International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 9 • Number 46 • 2016

A Study of Sustainability on Water Quality of North Indian Rivers using Data Mining Techniques

Ranjan Kumar Panda[a], A. Sai Sabitha[b] and Shweta Bhardwaj[c]

[a]*Correspondence author, Amity University, Uttar Pradesh, India. Email: [a]rp.pranjan@gmail.com*
[b,c]*Amity University, Uttar Pradesh, India. Email: [b]saisabitha@gmail.com, [c]sbhardwaj1@amity.edu*

Abstract: Water is an important natural resource for existence of living being on this planet. The water crisis is increasing day by day in many countries in the world due to over exploitation, industrial activities and rapid urbanization. Water quality and water crisis are important concerns for future sustainability. Water qualities of major north Indian rivers have been analyzed for water sustainability for future needs. The data mining techniques for water quality have been studied by analyzing the physical, chemical and biological parameters. Among the river stations of northern India those stations which have good water quality and those which have water quality below the acceptable limit have been identified based on the 8 water quality parameters. Clustering technique has been used to group the rivers depending on their similar characteristics of water quality parameters.

Keywords: Sustainability, Clustering, K-Mean, Water Pollution.

1. INTRODUCTION

Sustainable development is development which meets the needs of the present without compromising the ability of future generations to meet their own needs (World Commission on Environment and Development, 1987). It broadly constitutes three parts like (1) Environmental Sustainability (2) Economical Sustainability (3) Social Sustainability [21]. One of the important sustainability issues in India is water quality management. Water quality depends upon physical, chemical and biological factors of water. Water pollution, mainly happens due to poor sewage management [1]. Most of the surface water of rivers, ponds and lakes of India are polluted [2]. As per study conducted on Indian river (Pennar) considering physical, chemical, bacterial and fungal factors, it was found that the water of the polluted rivers are not safe for human use [6]. Water quality of rivers in India is being detoriated gradually as per finding in similar studies [7,8]. Major source of pollution of surface water is due to Municipal wastewater and industrial effluents. Surface runoff is a seasonal phenomenon which occurs due to climatic change inside the river basin [9]. The concentration of pollutants in the river water is also affected by seasonal variations of ground water flow, inflow, outflow and surface run off through the river [10]. Understanding the chemical, biological and hydrological characteristics of water are required for water quality management of river on a long term basis [11]. There have been several researches conducted on the sustainable water quality management using data mining techniques. In this paper K-Mean clustering technique is used to group major north Indian rivers using Rapid Miner tool. The rivers have been grouped into clusters depending on the similarity of water quality parameters. The best and the below average water quality river stations in sustaining aquatic life have been identified.

2. LITERATURE SURVEY

Water quality analysis is performed based on the physical, chemical and biological parameters of water. The different water quality parameters are temperature, PH, dissolved oxygen (DO), Biochemical oxygen demand (BOD), Nitrate (NO3), Fecal Coliform (FC) and Total Coli form (TC). The water quality standard desired by central pollution control board (CPCB), India for drinking water source after conventional treatment and disinfection, Propagation of Wildlife and Fisheries in river water has been considered for water quality comparison [12]. Below are some facts about these parameters in river water and their impacts on human and aquatic lives [13,14].

2.1. Water Quality Parameters

1. Temperature: The optimum temperature is the temperature at which the aquatic life will suitably survive. It is between 25 to 35 degree Celsius.
2. Dissolved Oxygen (DO): Oxygen enters into water from the surrounding air and is a product due to photosynthesis of aquatic plants. High levels of DO are healthy for the ecosystem. Desired level of DO in (mg/l) is > 4 mg/l.
3. PH: PH is a measure of a solution's (water) acidity level. Desired level is [6.5-8.5].
4. Conductivity: It is the capability of a solution like water in a stream to pass an electric current and indicates the concentration of dissolved electrolyte ions in the water. Increase in conductivity indicates an increased level of pollution of water. For Irrigation, Industrial Cooling, Controlled Waste disposal conductivity should be < 2250 micro mhos/cm. Freshwater streams should have optimal conductivity between 150 to 500 µS/cm for supporting different aquatic life.
5. Biochemical Oxygen Demand (BOD): Biochemical oxygen demand (BOD) is a measure of the amount of oxygen that bacteria consumes while decomposing the organic matters under aerobic conditions. Waste water treatment plants aim to reduce the BOD in the effluents discharged to water streams. More the BOD more is the pollution level. Desired limit of B.O.D. in (mg/l) is < 3 mg/l.
6. Nitrate: It enters into water streams from natural sources like decomposition of plants and animal wastes, human sources like sewage water and fertilizers. Nitrate is measured in mg/L. Desired limit of nitrate is < 50 mg/L.
7. Fecal coliform (FC): It is a kind of bacteria which causes micro biological contamination in waters. Desired limit of FC in (MPN/100ml) is < 2500 MPN/100ml. More the FC more is the pollution level of water and its presence causes illness in human and animals.
8. Total Coliforms (TC): TC and Escherichia coli (E. coli) are used as measures of the degree of pollution in water. E. coli is the member of the TC group of bacteria which is found in the intestines of mammals and humans. E. coli in water indicates the fecal contamination and presence of bacterias, viruses and parasites. Desired level of TC in (MPN/100ml) is < 5000 MPN/100ml.

2.2. Data Mining Techniques to Determine Water Quality

Different data mining techniques can be used to assess the water quality of both surface and underground water. Some important techniques used for water quality evaluation are discussed below in Table 1.

Table 1
Summary of Methodology/Techniques used for Water Quality Assessment

S.No.	*Author Names and Study id*	*Year*	*Key findings*	*Methodologies/ Techniques*
1	Liao et. al. [15]	2010	Used for water quality prediction. 1) The accuracy level of IDTL model is 85%. 2) The accuracy level of the decision tree model is 70%. 3) The model works well with clustered and interdependent parameters.	Back propagation neural network and improved decision tree (IDTL) method.
2	Qun et. al. [16]	2009	1.Used for evaluation of surface water quality. 2. River water quality standards are met in accordance with spatial and temporal distribution of pollutants.	Water quality index (WQI)
3	Zhang et. al. [17]	2011	Evaluation of temporal and spatial variations of river water quality.	Cluster analysis (CA) and Factor analysis (FA)
4	Camejo et. al. [18]	2013	For real time water quality classification, the KNN algorithm gives slightly better results than ANN and PART.	Artificial intelligence with Delphi method. PART, ANN and KNN.
5	Aghaarabi et. al. [19]	2013	Evaluation of uncertain water quality parameters. Fuzzy rule based results of microbial, Physio-chemical group is combined with fuzzy evidential Reasoning.	Fuzzy rule based system (Mamdani)
5	Ali et. al. [20]	2013	Hierarchical clustering Is more accurate unsupervised learning method for finding water quality index. Multilayer perceptron is more accurate supervised learning technique for classification.	K-means, Hierarchical clustering, Multilayer perceptron, support vector machines
6	Kumar et. al. [21]	2015	River water quality analysis with spatial variations.	Cluster analysis (CA) Principal component analysis (PCA)

2.3. Cluster Analysis (CA)

CA is a data mining technique for grouping of things based on similarity of features. Degree of association is strong among the members of the same cluster and weak among the members of different clusters. K-mean clustering technique is used to classify rivers by considering the 8 water quality parameters of north Indian rivers. The similarities between river stations are measured by Mixed Euclidian distance measurements. The distance between two river stations, i and j is represented (Refer eq. 1) as:

$$(d_{ij})^2 = \sum_{k=1}^{m} (z_{ik} - z_{jk})^2 \quad (1)$$

where, d_{ij} = the Euclidean distance, Z_{ik} and Z_{jk} = values of the variables.

m = total number of parameters.

2.4. Validity Measure

Determining the number of clusters in a data set, 'K' in the *K-Mean* algorithm, is an important criterion. Elbow method, cross validation and silhouette index are some of the methods used. Silhouette coefficient uses the average distance to elements in the same cluster with the average distance to elements in other clusters.

- Cohesion $a(x)$: average distance of x to all other vectors in the same cluster.
- Separation $b(x)$: average distance of x to the vectors in other clusters.

Silhouette coefficient is shown below (Refer equation 2):

$$s(x) = \{b(x) - a(x)\}/\max\{a(x), b(x)\} \quad (2)$$

The value of $s(x)$ lies between $[-1, +1]$: Objects with a high silhouette value are considered well clustered; objects with a low value may be outliers. This index works well with K-Mean clustering, and is also used to determine the optimal number of clusters.

3. METHODOLOGY

The data sets for water quality were collected for the year 2012 for major north Indian rivers from Indian government data base [3]. Below steps were performed before carrying out the experiments as shown in Figure 1.

1. Data collection
2. Data cleaning
3. Data integration
4. Data transformation
5. Data reduction

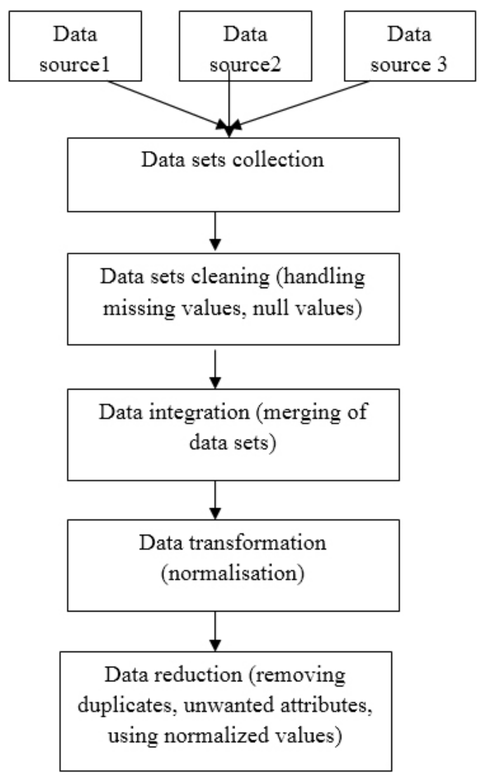


Figure 1: Data preprocessing steps

3.1. Data Collection

The data sets for water quality were collected for the year 2012 for major north Indian rivers from Indian government data base [3].

3.2. Data Cleaning

The missing and null values for individual parameter were filled with the average value of that parameter considering the values for the particular river and the state in which it is flowing.

3.3. Data Integration

After cleaning the data sets for different rivers, then those were merged into a single data set.

3.4. Data Transformation

The parameters were then normalized in the scale from 0 to 1 by using the below formula:

$$V' = [(V - V_{min})/(V_{max} - V_{min})][New_{max} - New_{min}] + New_{min} \quad (3)$$

where, V' = Normalized value of the parameter,

V = Current value of the parameter.

V_{min} = Minimum of all values for a particular parameter.

V_{max} = Maximum of all values for the parameter.

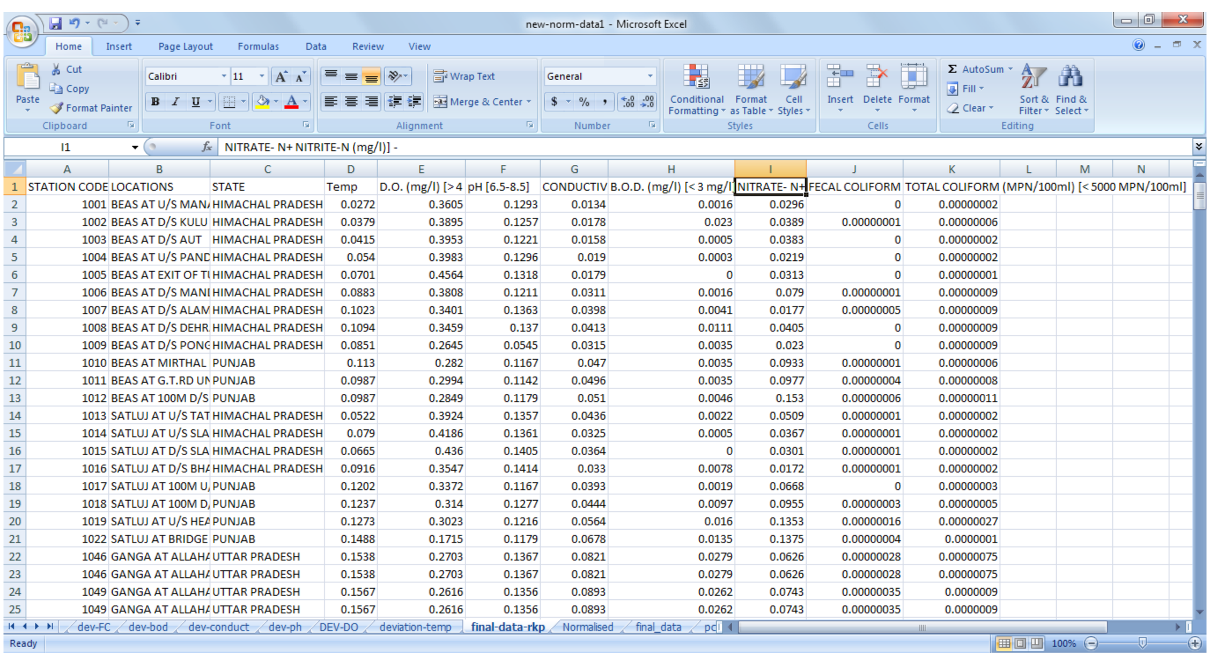
$New_{max} = 1$ and $New_{min} = 0$.

3.5. Data Reduction

Duplicate records were removed. Station code attribute is omitted from clustering attribute. Station code attribute is omitted from clustering attribute.

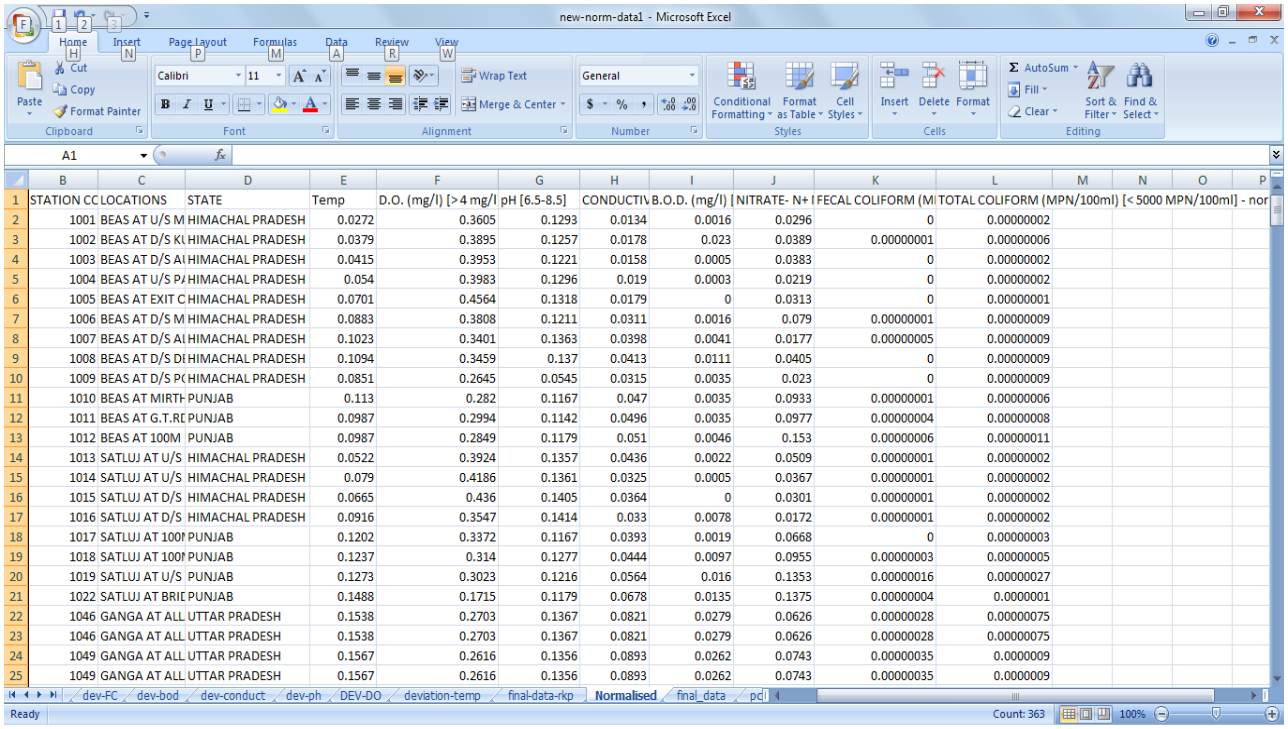
4. EXPERIMENTAL SET UP

The data sets after cleaning were normalized and the numerical attributes were replaced with normalized values. Below are snapshots in Figure 3 and Figure 4 of data sets before and after normalization.



	A	B	C	D	E	F	G	H	I	J	K
1	STATION CODE	LOCATIONS	STATE	Temp	D.O. (mg/l) [>4	pH [6.5-8.5]	CONDUCTIV	B.O.D. (mg/l) [<3 mg/l	NITRATE- N+	FECAL COLIFORM	TOTAL COLIFORM (MPN/100ml) [<5000 MPN/100ml]
2	1001	BEAS AT U/S MAN/	HIMACHAL PRADESH	0.0272	0.3605	0.1293	0.0134	0.0016	0.0296	0	0.00000002
3	1002	BEAS AT D/S KULU	HIMACHAL PRADESH	0.0379	0.3895	0.1257	0.0178	0.023	0.0389	0.00000001	0.00000006
4	1003	BEAS AT D/S AUT	HIMACHAL PRADESH	0.0415	0.3953	0.1221	0.0158	0.0005	0.0383	0	0.00000002
5	1004	BEAS AT U/S PAND	HIMACHAL PRADESH	0.054	0.3983	0.1296	0.019	0.0003	0.0219	0	0.00000002
6	1005	BEAS AT EXIT OF TI	HIMACHAL PRADESH	0.0701	0.4564	0.1318	0.0179	0	0.0313	0	0.00000001
7	1006	BEAS AT D/S MANI	HIMACHAL PRADESH	0.0883	0.3808	0.1211	0.0311	0.0016	0.079	0.00000001	0.00000009
8	1007	BEAS AT D/S ALAM	HIMACHAL PRADESH	0.1023	0.3401	0.1363	0.0398	0.0041	0.0177	0.00000005	0.00000009
9	1008	BEAS AT D/S DEHR.	HIMACHAL PRADESH	0.1094	0.3459	0.137	0.0413	0.0111	0.0405	0	0.00000009
10	1009	BEAS AT D/S PONG	HIMACHAL PRADESH	0.0851	0.2645	0.0545	0.0315	0.0035	0.023	0	0.00000009
11	1010	BEAS AT MIRTHAL	PUNJAB	0.113	0.282	0.1167	0.047	0.0035	0.0933	0.00000001	0.00000006
12	1011	BEAS AT G.T.RD UI	PUNJAB	0.0987	0.2994	0.1142	0.0496	0.0035	0.0977	0.00000004	0.00000008
13	1012	BEAS AT 100M D/S	PUNJAB	0.0987	0.2849	0.1179	0.051	0.0046	0.153	0.00000006	0.00000011
14	1013	SATLUJ AT U/S TAT	HIMACHAL PRADESH	0.0522	0.3924	0.1357	0.0436	0.0022	0.0509	0.00000001	0.00000002
15	1014	SATLUJ AT U/S SLA	HIMACHAL PRADESH	0.079	0.4186	0.1361	0.0325	0.0005	0.0367	0.00000001	0.00000002
16	1015	SATLUJ AT D/S SLA	HIMACHAL PRADESH	0.0665	0.436	0.1405	0.0364	0	0.0301	0.00000001	0.00000002
17	1016	SATLUJ AT D/S BH/	HIMACHAL PRADESH	0.0916	0.3547	0.1414	0.033	0.0078	0.0172	0.00000001	0.00000002
18	1017	SATLUJ AT 100M U,	PUNJAB	0.1202	0.3372	0.1167	0.0393	0.0019	0.0668	0	0.00000003
19	1018	SATLUJ AT 100M D,	PUNJAB	0.1237	0.314	0.1277	0.0444	0.0097	0.0955	0.00000003	0.00000005
20	1019	SATLUJ AT U/S HE/	PUNJAB	0.1273	0.3023	0.1216	0.0564	0.016	0.1353	0.00000016	0.00000027
21	1022	SATLUJ AT BRIDGE	PUNJAB	0.1488	0.1715	0.1179	0.0678	0.0135	0.1375	0.00000004	0.0000001
22	1046	GANGA AT ALLAH/	UTTAR PRADESH	0.1538	0.2703	0.1367	0.0821	0.0279	0.0626	0.00000028	0.00000075
23	1046	GANGA AT ALLAH/	UTTAR PRADESH	0.1538	0.2703	0.1367	0.0821	0.0279	0.0626	0.00000028	0.00000075
24	1049	GANGA AT ALLAH/	UTTAR PRADESH	0.1567	0.2616	0.1356	0.0893	0.0262	0.0743	0.00000035	0.0000009
25	1049	GANGA AT ALLAH/	UTTAR PRADESH	0.1567	0.2616	0.1356	0.0893	0.0262	0.0743	0.00000035	0.0000009

Figure 2: Non-Normalised data set



	B	C	D	E	F	G	H	I	J	K	L
1	STATION CC	LOCATIONS	STATE	Temp	D.O. (mg/l) [>4 mg/l	pH [6.5-8.5]	CONDUCTIV	B.O.D. (mg/l) [	NITRATE- N+ I	FECAL COLIFORM (MI	TOTAL COLIFORM (MPN/100ml) [< 5000 MPN/100ml] - nor
2	1001	BEAS AT U/S M	HIMACHAL PRADESH	0.0272	0.3605	0.1293	0.0134	0.0016	0.0296	0	0.00000002
3	1002	BEAS AT D/S KI	HIMACHAL PRADESH	0.0379	0.3895	0.1257	0.0178	0.023	0.0389	0.00000001	0.00000006
4	1003	BEAS AT D/S AI	HIMACHAL PRADESH	0.0415	0.3953	0.1221	0.0158	0.0005	0.0383	0	0.00000002
5	1004	BEAS AT U/S PA	HIMACHAL PRADESH	0.054	0.3983	0.1296	0.019	0.0003	0.0219	0	0.00000002
6	1005	BEAS AT EXIT C	HIMACHAL PRADESH	0.0701	0.4564	0.1318	0.0179	0	0.0313	0	0.00000001
7	1006	BEAS AT D/S M	HIMACHAL PRADESH	0.0883	0.3808	0.1211	0.0311	0.0016	0.079	0.00000001	0.00000009
8	1007	BEAS AT D/S AI	HIMACHAL PRADESH	0.1023	0.3401	0.1363	0.0398	0.0041	0.0177	0.00000005	0.00000009
9	1008	BEAS AT D/S DI	HIMACHAL PRADESH	0.1094	0.3459	0.137	0.0413	0.0111	0.0405	0	0.00000009
10	1009	BEAS AT D/S PC	HIMACHAL PRADESH	0.0851	0.2645	0.0545	0.0315	0.0035	0.023	0	0.00000009
11	1010	BEAS AT MIRTH	PUNJAB	0.113	0.282	0.1167	0.047	0.0035	0.0933	0.00000001	0.00000006
12	1011	BEAS AT G.T.RD	PUNJAB	0.0987	0.2994	0.1142	0.0496	0.0035	0.0977	0.00000004	0.00000008
13	1012	BEAS AT 100M	PUNJAB	0.0987	0.2849	0.1179	0.051	0.0046	0.153	0.00000006	0.00000011
14	1013	SATLUJ AT U/S	HIMACHAL PRADESH	0.0522	0.3924	0.1357	0.0436	0.0022	0.0509	0.00000001	0.00000002
15	1014	SATLUJ AT U/S	HIMACHAL PRADESH	0.079	0.4186	0.1361	0.0325	0.0005	0.0367	0.00000001	0.00000002
16	1015	SATLUJ AT D/S	HIMACHAL PRADESH	0.0665	0.436	0.1405	0.0364	0	0.0301	0.00000001	0.00000002
17	1016	SATLUJ AT D/S	HIMACHAL PRADESH	0.0916	0.3547	0.1414	0.033	0.0078	0.0172	0.00000001	0.00000002
18	1017	SATLUJ AT 100I	PUNJAB	0.1202	0.3372	0.1167	0.0393	0.0019	0.0668	0	0.00000003
19	1018	SATLUJ AT 100I	PUNJAB	0.1237	0.314	0.1277	0.0444	0.0097	0.0955	0.00000003	0.00000005
20	1019	SATLUJ AT U/S	PUNJAB	0.1273	0.3023	0.1216	0.0564	0.016	0.1353	0.00000016	0.00000027
21	1022	SATLUJ AT BRID	PUNJAB	0.1488	0.1715	0.1179	0.0678	0.0135	0.1375	0.00000004	0.0000001
22	1046	GANGA AT ALL	UTTAR PRADESH	0.1538	0.2703	0.1367	0.0821	0.0279	0.0626	0.00000028	0.00000075
23	1046	GANGA AT ALL	UTTAR PRADESH	0.1538	0.2703	0.1367	0.0821	0.0279	0.0626	0.00000028	0.00000075
24	1049	GANGA AT ALL	UTTAR PRADESH	0.1567	0.2616	0.1356	0.0893	0.0262	0.0743	0.00000035	0.0000009
25	1049	GANGA AT ALL	UTTAR PRADESH	0.1567	0.2616	0.1356	0.0893	0.0262	0.0743	0.00000035	0.0000009

Figure 3: Normalised data set

Determining the value of K: The value of K is taken as 7 as concluded from the average silhouette width of the whole dataset (0.636) by observing performance vector with the Rapid miner tool. The value of 'K' is shown in below Figure 4.

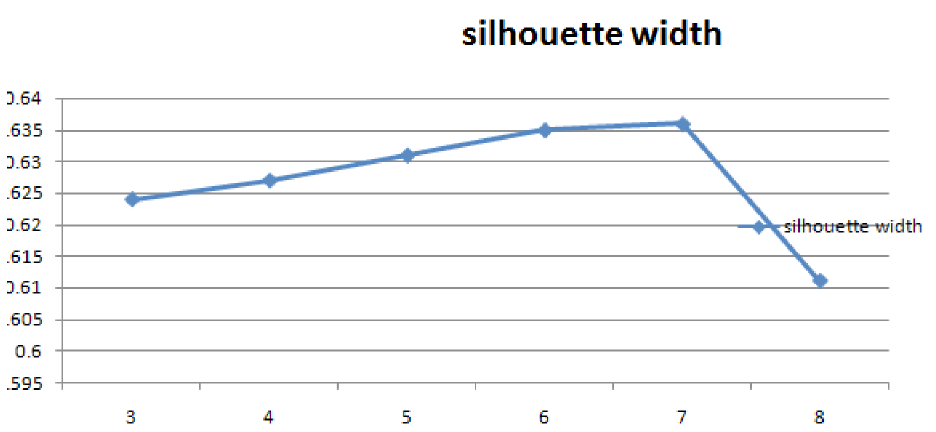


Figure 4: Performance vector

The normalized data set was then used as an input for K-Mean clustering using the rapid miner tool. Seven clusters of the rivers were chosen and mixed Euclidean distance measure for K-mean clustering. The below screenshot in Figure 4 shows the configuration for K-Mean clustering in the tool.

5. RESULTS

Below are the images of the clusters along with the individual water quality parameters.

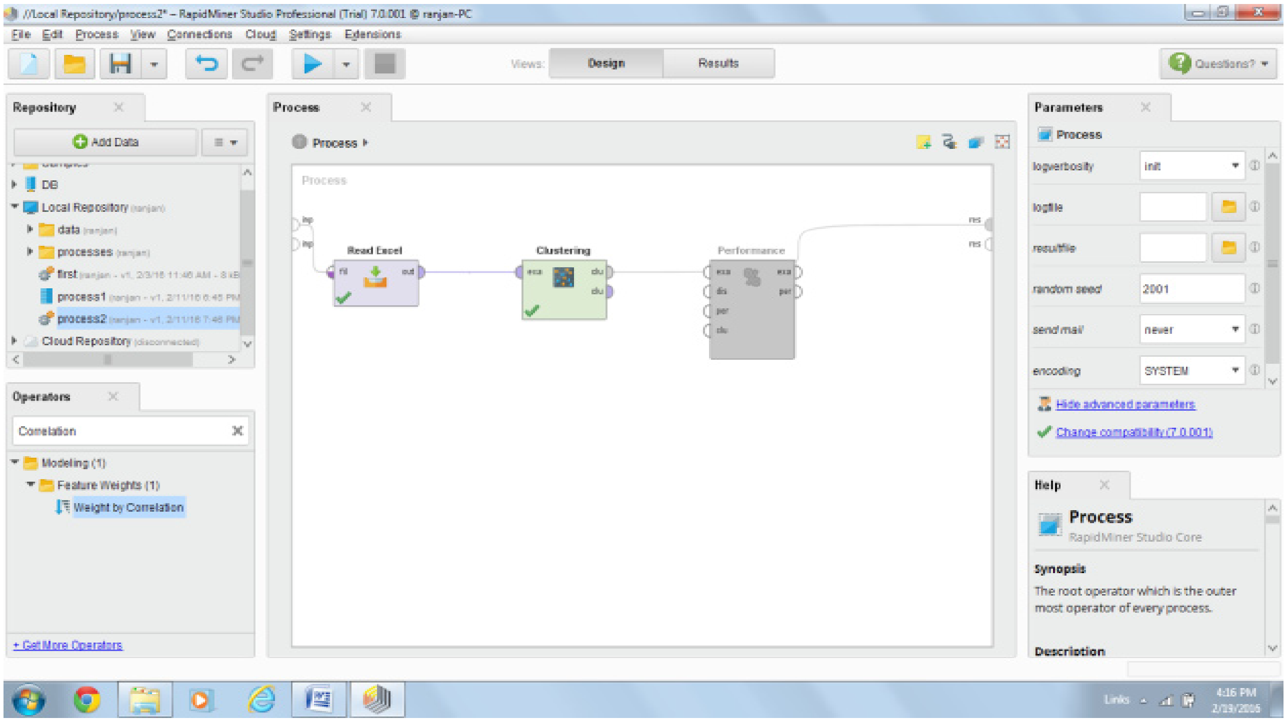


Figure 5: Configuration for K-Means clustering

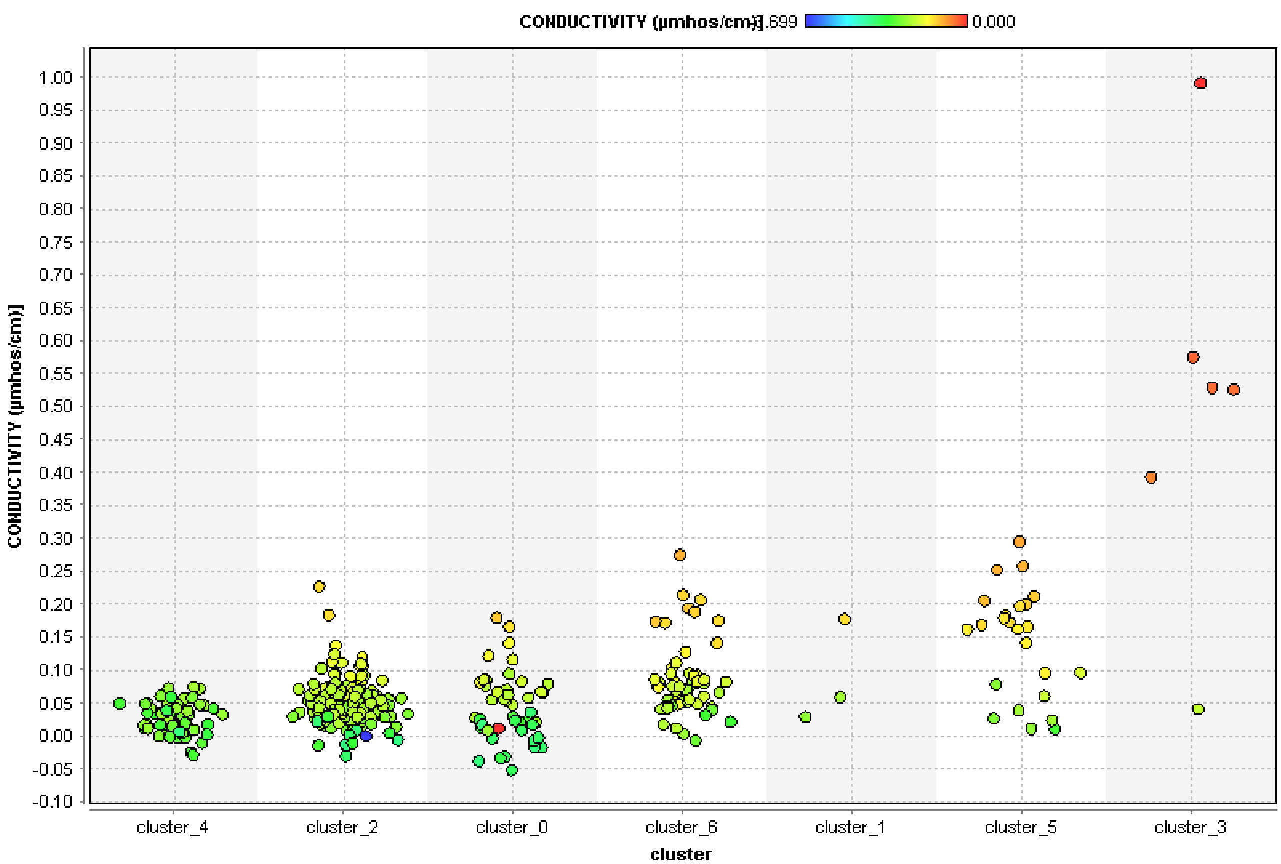


Figure 6: Clusters versus Conductivity

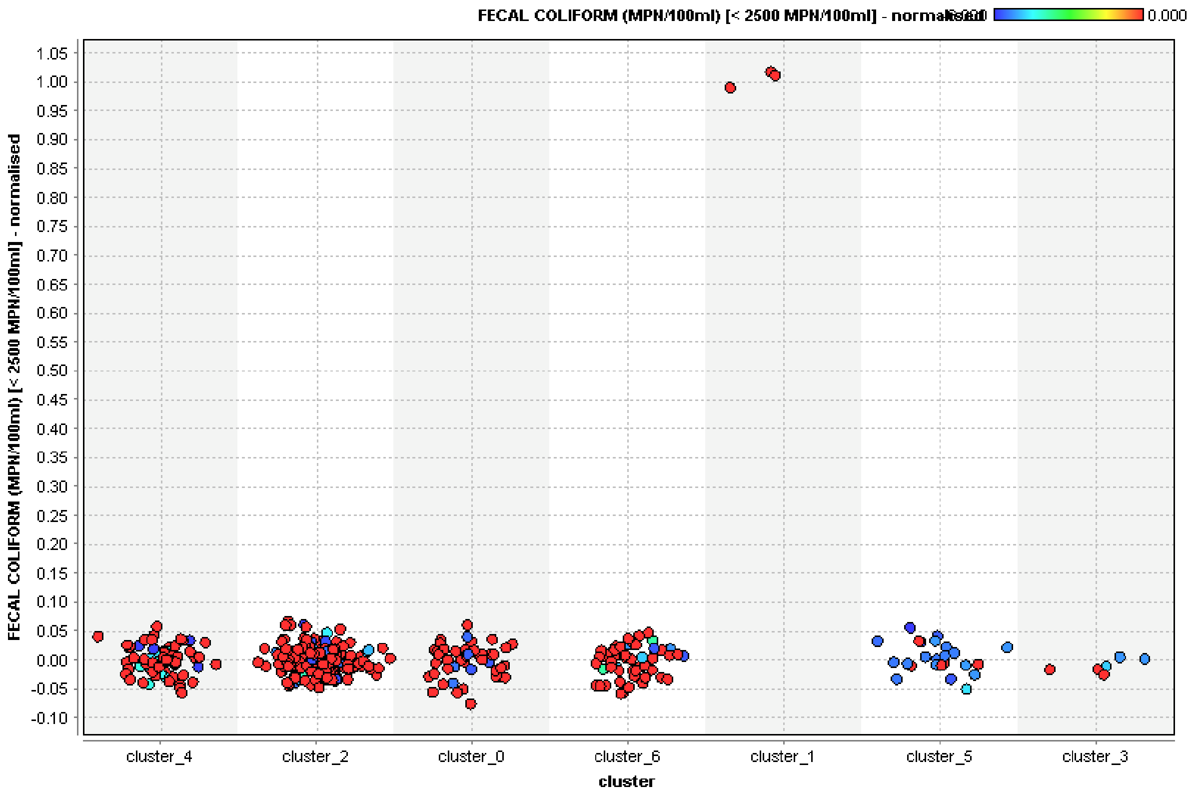


Figure 7: Clusters versus Fecal Coliform

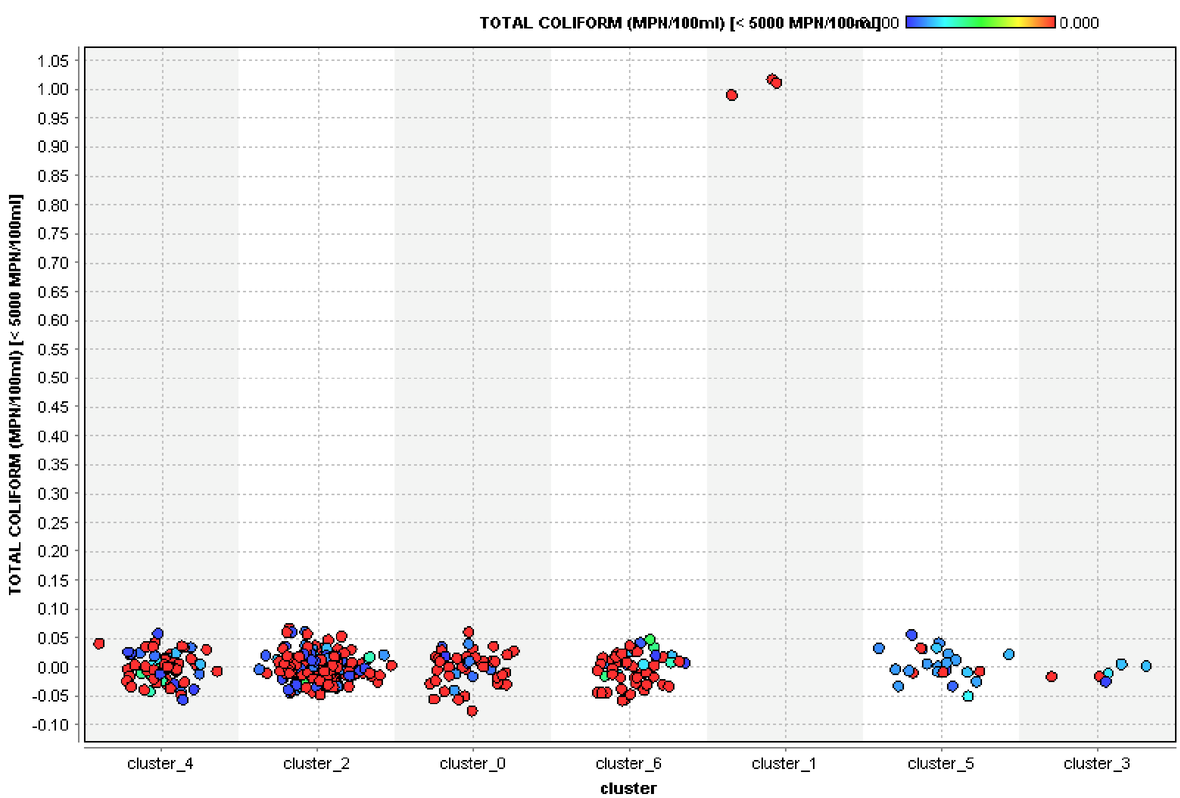


Figure 8: Clusters versus Total Coli form

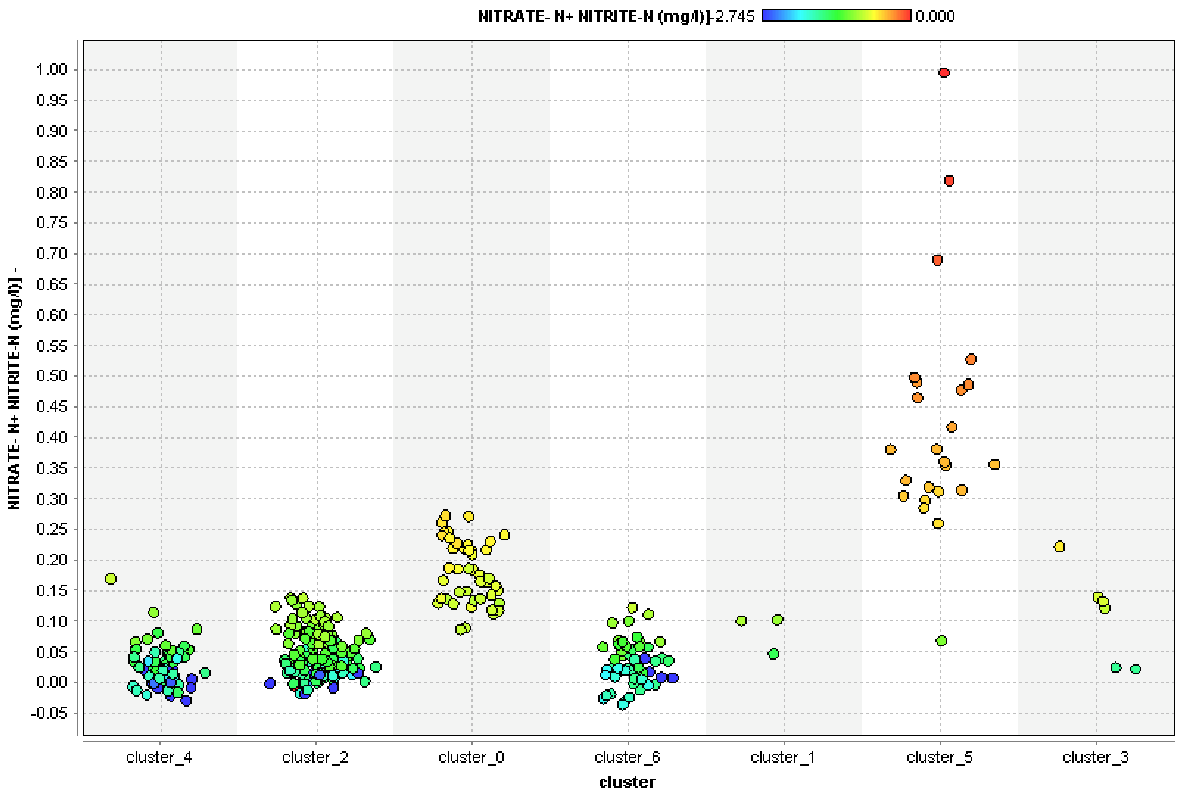


Figure 9: Clusters versus Nitrate

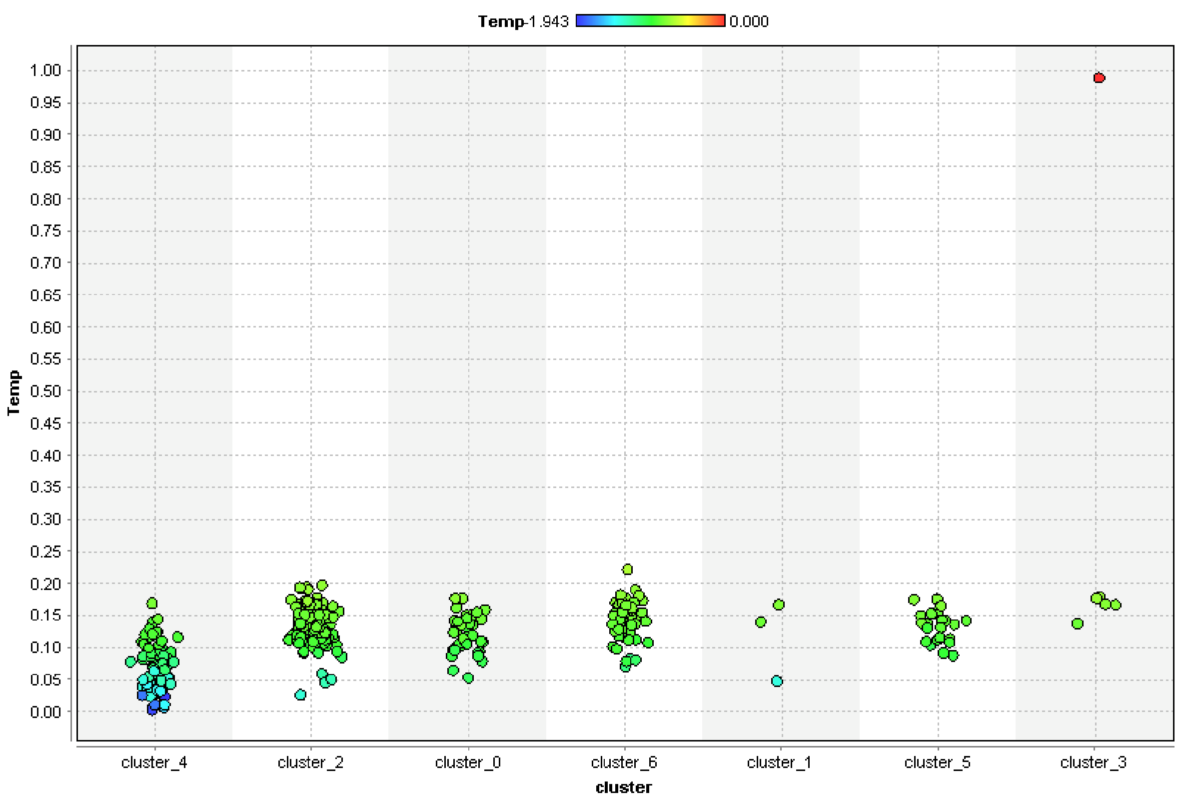


Figure 10: Clusters versus Temp.

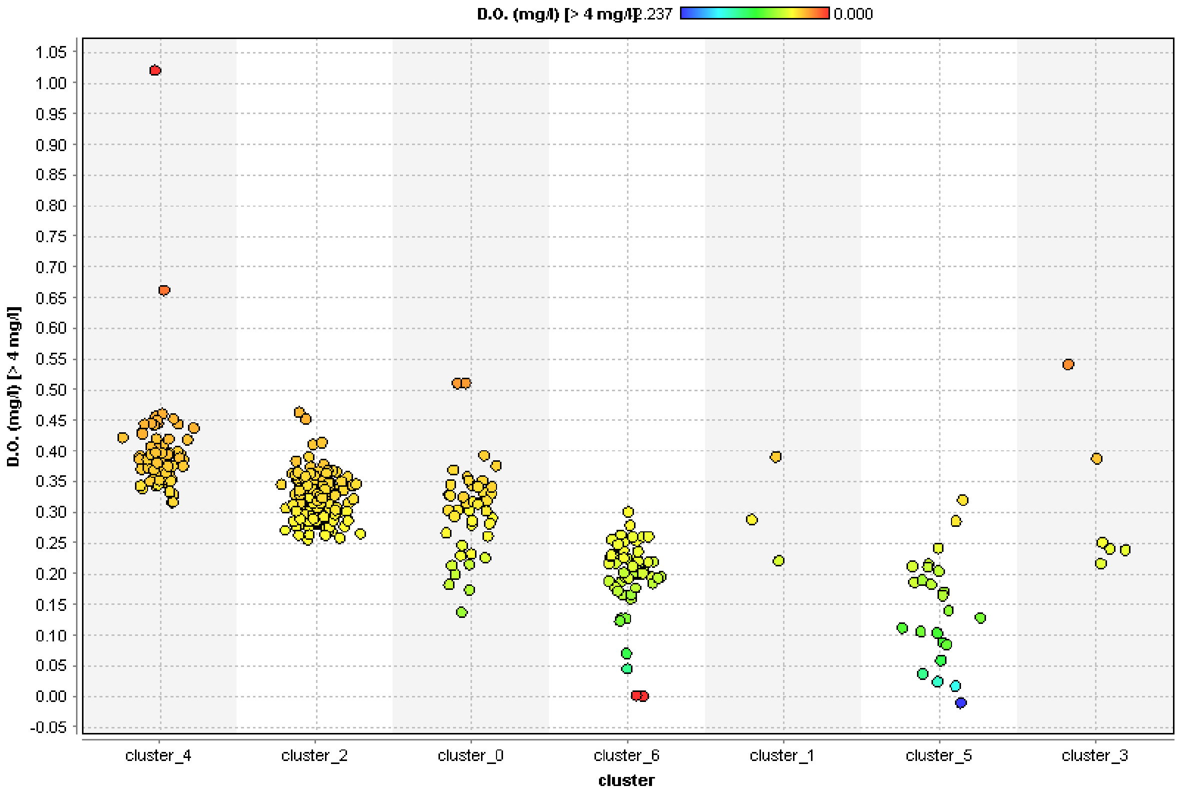


Figure 11: Clusters versus DO

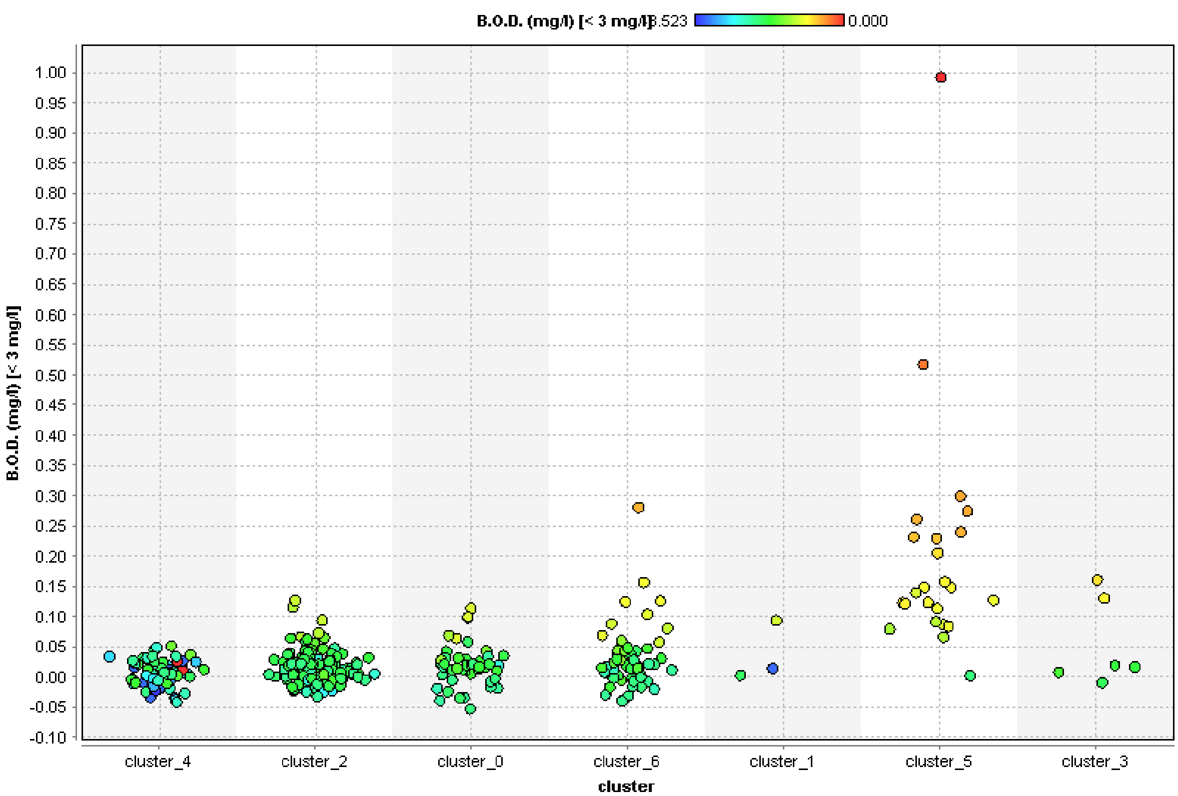


Figure 12: Clusters versus BOD

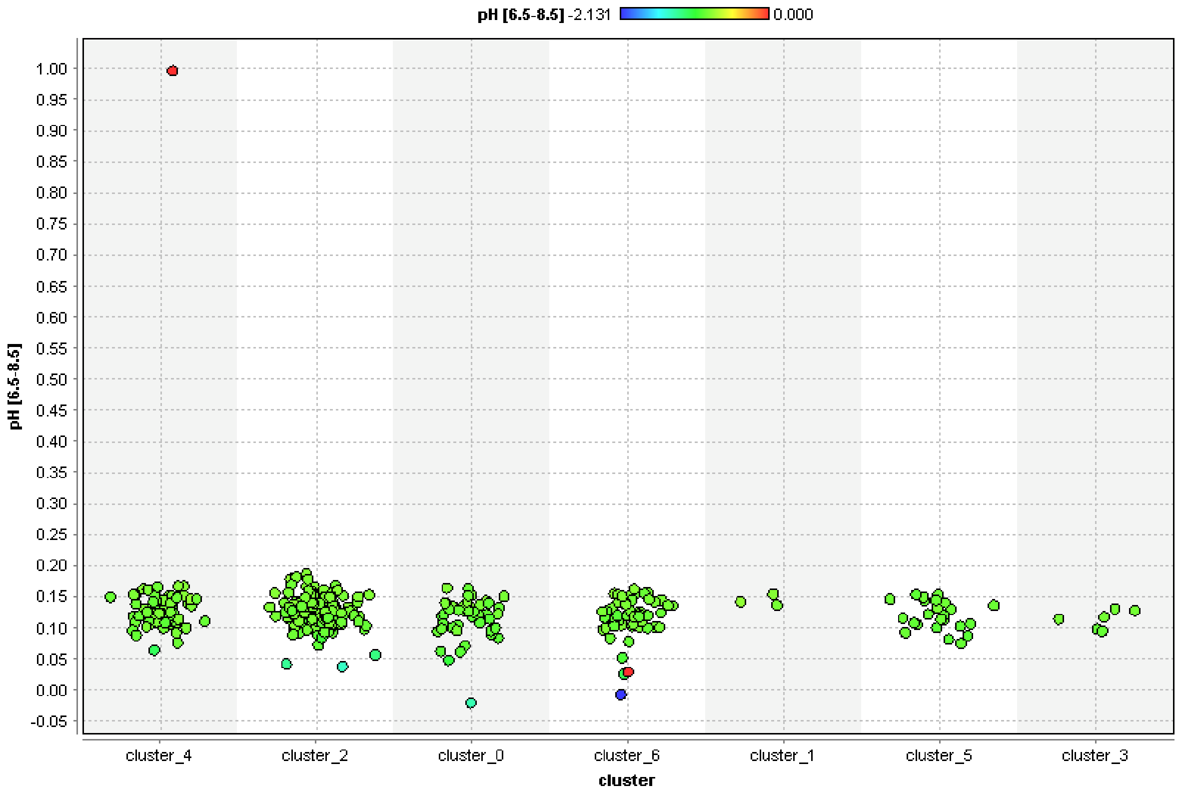


Figure 13: Clusters versus PH

From the above scatter plot it was found that some river stations in a cluster have individual water quality parameter values meeting the standard specifications of central pollution control board and some stations have values outside the standard specification values. The results have been shown in below (Refer Table 2)

Table 2
Summary of water quality parameters cluster-wise with respect to standard values

Water quality parameter	*No. of stations in Cluster 0*	*No. of stations in Cluster 1*	*No. of stations in Cluster 2*	*No. of stations in Cluster 3*	*No. of stations in Cluster 4*	*No. of stations in Cluster 5*	*No. of stations in Cluster 6*	*Remarks Meets the criteria (✓), Not meeting the criteria (✗)*
Temp.(25 C-35C)	2	1	35	5	16	4	19	✓
Temp.(<25C)	43	2	125	0	56	20	34	✗
Temp.>35 C	0	0	0	1	0	0	0	✗
DO>4 mg	45	3	160	6	72	16	49	✓
DO<4 mg	0	0	0	0	0	8	4	✗
PH[6.5-8.5]	41	2	155	6	69	24	50	✓
PH<6.5	2	0	4	0	1	0	3	✗
PH>8.5	2	1	1	0	2	0	0	✗
Conductivity<2250 µS/cm	45	3	160	1	72	24	53	✓
Conductivity>2250 µS/cm	0	0	0	5	0	0	0	✗
BOD<3 mg	30	2	109	1	63	2	25	✓
BOD>3 mg	15	1	51	5	9	22	28	✗

Water quality parameter	*No. of stations in Cluster 0*	*No. of stations in Cluster 1*	*No. of stations in Cluster 2*	*No. of stations in Cluster 3*	*No. of stations in Cluster 4*	*No. of stations in Cluster 5*	*No. of stations in Cluster 6*	*Remarks Meets the criteria (✓), Not meeting the criteria (✗)*
Nitrate< 50 mg	45	3	160	6	72	24	53	✓
Nitrate>50 mg	0	0	0	0	0	0	0	✗
FC<2500 MPN	31	0	85	2	48	5	39	✓
FC>2500 MPN	14	3	75	4	24	19	14	✗
TC<5000 MPN	30	0	72	2	46	4	38	✓
TC>5000 MPN	15	3	88	4	26	20	15	✗

Findings mentioned in Table 2 are summarized below:

Temperature: Optimal temperature >25 deg. Celsius (0.155 normalized value) and < 35 deg. Celsius. 43 stations from cluster0, 02 stations from cluster1, 125 stations from cluster2, 56 stations from cluster4, 20 stations from cluster 5, 34 stations from cluster6 fall below min. temp. level of 25 deg. Celsius and one station from clsuter3, station code 1021,satluj,Punjab which is 143 C which is highest among other rivers.

D.O: DO (mg/l) [> 4 mg/l] or normalized value> 0.10 is a good water quality indicator. It was found out that eight stations from cluster5, four stations from cluster6 fall below the mentioned level of DO range which indicates these stations have poor water quality. All other stations apart from these stations have DO level meeting the criteria.

PH: Level [6.5-8.5] or normalized value between 0.09 and 0.147 is a good water quality indicator. Four stations from cluster0, one station from cluster1, five stations from cluster2, three stations from cluster4, three stations from cluster 6 falls outside the mentioned level of PH range which indicates these stations have poor water quality. All other stations apart from these stations have a PH level meeting the criteria.

Conductivity: Five stations from cluster3 have conductivity > 2250 which is outside the mentioned range and all other stations have conductivity within expected range.

BOD: Permissible limit of B.O.D. (mg/l) is < 3 mg/l or normalized value < 0.015 for good water quality. 15 stations from cluster0, one station from cluster1, 51 stations from cluster2, 5 stations from cluster3, 9 stations from cluster 4, 22 stations from cluster5, 28 stations from cluster6 have a more BOD level than the mentioned level of BOD range which indicates these stations have poor water quality. All other stations apart from these stations have BOD level meeting the criteria.

Nitrate: Desired limit of nitrate is < 50 mg/L or normalized value <1. All the stations of all clusters fall within the limit of mentioned nitrate level.

FC: Permissible limit in (MPN/100ml) is < 2500 MPN/100ml or normalized value 0.00000025. 14 stations from cluster0, three stations from cluster1, 75 stations from cluster2, 4 stations from cluster3, 24 stations from cluster 4, 19 stations from cluster5, 14 stations from cluster6 have more FC level than the mentioned level which indicates these stations have poor water quality. All other stations apart from these stations have FC level meeting the range.

TC: Permissible limit of TC in (MPN/100ml) is < 5000 MPN/100ml or normalized value < 0.00000050. 15 stations from cluster0, three stations from cluster1, 88 stations from cluster2, 4 stations from cluster3, 26 stations from cluster 4, 20 stations from cluster5, 15 stations from cluster6 have more TC level than the mentioned level which indicates these stations have poor water quality. All other stations apart from these stations have TC level within the mentioned limit.

5.1. Water Quality Assessment

Case Study1: River stations with good water quality

By analyzing the 8 water quality parameters the best four stations with respect to overall water quality is mentioned in Table 3.

Table 3
River Stations with Best Water Quality

Station Code	*Station Name*	*River Name*	*State*
2121	Betwa Near Road Bridge, Bhojpur	Betwa	Madhya Pradesh
2122	Betwa Near W/S Intake Well Point Raisen	Betwa	Madhya Pradesh
2603	Beas at D/S Pandoh Dam	Beas	Himachal Pradesh
3380	Damring River (Krishnei River) at Resubelpara, East Garohills	Damring	Himachal Pradesh

Case Study 2: River Stations with lower level of water Quality

Similarly, four river stations having water quality, lower than the recommended level have been identified. The river stations Yamuna at Nizamuddin (station code, 1121), Yamuna at Okhla Bridge (station code, 1375), Yamuna at Okhla after meeting of Sahadra drain (station code, 1812) in Delhi and Hindon river in the Ghaziabad D/s (station code, 1358) in Uttar Pradesh have lower water quality levels.

6. CONCLUSION

Fresh water is a limited resource and hence its conservation and quality management is highly essential for sustainable development. Water crisis in India, especially in urban areas is growing day by day and is a future threat to water sustainability. In this paper, research has conducted experiments for assessing the water quality of the major rivers of north India. The rivers have been classified using K-Mean clustering technique based on the similarity of their water quality parameters. The four river stations having overall good water quality and four stations having lower water quality levels has been found out from the experiments. It is observed that Betwa river in Madhya Pradesh, Beas and Damring rivers in Himachal Pradesh have good water quality as compared to other northern Indian rivers. Similarly Yamuna river in Delhi, Hindon river in Uttar Pradesh have lower water quality levels. Due to the high density of population, untreated domestic and industrial wastewater flows to Yamuna river, it is becoming more polluted among all the north Indian rivers. Future studies should focus on water recycling and wastewater management in urban cities for water sustainability.

Acknowledgements

The authors are grateful to Indian government for providing the data repository and water quality specification standards to assess water quality of major north Indian rivers.

REFERENCES

[1] Evaluation of Operation and Maintenance of Sewage Treatment Plants in India. Series: Cups/68/2007.Central Pollution Control Board, India.2007. Available from: http://www.cpcb.nic.in/upload/NewItems/NewItem_99_NewItem_99_5.pdf.

[2] Annual Report (2008-2009).Central Pollution Control Board, India. Available from: http://cpcb.nic.in/upload/AnnualReports//AnnualReport_37_ANNUAL_REPORT-08-09.pdf.

[3] Available from: https://data.gov.in/.

[4] Standard Methods for the Examination of Water and Waste Water. American Public Health Association (APHA-AWWAWPCH).1998. 20th ed. APHA Washington D.C., 1270.

[5] Kumawat DM, Sharma MK. Quality status study of river Kshipra at Ujjain before its linkage with Narmada water. Journal of Global Biosciences. 2015;4(2):1508-16.

[6] Joseph PV, Jacob C. Bacterial, Fungal and Algal Population of Pennar River: A Fresh Water Wetland in Kottayam District, Kerala. Asian Journal of Chemistry. 2010 Jul 1;22(6):4286.

[7] Raja P, Amarnath AM, Elangovan R, Palanivel M. Evaluation of physical and chemical parameters of river Kaveri, Tiruchirappalli, Tamil Nadu, India. J. Environ. Biol. 2008 Sep 1;29(5):765-8.

[8] Ahipathy MV, Puttaiah ET. Ecological characteristics of vrishabhavathy River in Bangalore (India). Environmental geology. 2006 Apr 1;49(8):1217-22.

[9] Singh KP, Malik A, Mohan D, Sinha S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. Water research. 2004 Nov 30;38(18):3980-92.

[10] Vega M, Pardo R, Barrado E, Debán L. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water research. 1998 Dec 31;32(12):3581-92.

[11] Kebede YK, Kebedee T. Application of principal component analysis in surface water-quality monitoring. Principal Component Analysis-Engineering Applications, P. Sanguansat, Ed. 2012:83-100.

[12] Available from: http://www.cpcb.nic.in/Water_Quality_Criteria.php.

[13] Behar S. Definition of Water Quality Parameters. Testing the Waters: Chemical and Physical Vital Signs of a River. Montpelier, VT: River Watch Network.1997.

[14] American public health association (APHA-AWWAWPCH),1998,standard methods for the examination of water and waste water,20th ed. APHA Washington D.C.,1270.

[15] Liao H, Sun W. Forecasting and evaluating water quality of Chao Lake based on an improved decision tree method. Procedia Environmental Sciences. 2010 Dec 31;2:970-9.

[16] Qun M, Ying G, Zhiqiang L, Xiaohui T. Application of Comprehensive Water Quality Identification Index in Water Quality Assessment of River. In Intelligent Systems, GCIS'09. WRI Global Congress on 2009 May 19. Xiamen. P. 333-337.

[17] Zhang YH, Jia HF, Yang C. Assessment of spatial-temporal variations in surface water quality of Luzhi river system in Plain River-Net Areas, Suzhou, China. InRemote Sensing, Environment and Transportation Engineering (RSETE), 2011. International Conference on 2011 Jun 24.p. 2002-2005.

[18] Camejo J, Pacheco O, Guevara M. Classifier for drinking water quality in real time. In Computer Applications Technology (ICCAT), 2013 International Conference on 2013 Jan 20. p. 1-5.

[19] Aghaarabi E, Aminravan F, Sadiq R, Hoorfar M, Najjaran H, Rodriguez MJ. A fuzzy rule-based approach for water quality assessment in the distribution network. In IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS), 2013 Joint 2013 Jun 24.p.1149-1154.

[20] Ali M, Qamar AM. Data analysis, quality indexing and prediction of water quality for the management of rawal watershed in Pakistan. In Digital Information Management (ICDIM), 2013 Eighth International Conference on 2013 Sep 10.p. 108-113.

[21] Harris JM. Sustainability and sustainable development. International Society for Ecological Economics. 2003 Feb:1-12.