ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/269183046

Water quality analysis of Godavari river basin using multivariate analysis techniques

Article *in* Indian Journal of Environmental Health · January 2013

Source: PubMed

CITATIONS
0

READS
770

4 authors:

Indrani Gupta
National Environmental Engineering Research Institute
19 PUBLICATIONS 449 CITATIONS

SEE PROFILE

Abhaysinh Salunkhe
National Environmental Engineering Research Institute
11 PUBLICATIONS 96 CITATIONS

SEE PROFILE

Nanda Rohra
Institute of Chemical Technology, Mumbai
6 PUBLICATIONS 37 CITATIONS

SEE PROFILE

Rakesh Kumar
114 PUBLICATIONS 2,327 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project national ambient air quality monitoring View project

All content following this page was uploaded by Abhaysinh Salunkhe on 03 December 2015.

The user has requested enhancement of the downloaded file.

J Environ. Science & Engg. Vol. 55, No. 1, p. 65-80, January 2013

Water Quality Analysis of Godavari River Basin Using Multivariate Analysis Techniques

INDRANI GUPTA[+], ABHAYSINH SALUNKHE,
NANDA ROHRA AND RAKESH KUMAR

Multivariate statistical techniques, including cluster analysis, principal component analysis factor analysis and discriminant analysis, have been used to evaluate spatial variations and to interpret a large and complex water quality data set collected from the Godavari river basin. The data sets, containing 7 parameters, were generated during the 3-years (2007-2009) at 78 different sites along the river and its tributaries. Water quality indices based on four parameters (pH, DO, BOD and FC) calculated for all the sites were found to be medium to good, good to excellent and bad using modified NSF index. Three significant groups (cleaner, slightly and moderately polluted sites) were detected by CA method, and three latent factors were identified by PCA method. The results of DA revealed that only two parameters (i.e. pH and BOD) were necessary for analysis in spatial variation. 83.3% of the original sites were correctly classified using discriminant function developed from the analysis.

Key words: *Water Quality Index, Multivariate Statistical Analysis, Godavari River Basin*

1. Introduction

Rivers are among the most vulnerable water bodies to pollution because of their role in carrying municipal and industrial wastes and run-offs from agricultural lands in their vast drainage basins. The chemometrics methods help in explaining complex water quality monitoring data to get a better understanding of the ecological status of the studied systems[1]. The multivariate statistical analysis has been successfully applied in a number of studies, till date.[2-6] All the studies show that multivariate statistical analysis can help to interpret the complex data sets and assess the water quality. In addition, multivariate statistical analysis is useful in verifying temporal and spatial variations caused by natural and anthropogenic factors linked to seasonality. In the study, the large database analyzed, was obtained during a 3-years (2007-2009) monitoring program of the Godavari River Basin. It was subjected to multivariate statistical techniques cluster analysis (CA), principal components analysis (PCA), and discriminant analysis (DA) with a view to extract information about the similarities or dissimilarities among the sampling sites. Latent factors in river water quality were identified and the water quality variables responsible for spatial variations explained the structure of the data sets. Further, the hidden factors revealed the influence of possible sources on the water quality parameters and pollution levels. of sampling stations in the Godavari river basin.

2. Materials and methods

2.1. *Study area*

The Godavari, perennial river of India, is the second largest river in India. Godavari river rises near Trayambak (Nasik), northeast of Mumbai at an

Mumbai Zonal Laboratory, CSIR-National Environmental Engineering Research Institute(CSIR-NEERI), Nehru Marg, Nagpur-440 020 (India)
[i_gupta@neeri.res.in, abhay.salunkhe@rediffmail.com, nandarohra@ymail.com, r_kumar@neeri.res.in]
[+]Corresponding author: e-mail: i_gupta@neeri.res.in, Tel. : +91-22-24973521; Fax- +91-22-24936635

Dr. Indrani Gupta
Sr. Principal Scientist
CSIR-National Environmental
Engineering Research Institute (CSIR-NEERI)
Mumbai Zonal Laboratory
89/B, Dr. Annie Besant Road,
Worli, Mumbai – 400 018

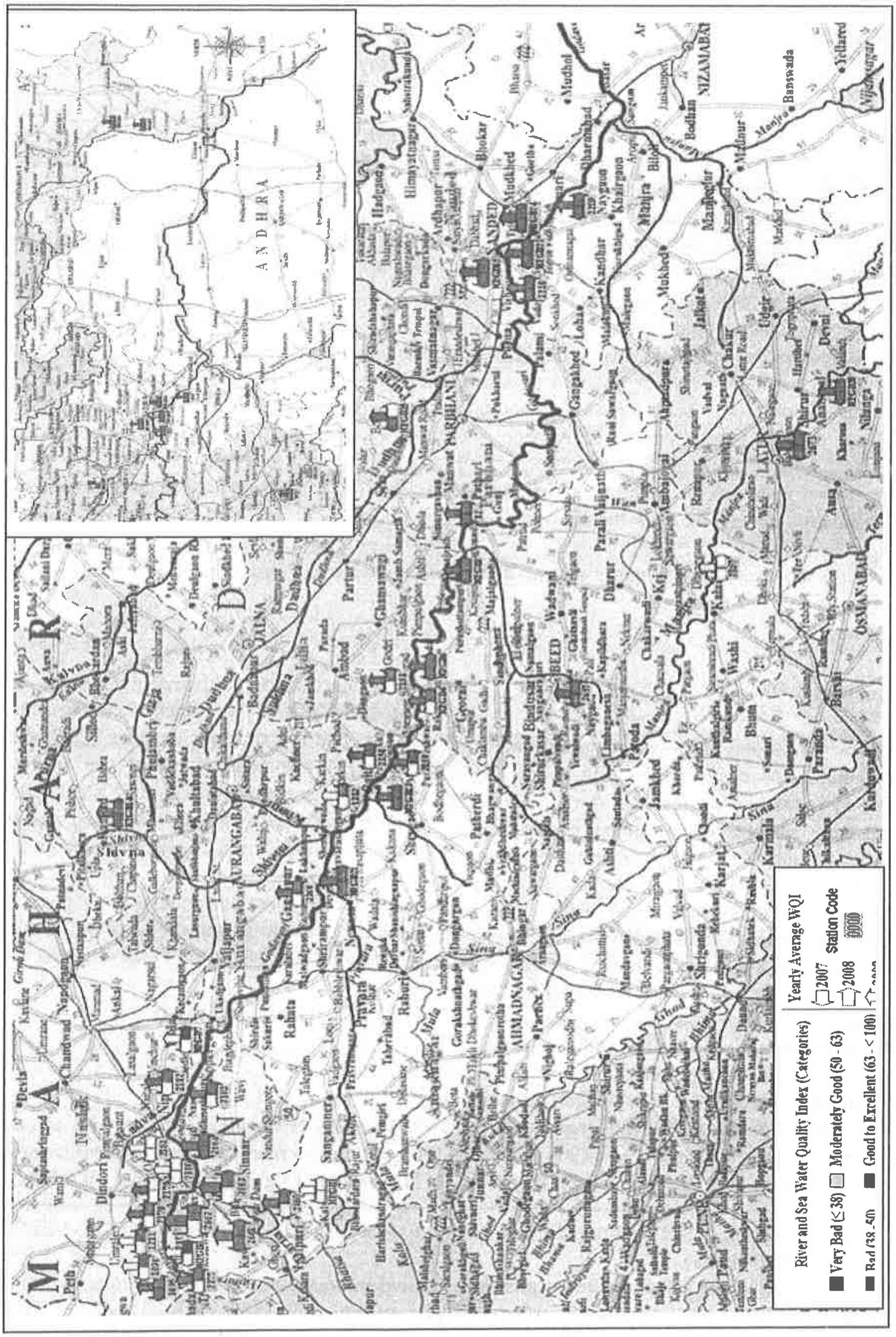

Fig. 1: WQI of Godavari River for 2007-2009

Table 1: Mean and standard deviations at monitoring locations on Godavari River

Sr. No.	St No.	St Name	pH		DO		BOD		FC		TC		NH_3-N		Temp	
			Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev
1	HP_GD_2	Dhalegaon	8.1	0.2	6.0	0.8	2.5	0.5	23.5	50.8	104.2	158.1			27.9	1.2
2	HP_GD_4	Hirapur	7.2	0.7	6.2	0.7	2.0	0.5	10.4	5.9	27.8	15.8			27.8	1.1
3	HP_GD_6	Killari	8.0	0.3	6.2	0.6	2.0	0.7	7.8	7.6	20.7	12.6			27.9	1.2
4	HP_GD_14	Pishor	8.0	0.2	5.9	0.4	2.1	0.4	2.1	3.8	10.9	7.2			27.0	1.3
5	HP_GD_15	Purnabridge	8.2	0.4	6.1	0.5	2.1	0.6	4.2	5.1	24.0	17.6			25.8	1.2
6	HP_GD_20	Takli	7.9	0.2	6.4	0.5	2.0	0.5	6.0	8.3	19.4	17.1			28.6	1.1
7	HP_GD_21	Taklidhangar	8.1	0.5	6.3	0.5	2.1	1.0	53.2	88.4	110.0	156.0	0.2	0.3	26.9	0.7
8	HP_GD_22	Toka	7.9	0.5	6.2	0.5	2.4	1.8	6.1	15.0	21.9	27.6	0.0	0.1	25.8	1.4
9	HP_GD_23	Wadvali	8.1	0.5	6.5	0.8	1.7	0.6	8.3	14.7	35.3	48.5	0.0	0.1	27.4	1.5
10	HP_GD_24	Yelli	8.2	0.6	6.0	0.6	2.1	0.6	6.3	10.1	16.4	15.7	0.2	0.4	27.3	1.2
11	HP_GD_25	Zari	7.9	0.3	6.1	0.4	2.0	0.6	4.4	5.6	12.8	7.4	0.0	0.0	27.8	1.3
12	HP_GD_1	Bhandardara	7.7	0.6	6.7	0.7	8.3	0.9	306.7	126.6	490.0	111.4	0.3	0.4	26.7	0.7
13	HP_GD_3	Gangapur Dam	8.1	0.8	7.7	3.7	13.1	3.6	253.3	171.0	1573.3	640.4	0.1	0.0	28.0	0.7
14	HP_GD_7	Kopargaon	7.7	0.7	6.9	0.5	15.0	15.9	83372.9	172580.9	1337942.9	3467043.5	0.1	0.0	27.0	1.6
15	HP_GD_26	Anantwadi	8.3	0.2	6.4	0.7	2.7	0.5	99.7	156.2	378.8	598.4	0.0	0.0	27.0	4.2
16	HP_GD_27	Damrencha	8.2	0.3	6.7	0.8	2.1	0.5	11.1	11.5	34.7	33.4	0.0	0.2	26.1	2.8
17	HP_GD_28	Deori	8.3	0.3	6.5	0.6	2.3	0.3	59.9	88.4	219.2	338.9	0.0	0.0	25.1	3.5
18	HP_GD_29	Dhaba	8.4	0.2	5.5	0.9	3.6	0.6	117.0	127.1	459.1	587.9	0.0	0.1	27.2	3.5
19	HP_GD_30	Drugwada	8.4	0.1	5.7	1.3	2.6	0.8	64.7	130.2	327.2	673.7	0.0	0.1	28.8	4.0
20	HP_GD_31	Kamtheekhairi	8.4	0.3	6.4	0.4	2.6	0.4	89.0	128.9	485.0	771.8	0.0	0.0	26.1	2.6
21	HP_GD_32	Kardha	8.3	0.2	6.3	0.6	2.5	0.5	53.8	70.0	184.0	239.1	0.0	0.0	28.0	4.6
22	HP_GD_33	Kolgaon	8.3	0.2	6.1	0.6	2.8		86.9	105.7	341.2	401.8	0.0	0.0	26.6	2.9
23	HP_GD_34	Mahagaon	8.4	0.2	6.5	0.6	2.5		59.3	108.3	215.1	375.9	0.0	0.0	27.5	3.3
24	HP_GD_35	Mathani	8.3	0.1	6.3	0.5	2.8		66.9	85.5	180.7	233.8	0.0	0.0	27.9	4.4
25	HP_GD_36	Petta	8.1	0.3	6.3	0.9	2.2		16.6	17.8	41.1	34.9	0.1	0.2	26.8	2.1
26	HP_GD_37	Saiphal	8.4	0.2	6.3	0.6	2.7		45.1	47.4	149.5	157.2	0.0	0.0	26.4	4.2
27	HP_GD_38	Dindora (Soit)	8.4	0.2	6.1	0.7	2.9		78.8	95.5	289.2	326.6	0.0	0.0	28.9	3.7
28	HP_GD_39	Tembhurdoh	8.3	0.2	6.3	0.6	2.6		93.8	171.3	395.4	604.7	0.0	0.0	27.4	3.4
29	HP_GD_40	Wadsa (Chincholi)	8.4	0.1	6.4	0.6	2.9		25.0	30.9	91.2	89.4	0.0	0.1	26.5	

Contd... Table 1: Mean and standard deviations at monitoring locations on Godavari River

Sr. No.	St No.	St Name	pH		DO		BOD		FC		TC		NH_3-N		Temp	
			Mean		Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev
30	HP_GD_41	Wagholi-Butti	8.4	0.0	6.5	0.7	2.5		23.0	25.9	89.4	93.8	0.0	0.1	28.3	4.6
31	HP_GD_42	Warud Bagaji	8.3	0.2	6.3	0.5	2.6		48.1	82.7	154.8	216.0	0.0	0.1	29.5	3.2
32	12	Dhalegaon. Parbhani.	7.9	0.4	6.4	1.1	4.2	0.8	7.2	2.1	263.9	49.8	1.0	1.6	26.6	2.1
33	1095	U/s of Gangapur Dam. Nashik	7.9	0.5	6.5	0.8	5.3	1.9	4.9	3.5	16.5	8.6	0.4	0.4	26.8	4.2
34	1096	Panchavati Nashik	7.8	0.5	5.9	1.0	8.3	3.5	36.8	91.1	202.4	418.8	0.6	1.1	27.3	4.3
35	1209	Raher.District -Nanded.	31.4	137.9	6.3	1.1	4.2	0.6	7.2	2.2	266.5	40.3	1.0	1.5	28.9	1.0
36	1210	Intake of pump house. Nanded	8.1	0.3	6.8	0.4	4.1	0.5	7.3	2.4	255.7	54.8	0.3	0.7	27.0	0.2
37	1211	D/s near Amardham, Nashik	7.8	0.5	5.3	1.1	10.9	6.1	32.4	45.6	229.6	431.2	0.8	1.5	27.5	4.5
38	1312	Jaikwadi Dam, Aurangabad.	7.8	0.5	6.3	1.1	4.1	0.9	7.1	2.3	264.1	46.3	1.0	1.4	27.4	5.3
39	2158	Paithan intake, Aurangabad.	7.7	0.4	6.0	1.3	4.6	1.1	6.5	1.3	287.5	23.1	1.7	1.7	28.5	5.0
40	2159	D/s of Paithan, Paithan, Aurangabad.	7.9	0.6	6.1	1.4	4.5	1.2	6.4	1.5	283.0	32.5	2.5	2.9	28.6	4.9
41	2160	U/s of Aurangabad Reservoir. Aurangabad.	7.7	0.5	6.2	1.1	4.0	1.1	6.6	1.6	291.0	37.7	1.7	1.6	28.2	5.1
42	2161	Jalna Intake, Shahabad. Ambad, Jalna.	7.7	0.4	5.9	1.3	4.0	0.9	6.6	1.0	283.9	26.4	1.1	1.3	26.8	2.5

Contd... Table 1 : Mean and standard deviations at monitoring locations on Godavari River

Sr. No.	St No.	St Name	pH		DO		BOD		FC		TC		NH_3-N		Temp	
			Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev
43	2177	Someshwar Temple. Nashik	7.9	0.4	6.0	1.0	6.1	2.0	8.7	6.2	23.8	17.9	1.0	1.2	25.3	3.6
44	2178	Chikhali nallah meets Godavari river. Nashik,	7.2	1.4	4.4	2.1	23.8	12.8	15.5	11.3	34.8	19.2	10.4	11.8	24.9	3.4
45	2179	Godavari river at Hanuman Ghat. Nashik	7.6	0.8	5.6	2.0	13.0	10.0	14.2	7.1	34.6	14.0	1.6	2.0	25.5	3.8
46	2180	Tapovan. Taluka-Nashik	7.7	0.6	4.6	1.5	13.6	6.9	18.2	9.9	42.9	18.8	3.4	5.1	25.3	3.6
47	2181	Kapila-Godavari confluence point.Nashik	7.8	0.6	4.6	1.5	13.5	8.1	17.8	10.1	39.7	17.1	4.8	7.3	25.6	3.7
48	2182	Saikheda. Nashik.	7.9	0.3	6.1	0.6	6.4	2.9	8.3	5.1	19.4	10.9	0.8	1.1	25.4	3.6
49	2183	Nandur-Madhmeshwar Dam.District-Nashik.	8.0	0.3	6.2	0.6	5.0	1.5	5.3	3.2	13.3	6.1	0.7	0.9	25.6	3.4
50	2657	Bindusara River at Dam at Paligaon, Beed.	7.8	0.0	6.1	0.0	7.7	0.0	7.4	0.0	300.0	0.0	0.6	0.0	25.0	0.0
51	2660	Darna river at Chehadi Pumping Station Nashik	8.0	0.2	5.4	1.4	10.4	5.8	8.6	4.2	27.2	15.2	2.4	1.5	26.2	3.3
52	2661	Darna river at Aswali, Igatpuri, Nashik	8.0	0.3	6.4	0.9	5.8	2.6	7.3	4.2	31.8	35.7	2.9	2.4	24.8	3.6

Contd... Table 1: Mean and standard deviations at monitoring locations on Godavari River

Sr. No.	St No.	St Name	pH		DO		BOD		FC		TC		NH_3-N		Temp	
			Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev
53	2662	Darna River MES Pumping st, Nashik	8.1	0.3	6.3	1.0	7.3	4.2	8.0	3.4	25.3	16.4	2.4	1.4	25.6	2.4
54	2663	Darna river Bhagur, Pumping st.	8.0	0.5	6.0	0.9	7.5	3.9	9.1	4.2	28.1	16.4	2.1	1.3	26.3	1.9
55	2664	Darna River at SanSari	7.9	0.3	6.0	1.3	7.0	3.9	10.6	3.6	32.2	25.2	2.1	1.3	26.5	2.5
56	1909	Kanhan Village-Agargaon, Nagpur.	7.7	0.3	5.5	0.7	7.6	3.1	226.3	275.9	511.0	588.1	1.2	1.2	25.6	3.0
57	2170	Kanhan river U/s Vid. Paper Mill. Nagpur.	7.8	0.5	6.0	0.6	6.1	1.9	29.6	24.1	87.7	78.3	0.5	0.4	21.9	3.0
58	2171	Kanhan river Vidarbha Paper Mill. Nagpur.	7.8	0.5	4.9	0.9	8.7	4.8	53.3	42.9	157.6	145.5	0.6	0.4	21.4	1.2
59	2673	Manjra river near Latur-Nanded bridge,	7.9	0.5	5.6	1.1	7.6	6.8	7.0	0.0	350.0	0.0	1.2	1.7	23.1	2.6
60	2697	Penganga river Umarkhed Yavatmal.	7.9	0.3	5.7	0.9	6.5	2.0	48.9	29.8	156.4	91.1	0.7	0.2	24.1	2.5
61	2698	Penganga river Pusad, Yavatmal	8.0	0.2	5.9	0.6	6.1	1.4	17.4	7.3	57.8	46.5	0.7	0.2	24.4	2.4
62	1913	Purna river Dhupeshwar Akola,	8.1	0.4	5.6	0.5	7.0	1.0	222.6	195.0	615.6	281.2	1.0	1.1	29.2	1.6
63	2155	Purna river at D/s of confluence of Morna and Purna, Akola.	7.7	0.5	5.1	0.9	9.4	4.1	308.9	274.2	990.0	637.9	1.9	3.4	28.3	4.0

Contd... Table 1: Mean and standard deviations at monitoring locations on Godavari River

Sr. No.	St No.	St Name	pH		DO		BOD		FC		TC		NH_3-N		Temp	
			Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev
64	2700	Purna river near Achalpur - Amravati.	8.0	0.5	4.3	1.1	11.2	2.8	80.0	52.0	172.0	43.8	2.4	2.7	29.0	2.4
65	11	Wainganga river Chandrapur.	8.1	0.5	5.6	0.6	6.9	2.4	94.0	99.0	231.2	257.6	1.2	2.1	25.0	3.3
66	1910	Wainganga, Nagpur.	7.8	0.4	5.5	1.0	9.6	8.0	235.6	282.9	709.4	633.9	0.9	1.0	24.1	3.8
67	2172	Wainganga river, Bhandara.	7.9	0.3	5.3	0.6	8.4	2.9	141.7	117.5	317.3	327.4	0.7	0.6	28.8	3.4
68	2173	Wainganga river Tumsar, Bhandara.	8.0	0.3	6.0	0.6	6.1	1.2	55.9	39.1	111.4	55.3	0.5	0.4	29.1	3.5
69	2175	Wainganga river at U/s of Gaurav Paper Mills,	8.2	0.2	6.1	0.8	6.4	2.1	91.5	67.5	155.1	68.8	0.5	0.3	24.9	3.0
70	1212	Wardha river at Rajura bridge. Chandrapur.	7.9	0.4	5.6	0.7	7.6	2.7	91.3	92.5	245.8	247.2	0.9	1.1	27.0	3.0
71	1315	Wardha river Pulgaon Railway Bridge	8.0	0.4	5.5	0.9	8.8	7.2	148.4	196.9	344.0	353.1	1.0	1.1	26.9	4.3
72	2156	Wardha river, confluence of Penganga & Wardha,	7.9	0.3	5.4	1.0	7.5	2.0	83.8	75.7	163.0	128.2	0.6	0.4	23.8	2.8
73	2174	Wardha river ACC Ghuggus. Chandrapur.	8.0	0.5	5.5	0.6	8.8	4.6	96.1	46.5	186.6	90.5	1.0	1.3	25.7	2.5

Contd... Table 1 : Mean and standard deviations at monitoring locations on Godavari River

Sr. No.	St No.	St Name	pH		DO		BOD		FC		TC		NH_3-N		Temp	
			Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev	Mean	st dev
74	2719	Wardha river Hadasti, Chandrapur.	7.9	0.5	5.3	0.8	9.5	5.4	116.3	91.8	357.8	485.9	1.2	1.4	26.2	2.4
75	2721	Wardha river, Chandrapur.	8.0	0.5	5.9	0.5	6.3	1.9	47.5	38.4	88.1	55.0	1.2	1.4	25.8	2.5
76	2720	Wardha river, Hadasti, Chandrapur.	7.9	0.4	5.7	0.6	7.3	2.8	58.3	45.2	169.0	163.6	1.3	1.4	26.5	2.4
77	2722	Wana River U/s of Mohata Mills, Wardha.	7.9	0.4	5.7	0.6	6.9	2.0	59.4	71.3	161.5	131.1	1.2	1.1	25.6	4.3
78	2723	Wena river, at D/s of Mohata Mills, Wardha.	8.2	0.3	5.2	0.7	8.9	2.4	114.2	117.9	508.0	348.0	1.6	1.5	26.3	5.1

st dev: Standard Deviation; St Name: Station Name

elevation of 1067m and flows for a length of about 1465 km before out falling into the Bay of Bengal. The Godavari river has many tributaries and distributaries. Some of them are Bindusara, Darna, Kanhan, Manjra, Painganga, Purna, Weinganga, Wardha and Wena. In the present study, an analysis has been carried out for 78 sampling stations (**Fig. 1**).

2.2 Data

The data sets of the 78 water quality monitoring stations, which comprised 7 water quality parameters monitored monthly over 3 years (2007-2009), were obtained from the Maharashtra Pollution Control Board. The selected water quality parameters included temperature; pH, dissolved oxygen, biochemical oxygen demand, ammonical nitrogen, Faecal Coliform and Total Coliform. **Table 1** gives the mean and standard deviation of these 7 parameters at 78 monitoring locations.

2.3 Water Qualiy Index (WQI)

An index is a mean device to reduce a large quantity of data down to a simplest form. The water quality indices help to evaluate the water quality profile of a river in its entire stretch as well as to identify the reaches where the gap between the desired and the existing water quality is significant enough to warrant urgent pollution control measures. In India, the National Sanitation Foundation WQI (NSFWQI) is being used by Central Pollution Control Board, with a slight modification in weights[7-8]. The NSF WQI is expressed mathematically as :

$$NSFWQI = \sum_{i=1}^{p} W_i I_i, \qquad ...(1)$$

where, I_i = sub index for ith water quality parameter , W_i = weight (in terms of importance) associated with water quality parameter, and p = number of water quality parameters.

The modified weights (W_i) and the equation for the sub indices (I_i) as per CPCB, are given in **Tables 2 and 3** respectively. The range of the NSF WQI corresponding to various designated best

Table 2: Original and modified weights for the computation of NSF WQI (CPCB 2001)

Water Quality parameters	Original Weights from NSF WQI	Modified Weights by CPCB
DO	0.17	0.31
FC	0.15	0.28
pH	0.12	0.22
BOD	0.1	0.19
Total	0.54	1.00

Table 3: Sub –Index equations for water quality parameters (NSF WQI)*

Water Quality Parameters	Range Applicable	Equation
DO (Percent saturation)	0-40% saturation	IDO = 0.18+0.66 x (% Saturation DO)
	40-100% saturation	IDO = -13.55+1.17 x (% Saturation DO)
	100-140% saturation	IDO = 163.34-0.62 x (% Saturation DO)
B.O.D.(mg/L)	0-10	IBOD = 96.67-7 (BOD)
	10-30	IBOD = 38.9-1.23 (BOD)
	> 30	IBOD = 2
pH	2-5	IpH = 16.1+7.35 x (pH)
	5-7.3	IpH = -142.67+33.5 x (pH)
	7.3-10	IpH = 316.96-29.85 x (pH)
	10-12	IpH = 96.17-8.0 x(pH)
	<2 , >12	IpH=0
Fecal Coliform (counts/100mL)	$1-10^3$	IFC = 97.2-26.6 x log(FC)
	10^3-10^5	IFC = 42.33 -7.75 x log(FC)
	$>10^5$	IFC = 2

* Abbasi 2002

Table 4: NSF WQI for various designated best use*

Serial No	NSF WQI	Description of quality (1978)	Class by CPCB	Remarks
1	63-100	Good to Excellent	A	Non polluted
2	50-63	Medium to Good	B	Non polluted
3	38-50	Bad	C	Polluted
4	38 & less	Bad to Very Bad	D, E	Heavily polluted

*CPCB 2001, Abbasi 2002

use classification is given in **Table 4**.

2.4 Multivariate statistical methods

River water quality and source identification were performed using several chemometrics methods, such as Cluster Analysis, Discriminant Analysis and Principle Component Analysis.

2.4.1 Cluster analysis

CA is an exploratory data analysis tool for solving classification problems. Its objective is to sort cases (monitoring pints) into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. Each cluster thus describes, in terms of the data collected, the class to which its members belong; and this description may be abstracted through use from the particular to the general class or type[9].

In this case prior to CA, the significant parameters were standardized to avoid any effect of scale of units on the distance measurements by applying the following equation

$$Z_{ji} = \left(X_{ji} - \overline{X_j}\right)/s_j \qquad ...(2)$$

where x_{ji} indicates the original value of measured parameter, z_{ji}-the standardized value, $\overline{X_j}$ -the average value of variable j, s_j-the standard deviation of j.

The similarities–dissimilarities were quantified through Euclidean distance measurements; the distance between two objects (monitoring point locations), i and j, is given as:

$$d_{ji^2} = \sum_{k=1}^{m}\left(Z_{ji} - Z_{jk}\right)^2 \qquad ...(3)$$

where d_{ij} denotes the Euclidean distance, z_{ik} and z_{jk} are the values of variable k for object i and j, respectively, and m is the number of variables. Squared Euclidian distances and the Ward's method were used to obtain dendrograms[10].

The procedure followed by cluster analysis at Stage 1 is to cluster the two cases that have the smallest squared Euclidean distance between them. Then the software will recompute the distance measures between all single cases and clusters (there is only one cluster of two cases after the first step and group the two with minimum distance. This process continues until all cases are clustered into a single group.

2.4.2 Discriminant analysis

The variance– covariance between the classes is maximized and the variance–covariance within the classes is minimized under simultaneous consideration of all analyzed features [9]. New objects can then be classified and the learning objects reclassified by means of the non-elementary discriminant functions. These discriminant functions are linear combinations on the basis of the optimum separation set of the original features. In this study, the DA has been used to classify and, thus, to confirm the groups found by means of the CA.

2.4.3 Principal Component Analysis

PCA starts by building the correlation matrix for the data. Diagonalization of this matrix provides its eigen values and eigen vectors. Since the variance explained by each eigen vector is proportional to its eigen value, only those eigen vectors with eigen values greater than 1 are selected as significant independent variables (components). Sum of eigen values is equal to the total number of variables. Correlation of principal components and original variables is called loadings. The eigen vectors or components are more easily interpretable if a VARIMAX rotation, which transfers the eigen vectors to make each of them representative of individuals.

3. Results and discussion

Fig. 2 gives the box plots of pH, DO, BOD, FC, TC and temperature. A box plot or a box-and-whisker diagram or plot is a convenient way of graphically depicting groups of numerical data through

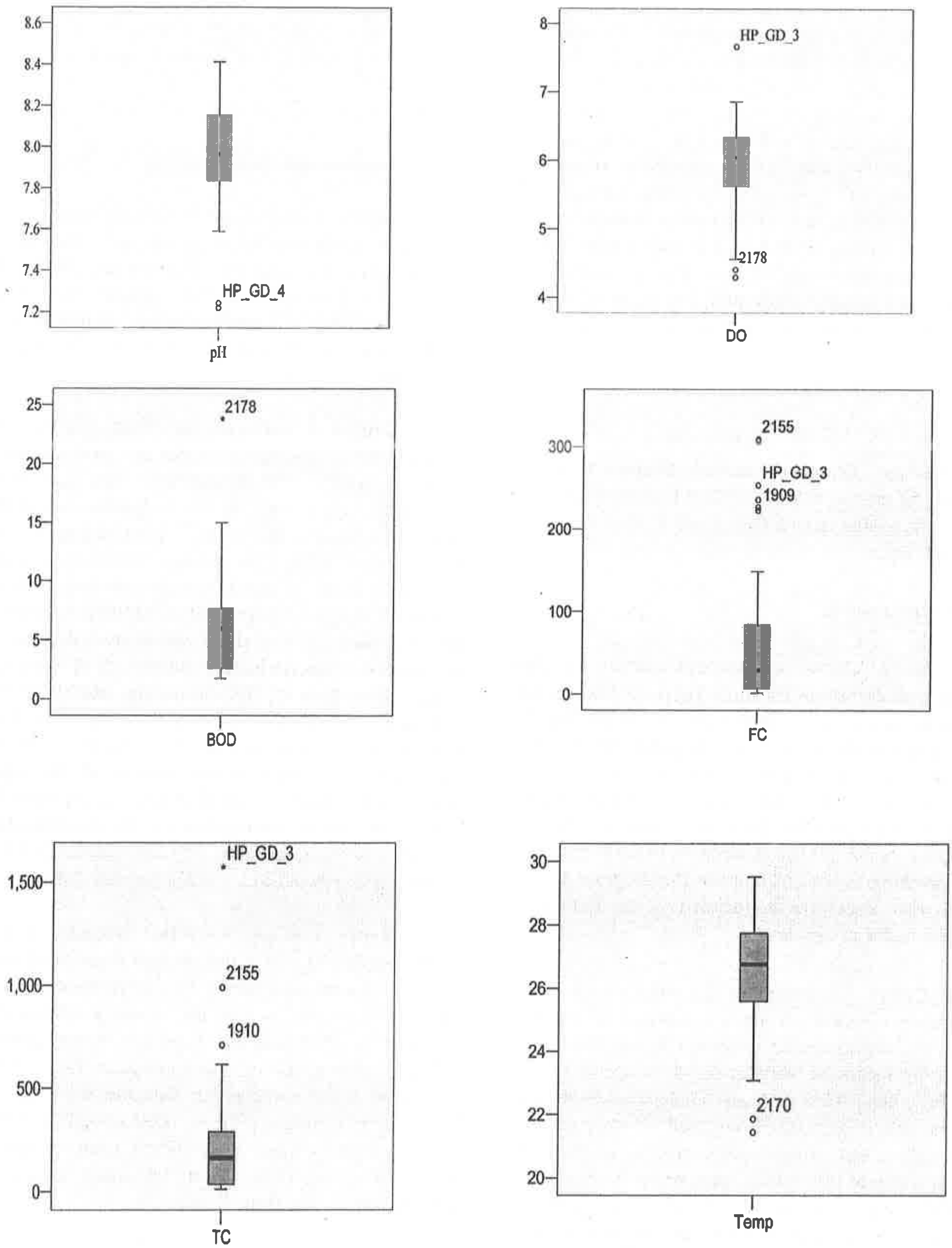

Fig. 2: Boxplots of pH, DO, BOD, TC and FC

their five-number summaries: the minimum, lower quartile, median, upper quartile, and sample maximum. A box plot may also indicate which observations, if any, might be considered outliers. The ends of the whiskers represent the minimum and maximum of all the data. The bottom and top of the box are the 25th and 75th percentiles. The centre represents the 50th percentile. Any data not included between the whiskers is plotted as an outlier with a small circle, or star. In box plot of DO in Fig. 2, the sampling stations HP_GD_3 and 2178 are outliers. If the data are normally distributed, the locations of the five marks on the box plot will be equally spaced.

3.1 Water Quality Index

Average Water Quality Indices of 78 sites varied between Good to Excellent, Medium to Good and Bad. 52 sites were found to have Good to Excellent, 23 sites had Medium to Good and 2 sites had Bad water quality.

3.2 Cluster analysis

Spatial cluster analysis was carried out with average concentrations for three years. Spatial cluster analysis produced a dendrogram with four groups (Fig. 3). Dendrogram gives an idea of how great the distance was between cases or groups that are clustered in a particular step, using a 0 to 25 scale along the top of the chart. The bigger the distances before two clusters are joined, the bigger the differences in these clusters. The branching-type nature of the Dendrogram allows you to trace backward or forward to any individual case or cluster at any level.

Group 1 consists of 24 sites which were moderately polluted. Group 2 comprised of 15 sites which were slightly polluted. Group 3 comprised of 12 monitoring locations while group 4 comprised of 27 sampling sites. Group 4 was cleaner as BOD and coliform values were low compared to other groups. The results of water quality index and those of cluster analysis more or less match since water quality index considers only 4 parameters, viz. pH, BOD, DO and FC while cluster analysis was performed using 5 parameters, viz. pH, BOD, DO, FC and TC. The water quality index of Purna river near Achalpur- Amravati bridge, Amravati (Station no 2700) and Godavari river at Kopargaon (Station no HP_GD_7) were Bad and belonged to Cluster 1. The reason being the very high BOD value (11.2 mg/L) at Purna river and very high TC value at Kopargaon (1337942.9 counts/100mL).

3.3 Discriminant function analysis

Table 5 indicates whether there is a statistically significant difference among the dependent variable means for each independent variable. In the table 'Tests of Equality of Group Means' the results of univariate ANOVA's, carried out for each independent variable, are presented. Here only FC and TC are not statistically significant.

Tables 6 indicates the first and second canonical linear discriminant functions. Because there are three groups' two discriminant functions were generated. Each function acts as projections of the data onto a dimension that best separates or discriminates between the groups. These eigen values are related to the canonical correlations and describe how much discriminating ability a function possesses. The magnitudes of the eigen values are indicative of the functions' discriminating abilities. % of variance is the proportion of discriminating ability of the continuous variables found in a given function. This proportion is calculated as the proportion of the function's eigen value to the sum of all the eigen values. In this analysis, the first function accounts for 73.6% of the discriminating ability of the discriminating variables and the second function accounts for 24.630% while the third function accounts for only 1.8%..

Table 7 gives the Wlks' lambda. Wilks' Lambda is the ratio of within-groups sums of squares to the total sums of squares. This is the proportion of the total variance in the discriminant scores not explained by differences among groups. A small lambda indicates that group means appear to differ. The associated significance value indicates whether the difference is significant. Here, the Lambda of 0.181, 0.559 and 0.951 have a significant value of 0.000, 0.000 and 0.159; thus, the group means appear to differ based on functions 1 and 2.

Table 8 provides an indication of the success rate for prediction of membership of the grouping

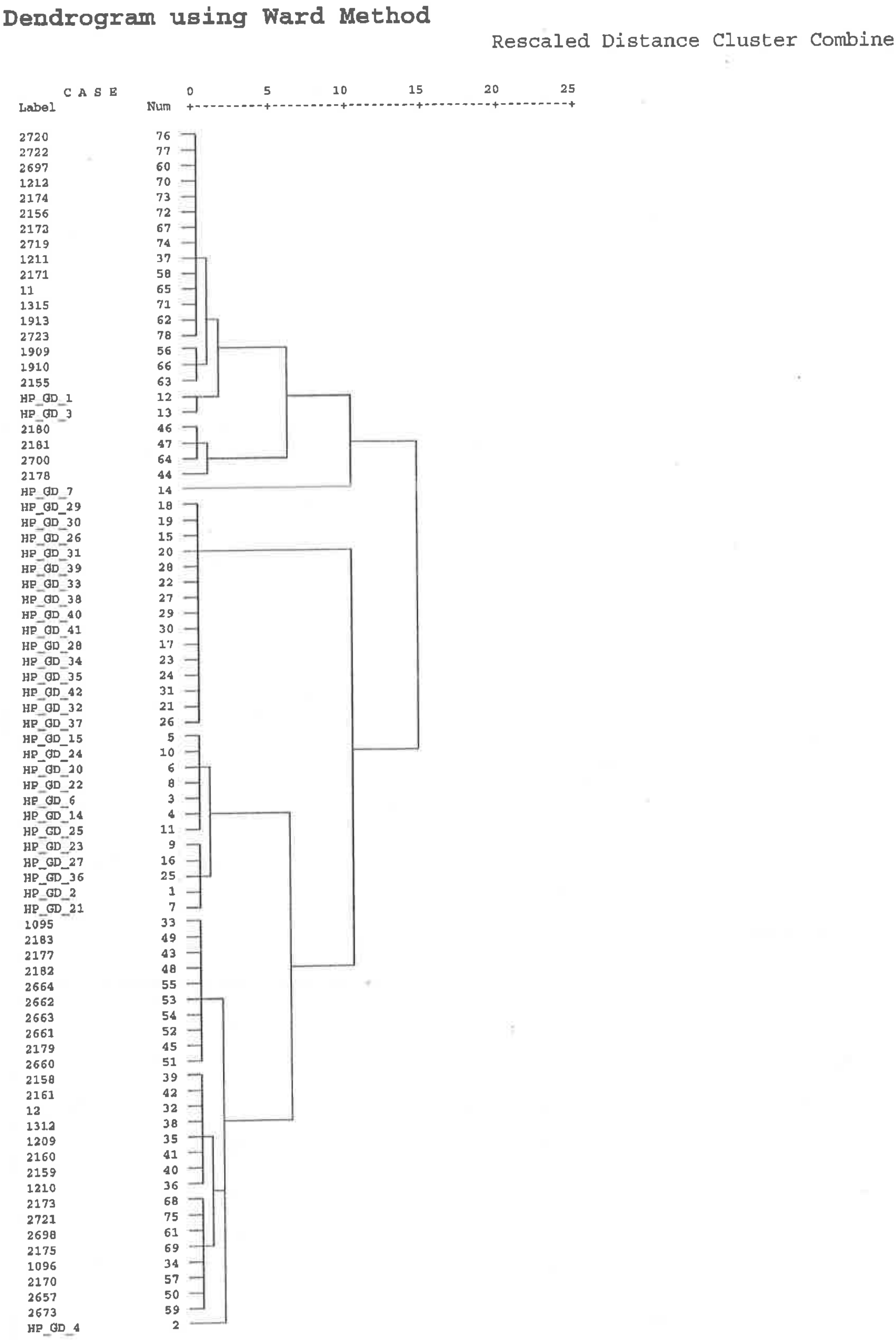

Fig. 3: Dendrogram of 78 monitoring locations

Table 5 : Tests of equality of group means

	Wilks' Lambda	F	df1	df2	Sig.
pH	.399	37.220	3	74	.000
DO	.691	11.052	3	74	.000
BOD	.390	38.615	3	74	.000
FC	.969	.783	3	74	.507
TC	.971	.748	3	74	.527

Table 6 : Eigen values

Function	Eigen value	% of Variance	Cumulative %	Canonical Correlation
1	2.095(a)	73.6	73.6	.823
2	.701(a)	24.6	98.2	.642
3	.052(a)	1.8	100.0	.222

First 3 canonical discriminant functions were used in the analysis.

Table 7 : Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 3	.181	124.931	12	.000
2 through 3	.559	42.461	6	.000
3	.951	3.678	2	.159

Table 8 : Classification results

Group		Predicted Group Membership				
		1	2	3	4	Total
Original Count	1	18	0	0	6	24
	2	0	15	0	0	15
	3	0	1	9	2	12
	4	3	0	0	24	27
%	1	75.0	.0	.0	25.0	100.0
	2	.0	100.0	.0	.0	100.0
	3	.0	8.3	75.0	16.7	100.0
	4	11.1	.0	.0	88.9	100.0

84.6% of original grouped cases correctly classified.

Table 9 : Rotated component matrix(a)

	Component		
Log transformed parameters	1	2	3
lgpH	.125	.795	.087
lgdo	.016	.807	.053
lgbod	.380	-.794	-.261
lgfc	.955	-.032	-.005
lgtc	.945	.008	.083
lgtemp	.065	.179	.978

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 4 iterations.

variable's categories using the discriminant function developed from the analysis. 83.3% of the original cases were correctly classified. The Predicted Group Membership are the predicted frequencies of groups from the analysis. The numbers going down each column indicate how many were correctly and incorrectly classified. For example, of the 24 cases in group I, 18 were correctly predicted, and 6 were incorrectly predicted.

3.4 Principal Component Analysis

PCA were applied to standardized log-transformed data. **Table 9** gives the rotated component matrix. Only the PCs with eigen values greater than 1 were considered essential. The first factor which explained 32.8% of the total variance had strong positive loadings on FC, TC and BOD. Thus, the first factor represented organic pollution from domestic wastewater and sewage treatment works. The second factor (32.4% of the total variance) had strong loading on pH, which may be interpreted as industrial pollution. The third factor which explains 17.3% of the variance had strong positive loading on temperature and it represents seasonal effects of temperature.

4. Conclusions

Average Water Quality Indices evaluated at 78 sites of Godavari basin for the period 2007-09 varied between Good to Excellent, Medium to Good and Bad out of which 52 sites had Good to Excellent, 23 sites had Medium to Good and 2 sites had bad water quality. The results of the spatial clustering showed a dendrogram with three groups. Group 1 consists of 24 sites which were most polluted. Group 2 comprised of 15 sites which were slightly polluted. Group 3 comprised of 39 monitoring locations which were almost clean. DA provided an important data reduction and used only two parameters (pH and BOD) affording 83.3% correct assignment in spatial analysis. Therefore, DA allowed a reduction in the dimensionality of large data set, delineating a few indicator parameters responsible for variations in water quality. PCA resulted in three factors explaining 82.5% of the total variance. The first factor obtained represents organic pollution. The second factor may be interpreted as industrial pollution. The third factor represented seasonal effect of temperature.

References

1. Vega, M., Pardo, R., Barrado, E.and Debán, L., Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis, *Water Research*, **32(12)**, 3581-3592 (1998).

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

3. Singh, K.P., Malik, A. and Sinha, S., Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: A case study, *Analytica Chimica Acta*, **538(1-2)**, 355-374 (2005).

4. Kowalkowskia T, Zbytniewskia, R., Szpejnab, J. and Buszewski B., Application of chemometrics in river water classification, *Water Research*, **40(4)**, 744-752 (2006).

5. H. Boyacioglu, Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey, *Environmental Geology*, **54 (2)**, 275-282 (2008).

6. Yang LI, Linyu XU and Shun LI, Water Quality Analysis of the Songhua River Basin Using Multivariate Techniques, *J. Water Res. and Protection*, **2**, 110-121 (2009).

7. Abbasi, S.A., Water Quality Indices State-of-the- Art, Pondicherry University, Centre for Pollution Control & Energy Technology , Pondicherry (2002).

8. *Environmental Atlas of India*, Central Pollution Control Board , New Delhi, 2001.

9. Einax, J.W., Truckenbrodt, D. and Kampe, O., River pollution data interpreted by means of chemometric methods. *Microchemical J.* **58**, 315-324 (1998).

10. Einax, J.W., Zwanziger, H.W. and Geiss, S., *Chemometrics in Environmental Analysis.* Wiley, Weinheim,1997

View publication stats