© 2021 The Authors
692

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

Water quality evaluation and apportionment of pollution sources: a case study of the Baralia and Puthimari River (India)

Kunwar Raghvendra Singh[a,*], Ankit Pratim Goswami[b], Ajay S. Kalamdhad[b] and Bimlesh Kumar[b]

[a] Department of Civil Engineering, GLA University, Mathura 281406, India

[b] Department of Civil Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, India

*Corresponding author. E-mail: s.kunwar@iitg.ac.in

Abstract

Water quality monitoring programs are indispensable for developing water conservation strategies, but elucidation of large and random datasets generated in these monitoring programs has become a global challenge. Rapid urbanization, industrialization and population growth pose a threat of pollution for the surface water bodies of the Assam, a state in northeastern India. This calls for strict water quality monitoring programs, which would thereby help in understanding the status of water bodies. In this study, the water quality of Baralia and Puthimari River of Assam was assessed using cluster analysis (CA), information entropy, and principal component analysis (PCA) to derive useful information from observed data. 15 sampling sites were selected for collection of samples during the period May 2016- June 2017. Collected samples were analysed for 20 physicochemical parameters. Hierarchal CA was used to classify the sampling sites in different clusters. CA grouped all the sites into 3 clusters based on observed variables. Water quality of rivers was evaluated using entropy weighted water quality index (EWQI). EWQI of rivers varied from 61.62 to 314.68. PCA was applied to recognise various pollution sources. PCA identified six principal components that elucidated 87.9% of the total variance and represented surface runoff, untreated domestic wastewater and illegally dumped municipal solid waste (MSW) as major factors affecting the water quality. This study will help policymakers and managers in making better decisions in allocating funds and determining priorities. It will also assist in effective and efficient policies for the improvement of water quality.

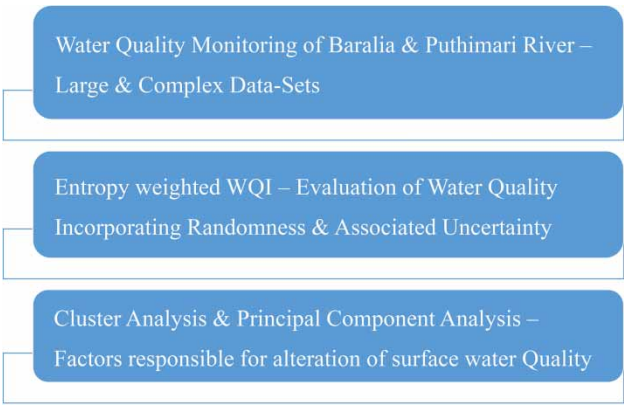
Key words: cluster analysis, information entropy, principal component analysis, water quality

Highlights

- Evaluation of water quality using entropy weighted WQI.
- Entropy provides valuable descriptions of random processes.
- Classification of sampling sites by using Hierarchal cluster analysis based on observed parameters.
- Application of principal component analysis to recognize the latent pollution sources.
- Study will assist policy makers to make better decisions for surface water quality management.

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

Graphical Abstract



INTRODUCTION

Rivers have always been the most significant freshwater source for life, social progress, and economic development, as ancient civilisations have prospered along with them (Varol *et al.* 2012). Anthropogenic activities and natural processes deteriorate surface water quality and adversely affect their importance (Singh *et al.* 2004; Dash *et al.* 2020; Prasad *et al.* 2020). Due to domestic and industrial wastewater discharges, agricultural runoff and uncontrolled dumping of municipal solid waste near the river banks, the surface water quality has been gravely affected (Singh *et al.* 2004, 2018, 2019; Shrestha & Kazama 2007; Dash *et al.* 2018; Zavaleta *et al.* 2021). In the past few decades, surface water quality has gained utmost importance, especially in developing countries like India, and has become a sensitive issue (Simeonov *et al.* 2003; Singh *et al.* 2019; Borah *et al.* 2020).

Assam, a north-eastern state of India, situated between Latitude 90° to 96° North and 24° to 28° East, has been forever a host to cultural diversity and ethnicity. An interesting quote from the 19th century writes 'The number and magnitude of rivers in Assam probably exceed those of any other country in the world of equal extent'. However, in the past few decades, there has been the rehabilitation of communities along with rapid urbanisation and industrial growth, which has taken a toll on the surface water quality of the state (Singh *et al.* 2019). There has been an increasing demand for water quality monitoring and policies to diminish the additional stresses on rivers. Reliable information on water quality and the identification of pollution sources is essential for preventing and controlling surface water pollution (Bu *et al.* 2010). Water pollution is defined as the presence of natural organic matter, which is a complex mixture of various organic molecules mainly originating from aquatic organisms, soil and terrestrial vegetation and toxic chemicals that exceed what is naturally found in the water and may pose a threat to the environment (Avsar *et al.* 2014; Avsar *et al.* 2015).

Pollutants compromising the health of river systems depend on the economic and social characteristics of the beneficiary/user societies (Lekkas *et al.* 2004). Environmental protection agencies monitor water quality based on comprehensive sets of indicators. In order to guard the ecological status, the Water Framework Directive declared that not only chemical concentrations of pollutants in rivers are to be used to assess water quality, but also its effects on trophic chains. However, chemical monitoring of parameters will continue to be an important data source. Monitoring of water resource is vital for reliable information about its quality and to prevent and control its pollution (Zavaleta *et al.* 2021). Major issues associated with water quality monitoring are handling huge and complex data sets generated due to many water quality parameters at different sampling locations and deriving useful information from them. Application of water quality indices (WQIs) and various multivariate statistical techniques (MSTs) such as cluster analysis (CA) and principal component analysis (PCA) offers a better understanding of data (Singh *et al.* 2017; Zavaleta *et al.* 2021).

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

In the present paper, Baralia and Puthimari River water quality has been assessed and the possible sources of pollution have been identified to understand the status of water quality. This will help in developing policies to reduce the additional stresses on these surface water resources. For the evaluation of the quality of river water, water quality is expressed in terms of entropy-weighted water quality index (EWQI). The concept of modern WQI was introduced in 1960 (Sutadian *et al.* 2016). Since then, many indices have been proposed, but there is no globally accepted WQI. There is a lot of subjectivity and uncertainty involved in WQI development. Water quality parameters are random variables, and their probability distribution affects the index's probability distribution (Landwehr 1979). Assignment of fixed weights of indices based on the indices' inherent information would reduce subjective disturbances (Li *et al.* 2011). Such information may be explained by Shannon or information entropy.

EWQI tries to provide an improved method for offering a cumulatively derived, numerical expression describing a certain level of quality of water based on information entropy (Li *et al.* 2010; Amiri *et al.* 2014; Fagbote *et al.* 2014; Gorgij *et al.* 2017; Karunanidhi *et al.* 2020). Information theory involves quantifying information and analyses the statistical structure of a series of numbers or symbol that builds a communication signal (Ozkul *et al.* 2000; Liu *et al.* 2012). Entropy refers to the randomness of a system and the concept of information entropy was introduced by Claude Shannon in 1948, which is also commonly known as Shannon entropy. Shannon entropy is the expected value of a random variable formed by information generated by any event or a particular set of events. The entropy concept of information theory has been successfully used in the various water resource and environmental engineering fields. In this study, the concept of entropy is used to determine water quality parameters' contribution to calculate the WQI.

Hierarchical cluster analysis (HCA) was applied to identify similar sites based on their characteristics. PCA is a useful tool for data reduction and explains inter-correlated variables' variance by transmuting them into a smaller set of independent variables (Yang *et al.* 2010; Dash *et al.* 2018). Over the last three decades, researchers have widely used these methods in surface water quality assessment. In this study, HCA and PCA methods have been used to identify the parameters responsible for water quality variations. The novelty of the present work reflects the combined use of WQI and MSTs in water quality monitoring and management. WQI evaluates the quality of water and MSTs recognise the unobservable, latent pollution sources of water bodies.

MATERIALS AND METHODS

Study area

Sample collection was done from Baralia River and Puthimari River during the period of May 2016 to June 2017. Baralia and Puthimari Rivers are northern bank tributaries of Brahmaputra River (Figure 1). Puthimari River originates from the foothills of the Himalayan Ranges in Bhutan. After crossing the Indo-Bhutan border, it bifurcates into Baralia and Puthimari Rivers near Bornodi Wildlife sanctuary, Arangajuli, Assam (26°43′24.82″N, 91°41′8.25″E), and possesses all the characteristics of a flashy river. Length of river Baralia is approximately 39.1 km and that of Puthimari River is 139 km. It meanders freely and has many loops, the slope being somewhat flatter in its lower reaches. Baralia River flows through the heart of Rangia. Rangia is a town in Kamrup rural district of Assam, whereas Puthimari River flows through the outside of the city. According to the provisional report of the 2011 Census of India, Rangia had a population of 26,389. Males account for 54% of the population and females for 46% of this population. The region has a humid subtropical climate with heavy rainfall, hot summer and high humidity. The average temperature varies from 12 to 38 °C during the year. The principal food crops produced in the region are rice (paddy) and vegetables. Heavy floods also characterise the region due to high rainfall during monsoon.

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

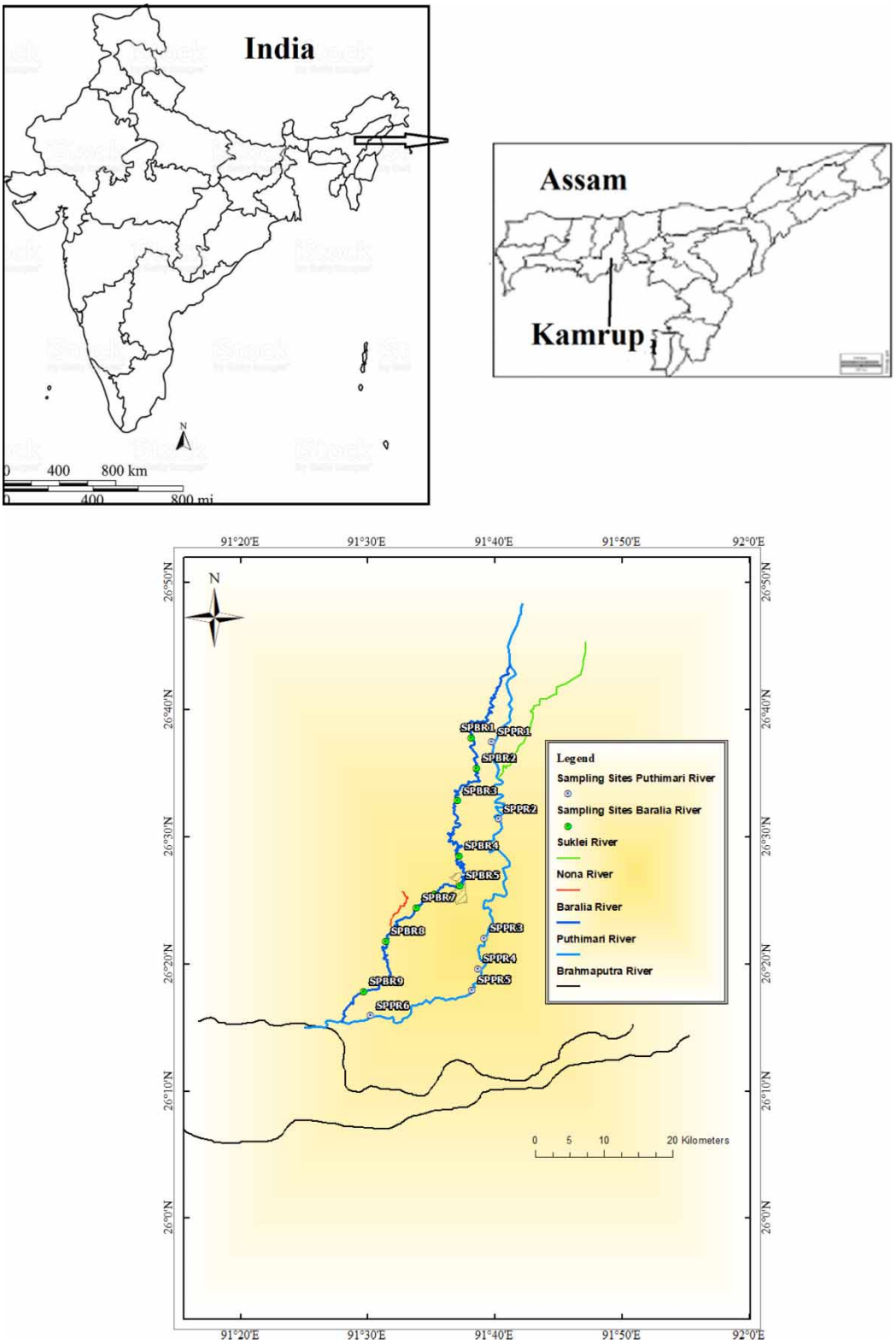


Figure 1 | Study area and location of sampling points.

Site sampling, preservation and analysis

Sampling was done from the two rivers from May 2016 to June 2017. Before sampling, a preliminary survey of the catchment area was carried out to decide the sampling sites' location and identify the

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

various point and nonpoint pollution sources. Prior information on the basic characteristics of the catchment area or basin is required before applying the mathematical or statistical tools on the measured parameters to validate and interpret the results judiciously (Alberto *et al.* 2001). Nine sites of the Baralia River and six sites of the Puthimari River were selected as sampling sites.

Water samplings were carried out in triplicate, from the well-mixed section of the rivers. Clean plastic bottles of 1 L capacity were used for collection of the water samples. Samples were collected in two forms, preserved samples (for the analysis of heavy metals) and non-preserved. For the preserved samples, HNO_3 (2 mL/L) was added to ensure pH ≤ 2. *Standard Methods for the Examination of Water and Wastewater* (APHA 2012) were adopted to analyse the samples. Temperature, pH, EC, DO and turbidity were determined in-situ. Quality control was maintained as recommended in the standard methods. Parameters such as pH, EC, and turbidity were analysed as early as possible in the laboratory since there is a change in the properties over time. Samples were protected from contamination and deterioration before their arrival in the laboratory. After collection, samples were immediately placed in a lightproof insulated box containing melting ice-packs to ensure rapid cooling. Reagents were prepared as recommended by standard methods (APHA 2012). Deionised water was used for carrying out the dilutions. Standard solutions were prepared by diluting the stock solutions. The water quality parameters analysed are shown in Table 1 along with the units, abbreviations and analytical methods used.

Entropy weighted water quality index (EWQI)

WQI is a single arithmetic number, based on a weighted average of selected parameters that express overall water quality. Assignment of weight to each selected parameter is an important and challenging task. Generally, assignment of weight to water quality parameters is a matter of opinion and hence subjective (Abbasi & Abbasi 2012). In this study, an entropy-based weight is assigned to each parameter. Entropy and related information measures offer valuable descriptions of the long term behaviour of random processes. Steps involved in the calculation of EWQI are as follows (Li *et al.* 2010; Amiri *et al.* 2014; Fagbote *et al.* 2014; Gorgij *et al.* 2017):

- A matrix 'U' was developed with all 'm' water samples (i = 1,2,…m) and 'n' measured parameters (j = 1,2,…n)

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & \cdots & u_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{m1} & u_{m2} & \cdots & \cdots & u_{mn} \end{bmatrix} \quad (1)$$

- Initial matrix was converted to the standard grade matrix 'V', to remove the error caused due to different scales. Matrix 'V' was obtained as

$$v_{ij} = \frac{u_{ij} - (u_{ij})_{\min}}{(u_{ij})_{\max} - (u_{ij})_{\min}} \quad (2)$$

where 'v_{ij}' is normalisation of an evaluated parameter (n) in a particular water sample (m).

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & \cdots & v_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{m1} & v_{m2} & \cdots & \cdots & v_{mn} \end{bmatrix} \quad (3)$$

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

Table 1 | Water quality parameters associated with their units, abbreviations, analytical methods and equipment used in this study

Parameters	Unit	Abbr.	Analytical methods	Equipment	Method (Standard Methods)
pH	–	pH	pH-meter	μ pH System 361 (Systronics)	4500-H^+ B.
Dissolved oxygen	mg/L	DO	DO meter	HQ30D Portable Dissolved Oxygen Meter (Hach)	4500-O G.
Total alkalinity	mg/L	TA	Titrimetric	————————	2320 B.
Total hardness	mg/L	TH	Titrimetric	————————	2340 C.
Turbidity	NTU	Tur	Nephelometric	Digital Nephelo Turbidity Meter 132 (Systronics)	2130 B.
Total dissolved solids	mg/L	TDS	Gravimetric		2540 B.
Electrical conductivity	μS/cm	EC	Electrometric	Microprocessor TDS/Cond/SAL/ Temperature Meter (MT-112TDS) (MANTI LAB SOLUTIONS)	2510 B.
Sodium	mg/L	Na^+	Flame photometer	μ Controller Based Flame photometer with Compressor (Type 128) (Systronics)	3500-Na B.
Potassium	mg/L	K^+	Flame photometer	μ Controller Based Flame photometer with Compressor (Type 128) (Systronics)	3500-K B.
Calcium	mg/L	Ca^{2+}	Flame photometer	μ Controller Based Flame photometer with Compressor (Type 128) (Systronics)	3111B,D
Magnesium	mg/L	Mg	Atomic absorption spectroscopy	iCE 3000 SERIES (Thermo Scientific)	3111 B.
Iron	mg/L	Fe	Atomic absorption spectroscopy	iCE 3000 SERIES (Thermo Scientific)	3111 B.
Lead	mg/L	Pb	Atomic absorption spectroscopy	iCE 3000 SERIES (Thermo Scientific)	3111 B.
Copper	mg/L	Cu	Atomic absorption spectroscopy	iCE 3000 SERIES (Thermo Scientific)	3111 B.
Zinc	mg/L	Zn	Atomic absorption spectroscopy	iCE 3000 SERIES (Thermo Scientific)	3111 B.
Manganese	mg/L	Mn	Atomic absorption spectroscopy	iCE 3000 SERIES (Thermo Scientific)	3111 B.
Fluoride	mg/L	F^-	Spectrophotometric	Spectro V-11D (MRC)	4500-F^- D.
Chloride	mg/L	Cl^-	Titrimetric	————————	4500-Cl– B.
Sulfate	mg/L	SO_4^{2-}	Turbidimetric	Digital Nephelo Turbidity Meter 132 (Systronics)	4500-SO_4^{2-} E.
Nitrate	mg/L	NO_3^-	Spectrophotometric	Cary 50 UV-Vis Spectrophotometer (Agilent Technologies)	4500-NO_3^- B.

- Information entropy was calculated by:

$$S_{ij} = \frac{v_{ij}}{\sum v_{ij}} \tag{4}$$

$$E_n = -\left(\frac{1}{\ln n}\right) \sum_{i=1}^{m} S_{ij} \ln S_{ij} \tag{5}$$

- Entropy weight of the parameter (j) was calculated by:

$$\mathrm{Wj} = \frac{1 - \mathrm{E_j}}{\sum_{\mathrm{j}=1}^{\mathrm{m}} 1 - \mathrm{E_j}} \tag{6}$$

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

- Quality rating scale for each parameter was assigned by:

$$Q_j = \left(\frac{C_j}{S_i}\right) \times 100 \tag{7}$$

- EWQI was calculated by:

$$EWQI = \sum_{j=1}^{n} W_j Q_j \tag{8}$$

where C_j = measured concentration of the parameter.

Classification of EWQI into five ranks is shown in Table 2 (Li *et al.* 2010; Amiri *et al.* 2014; Fagbote *et al.* 2014; Gorgij *et al.* 2017).

Table 2 | Classification of water quality index (Li *et al.* 2010)

Rank	EWQI	Water quality
1	< 50	Excellent
2	50–100	Good
3	100–150	Average
4	150–200	Poor
5	>200	Extremely poor

Multivariate statistical techniques (MSTs)

CA is an exploratory analysis that divides a large number of objects into a smaller number of different groups based on similarity. Clustering is unsupervised classification and its procedures may be hierarchical or non-hierarchical. A tree-like structure called dendrogram characterises a hierarchical CA (HCA). HCA can be agglomerative or divisive. In the present study, agglomerative HCA has been used to identify the similarity among sampling locations. HCA was performed on z transformed datasets using Ward's method of linkage. Ward's method of linkage begins with 'n' clusters, each containing a single observation and continues until all the observations are comprised into one cluster. This method is based on the error sum of squares. For the measure of similarity, Euclidean distance has been used. Euclidean distance measures the geometric distance between the two observations.

PCA was applied to transform the original variable into new and uncorrelated variables (Shrestha & Kazama 2007). It is a powerful data reduction technique used to reduce the variable numbers to explain the variance with fewer variables (Zhang *et al.* 2009; Dash *et al.* 2020). The following steps are involved in the PCA:

Step 1: Standardisation of the dataset (all the variables will be transformed to the same scale).

Step 2: Computation of covariance matrix (to observe how the variables are varying from the mean with respect to each other).

Step 3: Computation of eigenvalues and eigenvectors for the covariance matrix (to decide the principal components).

Step 4: Computation of the Principal Components (PCs).

Step 5: Reorientation the data from the original axes to the ones represented by the PCs.

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

The basic idea behind PCA is to ascertain patterns and correlations among observed variables. Based on a strong correlation between different variables, a final judgement is made about reducing the dimensions of the datasets in such a way that the substantial statistical information is still retained.

Statistical analysis was performed using IBM SPSS Statics 20 software.

RESULTS AND DISCUSSION

Descriptive statistics of the observed various water quality parameters of Baralia and Puthimari Rivers are shown in Tables 3–6. It has been observed that TDS and TUR have very high SD and Variance. This may be due to the influence of rainfall, surface runoff, river water flow and erosion from the river bed and banks. Erosion is more pronounced in both banks than the sedimentation (Baishya & Sahariah 2015).

In the present study, HCA was used to categorise the sampling sites and a Dendrogram was generated. HCA grouped the sampling locations into three different clusters. Grouped sampling sites under each cluster are shown in Figure 2. In the flow path, Baralia River encountered mostly agricultural, and forest areas in the upper reaches, a densely populated Rangia town in the middle reach and scattered population, forest areas and farming land in the lower reaches. But, Puthimari River encounters scattered population, forest areas and agriculture land in lower reaches throughout its flow length. Sampling sites located at the middle reach of the stream and near Rangia town were grouped under cluster 1. EWQI of all water samples with a value more than 150 indicated that the water quality was 'poor' or 'extremely poor' (Table 7). Higher EWQI was observed at sampling sites located near the densely populated market area of Rangia town. Sampling stations near the town receive pollutants from domestic wastewater. Wastewater from household activities was disposed of into open drains in front of the houses, which discharged this into the river without any treatment. There is no well-connected drainage system in the town. Baralia River is also used for washing clothes, bathing of pets, and fishing, which also contribute to the pollution (CPCB 2015). Another important factor contributing to pollution was municipal solid waste (MSW).

MSW is routinely dumped in town streets and along the banks of the rivers. MSW was found to be dumped about in thin, non-contiguous layers at numerous locations along the riverbank. Still, in many areas, thicker, contiguous fills existed on the river bank lying in contact with the flowing water. Water leaching through solid waste directly affects the water quality of the river (CPCB 2015). Sampling sites SPPR1 and SPBR1 were grouped in this Cluster 2 (Figure 2). These sites were located at the river's uppermost reach where inhabitant's density is significantly less, and human activities are minimal. EWQI of these two sampling locations were 69.65 and 61.62, respectively (Table 7), which indicate the water quality as 'good' (Table 2).

Cluster 3 consisted of sampling sites, namely SPPR2, SPBR2, SPBR3, SPBR8 and SPBR9. Sampling sites SPPR2, SPBR2 and SPBR3 were located upstream of Rangia town, at that part of the basin where inhabitant's density is low and agricultural activities and livestock breeding dominates land use pattern. The EWQI at those locations was in the range of 100–150, which indicated the water quality as 'average'. Sampling sites SPBR8 and SPBR9 were located in the river's downstream section, away from Rangia town. EWQI of SPBR8 was 149.45 and that of SPBR9 was 147.43, which indicated water quality as 'average'. Water quality of cluster 3 was better than the water quality of cluster 1. It indicates the self-assimilative process of the river.

PCA was performed on all observed water quality parameters collected from various sampling locations. For extraction of principal component, to explain the sources of variance in observed water quality parameters, an eigenvalue greater than one was taken as the criteria. PCA generated six useful factors which explained 87.98% of the total variance (Table 8). Factor 1, which explained 28.73% of the total variance associated with inorganic constituents. It had strong positive loading on

Table 3 | Statistical description of water quality parameters of Baralia River

Parameters	pH	DO	TDS	EC	TUR	TH	TA	Na^{+}	K^{+}	Ca^{+2}	Mg^{+2}	F^{-}	Cl^{-}	SO_4^{2-}	NO_3^{-}
Max	7.88	8.83	444.00	0.25	123.00	70.00	94.00	6.76	1.19	16.21	19.12	0.26	4.50	26.46	0.96
Min	7.20	6.20	112.00	0.21	21.50	60.00	82.00	4.36	0.93	6.77	14.32	0.00	0.50	14.80	0.03
Mean	7.49	7.19	186.33	0.22	85.20	67.00	90.78	5.03	1.08	12.47	16.29	0.09	2.06	19.05	0.49
Variance	0.04	0.56	9,999.50	0.00	1,114.25	16.50	13.44	0.58	0.01	12.06	1.71	0.01	1.65	15.51	0.11
Skewness	0.49	1.36	2.62	1.78	− 0.83	− 1.42	− 2.10	1.79	− 0.36	− 1.01	1.06	0.36	0.86	0.61	− 0.11
Kurtosis	1.52	2.53	7.38	4.19	− 0.02	0.41	4.53	3.01	− 0.78	− 0.41	2.77	− 1.35	0.19	− 0.20	− 1.40
SD	0.20	0.75	100.00	0.01	33.38	4.06	3.67	0.76	0.09	3.47	1.31	0.10	1.29	3.94	0.34
COV	0.03	0.10	0.54	0.05	0.39	0.06	0.04	0.15	0.08	0.28	0.08	1.04	0.63	0.21	0.69

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

Table 4 | Statistical description of heavy metal concentration in Baralia River

Parameters	Fe	Mn	Pb	Cu	Zn
Max	6.31	0.58	0.07	0.06	0.04
Min	0.03	0.02	0.00	0.01	0.01
Mean	1.97	0.18	0.02	0.02	0.02
Variance	3.23	0.03	0.00	0.00	0.00
Skewness	1.94	2.24	1.15	2.46	0.76
Kurtosis	5.14	6.12	0.73	6.85	0.18
SD	1.80	0.16	0.02	0.02	0.01
COV	0.91	0.87	1.12	0.70	0.54

EC, TH, TA, K^+, Ca^{2+} and Mg^{2+}. Conductivity in water is affected by the presence of inorganic dissolved solids such as Cl^-, SO_4^{2-}, Na^+, K^+ and Ca^{2+}. Ca^{2+} and Mg^{2+} dissolved in water are the two most common sources of hardness. This factor is associated with surface runoff (Goonetilleke *et al.* 2005). Factor 2 represented 17.5% of the total variance related to heavy metals such as Fe, Mn, Cu and Zn. This factor had strong positive loading on Fe, Mn and Cu and had a moderately strong positive loading on Zn. This heavy metal factor can be interpreted as metal pollution leaching from MSW, illegally dumped near the bank. Factor 3, which explained 12.7% of total variance had strong positive loading on DO and Cl^- and moderate loading on Na^+. This factor represents pollution sources mainly from municipal effluents (USGS 1999). Cl^- is a major constituent of municipal wastewater normally coming from kitchen wastewater. Salts such as table salt are composed of Na^+ and Cl^-. When table salt is mixed with water, its Na^+ and Cl^- ions separate as they dissolve. Chlorinated drinking water also increases chloride levels in the wastewater of a community (USGS 1999; Ha & Bae 2001). Factor 4 accounted for 11.89% of the total variance and had strong positive loading on pH and SO_4^{2-}. Sulfates naturally occur in minerals of some soil and rock formations (Al-Khashman & Shawabkeh 2006). This factor may be attributed to the physicochemical source of variability. Factor 5 had strong positive loading on TDS and moderate loading on NO_3^-. This factor can be attributed to pollution due to the use of fertilisers for agricultural activities. This can also occur with animal waste and manure finding their way into the river. Factor 6 explained 6.95% of the total variance and had strong negative loading on Pb and moderate positive loading on F^-. This factor may also be due to the physicochemical source of variability (Varol & Sen 2009).

CONCLUSION

In this study, water quality data for 20 physical and chemical parameters, collected from 9 sampling sites of Baralia River and 6 sampling sites of Puthimari River in Assam (India) during the period of May 2016 -June 2017 were analysed. EWQI was used to assess the water quality of rivers. HCA was applied to group the similar sites and it grouped all the monitored sites into 3 clusters based on pollution levels. PCA was applied to identify possible sources of pollution. The important conclusions from the study were drawn as follows:

- The analysis showed that domestic discharge coming from various household activities and runoff leaching from the illegally dumped municipal solid waste near the river bank are adversely affecting the water quality of Baralia and Puthimari River. Worst water quality has been observed near Rangia town.
- HCA grouped all the sampling sites into 3 clusters based on similarities in the water quality characteristics. This method can be used for the optimisation of sampling sites.

doi: 10.2166/wpt.2021.020

Table 5 | Statistical description of water quality parameters of Puthimari River

Parameters	pH	DO	TDS	EC	TUR	TH	TA	Na^{+}	K^{+}	Ca^{+2}	Mg^{+2}	F^{-}	Cl^{-}	SO_4^{2-}	NO_3^{-}
Max	7.76	7.55	244.00	0.69	135.00	156.00	168.00	4.60	4.10	22.69	30.30	0.34	2.00	38.75	0.78
Min	7.40	6.74	126.00	0.20	4.00	64.00	62.00	2.51	0.66	12.79	13.69	0.00	1.00	24.42	0.13
Mean	7.63	7.12	177.33	0.29	92.67	85.67	94.00	3.41	1.52	15.36	17.67	0.17	1.58	29.97	0.39
Variance	0.02	0.07	1,852.27	0.04	2,664.67	1,247.07	1,393.60	0.85	1.64	14.17	40.22	0.02	0.14	24.63	0.06
Skewness	− 1.05	0.44	0.56	2.44	− 1.32	2.21	2.12	0.51	2.33	2.01	2.20	− 0.18	− 0.31	1.21	0.77
Kurtosis	1.41	2.04	− 0.47	5.98	0.52	4.97	4.96	− 1.93	5.59	4.13	4.98	− 2.32	− 0.10	1.78	0.78
SD	0.13	0.26	43.04	0.20	51.62	35.31	37.33	0.92	1.28	3.76	6.34	0.15	0.38	4.96	0.24
COV	0.02	0.04	0.24	0.68	0.56	0.41	0.40	0.27	0.84	0.25	0.36	0.88	0.24	0.17	0.61

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

Table 6 | Statistical description of heavy metal concentration in Puthimari River

Parameters	Fe	Mn	Pb	Cu	Zn
Max	1.71	0.57	0.19	0.05	0.03
Min	0	0	0	0.001	0
Mean	0.79	0.14	0.04	0.02	0.01
Variance	0.32	0.05	0.01	0.00	0.00
Skewness	0.49	2.03	2.11	1.18	1.24
Kurtosis	1.18	4.23	4.46	0.50	− 0.29
SD	0.57	0.22	0.08	0.02	0.01
COV	0.72	1.61	1.90	0.99	1.60

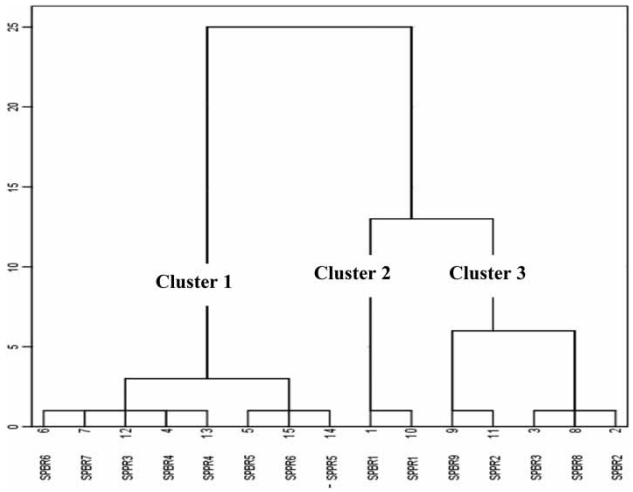


Figure 2 | Dendrogram showing cluster of sampling sites.

Table 7 | Sampling sites with their EWQI

Sampling site	EWQI	Sampling site	EWQI
SPBR1	69.65	SPPR1	61.62
SPBR2	128.18	SPPR2	113.72
SPBR3	145.23	SPPR3	231.61
SPBR4	192.67	SPPR4	152.33
SPBR5	314.68	SPPR5	235.48
SPBR6	177.02	SPPR6	161.54
SPBR7	182.52		
SPBR8	149.45		
SPBR9	147.43		

- The study demonstrated the importance of Shannon entropy and MSTs in water quality assessment. The study illustrated the utility of EWQI in evaluating surface water quality, the results of which were further reinforced by the application of PCA and HCA.
- The present work justifies the effectiveness of combined use of EWQI and MSTs in water quality monitoring and management.

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

Table 8 | Results of PCA for water quality parameters

	Factor					
	1	2	3	4	5	6
pH	0.051	−0.127	0.022	0.769	0.363	−0.016
DO	−0.165	−0.096	0.874	0.074	−0.319	0.091
TDS	0.004	−0.357	−0.281	−0.056	0.783	0.184
EC	0.959	−0.166	−0.121	0	−0.084	0.092
Tur	−0.482	0.496	0.151	0.342	0.399	−0.298
TH	0.97	0.007	−0.052	0.145	−0.071	0.081
TA	0.947	−0.074	−0.027	−0.237	−0.034	0.012
Na^+	0.108	−0.006	0.562	−0.729	0.18	0.229
K^+	0.974	−0.074	−0.125	−0.013	−0.044	0.098
Ca	0.751	−0.059	0.273	0.191	0.241	−0.108
Mg^{2+}	0.929	−0.035	−0.023	0.09	0.009	0.172
F^-	0.358	−0.443	−0.22	0.278	0.154	0.559
Cl^-	0.014	−0.078	0.946	−0.119	0.22	−0.065
SO_4^{2-}	0.094	−0.238	0.042	0.874	−0.248	0.184
NO_3^-	−0.19	0.195	0.465	0.012	0.736	0.301
Fe	−0.175	0.806	0.088	−0.227	−0.107	−0.106
Mn	0.119	0.855	−0.093	−0.04	0.147	0.017
Pb	−0.199	−0.329	−0.197	0.052	−0.254	−0.786
Cu	−0.093	0.905	−0.101	0	−0.079	0.171
Zn	−0.267	0.681	−0.158	−0.313	−0.37	0.202
Eigenvalues	5.746	3.51	2.543	2.379	2.027	1.391
% Total variance	28.731	17.551	12.714	11.896	10.135	6.955
Cumulative %	28.731	46.282	58.996	70.892	81.027	87.981

The study will help policymakers that take care of the water supply and water pollution control since these form a significant tool for easy understanding and thereby making their applicability uncomplicated. Indeed, these methodologies make the water quality datasets utilization enormously easy and lucid. This study will also assist in making decisions in allocating funds and determining priorities.

DISCLOSURE STATEMENT

The authors reported no potential conflict of interest.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Abbasi, T. & Abbasi, S. A. 2012 *Water Quality Indices*. Elsevier, Amesterdam, The Netherlands.

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

Alberto, W. D., del Pilar, D. M., Valeria, A. M., Fabiana, P. S., Cecilia, H. A. & de los Ángeles, B. M. 2001 Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Córdoba–Argentina). *Water Research* **35**(12), 2881–2894.

Al-Khashman, O. A. & Shawabkeh, R. A. 2006 Metals distribution in soils around the cement factory in southern Jordan. *Environmental Pollution* **140**(3), 387–394.

Amiri, V., Rezaei, M. & Sohrabi, N. 2014 Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran. *Environmental Earth Sciences* **72**(9), 3479–3490.

APHA 2012 *Standard Methods for the Examination of Water and Wastewater*. American Public Health Association, Washington, DC.

Avsar, E., Toröz, İ., Hanedar, A. & Yılmaz, M. 2014 *Chemical Characterization of Natural Organic Matter and Determination of Disinfection By-Product Formation Potentials in Surface Waters of Istanbul (Omerli and Buyukcekmece Water Dam)*. Turkey.

Avsar, E., Toroz, İ. & Hanedar, A. 2015 Physical characterisation of natural organic matter and determination of disinfection by-product formation potentials in İstanbul surface waters. *Fresenius Environmental Bulletin* **24**(9), 2763–2770.

Baishya, S. J. & Sahariah, D. 2015 A study of bank erosion and bankline migration of the Baralia River, Assam, using remote sensing and GIS. *International Journal of Current Research* **7**(11), 23373–23380.

Borah, M., Das, P. K., Borthakur, P., Basumatary, P. & Das, D. 2020 An assessment of surface and ground water quality of some selected locations in Guwahati. *International Journal of Applied Environmental Sciences* **15**(1), 93–108.

Bu, H., Tan, X., Li, S. & Zhang, Q. 2010 Water quality assessment of the Jinshui River (China) using multivariate statistical techniques. *Environmental Earth Sciences* **60**(8), 1631–1639.

CPCB 2015 *River Stretches for Restoration of Water Quality*. A Ministry of Environment, Forest and Climate change report, New Delhi, India.

Dash, S., Borah, S. S. & Kalamdhad, A. 2018 Monitoring and assessment of Deepor Beel water quality using multivariate statistical tools. *Water Practice & Technology* **13**(4), 893–908.

Dash, S., Borah, S. S. & Kalamdhad, A. S. 2020 Application of environmetrics tools for geochemistry, water quality assessment and apportionment of pollution sources in Deepor Beel, Assam, India. *Water Practice & Technology* **15**(4), 973–992.

Fagbote, E. O., Olanipekun, E. O. & Uyi, H. S. 2014 Water quality index of the groundwater of bitumen deposit impacted farm settlements using entropy weighted method. *International Journal of Environmental Science and Technology* **11**(1), 127–138.

Goonetilleke, A., Thomas, E., Ginn, S. & Gilbert, D. 2005 Understanding the role of land use in urban stormwater quality management. *Journal of Environmental Management* **74**, 31–42.

Gorgij, A. D., Kisi, O., Moghaddam, A. A. & Taghipour, A. 2017 Groundwater quality ranking for drinking purposes, using the entropy method and the spatial autocorrelation index. *Environmental Earth Sciences* **76**(7), 269.

Ha, S. R. & Bae, M.-S. 2001 Effects of land use and municipal wastewater treatment changes on stream water quality. *Water, Air, and Soil Pollution* **70**, 135–151.

Karunanidhi, D., Aravinthasamy, P., Deepali, M., Subramani, T., Bellows, B. C. & Li, P. 2020 Groundwater quality evolution based on geochemical modeling and aptness testing for ingestion using entropy water quality and total hazard indexes in an urban-industrial area (Tiruppur) of Southern India. *Environmental Science and Pollution Research* 1–16. doi: https://doi.org/10.1007/s11356-020-10724-0

Landwehr, J. M. 1979 A statistical view of a class of water quality indices. *Water Resources Research* **15**(2), 460–468.

Lekkas, T., Kolokythas, G., Nikolaou, A., Kostopoulou, M., Kotrikla, A., Gatidou, G., Thomaidis, N. S., Golfinopoulos, S., Makri, C., Babos, D. & Vagi, M. 2004 Evaluation of the pollution of the surface waters of Greece from the priority compounds of List II, 76/464/EEC Directive, and other toxic compounds. *Environment International* **30**(8), 995–1007.

Li, P., Qian, H. & Wu, J. 2010 Groundwater quality assessment based on improved water quality index in Pengyang County, Ningxia, Northwest China. *Journal of Chemistry* **7**(S1), S209–S216.

Li, X., Wang, K., Liu, L., Xin, J., Yang, H. & Gao, C. 2011 Application of the entropy weight and TOPSIS method in safety evaluation of coal mines. *Procedia Engineering* **26**, 2085–2091.

Liu, T. K., Yu, J. L., Chen, C. L. & Wei, P. S. 2012 Information theoretic perspective on coastal water quality monitoring and management near an offshore industrial park. *Environmental Monitoring and Assessment* **184**(8), 4725–4735.

Ozkul, S., Harmancioglu, N. B. & Singh, V. P. 2000 Entropy-based assessment of water quality monitoring networks. *Journal of Hydrologic Engineering* **5**(1), 90–100.

Prasad, B., Soni, A. K., Vishwakarma, A., Trivedi, R. & Singh, K. K. K. 2020 Evaluation of water quality near the Malanjhkhand copper mines, India, by use of multivariate analysis and a metal pollution index. *Environmental Earth Sciences* **79**, 1–23.

Shrestha, S. & Kazama, F. 2007 Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environmental Modelling & Software* **22**(4), 464–475.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, T. 2003 Assessment of the surface water quality in Northern Greece. *Water Research* **37**(17), 4119–4124.

Singh, K. P., Malik, A., Mohan, D. & 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 Research* **38**(18), 3980–3992.

Singh, K. R., Bharti, N., Kalamdhad, A. S. & Kumar, B. 2017 Surface water quality assessment of Amingaon (Assam, India) using multivariate statistical techniques. *Water Practice & Technology* **12**(4), 997–1008.

Water Practice & Technology Vol 16 No 2
doi: 10.2166/wpt.2021.020

Singh, K. R., Dutta, R., Kalamdhad, A. S. & Kumar, B. 2018 Risk characterisation and surface water quality assessment of Manas River, Assam (India) with an emphasis on the TOPSIS method of multi-objective decision making. *Environmental Earth Sciences* **77**(23), 780.

Singh, K. R., Dutta, R., Kalamdhad, A. S. & Kumar, B. 2019 Information entropy as a tool in surface water quality assessment. *Environmental Earth Sciences* **78**(1), 15.

Sutadian, A. D., Muttil, N., Yilmaz, A. G. & Perera, B. J. C. 2016 Development of river water quality indices – a review. *Environmental Monitoring and Assessment* **188**(1), 58.

USGS 1999 *U.S. Geological Survey, The Quality of our Nation's Waters-Nutrients and Pesticides*. U.S. Geological Survey Circular 1225.

Varol, M. & Şen, B. 2009 Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. *Environmental Monitoring and Assessment* **159**(1–4), 543.

Varol, M., Gökot, B., Bekleyen, A. & 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.

Yang, Y. H., Zhou, F., Guo, H. C., Sheng, H., Liu, H., Dao, X. & He, C. J. 2010 Analysis of spatial and temporal water pollution patterns in Lake Dianchi using multivariate statistical methods. *Environmental Monitoring and Assessment* **170**(1), 407–416.

Zavaleta, M. A. J., Alcaraz, M. R., Peñaloza, L. G., Boemo, A., Cardozo, A., Tarcaya, G., Azcarate, S. M. & Goicoechea, H. C. 2021 Chemometric modeling for spatiotemporal characterization and self-depuration monitoring of surface water assessing the pollution sources impact of northern Argentina rivers. *Microchemical Journal* **162**, 105841.

Zhang, Q., Li, Z., Zeng, G., Li, J., Fang, Y., Yuan, Q., Wang, Y. & Ye, F. 2009 Assessment of surface water quality using multivariate statistical techniques in red soil hilly region: a case study of Xiangjiang watershed, China. *Environmental Monitoring and Assessment* **152**(1), 123–131.

First received 14 September 2020; accepted in revised form 12 February 2021. Available online 24 February 2021