Society, irrigation, human consumption and industry, rivers are the most important source of water. Information on the quality and fluctuation of river water is critical

for the successful operation of these water resources. This is especially true in semi-arid places, where it is getting increasingly difficult to replenish waterways due to overuse by a fast increasing population. Natural cycles can degrade the quality of water and, more recently, human-made activities like the release of industrial effluents and household wastewater, in addition to agricultural drainage. The majority of river contamination originates through commercial and residential wastewater, along with drainage in farming (Carpenter et al., 1998; Jarvie et al., 1998). It is crucial in the assimilation and removal of urban and industrial wastewater along with runoff from agricultural land. Surface runoff is a periodical event influenced greatly by the climate of the basin, whereas urban and industrial wastewater outflow is a continual pollutant source. Rainfall patterns, surface runoff, interflow, groundwater flow, and pumped in and outflows all have an impact on river discharge, leading to higher pollution levels (Vega et al., 1998). Considering rivers are the primary supply of inland water for the home, industrial, and agricultural needs, it is critical to avoid and control river pollution, as well as to have trustworthy information on water quality for effective management. Because of the simplicity with which sewage can be disposed of, surface waters are the most polluted. The quality of rainwater has always been driven by environmental processes in a given vicinity meteorological contributions, erosion, and degradation of volcanic activity elements, and even some manmade impacts such as increased water resource exploitation and city, commercial, and farming activity (Carpenter et al., 1998; Jarvie et al., 1998). Environmental influences (temperature fluctuations, soil degradation and rainfall), as well as human involvement, influence the quality of a region's surface water (over-exploitation of waterways from residential and industrial sources) (Twine et al., 2005). Because the disposal of metropolitan sewage and industrial wastewater is a constant threat to the environment, proper sewage discharge control is essential for enhancing water quality (Younis et al., 2015; Ismail et al., 2015).

Rainwater is a monsoonal activity influenced mostly by the catchment's weather. River flow is affected by seasonal variations in rainfall, streamflow, interflow, and groundwater flow, along with pumping into and out of, as a result, pollution accumulation (Gashi et al., 2016). Seasonal changes in both human and natural processes that affect the river's hydrology that result in multiple functionalities between seasons which are influenced by fluctuations due to natural and human mechanisms (Vega et al., 1998). As a result, for an integrated approach of these water supplies, continual monitoring and evaluation of the river's water quality are essential (Singh et al., 2005). Human activities (Kumar, Sharma, and Taxak, 2017b) and climate change (surface runoff, land erosion, rock weathering) have an impact on river quality, particularly in metropolitan areas and crop production in rural areas (Ayeni, 2010; Kankaland Wate, 2012; Raj and Azzez, 2009). Sampling networks appear to be a useful data source for determining the river's water quality on a local and temporal perspective, as well as monitoring its quality. These connections provide a snapshot of the ecosystem's temporal state, as well as its seasonal and spatial change (Berzas et al., 2000; Simeonov et al., 2003). Multivariate statistics are the most appropriate and extensively used ways for assisting during the procedure and a review of various data sets that have been growing in size over time, and sampling networks appear to be a terrific proportion of specific information (Lopez-Lopez et al., 2014; Wang et al., 2014; Osman et al., 2012; Garata et al., 2011; Bouza and Deano 2008; Idris et al., 2008; Mendiguchia et al., 2004). The quality of Indian rivers' water, especially the Mahanadi, has deteriorated over the last several decades as a result of the continual point and non-point outlets release partially/untreated sewage

water, urban runoffs, and sewages (Duran and Suicmez, 2007; Mishra and Kumar, 2020; Mishra and Shukla, 2020; Elangovan and Palanivel, 2008; Sivakumar and Azzez, 2000; Byrappa and Ramaswamy, 2007). Some of the river's tributaries and distributaries are likewise heavily polluted. Several analyses indicated the above discharges as the primary source of pollution in the waterway (Annalakshmi and Amsath, 2012; Hema and Elango, 2010; Jameel and Hussain, 2005; Jeenaand Kalavathy, 2012; Kalavathy and Sureshkumar, 2011; Kathiravan, and Natesan, 2010, Shanti and Lakshmanaperumalsamy, 2010; Varunprasath and Daniel, 2010; Venkatachalapathy and Karthikeyan, 2013; Vimala et al., 2006). Multivariate approaches are useful for lowering WQ parameters and determining relationships between them, as well as grouping samples (Praus, 2007). PCA is used by Bhardwaj and Singh (2010) to determine the assembly of WQ and perhaps even the source of pollutants emitted by agricultural and domestic activities. PCA demonstrated that all physicochemical factors in the river Mahanadi basin contributed equally and strongly to WQ changes, while various connections seen between stations were shown by CA, indicating WQ features (Venkatesharaju and Prakash, 2010; Yerel and Ankara, 2012). Tharejaand Trivedi (2011) also illustrated that the PCA approach proved to be an effective tool for identifying crucial river WQ monitoring and indicators. Only after an accurate determination of the cause is feasible to prevent and remediate polluted water. As a result, determining the emission level and its origin becomes a prerequisite for taking further action. Simple methods for assessing WQ for stream hydrology and environmental sustainability include multivariate analysis, such as principal component analysis (PCA) and hierarchical cluster analysis (HCA).

Due to its ability to treat huge volumes of geographical and temporal data from a range of monitoring stations, the multivariate statistical technique has recently become popular for a clearer appreciation of drinking water and ecological state. Statistical approaches have all been used in academic papers for this sort of investigation because they can predict the new level of pollutants by assessing spatiotemporal variability in river water quality (Phung et al., 2015; Khan et al., 2016; Sharma et al., 2015; Varekar et al., 2015; Kumarasamy et al., 2014; Thuong et al., 2013; Razmkhah et al., 2010b; Kazi et al., 2009; Kumar and Dua, 2009; Varol and Sen, 2009; Zhang et al., 2009). Using CA, PCA, FA, and DA, Phung et al. (2015) analysed temporal/spatial alterations in the quality of surface water in Can Tho City, a Mekong Delta location of Vietnam. To investigate the hydrochemistry of the Tamiraparani river basin in Southern India, Kumarasamy et al. (2014) used CA and PCA/FA. Using hydro-chemical data, Khan et al. (2016) adopted CA to explore the geographic variance of River Ramganga and its sources' water quality (Ganga Basin, India). Sharma et al. (2015) conducted PCA and CA components, as well as correlation analysis, to assess seasonal patterns, possible pollution causes, and the clustering of Ganga and Yamuna River monitoring locations in Uttarakhand (India). PCA and CA were used in other projects for the analysis of quality variations (Simeonov et al., 2003, Bouza and Deano, 2008, Kazi et al., 2009, Hai et al., 2009, Razmkhah et al., 2010b). PCA and CA were employed in several research works to classify sampling sites and estimate the underlying source of pollution (Boyacioglu and Boyacioglu, 2007; Zhang et al., 2009; Zhou and Guo, 2007).

The quantitative techniques like PCA and CA were found to be useful in determining underlying correlations between water qualities data, identifying pollution sources, and forming clusters of identical monitoring stations

with similar qualities in all of the research mentioned above. For reliable data minimization and comprehension of multi-constituent chemical and physical discoveries, exploratory research and multivariate statistical approaches are useful tools. (Massart et al., 1988). Multivariate statistical strategies such as cluster analysis (CA), factor analysis (FA), including principal component analysis (PCA) have been regularly designed to obtain valuable knowledge from water quality data (Brown et al., 1996; Vega et al., 1998; Helena et al., 2000; Bengraine and Marhaba, 2003; Voncina et al., 2002; Lieu et al., 2003; Raghunath et al., 2002; Wunderlin et al., 2001; Simeonov et al., 2003). Surface and freshwater quality are widely depicted and evaluated using multivariate data processing, which is especially useful for displaying seasonal temporal and spatial cycles induced by natural and anthropogenic influences(Vega et al., 1998; Reisenhofer et al., 1998; Helena et al., 2000). Cluster analysis is a technique for sorting entities (cases) into categories (clusters) based on similarities within that class and differences between classes. The properties of the classes are unknown in advance, but they can be determined through analysis. The CA results to aid in the analysis of data and the detection of trends (Vega et al., 1998). PCA is a robust technique for limiting the quantity of the components of a data set with several relevant variables while keeping as much diversity as desirable. The amount of data collected is decreased by converting it into a fresh range of orthogonal (non-correlated) variables termed principal components (PCs), which are then sorted in declining order of importance. To obtain Eigen values and Eigen vectors using covariance or another cross-product matrix that characterizes the distribution of the measured data, the PCs are determined analytically. Linear combinations of the dependent data and Eigenvectors make up the major component (Wunderlin et al., 2001). Varifactors (VF s) are new variables constructed by rotating the PCA axis. Varimax rotation distributes the PC loadings in such a way that their dispersion is maximized by limiting the handful of big and moderate coefficients (Richman, 1986). Apart from significant data reduction, without sacrificing too much content, the whole data set variability can be described using only a few VFs/PCs. In addition, the use of VFs to group the variables studied according to their specific traits promotes data interpretation (Vega et al., 1998; Morales et al., 1999; Helena et al., 2000; Simeonov et al., 2003). As a result, reliable identification of available scenarios of surface water quality deterioration is crucial for the regulation of water quality. Multivariate statistical analysis is a sophisticated analysis method that evolved from classical statistics (Singh et al., 2004; Muangthong et al., 2015). CA, DA, PCA, and FA are examples of statistical rules that can be applied to many objects and indicators when they are connected (Corporal and Lodango et al. 2014; Oda et al., 2020). Multivariate statistical analysis is a widely accepted solution for assessing and comprehending crucial details in multi-component physicochemical investigations (Singh et al., 2011).

It's a useful tool for identifying factors and sources that could have an impact on water systems and cause water quality changes (Chattopadhay et al., 2012). Researchers employed the Mahanadi River as a research subject for the first time in this study, establishing 19 primary detection stations across the stream and detecting and assessing 20 physicochemical water samples' attributes.The detection time was 18 years long. To explore the semantic relatedness between monitored duration and checkpoints, and to determine the water quality causes driving observed changes in waterbodies, and discuss the effect of water sources, a plethora of multivariate statistical methods were used (natural and anthropogenic factors). To overcome these challenges, Multivariate techniques such as Pearson correlation strategy, PCA, and CA were applied to

1. Compare its commonalities in between the sampling sites to appraise the Mahanadi River's water quality.
2. Review the influence of water quality indicators on surface water quality variations throughout the duration.
3. Discover the sources of contamination that are influencing water quality.
4. Determine whether the water is suitable for consumption, cultivation, or economy.

The findings could assist people to resolve the space-time evolution of Mahanadi Watersheds, allowing them to better summarize the key sources of pollution in different areas of the river system.

MATERIAL AND METHODS

Study Area

The Mahanadi River seems to be a small pool about 6 km from PharsiyaTownship in the Amarkantak Mountains of the Bastar plains, which is located in Chhattisgarh's Raipur district in the extreme south-east. It runs for 494 km in Odisha, out of a total length of 851 km. In Odisha, the main tributaries are Ib, Ong, Tel, Hariharjore, and Jeera, while the primary distributaries are Kathajodi, Kuakhai, Devi, and Birupa. Nearly 400 kilometres from the mouth of the Mahanadi, the multipurpose Hirakud Dam in Sambalpur is situated directly in the middle of the mainstream. Near Bagra, the river Ib joins the Mahanadi and flows to the left of the Hirakud reservoir. The river travels south from Sambalpur until it meets Ong and Tel. The river travels eastward again at Sonepur, the Mahanadi's greatest tributary, Tel, finally joining the Bay of Bengal. The river runs through the Eastern Himalayas before reaching the coastal region and forming the delta, cutting through a 60 km long "Satkosia Gorge" bordered by high hills and tropical vegetation. Finally, around Naraj, about 10 km west of Cuttack, the river emerges from the Eastern Ghats. Near Naraj, the riverbed splits into two large distributaries, the Mahanadi on the north and the Kathjodi on the south, kicking off the deltaic movement. It has created a massive delta by passing through the towns of Cuttack and Puri and a variety of distributaries from westward. The Mahanadi River receives minimal external rainfall replenishment and is contaminated by solid wastewater in its main body. It gets its water primarily from sewer systems in cities and wastewater treatment plants in paper mills. It is the primary inland water supply used for home, industrial, and irrigation uses, as well as incorporating or eliminating urban and industrial pollutants and farming runoff (Yidana et al., 2011). As a result, river pollution must be avoided and controlled, and accurate water quality data must be available for successful management. Because of the geographical and temporal fluctuations in river water chemistry, quality assurance initiatives are essential to accurately evaluate water quality(Sinha et al., 2005). As a result, big, complicated data arrays including a significant number of physiochemical elements are difficult to read, making it difficult to draw definitive information (Jebbet al., 2017). Spatial representation of River Mahanadi is being performed using GIS 10.3 software as shown in (Fig. 1, 2).

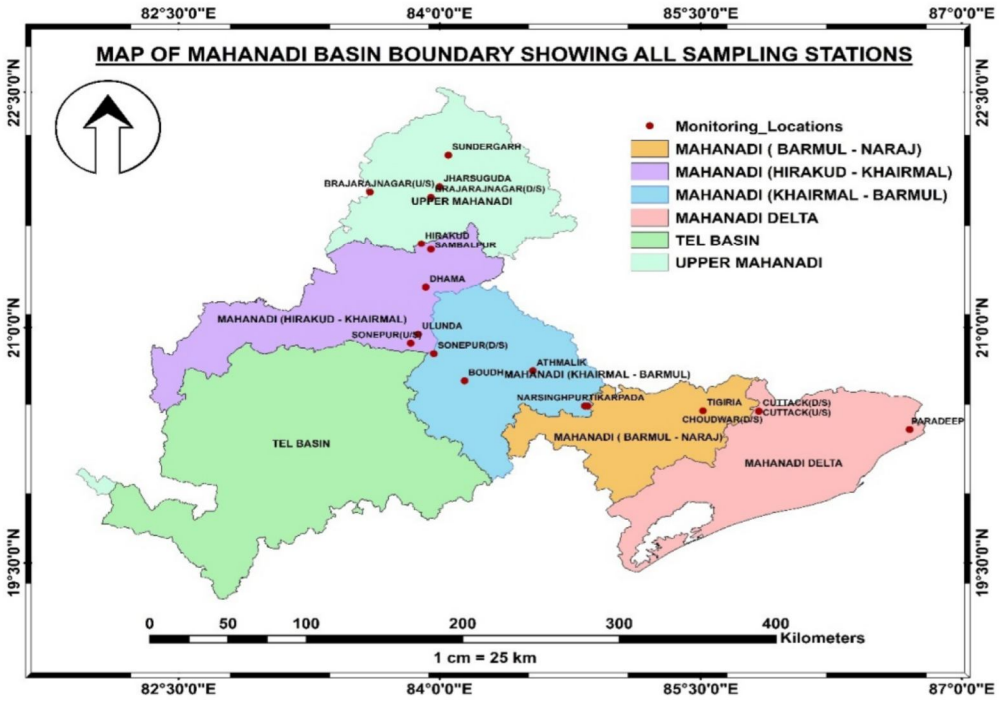


Fig 1. Map of Mahanadi basin showing all sampling or monitoring sites

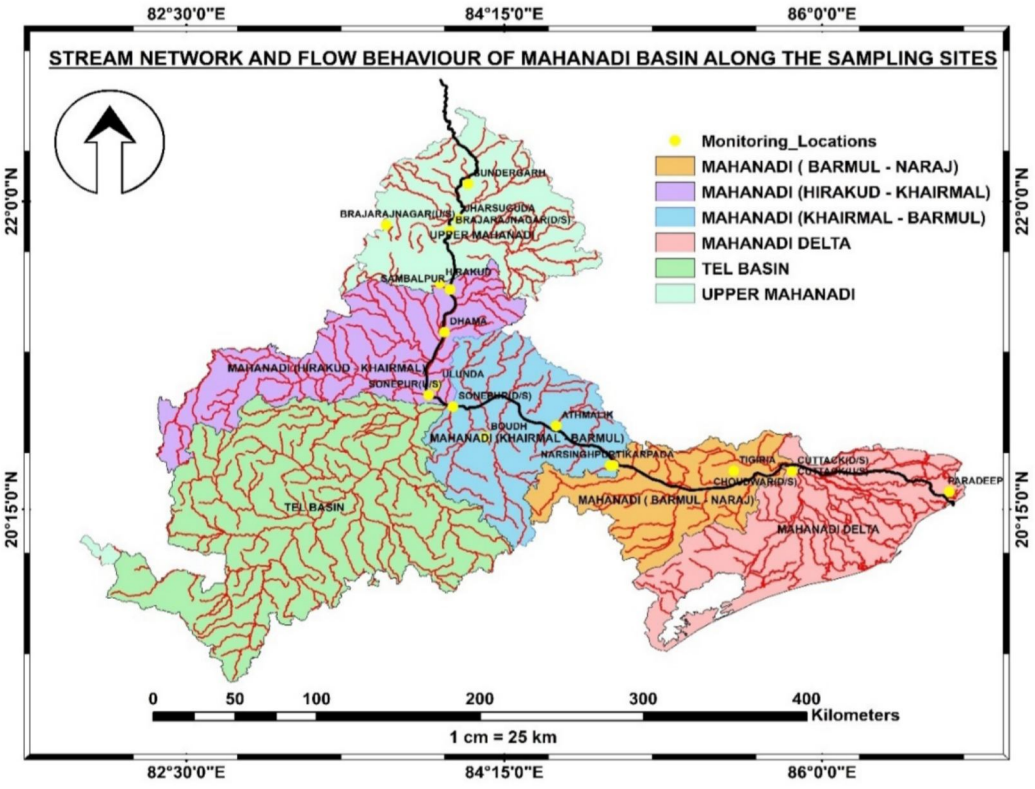


Fig 2. Stream network and flow behaviour of Mahanadi Basin along with the sampling sites

Monitored Parameters and Analytical Methods

The sample network's purpose was to find a variety of significant sites, such as tributaries and freshwater sources, that have a stronger impact upon that river, as well as to accurately reflect the river basin's water resources as a whole. The sampling method and monitoring network were designed to make a variety of variables at critical locations that accurately depict the river's water quality system, taking into account tributaries and inputs that affect downstream water quality. Water samples were collected, preserved, and transferred to the laboratory according to standard methods for the Mahanadi River's water quality monitoring programme (APHA, 1992). A mercury thermometer was used to test the temperature of the water on the spot. All of the other parameters were evaluated in the lab using standard techniques (APHA, 1992). The data collection included 20 water quality metrics that were observed over 18 years by sampling 19 test sites (2008 - 2018). This investigation looked at several elements like pH, dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), ammonia-nitrogen (NH3-N), free ammonia, total Kjeldahl nitrogen (TKN), electrical conductivity (EC), total dissolved solids (TDS), boron (B), sodium absorption ratio (SAR), total hardness (TH), chloride (Cl-), sulphate (SO42-), fluoride (F-), nitrate (NO3-) and iron (Fe). Careful standardization, systematic blank observations, spiked and duplicate samples were used to assure the analytical data quality. A typical open cartridge sampler was used to collect water samples for water quality testing (1.5-litre capacity). This sampler may take water samples from various depths of water to guarantee that the data is representative. The 2-litre polythene plastic bottle was cleansed with metal-free solution, rinsed numerous times with distilled water, steeped in 10% nitric acid for 24 hours, and then rinsed with ultrapure water before collecting the water sample. Before being transported to the lab for analysis, the water samples were maintained in an insulated chiller and stored in a refrigerator at 4 degrees Celsius on the day they were obtained. All reagents used for ease of understanding were analytical grade for greaterefficiency and correctness (Merck, India). Some samples were sent to the water quality laboratory of Central Water Commission, Bhubaneswar for analysis of physicochemical parameters like pH, DO, BOD, COD, NH3-N, Free ammonia, TKN, EC, TDS, B, SAR, TH, Cl-, SO42-, F-, NO3- and Fe for examination and proper treatment. All these verification purposes are being carried out with extra care and precautions so that accuracy should be maintained. Data has been collected for 10 years from nineteen water sampling locations as discussed below in (Table 1) to carry out the analysis.

Table 1. Water Sampling Locations

Sampling Symbol	Stations
R1	Hirakud
R2	Sambalpur
R3	Sonepur(U/S)
R4	Sonepur(D/S)
R5	Tikarpada
R6	Narsinghpur
R7	Cuttack(U/S)
R8	Cuttack(D/S)
R9	Paradeep

R10	Sundergarh
R11	Jharsuguda
R12	Brajarajnagar(U/S)
R13	Brajarajnagar(D/S)
R14	Dhama
R15	Ulunda
R16	Boudh
R17	Athmalik
R18	Tigiria
R19	Choudwar(D/S)(Birupa)

Data Treatment and Multivariate Statistical Methods

To account for the non-normal distribution of gathered water quality indicators, the Pearson coefficient, a non-parametric measurement of the correlation between the variables derived across ranking data, was employed to investigate the interrelationship between the elements (Wunderlin et al., 2001). In the current study, the differences in river water indicators were first explored using a correlation matrix and the Pearson non-parametric correlation coefficient. Although water sampling was accomplished at all locations every year, due to adverse weather, certain locations were unable to be surveyed, and the missing data were filled by the average value. As revealed in the analysis, the fundamental statistics of the eighteen-year water quality data frame were conducted and executed well (Table 2).

Table 2. Statistical description of all nineteen stations' water quality metrics (min, max, mean, Stdev, skewness, and kurtosis)

Sl No	Parameters	Minimum	Maximum	Mean	Stdev	Skewness Coeff.	Kurtosis Coeff.
1	PH	7.74	7.92	7.82	0.05	0.02	-0.84
2	DO	7.26	7.83	7.68	0.14	-1.87	3.66
3	BOD	1.05	2.40	1.36	0.34	1.84	4.19
4	TC	1212.40	42529.20	5151.25	9193.97	4.15	17.64
5	TSS	28.63	74.90	39.33	11.56	1.89	4.06
6	Total Alkalinity	70.40	100.90	85.71	8.24	0.07	-0.61
7	COD	6.76	21.88	11.25	3.96	1.63	2.52
8	NH3-N	0.51	1.93	0.66	0.31	4.03	16.97
9	Free Ammonia	0.02	0.06	0.03	0.01	2.04	5.71
10	TKN	3.28	11.80	5.73	2.07	1.49	2.94
11	EC	138.10	7779.35	580.86	1743.33	4.36	18.99
12	TDS	82.30	13230.60	812.80	3007.19	4.36	19.00
13	B	0.03	0.55	0.08	0.12	4.02	16.85
14	SAR	0.41	16.59	1.34	3.69	4.36	18.99
15	TH	51.20	2195.20	186.45	486.60	4.35	18.97
16	CL2	9.65	4904.91	269.23	1122.58	4.36	19.00
17	$S04^{2-}$	4.97	376.07	26.43	84.68	4.36	18.99
18	F	0.26	1.00	0.37	0.17	3.26	11.13
19	NO3	1.29	2.70	2.00	0.41	-0.22	-0.66
20	FE	0.60	2.61	1.31	0.46	1.04	2.41

Because the bell curve is frequently the input for statistical methods, the skewness and kurtosis statistics were used to check for conformance to the normal distribution before proceeding with the multivariate statistical analysis. According to the findings of the tests, all constituents fall into a range that is close to or equal to the normal curve. Skewness and kurtosis were calculated in the ranges of 1.87 to 4.36 and -0.84 to 19, respectively. To accommodate for the varied orders of magnitude and intensity values of numerous water quality indicators, all feature selections were z-scaled standardized with average 1 and variability 0 for CA and PCA.To find the parameters that produced sources of contamination across different timelines, the data was studied utilizing a range of multivariate statistical analytic approaches. For strict pollution management and water maintenance, a tremendous amount of water quality data must be addressed. CA, PCA, and FA are often used to limit river pollution and create trustworthy river water quality datasets. Experimental data were normalized employing z-scaled transformation to reduce misinterpretation owing to huge variances in data distributions. Small-variance variables get more influence as a result of standardization, while large-variance variables lose effect. Furthermore, a more thorough standardization technique minimizes the impact of multiple measurement units and renders the data dimensions. Factor analysis was used to the Pearson correlation of rearranged data to grasp the fundamental collection of data's structure. The correlation coefficient matrix shows how well each item's variance may be predicted by its relationships with others (Lieu et al., 2003). The variance, covariance, and correlation coefficients of the factors are thus calculated. Excel 2013 and Matlab were used to do all computations (mathematical and statistical).

Pearson's Correlation

Pearson's correlation analysis is a valuable numerical method for showing how closely two variables are linked (Belkhiri et al., 2011). In reality, the coefficient of correlation is used to quantify the interdependencies and level of correlations between variables. The number +1 denotes a perfect association between the two at a significant threshold of $p < 0.05$, whereas 1 denotes a perfect link between the variables despite their inverted fluctuations (Mudgal et al., 2009), and a value of zero implies that the variables have no link (Mudgal et al., 2009). In general, these coefficient values of $r > 0.7$ are regarded as high correlation, r values of 0.5 to 0.7 are considered moderate correlation, and r values of less than 0.5 are considered weak correlation. The interconnectedness of variables is demonstrated.

Cluster Analysis

CA is an unsupervised classifier that unveils the structural information or fundamental behavioural patterns of a data frame without attempting to make any approximations to identify or conglomerate the system's objects based on their relative position or similarity without making any predictions (Vega et al., 1998). It's a multivariate data analysis strategy for categorizing items based on their closeness or range (Cruz et al., 2017). The network elements can be categorized into groups called classes based on how identical or distinct their things are.The hierarchical CA strategy employed in this research is the most often utilized clustering method. Through sequential agglomeration, this strategy categorizes the data that is the closest or most identical and then organizes

these clusters into coherent groups. The Euclidean distance is used to determine whether two samples are comparable, and the "distance" is alluded to as the "difference" between the two samples' analytical results (Zhang et al., 2011). The Ward methodology was used in this study, and a hierarchical aggregate CA on a scaled data frame was performed using the squared Euclidean distance as a metric of similarity. The size of clusters was calculated using analysis of variance, and the sum of squares of the two major groups generated in each step was reduced. CA groups geographical and temporal differences in river water quality datasets using homology, yielding a spatiotemporal structure of samples. It creates a dynamic depiction, displaying a snapshot of each unit and those nearby, even though the original data's dimensions have been drastically reduced. To simplify the connection distance on the y-axis in a hypothetical situation, the link length is expressed as Dlink/Dmax, which is the ratio of the link distance divided by the maximum distance, multiply by 100. The Ward methodology and squared Euclidean distance were used to categorize the descriptive data (Zhao et al., 2019).

Principal Component Analysis (PCA)/ Factor Analysis (FA)

The PCA technique is used to extract the eigenvalues and eigenvectors of original variables from the covariance matrix. PCs are uncorrelated (orthogonal) variables produced by multiplying correlated variables by the eigenvector (a set of coefficients) (loadings or weightings). As a result, the PCs are original variables in linear combinations that are weighted. PC provides details on the most crucial aspects of the data gathering, enabling a reduction in data with minimal original data loss (Vega et al.,1998; Helena et al., 2000). It's a robust image recognition strategy that seeks to explain the variation of a wide number of interrelated parameters by reducing them to a smaller number of unrelated variables (principal components). It's a dimensionality reduction methodology. Its main goal is to reduce the number of variables required to explain the majority of the variation in the original data, as well as to convert a huge proportion of closely related variables into independent or unrelated variables (Liu et al., 2017). Primary components are usually used to characterize the comprehensive index of the data since they are less numerous than the estimated parameters and may explain the variation in most of the data. The basic principle behind the principal component analysis is to create the "optimal" fitting line for n points, with the minimum squared of the distance away between these n points and the line, and this line is known as the first principal component (Gonzalez et al., 2018). The second major constituent, which is unrelated to the first and has the smallest square sum of vertical distances between n locations, is discovered next. Similarly, until m fundamental components are detected, the number of m is frequently adjusted so that the dispersion of the first few major components accounts for about 85% of the entire variance (Mohammad et al. 2020). Factor analysis (factor analysis) is a multivariate prediction model that uses a few potential random variables to represent the covariance connection between multiple aspects (factors) (Ebrahimiyanet al. 2013). PCA's contribution to the involvement with fewer key factors is further decreased by twisting the axis designated by PCA, and a new set of variables known as varifactors is recovered. A VF can include theoretical key variables that are not observable, whereas a linear combination of evident water-quality indices is used to calculate PC (Vega et al., 1998; Helena et al., 2000; Wunderlin et al., 2001). To find significant PCs from scaled variables, PCA was utilised (water-quality data frame) and to limit the contribution of low-value variables even more; these PCs were then rotated (raw) to generate VFs. The factor in this study is a good approximation of the initial

variables, which were formed by adopting the maximum variance criterion (Ding et al., 2011). To achieve multivariate data dimensionality reduction, the raw data are described as correctly as possible under the assumption of assuring the least amount of information loss. In general, only factors with eigenvalues greater than one are considered in the analysis.

RESULTS AND DISCUSSION

Over the course of eighteen years (2008-2018), water quality monitoring of the Mahanadi River was examined out at nineteen different locations. All of the samples were evaluated for a variety of factors (twenty in all), and the statistical descriptive results for each location are reported in the table below (Table 2).The Pearson correlation matrix is used to identify the association between the variables and to estimate the goodness of fit. It gathers details on not only the magnitude but also the path coefficients. To quantify the correlation between different variables, the Pearson Correlation Matrix is used. These correlations show that wastewater from both home and industrial sources, as well as its organic load, is discharged into the river. The correlation empirical findings are taken into account in the ensuing interpretation. A high correlation coefficient (almost 1 or -1) indicates a strong association between two components, while a correlation coefficient approaching 0 indicates no such relationship. These numbers imply a positive association, whilst negative numbers indicate the opposite.

The actual values of the variables pH, DO, BOD, chemical COD, NH3-N, free ammonia, TKN, EC, TDS, B, SAR, TH, Cl-, SO42-, F-, NO3- and Fe were acquired for statistical purposes. Correlation matrix analysis as shown in (Table 3)represents the strong positive correlations that exist between TC-BOD, TKN-TSS, EC-TSS, TDS-TSS, B-TSS, TH-TSS, Cl-TSS, Sulphate-TSS, Iron-TSS, NH3-N-COD, Free Ammonia-COD, Fluoride-COD, Free Ammonia - NH3-N, Fluoride- NH3-N, (TDS, B, TH, Chloride, Sulphate)-EC, (TH, Cl, Sulphate, B)-TDS, (TH, Cl, Sulphate)-B, F-SAR, (Cl, Sulphate)-TH and Sulphate-Chloride. The values with red colour were found to exist strong (positively or negatively) correlated, values with blue colour signifies moderate (positively or negatively correlated) and ultimately, the values less than 0.5 depicts weakly correlated. This denotes that the parameters are changing in a straight proportional manner. Some parameters were discovered to have a strong negative correlation, indicating that they vary in inverse proportionality. These connected characteristics might be viewed as the primary cause of water quality fluctuations throughout time. The non-significant association suggests that anthropogenic sources play a role in the catchments.

The data on water quality were then applied to a variety of multivariate statistical techniques to investigate rends in time and space. Cluster analysis produced a dendrogram, dividing the river's nineteen sampling sites into three statistically significant groupings. Because the sites in these groups have comparable characteristics and natural background source types, the clustering technique yielded three distinct groupings of sites, each of which is highly persuasive. Cluster 1 (Hirakud, Sonepur (U/s), Tikarpada, Cuttack (U/s), Sundergarh, Jharsuguda, Brajarajnagar (U/s), Ulunda, Boudh and Tigiria), Cluster 2 (Sambalpur, Sonepur (D/s), Narsinghpur, Brajarajnagar (D/s), Dhama, Athamalik and Choudwar D/s) and Cluster 3 (Cuttack D/s and Paradeep) correspond to relatively less polluted areas, moderately polluted areas and highly polluted areas respectively. These cluster diagrams (Fig. 3) and grouping of clusters is being represented in (Table 4).

Table 3. Pearson correlation matrix of water quality parameters

PARAMETERS	PH	DO	BOD	TC	TSS	TOTAL ALKALINITY	COD	NH3-N	FREE AMMONIA	TKN	EC	TDS	B	SAR	TH	CL_2	SO_4^{2-}	F	NO_3	FE
PH	1.00																			
DO	0.05	1.00																		
BOD	0.49	-0.59	1.00																	
TC	0.36	-0.77	0.77	1.00																
TSS	-0.23	-0.54	0.04	0.25	1.00															
TOTAL ALKALINITY	0.45	-0.15	0.34	0.14	-0.04	1.00														
COD	0.57	-0.39	0.66	0.62	0.17	0.50	1.00													
NH3-N	0.29	0.13	0.15	0.12	0.07	0.06	0.72	1.00												
FREE AMMONIA	0.53	0.08	0.39	0.31	0.04	0.02	0.75	0.88	1.00											
TKN	-0.18	-0.54	0.13	0.15	0.73	0.23	0.16	-0.14	-0.10	1.00										
EC	0.05	-0.42	0.03	0.08	0.74	0.46	0.21	-0.06	-0.15	0.71	1.00									
TDS	0.05	-0.42	0.02	0.08	0.74	0.45	0.20	-0.07	-0.16	0.71	1.00	1.00								
B	0.10	-0.41	0.07	0.11	0.71	0.48	0.33	0.10	-0.01	0.67	0.98	0.97	1.00							
SAR	0.16	0.14	0.02	0.03	0.14	0.12	0.67	0.97	0.78	-0.08	0.04	0.03	0.20	1.00						
TH	0.06	-0.42	0.03	0.08	0.74	0.46	0.22	-0.05	-0.14	0.70	1.00	1.00	0.98	0.05	1.00					
CL_2	0.04	-0.42	0.02	0.08	0.75	0.45	0.20	-0.06	-0.16	0.71	1.00	1.00	0.98	0.04	1.00	1.00				
SO_4^{2-}	0.06	-0.42	0.03	0.08	0.74	0.45	0.21	-0.05	-0.15	0.71	1.00	1.00	0.98	0.04	1.00	1.00	1.00			
F	0.28	0.01	0.12	0.06	0.30	0.36	0.74	0.88	0.70	0.14	0.37	0.36	0.51	0.93	0.38	0.37	0.38	1.00		
NO_3	-0.10	-0.26	0.15	-0.08	0.08	0.52	0.25	0.09	-0.05	0.47	0.42	0.41	0.47	0.18	0.42	0.41	0.41	0.33	1.00	
FE	0.19	-0.20	-0.05	0.14	0.75	0.02	0.17	0.16	0.24	0.54	0.68	0.68	0.69	0.17	0.68	0.68	0.68	0.36	0.01	1.00

Table 4. Cluster membership differentiate into three clusters based on distance or similarity in regards to pollution factor

Cluster Membership				
Symbols	**Cluster Case**	**CLUSTER 1**	**CLUSTER 2**	**CLUSTER 3**
R1	HIRAKUD	HIRAKUD		
R2	SAMBALPUR		SAMBALPUR	
R3	SONEPUR(U/S)	SONEPUR(U/S)		
R4	SONEPUR(D/S)		SONEPUR D/S	
R5	TIKARPADA	TIKARPADA		
R6	NARSINGHPUR		NARSINGHPUR	
R7	CUTTACK(U/S)	CUTTACK(U/S)		
R8	CUTTACK(D/S)			CUTTACK (D/S
R9	PARADEEP			PARADEEP
R10	SUNDERGARH	SUNDERGARH		
R11	JHARSUGUDA	JHARSUGUDA		
R12	BRAJARAJNAGAR(U/S)	BRAJARAJNAGAR(U/S)		
R13	BRAJARAJNAGAR(D/S)		BRAJARAJNAGAR(D/S)	
R14	DHAMA		DHAMA	
R15	ULUNDA	ULUNDA		
R16	BOUDH	BOUDH		
R17	ATHMALIK		ATHAMALIK	
R18	TIGIRIA	TIGIRIA		
R19	CHOUDWAR(D/S)(BIRUPA)		CHOUDWAR(D/S)	

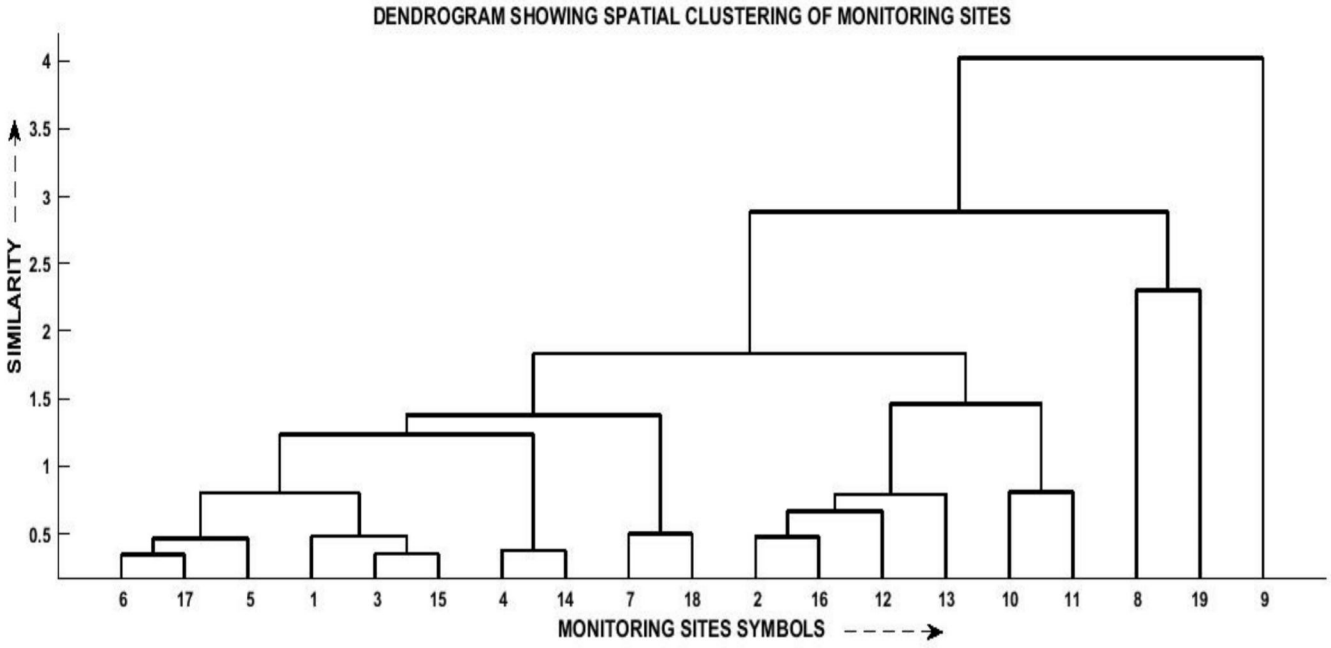


Fig 3. Dendrogram depicting monitored site clustering based on Mahanadi River basin surface water quality variables

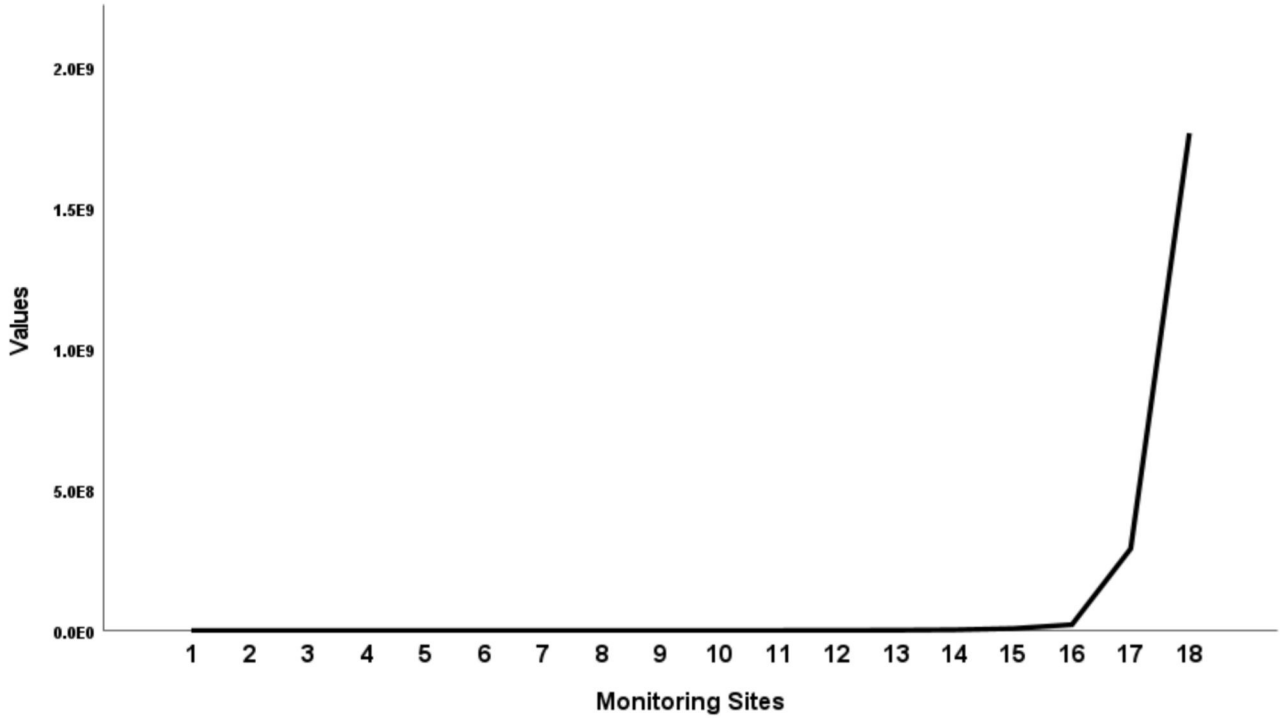


Fig 4. Distance between the clusters of all sampling sites

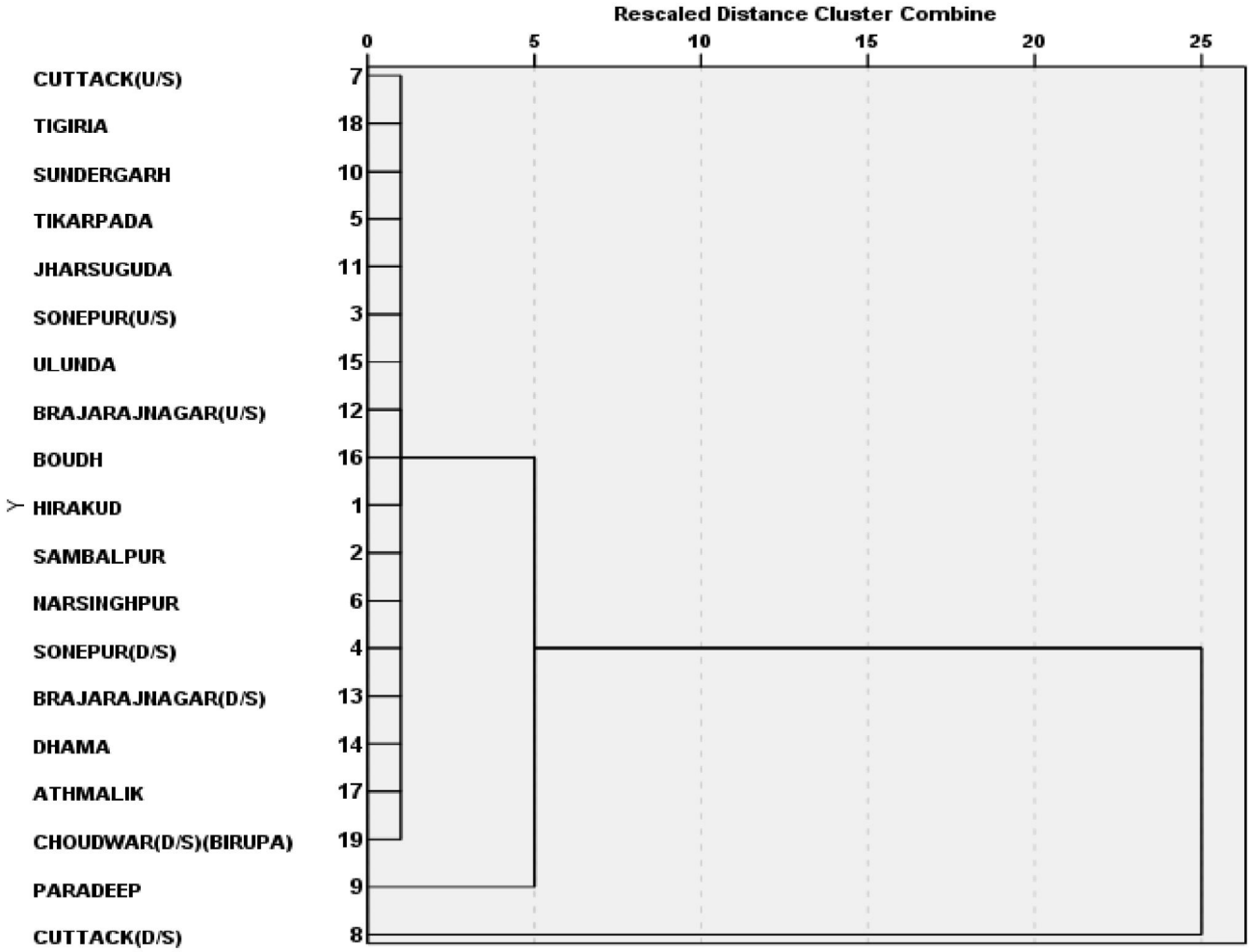


Fig 5. Dendrogram of the sampling sites

As seen in the diagram above, (Fig. 4, 5), the distance between the cluster up to seventeen sampling sites is visible and placed closer to each other, but the last two sites, 18 and 19, carry a greater distance and, depending on the pollution factor criteria, The main cause of contamination in high-pollution areas was the direct dumping of neighbouring rural residential sewage into rivers and streams. As urban living standards have improved, household water demand has outpaced sewage treatment plant capacity, leads to poor water quality in the Mahandi Basin's delta zone, which is currently being created, which includes Cuttack D/s and Paradeep U/s and D/s. The industrial effluent released into the river primarily polluted the medium pollution zones. There was no exterior contamination in the low-pollution region, which was located in the lower parts. The waterway has surpassed the cleansing of manmade wetlands and even its self-purification potential, indicating that it is in the low-pollution zone. In terms of spatial evaluation, this means that only a single site from every cluster may be as useful as the complete network for quick water quality testing. It is obvious that the CA technique is effective in delivering accurate surface water classification across the region, and that it will allow for the most efficient construction of a future spatial sampling strategy. As a result, the majority of sampling points in monitoring networks was reduced, as was the cost, without compromising the significance of the results. In other publications, this strategy has been successfully used in water quality programmes(Wunderlin et al., 2001; Simeonov et al., 2003).

Description of Each Cluster

Cluster 1 (Hirakud, Sonepur (U/s), Tikarpada, Cuttack (U/s), Sundergarh, Jharsuguda, Brajarajnagar (U/s), Ulunda, Boudh and Tigiria) - It is made up of low distance groups (Figure 6, 7) that represent fewer contaminants and correspond to ten stations with similar water properties. At these sites, they transport less pollution to the reservoir. Rajnandagaon, Bhillai, Durg, Shimoga, Raipur, Bilaspur, and Korba are just a few of the major sites and sectors along the Mahanadi River's banks, and these alone contribute a large pollution load to the reservoir. The river in Odisha discharges its impurities into the reservoir. Despite this, the water at Hirakudreservoir practically meets Class B standards, except TC levels. The river runs through a region near Sonepur (U/s) that is devoid of large urban communities or wastewater outfalls. The Mahanadi and two of its major right bank branches, Ong and Tel, meet at this point. As a result, the water quality at Sonepur U/s, which is located directly down of the Ong convergence, is excellent. Even thoughSonepur is the district headquarters and hosts all of the district's operations, the water quality has not deteriorated as much as one might think. There is no industry or urban habitation on the river's 102 km run from Sonepur D/s to Tikarpada (two minor towns in sub-divisions - Boudh and Athamalik), and there seem to be no large wastewater outfalls in this area. In the late 1990s, a paper mill operated in Sundergarh, Jharsuguda, and Brajarajnagar U/s, but it has been shut down since December 1998, andbecause none of the three municipalities is a big metropolitan center and there is little organized residential wastewater disposal to the river, water quality is normally Class C. Jharsuguda's role as the state's industrial hub has lately increased. Industrial activity, on the other hand, has a minor impact on the Ib River's water quality. The remaining sites are pollution-free, and all of the sampling locations are considered low-pollution. Total Coliform is the metric that causes the water quality to deteriorate in all of these sample sites (TC). Dendrogram and scree plot is being drawn for Cluster I illustrated in (Fig. 6, 7).

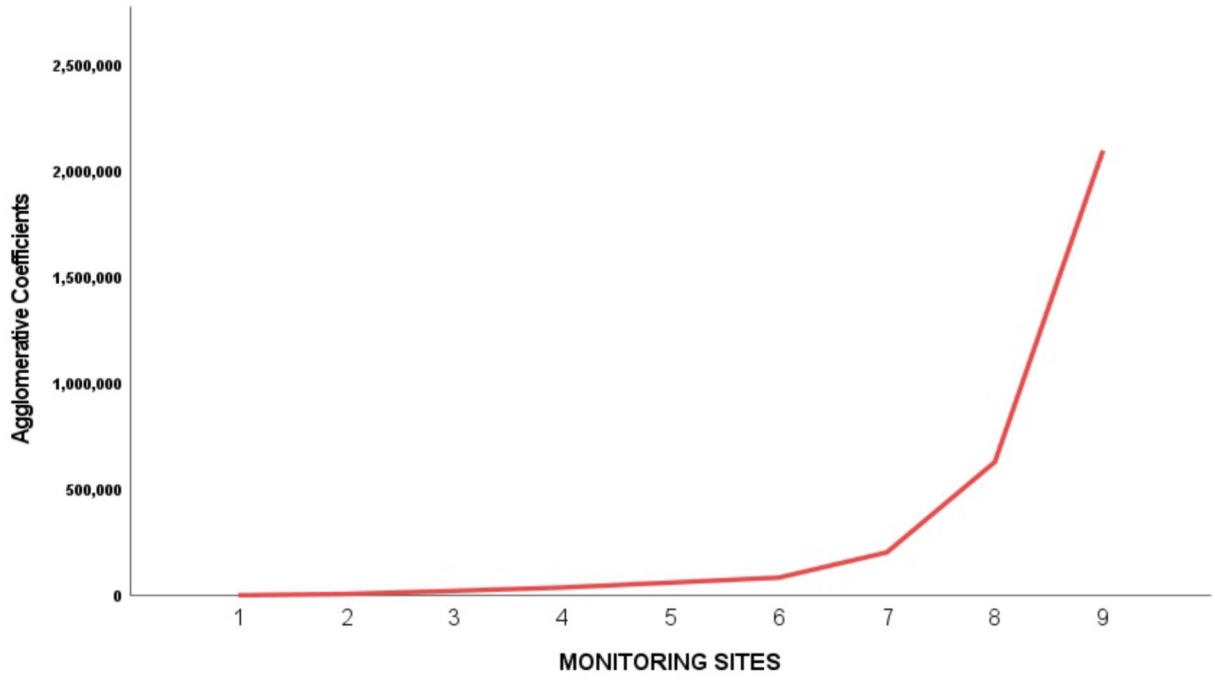


Fig 6. Distance between the clusters of all sampling sites present in Cluster 1

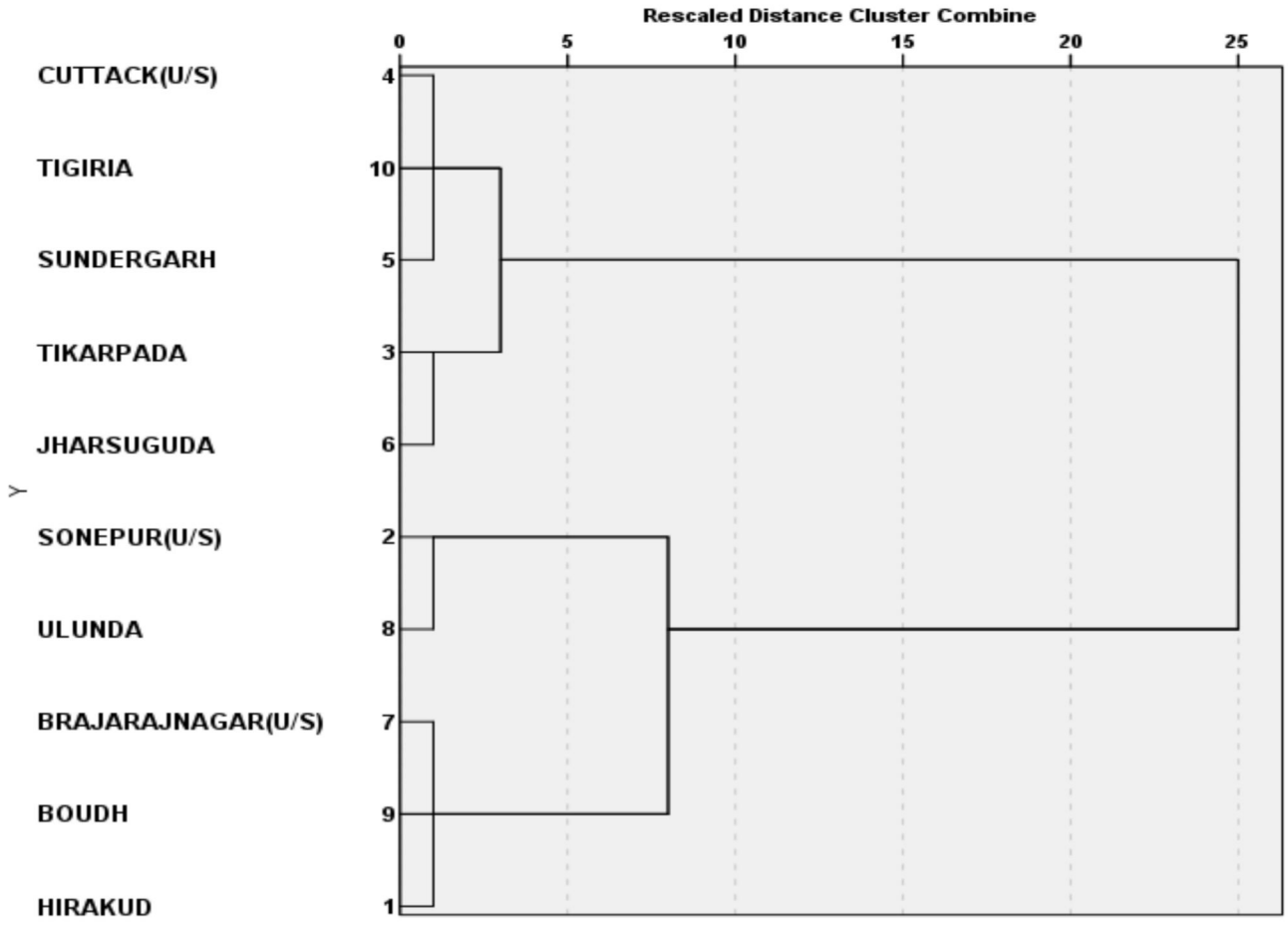


Fig 7. Cluster Ihas a dendrogram showing the CA of the sampling stations

Cluster 2 (Sambalpur, Sonepur (D/s), Narsinghpur, Brajarajnagar (D/s), Dhama, Athamalik and Choudwar D/s) – With a population of over 1.5 million people, Sambalpur is the state's largest city.The district, division, and lakhs headquarters are all located directly downstream of the Hirakud reservoir (about 5 km). The Mahanadi in Sambalpur is employed for bathing and wastewater (untreated) discharge in addition to providing drinking water, which is prompting the water supplies in Sambalpur D/s to deteriorate. Other locations, such as Dhama and Athamalik, encounter or tolerate moderate pollution loads, while Brajarajnagar D/s contributes less to water quality and conforms to Class C due to the shutdown of the paper mill. Sonepur D/s is connected to the Tel basin, which has a large yearly total flow and relatively low load. Furthermore, although being the administrative centre, it is still a tiny town with minimal evidence of urban expansion (population around 19000). Choudwar D/s, for example, was once home to substantial industrial enterprises such as a textile mill, a big pulp and paper mill, and a charge chrome firm with its thermal power plant. Only the charge chrome industry is now functioning, and water contamination is only a minor concern. In general, the water quality is Class C. The remaining locations, Narshingpur, Dhama, and Athamalik, have a moderate level of pollution. Except for Sambalpur D/s, where BOD and TC threaten water quality, TC is the dominant water quality degrading metric at all sites. Dendrogram and scree plot is being drawn for Cluster I illustrated in (Fig 8, 9).

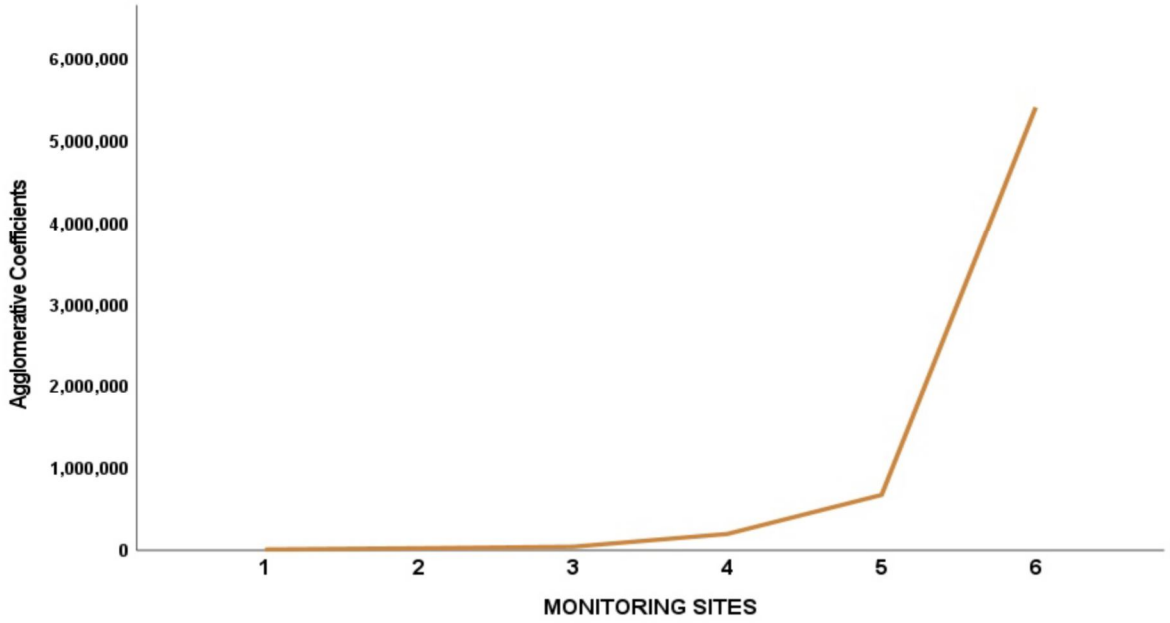


Fig 8. Distance between the clusters of all sampling sites present in Cluster 2

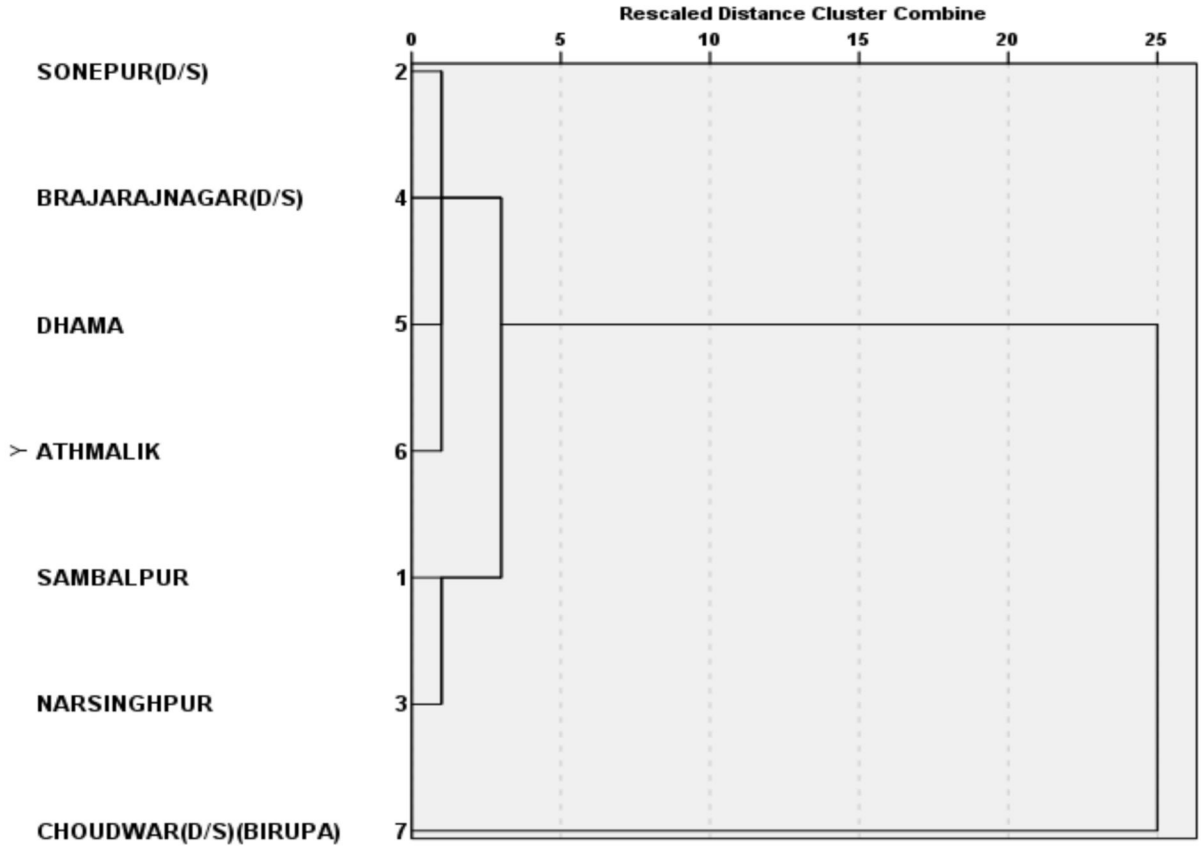


Fig 9. Cluster II has a dendrogram showing the CA of the sampling stations

Cluster 3 (Cuttack D/s and Paradeep) –This cluster is classified as extremely contaminated. The river enters its deltaic zone between Narsinghpur and Cuttack (approximately 56 km), which is marked by high population density and intensive agricultural activity. As a result, the river water ingress into Cuttack U/s has deteriorated, in terms of TC, but it still passes Class C standards. The river receives a large amount of dirty water from the city (population: 5.35 lakhs), causing the water condition at Cuttack D/s to degrade even more. Many industrial clusters in Paradeep, such as oil refineries, iron and steel, thermal power, fertilizer, and breweries, have influenced and thus exerted impacts on water and riverine discharges, which have contributed pollutants carried by agricultural runoff, factory flush outs, mine discharges, and state and city discharges from its catchment to the estuarine sea. The seashore is constantly under strain from both natural and anthropogenic causes due to the significant pollution impact. The system pressure and possible coast degradation are depicted, which could be caused by overfishing, sporadic fish culture, pollution from industries, port activities such as discharge and dredging, marine transport, associated barge discharges and accidents, and sediment drifting from riverine systems. For Cuttack D/s, the key parameters responsible for downgrading are FC, TKN, and EC, whereas, for Paradeep U/s and D/s, multiple parameters prevent it from even qualifying for Class E. (TC, EC, SAR, and Chloride). With reference to the above discussion, use-based classification categorizes the status of quality as Class A denotes a supply of drinking water that has not been treated conventionally but has been disinfected, Organized outdoor bathing is represented by Class B, Drinking waterways with conventional treatment and sterilization are classified as Class C, Fish culture and wildlife proliferation are included in Class D and finally Irrigation, industrial cooling, and monitored waste disposal are all classified as Class E.Clustering analysis and its summary of observations is being represented in spatial map as shown in (Fig 10).

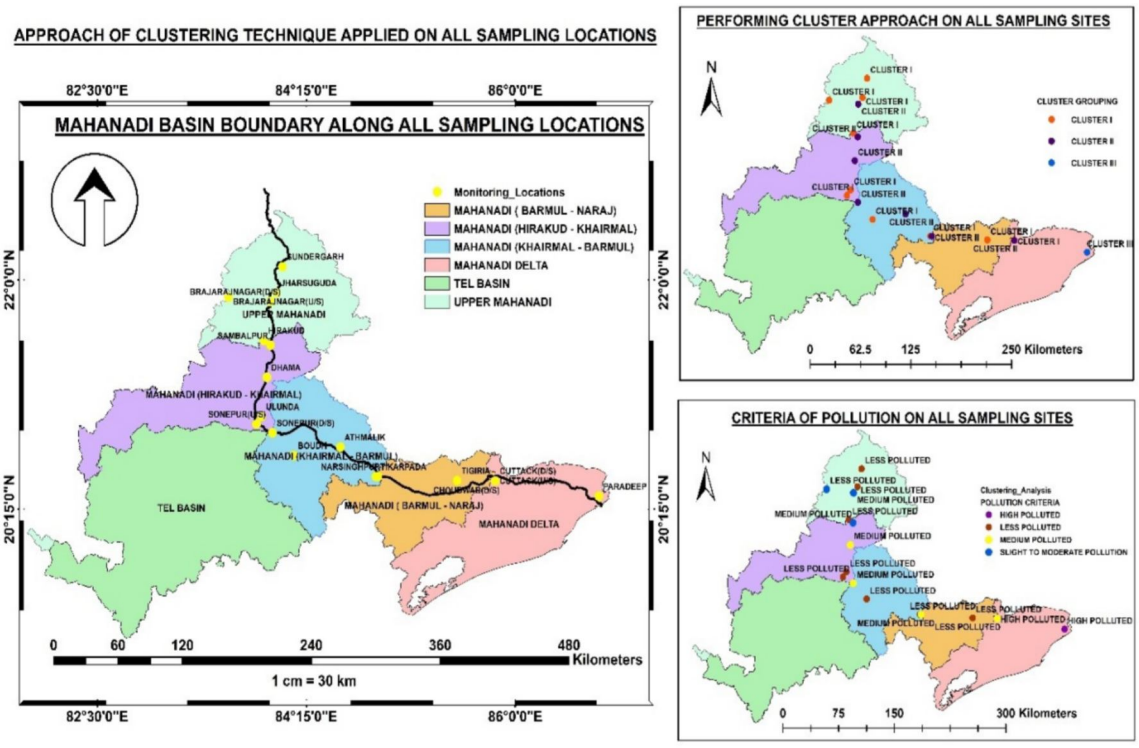


Fig 10. Overall summary of clustering approach performed on all monitoring sites

To examine the constituent trends among the studied water samples and find the variables that affect each one, FA/PCA was applied to the entire dataset. To test whether the data were accurate and valid, the Kaiser-Meyer-Olkin (KMO) and Bartlett's sphericity tests were employed to examine the association and partial link between variables. The KMO statistic (LoF, Hong et al. 2012) appears to have a value between 0 and 1. The factor analysis data is judged to have a superior effect when the KMO value is greater than 0.7. The KMO outcome is 0.716, and Bartlett's sphericity results as 1321.21 ($p < 0.05$), demonstrating that PCA may potentially reduce complexity.

Table 5. Calculation of Eigenvectors/ Component loadings

PARAMETERS	Eigenvectors (Aij)/ Component loadings			
	PC1	PC2	PC3	PC4
PH	0.044958	0.262974	-0.16739	0.209553
DO	-0.178	0.015116	0.440689	0.188453
BOD	0.065956	0.21925	-0.48907	0.047021
TC	0.083131	0.177423	-0.49602	-0.22461
TSS	0.265601	-0.06833	0.023705	-0.4221
TOTAL ALKALINITY	0.166287	0.094447	-0.11207	0.574911
COD	0.146677	0.394302	-0.15986	0.055321
NH3-N	0.04513	0.416128	0.231459	-0.0998
FREE AMMONIA	0.022544	0.431743	0.065801	-0.17034
TKN	0.256681	-0.12008	-0.07234	-0.10416
EC	0.327437	-0.10341	0.046432	0.047694
TDS	0.326566	-0.10893	0.044765	0.043508
B	0.331731	-0.03208	0.074009	0.05536
SAR	0.071776	0.374957	0.299758	-0.0609
TH	0.328094	-0.09741	0.047563	0.052329
CL2	0.327113	-0.10563	0.048588	0.040887
S04 $^{2-}$	0.327615	-0.10119	0.047345	0.043179
F	0.177195	0.335938	0.263006	0.038964
NO3	0.160128	0.010491	0.017072	0.422752
FE	0.242445	0.000805	0.122306	-0.32971
EIGEN VALUE	8.627	4.611	2.573	1.721
% VARIANCE	43.133	23.055	12.866	8.603
CUMMULATIVE %	43.133	66.189	79.054	87.658

Table 6. Calculation of Factor loadings/ Component matrix

Component Matrix/Factor Loadings					
PARAMETERS	Component				
	PC1	PC2	PC3	PC4	PC5
PH	0.132	**0.565**	-0.269	0.275	**0.677**
DO	**-0.523**	0.032	**0.707**	0.247	0.332
BOD	0.194	0.471	**-0.785**	0.062	-0.023
TC	0.244	0.381	**-0.796**	-0.295	-0.002
TSS	**0.780**	-0.147	0.038	**-0.554**	-0.101
TOTAL ALKALINITY	0.488	0.203	-0.180	**0.754**	0.092
COD	0.431	**0.847**	-0.256	0.073	-0.072
NH3-N	0.133	**0.894**	0.371	-0.131	-0.139
FREE AMMONIA	0.066	**0.927**	0.106	-0.223	0.083

TKN	**0.754**	-0.258	-0.116	-0.137	-0.274
EC	**0.962**	-0.222	0.074	0.063	0.090
TDS	**0.959**	-0.234	0.072	0.057	0.093
B	**0.974**	-0.069	0.119	0.073	0.048
SAR	0.211	**0.805**	0.481	-0.080	-0.226
TH	**0.964**	-0.209	0.076	0.069	0.091
CL2	**0.961**	-0.227	0.078	0.054	0.088
S04 $^{2-}$	**0.962**	-0.217	0.076	0.057	0.095
F	**0.520**	**0.721**	0.422	0.051	-0.110
NO3	0.470	0.023	0.027	**0.555**	**-0.562**
FE	**0.712**	0.002	0.196	-0.433	0.360
EIGEN VALUE	8.627	4.611	2.573	1.721	1.248
% VARIANCE	43.133	23.055	12.866	8.603	6.241
CUMMULATIVE %	43.133	66.189	79.054	87.658	93.899

The FA/PCA section studied and evaluated the compositional properties of water samples using standardized data to identify aspects that influenced individual water samples(Koutsias N et al., 2009). With Eigenvalues > 1, the PCA of all datasets (Table 5, 6, 7) generated five (principal component and its scores) PCs that explained 93.899 percent of total variance. TSS, TKN, EC, TDS, B, SAR, and Fe were all correlated with (loading > 0.7) in the first PC, which accounted for 43.133 percent of the total variance. The second PC was associated with COD, NH3-N, Free ammonia, and fluoride, accounting for 23.055 percent of total variance, whereas the third, fourth, and fifth PCs, accounting for 12.866 percent, 8.603 percent, and 6.241 percent of the total variance, respectively, were unrelated to any of the parameters (loading > 0.70) (Fig. 11). Principal component scores are calculated with high clarity by incorporating the native industry base and dispersal, and they foresee which station has the harmful parameter that portrays generated from industrial industries like paper manufacture, coking, synthetic raw materials and product fabrication, and metal products are all compliant with the EPA's current criteria

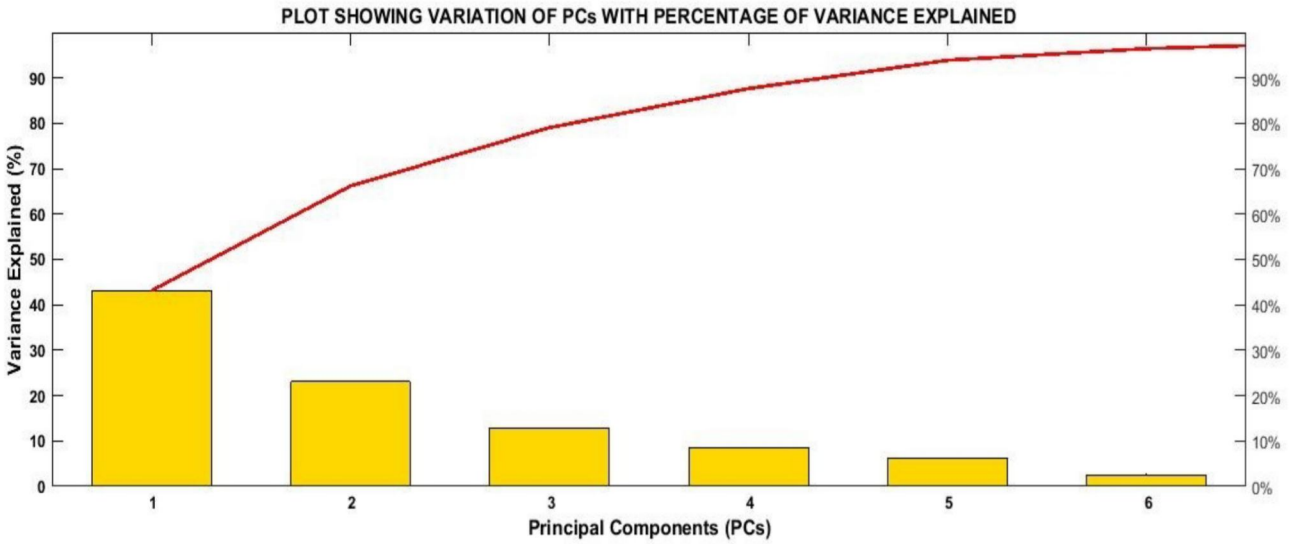


Fig 11. Scree plot for all sampling sites showing the percentage of variance explained

For all sampling sites, a scree plot depicting the proportion of variation is explained. The scree plot can be used to understand the underlying data structure. (Fig 12) indicates the number of PCs to preserve (Rayner et al., 2010). After the eighth Eigenvalue, the scree plot indicated a considerable shift in slope. On the PC subspace, the loading will be the first variable, and it matched the coefficient of correlation between PC and the factor.

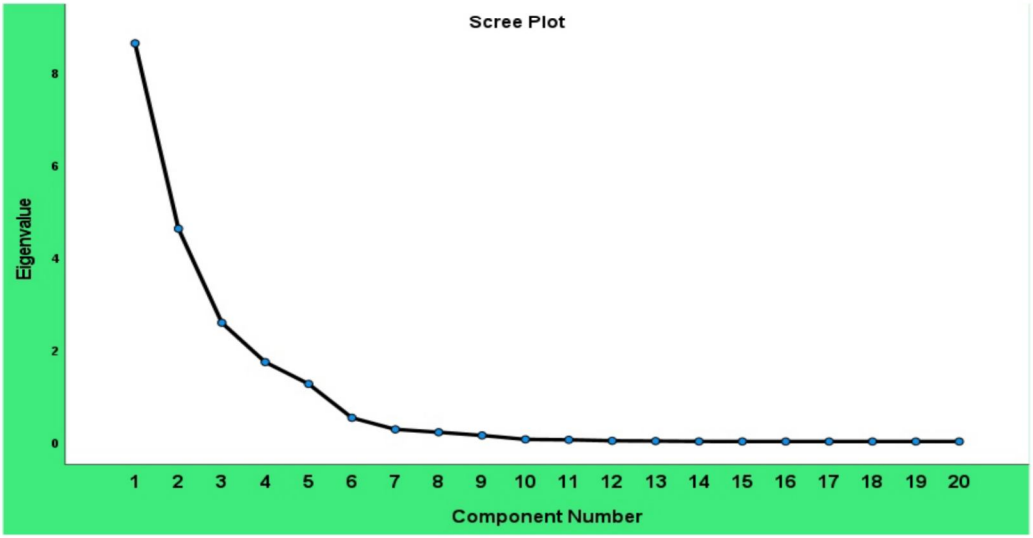


Fig 12. Scree plot for all sampling sites

The component plot (Fig. 13) reveals five Eigenvalues greater than one, indicating that they are significant, explaining 93.899 percent of the total variance. PC1 is responsible for 43.133 percent of the total variance, PC2 is responsible for 23.055 percent of the total variance, and PC3, PC4, and PC5 are responsible for 12.866 percent, 8.603 percent, and 6.241 percent of the total variance, respectively.

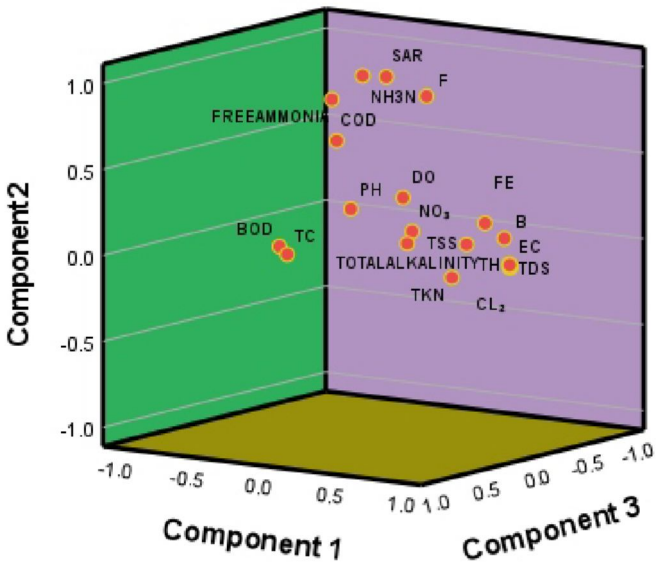


Fig 13. Component loading plot of water quality for eighteen years

Table 7. Calculation of principal component scores

PARAMETERS	Principal Component Scores			
	P1	P2	P3	P4
1	0.190	0.224	0.227	0.478
2	0.414	0.369	0.036	0.591
3	0.118	0.285	0.241	0.714
4	0.403	0.742	-0.021	0.891
5	0.314	0.360	0.255	0.411
6	0.338	0.425	0.181	0.695
7	0.186	0.668	0.274	0.252
8	0.805	1.240	-1.196	0.088
9	3.232	0.058	0.228	0.586
10	0.317	0.107	0.444	-0.363
11	0.079	0.065	0.330	0.041
12	0.298	0.051	0.349	0.374
13	0.438	0.228	0.005	0.056
14	0.426	0.788	-0.166	1.020
15	0.171	0.310	0.141	0.837
16	0.433	0.354	0.114	0.662
17	0.364	0.415	0.126	0.722
18	0.260	0.667	0.218	0.404
19	0.866	2.225	0.838	0.401

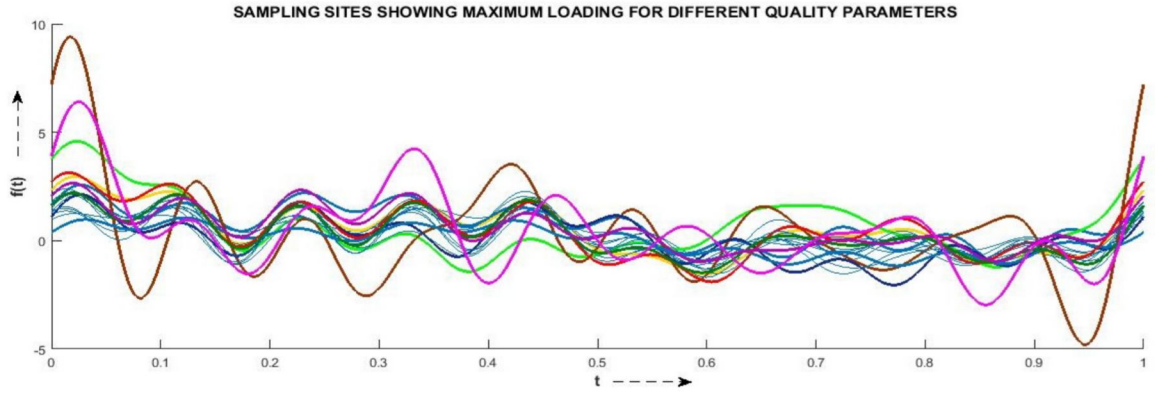


Fig 14. Andrew plot of all loadings for different water quality parameters

The above graph (Fig. 14) illustrates how higher-intensity loadings are more susceptible to pollution and can impede human and animal growth and lifestyle. The higher loading represented in brown colour indicates that the primary downgrading parameters responsible for station Paradeep are TC, EC, SAR, Cl), while the key downgrading parameters responsible for Cuttack (D/s) are FC, TKN, TC, which demonstrates slight to moderate pollution. This map explains the loading of each parameter and the amount to which it can influence the position of all sampling sites in this way. To put it another way, this graphic encourages the use of the varimax rotation technique, which produces new indices that are primarily made up of a subset of the variables used initially.

The PCA rotation axis will establish a new sequence of factors, each of which has a different effect,principally comprises similarity measure variables and has as little duplication as feasible, dividing the original variables into multiple independent groups (Dominick D et al., 2010). As a result, factor analysis (FA) of the present Mahanadi River database lowers the impact of non-significant factors contributing by PCA. The PC (original) maximum variance rotation explained five separate VFs with Eigenvalues > 1, accounting for 93.899 percent of the total variance. After implementing the maximum variance approach to rotate the data, the value of PC was exposed even more, and the original variable's involvement in VF became evident. Lieu et al. (2003) (Liu et al., 2003) categorized the factor loadings classified as "strong," "moderate," and "weak," with comparative loading values of > 0.75, 0.75-0.5, and 0.50-0.30, respectively. To assess the fundamental database schema, a scree plot (Fig. 15) with varimax rotation is employed to compute the proportion of PCs to be maintained (Jackson, 1991; Vega et al., 1998). After the eighth Eigenvalue, it exhibited a considerable shift in slope.

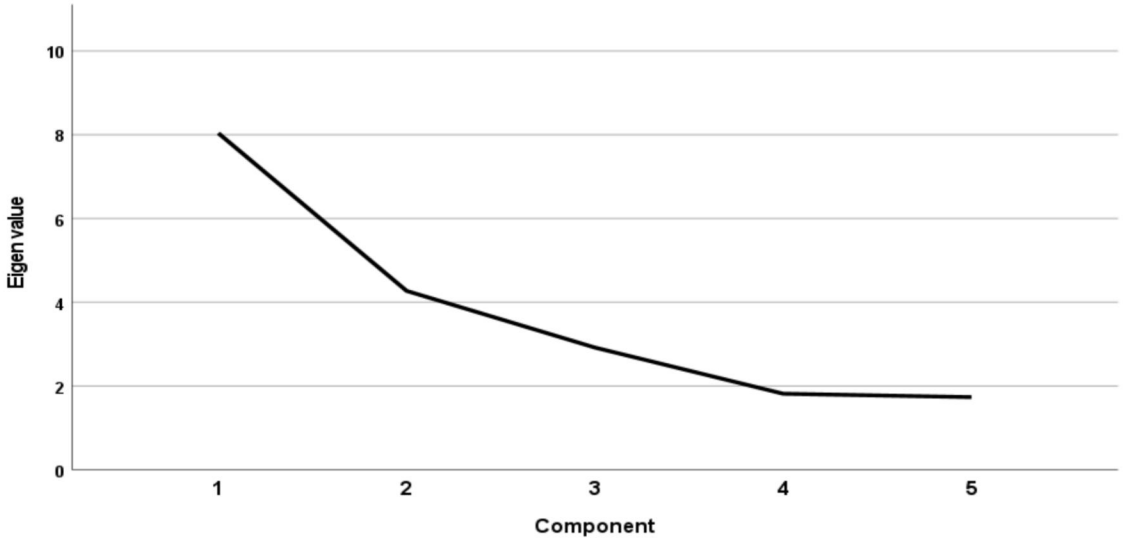


Fig 15. Scree plot for all sampling sites after performing rotation

VF1 (40.191 percent of the entire variability) had positive components that are quite robust on chloride, total dissolved solids (TDS), sulphate, EC, TH, B, TSS, Fe, and TKN, indicating metal content and solid waste decomposition (Table 8**).** This is because the primary source of river drainage is residential sewage facilities, the primary contaminant is domestic pollution sources, and the major source of heavy metal contamination is coal and metal factories upstream of the tributary. Ammonia is a natural biological disintegration product of nitrogen that uses organic components that can be found in most streams. It is found in some water bodies. It poisons fish, and the degree of toxicity is reliant on the pH of the water. It also represents the hydrochemistry of the river. The source was industrial wastewater contamination since VF2 (21.358 percent of total variation) exhibited substantial positive loadings on NH3-N, SAR, F, Free Ammonia, and COD. This solids-laden component (factors loaded) can be traced back to waste collection practices and emission from fields with substantial solids loading. This is related to the main paper business in Brajarajnagar D/s, and while the mill is closed, some paper-making activities continue on a lesser scale, degrading the quality and posing a health risk. Some metal goods and steel casting manufacturing industries near the river also contribute to water quality degradation, which can harm or endanger human health. High loadings

(significant) on TC and BOD and strong loadings (negative) on DO were seen in VF3 (14.593 percent of total variance), which were primarily manifested as pollution from heavy industries, fertilizers, thermal power plants, agricultural runoff, industrial flush outs, mine discharges, municipal discharges, and other sources that contain organic substances and encourage the growth of decomposers that consume large amounts of them. This clustering suggests that the contamination came from wastewater discharged by the chemical sector, which is linked to several chemical and metal-producing businesses across the river. VF4 showed a high nitrate loading and a moderate total alkalinity (TA) loading, indicating that the contamination emanated from municipal, industrial, and sewage discharges, accounting for 9.082 percent of the entire variance. High nitrate levels promote eutrophication because nitrate can enter water bodies. The intestinal mucosa produces less nitrite, which subsequently interacts with the circulation to produce methemoglobin, hindering oxygen delivery, high nitrate intake or consumption offers significant risk. VF5 (8.675% of entire variation) has high positive factors on PH, making it mildly alkaline. It's because home sewage contains detergents and personal hygiene products.

Table 8. Principal components (with varimax rotation)

Rotated Component Matrix					
PARAMETERS	**Component**				
	VF1	**VF2**	**VF3**	**VF4**	**VF5**
PH	0.012	0.258	0.241	-0.086	0.900
DO	-0.383	0.127	-0.843	-0.113	0.242
BOD	-0.053	0.122	0.861	0.157	0.309
TC	0.054	0.098	0.936	-0.139	0.128
TSS	0.817	0.132	0.237	-0.263	-0.371
TOTAL ALKALINITY	0.323	0.071	0.123	0.673	0.558
COD	0.136	0.713	0.560	0.213	0.303
NH3-N	-0.062	0.992	0.018	-0.007	0.036
FREE AMMONIA	-0.128	0.879	0.231	-0.199	0.224
TKN	0.735	-0.065	0.267	0.174	-0.311
EC	0.979	-0.005	0.038	0.167	0.061
TDS	0.980	-0.016	0.037	0.161	0.057
B	0.951	0.154	0.052	0.202	0.072
SAR	0.032	0.981	-0.084	0.087	-0.052
TH	0.977	0.006	0.039	0.173	0.068
CL2	0.980	-0.006	0.036	0.161	0.052
$S04^{2-}$	0.980	0.000	0.039	0.160	0.063
F	0.340	0.905	-0.045	0.189	0.102
NO3	0.307	0.129	0.050	0.842	-0.153
FE	0.790	0.220	0.014	-0.431	0.068
EIGEN VALUE	8.038	4.272	2.919	1.816	1.735
% VARIANCE	40.191	21.358	14.593	9.082	8.675
CUMMULATIVE %	40.191	61.548	76.142	85.223	93.899

Human impacts are changing and decreasing stream water quality measures directly or indirectly, according to the study's findings, and that these changes are linked to spatiotemporal variability. The river is increasingly being transformed into an industrial and municipal drainage system. The rate of deterioration could result in the extinction of biological communities of numerous organisms in the aquatic environment, making the ecosystem more vulnerable. Information on changes in physiochemical aspects of water quality through time, along with possible pollutant source identification, is thought to provide a foundation for a better understanding of the river's ecology, management, and future ecological assessment. The pollutants in the water, according to the research, devoured a large amount of oxygen. The majority of the changes are explained by a combination of soluble salts and organic contaminants, according to the study (anthropogenic). The FA can figure out which variables are most attributable to changes in river water. In an early phase, the approach of discovering temporal and spatial sources of contamination using FA/PCA was used fora water qualitative study.

CONCLUSIONS

It generates complicated textual information for water quality monitoring programmes, which needs multivariate statistical processing to evaluate and grasp the data's vital information. A multitude of multivariate statistical approaches was used in this work to examine the geographical and temporal fluctuations in the Mahanadi River's surface water quality. These methods were found to be useful in explaining synoptic fluctuations in water quality metrics that contributed appreciably to alterations in stream water chemistry and quality. The study discovered distinct geographical and temporal variability in water quality measures, in addition to the fact that surface water pollution is spatially heterogeneous due to anthropogenic influences. According to the similarities of river basin features and pollutants, cluster analysis (CA) categorized the 19 data points into three categories. It provides a suitable foundation for describing surface water in the examined area and, by eliminating data loss, it can significantly reduce the number of sample sites required to investigate the riverbed. In comparison to other strategies used for historical and geographical investigations, it supports the optimum data acquisition and pattern identification method. Only a little quantity of data was decreased, even though the FA/PCA identified the five factors required to explain 93.899 percent of fluctuations in the data frame.The six VFs procured, which explained VF1 (40.191 percent of total variance), VF2 (21.358 percent of total variance), VF3 (14.593 percent of total variance), VF4 (9.082 percent of total variance), and VF5 (8.675 percent of the total variance, indicated that river water quality parameters were primarily divided into natural (dissolved salts) and manmade (organic substances) attribution.Local contaminant intakes from unregulated sewage sludge, as well as farmland runoffs, appeared to be linked to the water shortages. To reduce the accumulation of pollutants in water and soil, as well as to prevent environmental degradation, proper treatment of home and industrial wastes is essential. This can be accomplished by constructing municipal solid waste landfills, before releasing it into the environment, residential and industrial effluent is treated and improving agricultural practices. This data also revealed advancements in farming methods. The findings also demonstrated the need of performing multivariate statistical analyses on huge datasets to obtain more information about river water. This can help environmental managers make better awareness campaign judgments. As a result, multivariate statistical algorithms are an ideal exploring tool for evaluating and perceiving complicated data sets linked to water quality, as well as understanding how they evolve and space.

REFERENCES

Acar, O. 2012. Evaluation of cadmium, lead, copper, iron and zinc in Turkish dietary vegetable oils and olives using electrothermal and flame atomic absorption spectrometry. GRASAS ACEITES.; 63(4):383–93.

Amir, M., Khan, A., Mujeeb, M., Ahmad, A., Usmani, S., Akhtar, M. 2011. Phytochemical Analysis and in vitro Antioxidant Activity of Zingiber officinale. Free Radicals and Antioxidants. 1(4):75–81.

APHA, 1992. Standard Methods for the Examination of Water and Wastewater, eighteenth ed. American Public Health Association, Washington, DC.

Aristizabal, J., Giraldo, R. and Mateu, J. 2019. Analysis of variance for spatially correlated functional data: Application to brain data. SPAT STAT-NETH; 32:100381.

Bengraine, K. and Marhaba, T.F. 2003. Using principal component analysis to monitor spatial and temporal changes in water quality. J. Hazard. Mater. B 100:179–195.

Bouza-Deano, R., Ternero-Rodrıguez, M. and Fernandez-Espinosa, A.J. 2008. Trend study and assessment of surface water quality in the Ebro River (Spain). J Hydrology. 361(3–4): 227–39.

Brown, S.D., Sum, S.T. and Despagne, F. 1996. Chemometrics. Anal. Chem. 68: 21R–61R.

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N. and Smith, V. H. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol. Appl. 8:559–568.

Chapman, D. 1992. Water quality assessment. In: Chapman, D. (Ed.), on behalf of UNESCO, WHO and UNEP, Chapman & Hall, London, 585.

Chattopadhyay, T., Sharina, M., Davoust, E., De, T. and Chattopadhyay, A.K. 2012. Uncovering the Formation of Ultracompactdwarf Galaxies by Multivariate Statistical Analysis. The Astro- physical Journal, 750(2): 91.

Chow, M.F., Shiah, F.K., Lai, C.C., Kuo, H.Y., Wang, K.W., Lin, C.H., et al. 2016. Evaluation of surface water quality using multivariate statistical techniques: a case study of Fei-Tsui Reservoir basin, Taiwan. Environ Earth Sci.75(1).

Conaway, C.H., Ross, J.R.M., Looker, R., Mason, R.P. and Flegal, A.R. 2007. Decadal mercury trends in San Francisco Estuary sediments. Environ Res.; 105(1):53–66.

Corporal-Lodangco, I.L., Richman, M.B., Leslie, L.M. and Lamb, P.J .2014. Cluster Analysis of North Atlantic Tropical Cyclones. Procedia Computer Science. 36:293–300.14.

Cruz, R.C.D. and Eler, M.M. 2017. Using a cluster analysis method for grouping classes according to their inferred testability: An investigation of CK metrics, code coverage and mutation score, IEEE. 1(1):1–11.

Ding, S., Jia, W., Su, C., Zhang, L. and Liu, L. 2011. Research of neural network algorithm based on factor analysis and cluster analysis. Neural Computing and Applications. 20(2):297–302.

Dixon, W. and Chiswell, B. 1996. Review of aquatic monitoring program design. Water Res. 30: 1935–1948.

Dominick, D., Juahir, H., Latif, M.T., Zain, S.M. and Aris, A.Z. 2012. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. Atmos Environ. 60:172–81.

Duan, W., He, B., Nover, D., Yang, G., Chen, W., Meng, H., et al. 2016. Water Quality Assessment and Pollution Source Identification of the Eastern Poyang Lake Basin Using Multivariate Statistical Methods. Sus- Tainability-Basel, 8(2):133.

Ebrahimiyan, M., Majidi, M.M. and Mirlohi, A. 2013. Genotypic variation and selection of traits related to forage yield in tall fescue under irrigated and drought stress environments. Grass Forage Sci, 68(1):59– 71.

El-Ashtoukhy, E.S.Z. and Fouad, Y.O. 2015. Liquid-liquid extraction of methylene blue dye from aqueous solutions using sodium dodecyl benzenesulfonate as an extractant. Alexandria Engineering Journal, 54 (1):77–81.

El-Sheikh, A.H., Al-Quse, R.W., El-Barghouthi, M.I. and Al-Masri, F.A.S. 2010. Derivatization of 2-chlorophenol with 4- amino-anti-pyrine: A novel method for improving the selectivity of molecularly imprinted solid phase extraction of 2-chlorophenol from water. Talanta. 83(2):667–73.

Friedel, M.J. 2006. Predictive streamflow uncertainty in relation to calibration-constraint information, model complexity, and model bias. International journal of river basin management, 4 (2):109–23.

Gashi, S., Bajmaku, Y. and Drini, P. 2016. Comparison of some Chemical Parameters of Urban and Industrial Water Discharges Onto the River Lumbardh of Prizren. IFAC-PapersOnLine, 49(29):129–32.

Gonzalez, M.J.G. and Vallejo-Pascual, M. 2018. The Application of Principal Component Analysis (PCA) for the Study of the Spanish Tourist Demand. Questioners geographical, 37(4):43–52.

Guo, L., Zhao, Y. and Wang, P. 2012. Determination of the principal factors of river water quality through cluster analysis method and its prediction. Front Env Sci Eng., 6(2):238–45.

Hagedorn, C., Robinson, S.L., Filtz, J.R., Grubbs, S.M., Angier, T.A. and Reneau Jr., R.B. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Appl. Environ. Microbiol. 65:5522–5531.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M. and Fernandez, L. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. Water Res. 34:807–816.

Huang, M.D., Becker-Ross, H., Florek, S., Heitmann, U. and Okruss, M. 2005. Direct determination of total sulfur in wine using a continuum-source atomic-absorption spectrometer and an air–acetylene flame. Anal Bioa- Nal Chem., 382(8):1877–81.

Jackson, J.E. 1991. A User's Guide to Principal Components. Wiley, New York. Jarvie, H.P., Whitton, B.A., Neal, C., 1998. Nitrogen and phosphorus in east-coast British rivers: speciation, sources and biological significance. Sci. Tot. Environ. 210–211: 79–109.

Jebb, A.T., Parrigon, S. and Woo, S.E. 2017. Exploratory data analysis as a foundation of inductive research. Hum Resourmanage R., 27(2):265–76.

Johnson, R.A. and Wichern, D.W. 1992. Applied Multivariate Statistical Analysis, ed. Prentice-Hall International, Engle- wood Cliffs, NJ, USA, 642.

Kazi, T.G., Arain, M.B., Jamali, M.K., Jalbani, N., Afridi, H.I. and Sarfraz, R.A. 2008. Assessment of water quality of polluted lake using multivariate statistical techniques: A case study. Ecotox Environ Safe, 72(2):301–9.

Koutsias, N., Mallinis, G. and Karteris, M. 2009. A forward/backward principal component analysis of Landsat-7 ETM + data to enhance the spectral signal of burnt surfaces. Isprs J Photogramm, 64(1):37–46.

Li, H., Liu, J., Liu, R., Xiong, N., Wu, K. and Kim, T. 2017. A Dimensionality Reduction-Based Multi-Step Clustering Method for Robust Vessel Trajectory Analysis. Sensors-Basel. 17(8):1792.

Liu, C., Lin, K. and Kuo, Y. 2003. Application of factor analysis in the assessment of groundwater quality in a black-foot disease area in Taiwan. Sci Total Environ. 313(1–3):77–89.

Liu, G., Ma, F., Liu, G., Zhao, H., Guo, J. and Cao, J. 2019. Application of Multivariate Statistical Analysis to Identify Water Sources in A Coastal Gold Mine, Shandong, China. Sustainability-Basel, 11(12):3345.

Lo, F., Hong, J., Lin, M. and Hsu, C. 2012. Extending the Technology Acceptance Model to Investigate Impact of Embodied Games on Learning of Xiao-zhuan. Procedia—Social and Behavioral Sciences, 64:545–54.

Massart, D.L., Vandeginste, B.G.M., Deming, S.N., Michotte, Y. and Kaufman, L. 1988. Chemometrices: A Textbook. Elsevier, Amsterdam.

Mohamad Asri, M.N., Mat Desa, W.N.S. and Ismail, D. 2020. Combined Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA): an efficient chemometric approach in aged gel inks discrimination. Aust J Forensic Sci, 52(1):38–59.

Morales, M.M., Marti, P., Llopis, A., Campos, L. and Sagrado, S. 1999. An environmental studybyfactor analysis of surface seawaters in the gulf of Valencia (Western Mediterranean). Anal. Chim. Acta, 394: 109–117.

Muangthong, S. and Shrestha, S. 2015. Assessment of surface water quality using multivariate statistical techniques: case study of the Nampong River and Songkhram River, Thailand. Environ Monit Assess., 187(9).

Nowak, B., Nadolna, A. and Stanek, P. 2018. Evaluation of the potential for the use of lakes in restoring water resources and flood protection, with the example of the NotećZachodnia River catchment (Gniezno Lakeland, Poland). Meteorology Hydrology and Water Management.

Oda, R., Suzuki, Y., Yanagihara, H. and Fujikoshi, Y. 2020. A consistent variable selection method in high-dimensional canonical discriminant analysis. J Multivariate Anal. 175:104561.

Otto, M. 1998. Multivariate methods. In: Kellner, R., Mermet, J.M., Otto, M., Widmer, H.M. (Eds.), Analytical Chem- istry. Wiley-VCH, Weinheim, Germany 916.

Paıga, P., Mendes, L., Albergaria, J. and Delerue-Matos, C. 2012. Determination of total petroleum hydrocarbons in soil from different locations using infrared spectrophotometry and gas chromatography. Chem Pap, 66(8).

Parveen, S., Portier, K.M., Robinson, K., Edmiston, L. and Tamplin, M.L. 1999. Discriminant analysis of ribotype profiles of Escherichia coli for differentiating human and nonhuman sources of fecal pollution. Appl. Environ. Microbiol. 65: 3142–3147.

Pehlivan, E. and Arslan, G. 2007. Removal of metal ions using lignite in aqueous solution—Low cost biosorbents. Fuel Process Technol, 88(1):99–106.

Rayner, J., Williams, H.M., Lawton, A. and Allinson, C.W. 2010. Public Service Ethos: Developing a Generic Measure. J Publadmres Theor, 21(1):27–51.

Reghunath, R., Murthy, T.R.S. and Raghavan, B.R 2002. The utility of multivariate statistical techniques in hydro-geochemical studies: an example from Karnataka, India. Water Res. 36: 2437–2442.

Reisenhofer, E., Adami, G. and Barbieri, P.1998. Using chemical and physical parameters to define the quality of karstic freshwaters (Timavo River, North-eastern Italy): a chemometric approach. Water Res. 32: 1193–1203.

Rezaee, F. and Jafari, M. 2015. The effect of marketing knowledge management on sustainable competitive advantage: Evidence from banking industry. Accounting, 69–88.

Richman, M.B. 1986. Rotation of principal components. J. Climatol. 6: 293–335.

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. and Kouimtzis, Th. 2003. Assessment of the surface water quality in Northern Greece. Water Res., 37: 4119–4124.

Singh, C.K., Shashtri, S. and Mukherjee, S. 2011. Integrating multivariate statistical analysis with GIS for geochemical assessment of groundwater quality in Shiwaliks of Punjab, India. Environ Earth Sci., 62 (7):1387–405.

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):3980–92.

Singh, K.P., Malik, A. and Sinha, S. 2005. Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. Anal Chim Acta., 538(1–2):355–74.

Toman, R., Hluchy, S., Massanyi, P., Lukac, N., Adamkovicova, M. and Cabaj, M. 2014. Selenium and Cadmium Tissue Concentrations and the CASA Sperm Motility Analysis after Administration to Rats. American Journal of Animal and Veterinary Sciences, 9(4):194–202.

Tri Wahyuni, E. 2016. Photodegradation of Detergent Anionic Surfactant in Wastewater UsingUV/ Processes. American Journal of Applied Chemistry, 4(5):174.

Twine, T.E., Kucharik, C.J. and Foley, J.A. 2005. Effects of El Niño–Southern Oscillation on the Climate, Water Balance, and Streamflow of the Mississippi River Basin. J Climate, 18(22):4840–61.

Vargas, F.D., Hoffmeister, F.X., Prates, P.F. and Vasconcellos, S.J.L. 2015. Depressão, ansiedade e psicopatia: um estudocorrelacional com indivı´duos privados de liberdade. JornalBrasileiro de Psiquiatria, 64 (4):266–71.

Varol, M., Gokot, B., Bekleyen, A. and Şen, B. 2012. Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. Catena, 92:11–21.

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.

Voncina, D.B., Dobcnik, D., Novic, M. and Zupan, J. 2002. Chemometric Characterisation of the quality of river water. Anal. Chim. Acta, 462: 87–100.

Wiggins, B.A., Andrews, R.W., Conway, R.A., Corr, C.L., Dobratz, E.J., Dougherty, D.P., EppardKnupp Jr., S.R., Limjoco, M.C., Mettenburgm, J.M., Rinehardtm, J.M., Sonsinom, J., Torrijosm, R.L. and Zimmerman, M.E. 1999. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. Appl. Environ. Microbiol. 65: 3483–3486.

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.

Yang, J.S., Hu, X.J., Li, X.X., Li, H.Y. and Wang, Y. 2012. Application of Principal Component Analysis (PCA) for the Estimation of Source of Heavy Metal Contamination in Sediments of Xihe River, Shenyang City. Advanced Materials Research, 610–613:948–51.

Yidana, S.M., Sakyi, P.A. and Stamp, G. 2011. Analysis of the Suitability of Surface Water for Irrigation Purposes: The Southwestern and Coastal River Systems in Ghana. Journal of Water Resource and Protection; 03(10):695–710.

Yongchao, L., Yongjiang, D., Ersi, K., Quanjie, M.A. and Jishi, Z. 2003. The relationship between ENSO cycle and high and low—flow in the upper Yellow River, Journal of Geographical Sciences, 13(1):105–11.

Younis, A.M., Ismail, I.S., Mohamedein, L. and Ahmed, S.F. 2015. Spatial Variation and Environmental Risk Assessment of Heavy Metal in the Surficial Sediments along the Egyptian Red Sea Coast. Catrina: The International Journal of Environmental Sciences, 10(1):45–52.

Zhang, W., Li, X., Liu, T. and Li, F. 2012. Enhanced nitrate reduction and current generation by Bacillus sp. in the presence of iron oxides. J Soil Sediment, 12(3):354–65.

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(1–4):17–27.

Zhao, X., Liu, D., Huang, H., Zhang, W., Yang, Q. and Zhong, C. 2014. The stability and defluoridation performance of MOFs in fluoride solutions. Micropormesopor Mat., 185:72–8.

❑❑❑