

# Prediction of Water Quality Using Principal Component Analysis

S.S. Mahapatra · Mrutyunjaya Sahu · R.K. Patel ·  
Biranchi Narayan Panda

Received: 12 December 2011 / Revised: 1 April 2012 / Accepted: 5 April 2012 / Published online: 4 May 2012  
© Springer Science+Business Media B.V. 2012

**Abstract** The groundwater is contaminated heavily with acidity, alkalinity, toxicity, heavy minerals, and microbes throughout the world due to population growth, urbanization and industrialization. Hence, evaluation of water quality of groundwater is extremely important to prepare for remedial measures. This paper presents application of an empirical approach for classification of water samples based on 10 quality parameters of water. In this research work, water samples from 10 sources in three different years and seasons have been collected to assess the quality of water. Q-mode principal component analysis has been applied to classify the water samples into four different categories considering parameters such as pH, DO, turbidity, TDS, hardness, calcium ion ( $\text{Ca}^{++}$ ), chloride ion ( $\text{Cl}^-$ ), BOD, iron ( $\text{Fe}^{++}$ ), sulfate ( $\text{SO}_4^{--}$ ). This classification will be useful for the planners and field engineers for taking ameliorative measures in advance for preventing the contamination of groundwater. The non-parametric method proposed here

efficiently assesses water quality index for classification of water quality. The model can also be used for estimating water quality on-line but the accuracy of the model depends upon the judicious selection of parameters.

**Keywords** Water quality index · Principal component analysis · Correlation · Classification

## Introduction

Groundwater makes up about 20 % of the world's fresh water supply, which is about 0.61 % of the entire world's water including oceans and permanent ice. In India, almost 80 % of the rural population depends on untreated groundwater for domestic and agriculture purpose. In recent years, the rate of discharge of pollutants into the environment is continuously increasing due to rapid growth of population, urbanization and accelerated pace of industrialization. It causes contamination of both fresh water and groundwater. Groundwater pollution can occur where industrial waste is discharged into pits, ponds, lagoons, rivers enabling the waste to percolate to water table. The polluted water endangers not only the valuable human life but also causes considerable biological disorder in the organisms. In addition, they cause serious environmental pollution responsible for health hazards. Groundwater contamination and its management have become important because of far reaching impact on human health. Therefore, knowledge of the water quality and evaluation of water quality index (WQI) plays a significant role in water quality control and management. WQI is generally considered as a means of summarizing the various water quality parameters into a simple index. The index helps in interpreting the quality of water in a single numerical value (Horton 1965; Brown et al. 1970; Dinius 1972; Lohani and Todino 1984).

---

S.S. Mahapatra (✉)  
Department of Mechanical Engineering,  
National Institute of Technology, Rourkela, India  
e-mail: [mahapatrass2003@yahoo.com](mailto:mahapatrass2003@yahoo.com)

M. Sahu  
Department of Civil Engineering,  
National Institute of Technology, Rourkela, India  
e-mail: [mrityunjay.nitrkl@gmail.com](mailto:mrityunjay.nitrkl@gmail.com)

R.K. Patel  
Department of Chemistry, National Institute of Technology,  
Rourkela, India  
e-mail: [rkpatel@nitrkl.ac.in](mailto:rkpatel@nitrkl.ac.in)

B.N. Panda  
Department of Manufacturing Science,  
Veer Surendra Sai University, Burla, India  
e-mail: [biranchi.panda3@gmail.com](mailto:biranchi.panda3@gmail.com)

Classification of WQI is useful for quantifying the heterogeneity existing in the system. Researchers such as Bhargava (1983), Swamee and Tyagi (2000), Sarkar and Abbasi (2006) and Tiwari and Mishra (1985) have worked in this direction and proposed mathematical and statistical models for classification of groundwater. Recently, Lumb et al. (2011) have reviewed various water quality classification models based on WQI. Various WQIs differ in the manner in which statistical integration and interpretation of parameter values are made. A totally different approach was adopted in the Canadian Water Quality Index also known as Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) (Lumb et al. 2011, 2006). CCME WQI is also being used by many countries all over the world and has also been endorsed by United Nations Environmental Program (UNEP) in 2007 as a model for Global Drinking Water Quality Index (GDWQI). The most commonly used parameters in this model are dissolved oxygen, pH, turbidity, total dissolved solids, nitrates, phosphates, and metals, among others.

The present study proposes an empirical method of classification scheme for assessing water quality of groundwater. The correlation of water quality parameters sometimes results in unjustified classifications if water quality is expressed by an index. The proposed method is quite efficient and prediction quality is reasonably good. The method first calculates similarity coefficients for all the members in the data set using Euclidean distance as a similarity measure. The Euclidean distance matrix is used as input to the Q-mode Principal Component Analysis (PCA) to classify the data (Albadawi et al. 2005). The factor loadings of each member of data set on principal component are taken into account to cluster the water sample into appropriate group. The number of clusters is decided on the basis of percentage variation explained by the principal components.

## Literature Review

The quality of groundwater has been studied earlier by various researchers. Among them, Karnchanawong and Ikeguchi (1993) have evaluated quality of well water near the Mae-Hia waste disposal site. Zhang et al. (1996) have predicted water quality index in 14 cities of China where the groundwater is polluted due to use of fertilizers for agricultural purposes. Lind et al. (1998) have studied the impact of mining activity on pH of groundwater and its effect on water quality. Maticie (1999) have observed the impact of agriculture on groundwater quality in Slovenia. Shamruck et al. (2001) have studied the impact of environmental parameters on quality of Nile Valley aquifer. Ammann et al. (2003) have evaluated groundwater pollution and its impact on water quality by run-off. Almasri and Kaluarachchi (2004) have

reported on occurrence of nitrate in the groundwater in agricultural watersheds in Whatcom County, Washington. WQI is strongly dependent on various correlated parameters taken for the study.

However, identification of the suitability of the parameters is critical for accurate evaluation of WQI. Water quality is generally ascertained based on physical, chemical, and biological indicators using indicators such as pH, electrical conductivity (EC), total dissolved solids (TDS), total suspended solids (TSS), hardness, turbidity and contaminant concentrations based on guidelines provided by agencies such as the World Health Organization WHO (2006) and the Bureau of Indian Standards (BIS) (1991). Nagarajan and Priya (1999) have studied the groundwater quality deterioration in Tiruchirapalli, Tamil Nadu and found that TSS, iron and magnesium values are beyond the permissible limit. Singh and Parwana (1999) have studied the pollution load in the groundwater in Punjab state due to industrial waste water and found the presence of chromium and cyanide in groundwater beyond permissible limit of drinking water standards. Jha and Verma (2000) have studied the physicochemical properties of drinking water in town area of Godda district under Santal Pargana (Bihar). They have reported that most of the surface water quality parameters are within the limit of drinking water standards. However, well water is characterized by a very high concentration of chloride, chromium and selenium. Srinivas et al. (2000) have studied the groundwater quality of Hyderabad taking 32 tube well water samples and reported that electrical conductivity, TDS, total alkali, hardness, calcium, magnesium, sodium and chlorides are above the permissible limit according to WHO and Indian Standards. Chaudhari et al. (2004) have studied the quality of groundwater near an industrial area at Jalgaon (Maharashtra) and WQI of samples suggests that the water is not suitable for direct consumption. Shaji et al. (2009) have studied quality of water in mineral and industrial area of Kerala in India. Waste materials near the factories are subjected to reaction with percolating rain water and reach the aquifer system and, hence, degrade the groundwater quality (Tyagi et al. 2003). Heavy metals constitute a very heterogeneous group of elements widely varied in their chemical properties and biological functions. They are persistent in nature; therefore, they get accumulated in soil and plants. Dietary intake of many heavy metals through consumption of plants and drinking water has long-term detrimental effect on human health (Sharma and Agarwal 2005; Ubala et al. 2001; Sabal and Khan 2008).

However, quality is a vague term that cannot easily be described using crisp data set e.g. good quality water cannot simply be described as having a pH value of 7.0 or above. Instead, water quality can best be described based on its degree of potability and potential usages rather than expressing its constituents in numerical terms. Fuzzy reasoning technique

has also been applied in groundwater and surface water quality forecasting (Dahiya et al. 2007; Parinet et al. 2004; Singh et al. 2008). Some of the artificial neural network as well as multivariate analysis methods are also found to be more useful for determination of water quality based on fuzzy and principal component analysis (PCA) techniques (Isceen et al. 2008; Shrestha and Kazama 2007).

A number of indices have been developed to summarize water quality data in an easily expressible and easily understood format (Couillard and Lefebvre 1985). Horton (1965) proposed the first water quality index (WQI), where a great deal of consideration has been given to the development of index methods. The basic differences among these indices are the way their sub-indices were developed. Wal-ski and Parker (1974) used an exponential function to represent the sub-indices of various quality variables. Landwehr (1979) suggested the Pearson type 3-distribution function to represent the sub-indices of all the quality variables. Bhargava (1987) modified the exponential formula; Dinius (1987) used power function for the majority of sub-indices. Nives (1999), Swamee and Tyagi (2000) proposed aggregate index for water quality description. In addition, Harrison et al. (2000), Faisal et al. (2003), Ahmed et al. (2004) and Shio-Mey et al. (2004), each have recently modified a water quality index. Some of the sub-indices have since been incorporated into water quality indices used by agencies such as the National Sanitation Foundation (NSF) (Ahmed et al. 2004). The most important WQIs belonging to environmental departments or agencies are the National Sanitation Foundation (NSF), British Columbia Water Act Quality Index, Oregon Water quality Index, Stream Watch (Southern Indiana), Malaysian Water Quality Index, France Water Quality Index, French Creek Quality Index, Florida Stream Water Quality Index, British Columbia Water Quality Index, Canadian Water Quality Index, Taiwan Water Quality Index and Washington State Water Quality Index. Comparison among several WQI systems currently in use showed that none of them describes quality of water from mining effluent because most of sub-indices in current WQIs are not relevant to indicating changes in water quality brought about by mining activities.

### Study Area

Odisha is a state in India located at an elevation of about 219 meters above mean sea level. In this study, groundwater quality of wells of urban area of Rourkela in Sundergarh district which is located at 84.54E longitude and 22.12N latitude is considered. Rourkela comes under tropical monsoon climate and is more like that of the Deccan Plateau. Being in the north-eastern corner of the Deccan Plateau, the climate is milder than the climate of the main Deccan region. The

area of Rourkela is 200 square kilometers approximately. Red and laterite soils are found here which are quite rich in minerals. The area near Rourkela is rich in iron ore; hence a steel plant and other iron and steel industries are situated in the region. These industries are polluting the surrounding areas including groundwater resources. Large number of motor vehicles may also contribute in the release of heavy metals into surrounding environment. The climate is hot and dry during summer season. Normally, there is heavy rainfall due to south-west monsoon and light rainfall during the pre-monsoon seasons. The south-west monsoon usually onsets during second week of June and retreats by mid September. The humidity is generally high mostly in the monsoon and post-monsoon periods. The relative humidity is low during summer season. The mean values of the humidity, however, in a year range from 35 to 85 % and the annual average is 66 %. The Koel and Sankha Rivers meet at Vedvyas, Rourkela and flow as a single river called Brahmani. Hence, Rourkela is the confluence of three rivers: Koel, Sankha and Brahmani. The geographical location of study area is shown in Fig. 1.

### Methodology for Sampling

In order to classify water quality into different clusters, a number of water samples were collected from 10 wells and their locations are shown in Table 1. Water samples of three years have been taken into consideration for the study. Also season-wise data for each corresponding year was taken. Water samples of all the places and seasons were not included due to some seasonal impact like heavy rainfall, encroaching heat, etc. Water samples from different sampling stations are collected in standardized PET (polyethylene terephthalate) bottles, which are thermostated bottles. The PET bottles of 1.5 liter capacities with stopper were used for collecting samples. The PET bottles can be used for collection of samples to analyze both organic and inorganic constituents in water. The bottles were washed thoroughly with 2 % nitric acid and subsequently rinsed with distilled water. The bottles were then preserved in a clean place. Before taking the water samples, all the supply bottles are rinsed with sample water 2–3 times. As all the physicochemical parameters are measured within 24 hours of sample collection, there is very little possibility of changing concentration of any parameters including heavy metals. The sampled bottle is made watertight by air tightening it inside water. Precautions have been taken to remove any air bubble present. Each container was clearly marked with the name and date of sampling. Various physicochemical parameters such as pH, turbidity, total dissolved solids (TDS), hardness, biochemical oxygen demand (BOD), dissolved oxygen (DO), chloride, sulfate, iron, calcium have been taken for analysis.

**Fig. 1** The study area**Table 1** The location of wells used for collection of water samples

Well number	Location
W <sub>1</sub>	Jagda
W <sub>2</sub>	Koelnagar
W <sub>3</sub>	Shaktinagar
W <sub>4</sub>	NIT Campus
W <sub>5</sub>	Jiripani
W <sub>6</sub>	Chhend
W <sub>7</sub>	Civil Township
W <sub>8</sub>	Hamirpur
W <sub>9</sub>	Sector 6
W <sub>10</sub>	Sector 7

Physicochemical parameters such as pH, turbidity, dissolved oxygen (DO), and total dissolved solids (TDS) were measured using water analysis kit model 191 E. The methodologies adopted for determination of water quality parameters of the collected samples are shown in Table 2.

Water samples were collected from groundwater sources on monthly basis. However, study of physicochemical characteristics was made through seasonal observations. The seasons are broadly divided into three seasons such as summer (March to June), rainy (July to October) and winter (November to February). The summer is too hot (max temperature 48 °C) in this part of the country; hence, data have not been collected in summer season. The procedure for estimating 10 quality parameters such as pH, DO (dissolved oxygen), turbidity, TDS (total dissolved solids), hardness, calcium ions, chloride ions, BOD (biological oxygen demand), iron and sulfate is shown in Table 2. The samples were collected in three years: 2008, 2009, and 2010. The physicochemical characteristics of water samples are given in Tables 3a and 3b, respectively, for rainy and winter seasons. The notation used to denote samples is as follows: *W* stands for the well and the number following *W* denotes

the well number. The letters *r* and *w* respectively denote the rainy and winter season. The numbers 01, 02, and 03 respectively denote the year of sampling, 2008, 2009, or 2010.

The average values of water quality parameters in two different seasons are verified with the permissible limits prescribed in IS:10500 shown in Table 4. From Tables 3a, and 3b it can be observed that most of the parameters are within permissible limit. However, parameters such as TDS, hardness, calcium, chloride, and iron lie below the permissible limit whereas, while DO lies above the limit. It may be noted that the value of turbidity decreases to almost half during winter season as compared to rainy season. Pearson correlation coefficients for parameters in rainy and winter seasons are shown in Tables 5a and 5b, respectively. The correlation coefficient of 0.5–0.75 is considered as moderate correlation between two variables (Montgomery and Runger 1999). In rainy season, highest correlation is observed for parameters turbidity and pH (0.731), followed by calcium and hardness (0.699). In winter season, the strongest correlation is observed between calcium and pH (0.529). The outliers in the data are removed through examination of scatter plots. The changes in average parameter values and the cor-

**Table 2** Test methods for various parameters of water samples

Parameters	Methods adopted
pH	Systronic-361 pH meter
DO (mg/l)	Winkler’s method
Turbidity (NTU)	Nepheloturbidity method
TDS (mg/l)	Water analysis kit model 191 E
Hardness (mg/l)	EDTA Titration by using Erichrome Black-T Indicator
Calcium ion (Ca <sup>++</sup> ) (mg/l)	EDTA Titration by using murexide indicator
Chloride (Cl <sup>-</sup> ) (mg/l)	Titrating against N/50 solution of silver nitrate using potassium chromate as indicator
BOD (mg/l)	Standard method
Iron (Fe <sup>++</sup> ) (mg/l)	Spectrophotometrically using the standard procedure recommended by APHA, AWWA, WPCF (1985)
Sulfate (SO <sub>4</sub> <sup>-</sup> ) (mg/l)	Spectrophotometrically using the standard procedure recommended by APHA, AWWA, WPCF (1985)

**Table 3a** Values of physicochemical parameters in rainy season

Sample number	pH	DO	Turbidity	TDS	Hardness	Calcium	Chloride	BOD	Iron	Sulfate
W <sub>1</sub> .r.01	6.2	8	3.5	150	118	2.8	18	2	0.034	115
W <sub>1</sub> .r.02	6.4	7	4	138	107	3.2	20	3.5	0.028	117
W <sub>1</sub> .r.03	6.1	7.4	3	155	115	4	24	2	0.039	111
W <sub>2</sub> .r.01	6.5	8	3.2	152	57	17	23.5	1.5	0.078	167
W <sub>2</sub> .r.02	6.6	6	3.5	162	59	16	21.5	1.5	0.02	172
W <sub>2</sub> .r.03	6.8	5	4.4	164	62	19	23	1.5	0.01	162
W <sub>3</sub> .r.01	7.2	7.5	3.2	171	138	14.9	50	1.8	0.02	163
W <sub>3</sub> .r.02	6.9	8	3.7	169	143	15	40	2	0.06	153
W <sub>3</sub> .r.03	6.7	6.5	3.8	173	129	18	31	2.2	0.05	172
W <sub>4</sub> .r.01	7	7.2	7	172	78	20	65	4	0.23	170
W <sub>5</sub> .r.02	6.9	7.2	4.2	156	83	15	62	2.7	0.04	178
W <sub>6</sub> .r.01	8.1	7	11.2	229	111	29	32	0.8	0.17	145.6
W <sub>6</sub> .r.03	8	8.5	10.8	229.5	114	38	37	0.2	0.18	150.5
W <sub>7</sub> .r.03	7.8	8.2	8	220	300	85	11	0.2	0.121	85
W <sub>9</sub> .r.01	9	9	17.2	120	140	16	31	3	0.011	136
W <sub>9</sub> .r.03	7.5	7	7	120	170	27	89	6	0.009	189
W <sub>10</sub> .r.01	7.9	7.5	7.9	163	112	25	57	0.8	0.089	42
W <sub>10</sub> .r.02	7.3	7	9.6	156	115	26	64	0.9	0.076	22
W <sub>10</sub> .r.03	7.5	8	8.2	172	114	34	72	1.02	0.081	19
Average	7.18	7.37	6.49	166.92	119.21	22.36	40.58	1.98	0.07	129.95

relation coefficients in different seasons are observed due to human and industrial activities.

**Determination of Water Quality Index (WQI)**

In the formulation of water quality index, the importance of various parameters depends on the intended use of water and water quality parameters are studied from the point of view of suitability for human consumption. The ‘standards’ (permissible values of various pollutants) for the drinking

water are recommended by the Indian Council of Medical Research (ICMR). When the ICMR standards are not available, the standards of United States Public Health Services (USPHS), World Health Organization (WHO), Indian Standards Institution (ISI) and European Economic Community (EEC) are being quoted.

The quality rating  $q_i$  for the  $i$ th water quality parameter is obtained from the relation

$$q_i = 100(v_i/s_i) \tag{1}$$

**Table 3b** Values of physicochemical parameters in winter season

Sample number	pH	DO	Turbidity	TDS	Hardness	Calcium	Chloride	BOD	Iron	Sulfate
W <sub>1</sub> .w.01	6.8	7.00	2.50	158.00	98.00	3.40	22.00	2.00	0.03	102.00
W <sub>1</sub> .w.02	6.6	7.50	2.70	156.00	110.00	2.30	17.00	2.00	0.04	109.00
W <sub>1</sub> .w.03	7.1	7.80	1.50	130.00	110.00	3.00	17.00	2.00	0.04	110.00
W <sub>2</sub> .w.02	6.5	8.00	3.00	150.00	53.00	12.00	21.00	2.00	0.03	168.00
W <sub>2</sub> .w.03	6.6	6.00	3.90	145.00	58.00	16.00	22.00	1.50	0.01	159.00
W <sub>3</sub> .w.02	6.7	7.50	2.50	151.00	136.00	18.00	42.00	2.00	0.04	150.20
W <sub>3</sub> .w.03	6.2	6.00	2.60	157.00	138.00	17.00	20.00	2.30	0.04	155.00
W <sub>4</sub> .w.02	7.1	8.20	6.00	162.00	75.00	16.40	65.00	4.80	0.24	172.00
W <sub>4</sub> .w.03	7.1	7.60	6.20	165.00	68.30	20.00	62.00	3.90	0.25	184.00
W <sub>7</sub> .w.01	7.9	8.00	3.20	220.00	189.00	80.00	20.00	0.20	0.15	70.00
W <sub>8</sub> .w.01	7.5	7.00	2.30	52.00	80.00	21.00	50.00	1.20	0.00	120.00
W <sub>8</sub> .w.02	7.9	9.00	4.00	71.00	73.00	14.00	51.00	2.00	0.01	152.00
W <sub>8</sub> .w.03	7	9.00	3.00	118.00	92.00	25.00	59.00	2.00	0.01	150.00
W <sub>9</sub> .w.01	8.8	9.00	7.20	80.00	115.00	12.00	22.00	3.00	0.01	144.00
W <sub>9</sub> .w.02	8.9	8.00	4.00	110.00	100.00	13.00	25.00	3.00	0.00	100.00
W <sub>10</sub> .w.01	7.4	7.00	4.40	142.00	108.00	28.00	53.00	0.86	0.02	2.47
W <sub>10</sub> .w.03	7.3	7.00	4.20	150.00	109.00	32.00	48.00	0.80	0.02	4.90
Average	7.26	7.62	3.72	136.29	100.72	19.59	36.24	2.09	0.06	120.74

**Table 4** Permissible limits of parameters (IS:10500)

Parameters	Permissible limit IS:10500
pH	6.5–8.5
DO (mg/l)	5.0
Turbidity (NTU)	5.0–10
TDS (mg/l)	500–2000
Hardness (mg/l)	300–600
Calcium ion (mg/l)	75–200
Chloride (mg/l)	250–1000
BOD (mg/l)	0–5
Iron (mg/l)	0.3–1
Sulfate (mg/l)	400

where  $v_i$  = value of the  $i$ th parameter at a given sampling station and  $s_i$  = standard permissible value of the  $i$ th parameter. This equation ensures that  $q_i = 0$  when a pollutant (the  $i$ th parameter) is absent in the water while  $q_i = 100$  if the value of this parameter is just equal to its permissible value for drinking water. Thus, the larger the value of  $q_i$  the more polluted is the water with the  $i$ th pollutant. However, quality ratings for pH and DO require special handling. The permissible range of pH for the drinking water is 7.0 to 8.5. Therefore, the quality rating for pH may be

$$q_{pH} = 100[(v_{pH} - 7)/(8.5 - 7.0)] \tag{2}$$

where  $v_{pH}$  = value of pH  $\sim 7$ , it means the numerical difference between  $v_{pH}$  and 7.0 ignoring algebraic sign. Equation

(2) ensures the  $q_{pH} = 0$  for pH = 7.0. In contrast to other pollutants, the case of DO is slightly complicated because the quality of water is enhanced if it contains more DO. Therefore, the quality rating  $q_{DO}$  has been calculated from the relation

$$q_{DO} = 100[(14.6 - v_{DO})/(14.6 - 5)] \tag{3}$$

where  $v_{DO}$  = value of DO.

In Eq. (3), 14.6 is the solubility of oxygen (mg/l) in distilled water at 0 °C and 5.0 mg/l is the standard for drinking water. Equation (3) gives  $q_{DO} = 0$  when DO = 14.6 mg/l and  $q_{DO} = 100$  when  $v_{DO} = 5.0$  mg/l. The more harmful a given pollutant is, the smaller is its permissible value for drinking water. So the ‘weights’ for various water quality

**Table 5a** Correlation matrix of parameters (rainy season)

	pH	DO	Turbidity	TDS	Hardness	Calcium	Chloride	BOD	Iron	Sulfate
pH	1									
DO	0.447	1								
Turbidity	0.731	0.303	1							
TDS	-0.017	-0.256	0.333	1						
Hardness	0.352	0.270	0.230	0.242	1					
Calcium	0.464	0.126	0.340	0.432	0.699	1				
Chloride	0.132	0.201	0.052	-0.300	-0.215	-0.087	1			
BOD	-0.049	-0.138	0.404	0.337	-0.009	-0.088	0.196	1		
Iron	0.079	0.075	-0.057	0.313	-0.084	0.229	0.067	-0.036	1	
Sulfate	-0.230	-0.111	-0.338	-0.259	-0.331	-0.372	0.074	0.106	0.073	1

**Table 5b** Correlation matrix of parameters (winter season)

	pH	DO	Turbidity	TDS	Hardness	Calcium	Chloride	BOD	Iron	Sulfate
pH	1									
DO	0.421	1								
Turbidity	-0.478	-0.022	1							
TDS	0.256	-0.191	0.306	1						
Hardness	0.079	0.362	-0.153	-0.344	1					
Calcium	0.529	0.326	-0.294	-0.022	0.262	1				
Chloride	-0.057	0.103	-0.115	-0.285	0.244	0.122	1			
BOD	-0.289	-0.070	-0.127	-0.409	0.633	0.15	0.234	1		
Iron	-0.086	0.443	0.501	0.015	0.485	0.109	0.144	0.202	1	
Sulfate	0.281	-0.017	0.310	0.606	-0.371	0.026	0.204	-0.285	0.206	1

parameters are assumed to be inversely proportional to the recommended standards for the corresponding parameters, i.e.

$$W_i = \frac{K}{S_i} \tag{4}$$

where  $W_i$  = unit weight for the  $i$ th parameter ( $i = 1, 2, 3, \dots, 10$ ),  $k$  = constant of proportionality which is determined from the condition and  $k = 1$  for the sake of simplicity:

$$\sum_{i=1}^{10} W_i = 1 \tag{5}$$

To calculate the WQI, first the sub-index  $(SI)_i$  corresponding the  $i$ th parameter is calculated. These are given by the product of the quality rating  $q_i$  and the unit weight  $W_i$  of the  $i$ th parameter, i.e.

$$(SI)_i = q_i W_i \tag{6}$$

The overall WQI is then calculated by aggregating these sub-indices (SI) linearly. Thus, WQI can be written as

$$WQI = \left[ \frac{\sum_{i=1}^{12} q_i W_i}{\sum_{i=1}^{12} W_i} \right] \tag{7}$$

which gives

$$WQI = \sum_{i=1}^{12} q_i W_i \tag{8}$$

Water quality can be categorized into five classifications depending on WQI values. Water quality can be treated as excellent, good, poor, very poor, and unsuitable for drinking if WQI lies in the ranges of 0–25, 26–50, 51–75, 76–100, and above 100, respectively.

The water quality index values for all the data shown in Tables 3a and 3b are obtained using Eqs. (1) through (8) and are shown in Table 6. It should be noted that WQI for all the data lies in the range from excellent to very poor for human consumption.

**Generation of Euclidean Distance Matrix**

The parameters of water samples shown in Tables 3a and 3b possess different measuring scales. Therefore, they need to be normalized to reduce the scaling effect. A simple normalization procedure of dividing selected variables by their maximum value is adopted here. After normalization, all

**Table 6** The water quality index (WQI) values for all samples

Sample number	WQI in rainy season	Comment	Sample number	WQI in winter season	Comment
W <sub>1</sub> .r.01	13.978	Cluster 1, Excellent	W <sub>1</sub> .w.01	15.667	Cluster 1, Excellent
W <sub>1</sub> .r.02	14.932	Cluster 1, Excellent	W <sub>1</sub> .w.02	17.757	Cluster 1, Excellent
W <sub>1</sub> .r.03	15.320	Cluster 1, Excellent	W <sub>1</sub> .w.03	17.104	Cluster 1, Excellent
W <sub>2</sub> .r.01	26.409	Cluster 2, Good	W <sub>2</sub> .w.02	13.514	Cluster 1, Excellent
W <sub>2</sub> .r.02	11.666	Cluster 1, Excellent	W <sub>2</sub> .w.03	8.981	Cluster 1, Excellent
W <sub>2</sub> .r.03	10.120	Cluster 1, Excellent	W <sub>3</sub> .w.02	16.947	Cluster 1, Excellent
W <sub>3</sub> .r.01	12.561	Cluster 1, Excellent	W <sub>3</sub> .w.03	16.854	Cluster 1, Excellent
W <sub>3</sub> .r.02	23.004	Cluster 1, Excellent	W <sub>4</sub> .w.02	76.708	Cluster 4, Very poor
W <sub>3</sub> .r.03	20.776	Cluster 1, Excellent	W <sub>4</sub> .w.03	78.966	Cluster 4, Very poor
W <sub>4</sub> .r.01	73.682	Cluster 3, Poor	W <sub>7</sub> .w.01	79.237	Cluster 4, Very poor
W <sub>5</sub> .r.02	28.701	Cluster 2, Good	W <sub>8</sub> .w.01	28.243	Cluster 2, Good
W <sub>6</sub> .r.01	57.630	Cluster 3, Poor	W <sub>8</sub> .w.02	29.644	Cluster 2, Good
W <sub>6</sub> .r.03	58.734	Cluster 3, Poor	W <sub>8</sub> .w.03	38.424	Cluster 2, Good
W <sub>7</sub> .r.03	41.550	Cluster 2, Good	W <sub>9</sub> .w.01	15.513	Cluster 1, Excellent
W <sub>9</sub> .r.01	28.236	Cluster 2, Good	W <sub>9</sub> .w.02	12.674	Cluster 1, Excellent
W <sub>9</sub> .r.03	24.717	Cluster 1, Excellent	W <sub>10</sub> .w.01	52.265	Cluster 3, Poor
W <sub>10</sub> .r.01	73.513	Cluster 4, Very Poor	W <sub>10</sub> .w.03	58.196	Cluster 2, Good
W <sub>10</sub> .r.02	79.258	Cluster 4, Very poor			
W <sub>10</sub> .r.03	80.409	Cluster 4, Very poor			

data vary from zero to one. Implementation of factor analysis requires the correlation matrix of the initial data set. The correlation matrix is obtained in the form of a Euclidean distance matrix (Hair et al. 2009). Euclidean distance is taken as the similarity measure and is defined as the sum of the squares of the difference between the values of attributes of two water samples. Mathematically, it may be given as

$$d(x, y) = \sqrt{\sum_i (x_i - u_i)^2} \quad (9)$$

where  $d(x, y)$  = Euclidean distance,  $x = x_1, x_2, \dots, x_m$  and  $y = y_1, y_2, \dots, y_m$  represent  $m$  attribute values of two samples. If the distance is zero, both the coal samples are similar. If it is above zero, the Euclidean distance indicates the intensity of dissimilarity between two water samples. The Euclidean distance matrix is generated considering all the water samples. An entry in the matrix denotes Euclidean distance between the  $p$ th row and the  $(p + 1)$ th row of the water samples. The Euclidean distance matrix is thus generated for the use in Q-mode PCA.

### Q-mode Principal Component Analysis

PCA is the most widely used, straightforward and quantitatively involved method for transforming a given set of inter-related variables into a new set of variables called the prin-

cipal components (corresponding to factors in factor analysis). The set of principal components generated presents uncorrelated linear combinations of the original variables and accounts for the total variance of the original data. In this method, all the principal components are generated in such a way that they are orthogonal to each other; hence, correlation between them is zero. The principal components are generated in a sequentially ordered manner with decreasing contributions to the variance, i.e. the first principal component explains most of the variations present in the original data, and successive principal components account for decreasing proportions of the variance. This property means that the data points can be rigorously separated into distinct clusters when projected into a space spanned by the first few principal components, which are called factors. This achieves the dimensionality reduction objective of factor analysis. PCA can be broadly classified into two categories, viz., R-mode and Q-mode, based on application. If PCA is used to develop a structure among variables, it is referred to as an R-mode PCA. When PCA analysis is used to group cases, it is called a Q-mode PCA. It is customary to use rotation methods to transform the factors to simpler and more interpretable constructs. After rotation, each variable will be only related to one of the factors and each factor will have high correlation with only a small set of variables. In recent years, Q-mode PCA has been widely adopted by the researchers for classification of groundwater quality, coffee



preference, gene regulatory process, and machines in cellular manufacturing (Albadawi et al. 2005; Dijksterhuis 1998; Park et al. 2001; Singh et al. 2008; Woocay and Walton 2008).

Assuming the sample parameters as the original set of variables, and the Euclidean distance matrix as an estimate of the correlation matrix explaining the correlations between each pair of samples, we proceed to use the PCA framework for grouping the samples into separate independent clusters. In the PCA method, the initial clusters are extracted out by the eigenvalue-eigenvector analysis of the similarity coefficient matrix as presented in Eq. (10):

$$(S - I\lambda_i)Y_i = 0, \quad i = 1, 2, \dots, P \tag{10}$$

where  $S$  is a  $P \times P$  Euclidean distance matrix,  $I$  is the identity matrix,  $\lambda_i$  are the characteristic roots (eigenvalues), and  $Y_i$  are the corresponding eigenvectors.

Equation (10) is an eigenvalue-eigenvector equation,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$  are the real, nonnegative roots of the determinant polynomial of degree  $P$  given as

$$I(S - I\lambda_i)I = 0 \tag{11}$$

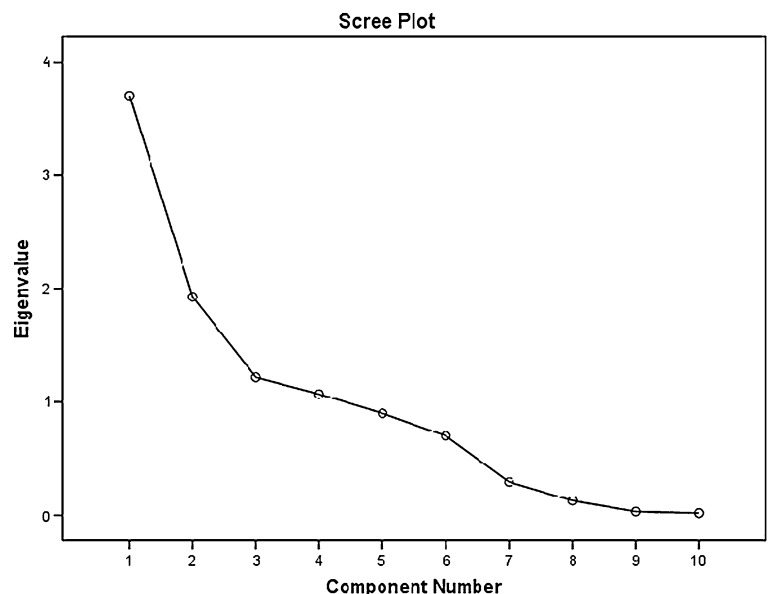
This equation is solved for  $\lambda_i$  and then  $Y_i$  can be calculated, using the values of  $\lambda_i$  in Eq. (10). It is proven that the eigenvectors thus computed represent the unique set of  $P$  independent principal components (factors) of the data set, which maximize the variance (Basilevsky 1994). According to the PCA method, each of the  $P$  independent principal components (factors) can be written as a linear combination of the original variables (water samples), with the elements of the  $P$  eigenvectors as the coefficients of these linear combinations. Furthermore, the elements of these eigenvectors reflect the degree of association between each principal

component (factor) and the sample, and are called the ‘factor loadings’ of the samples on the  $i$ th principal component in factor analysis terminology. Each of the  $P$  independent principal components represents a cluster. There should be low similarities among samples that are associated with different clusters and high similarities among samples strongly associated with the same cluster. In regard to the number of sample size, Basilevsky’s assumed data set should be three to four times the number of variables.

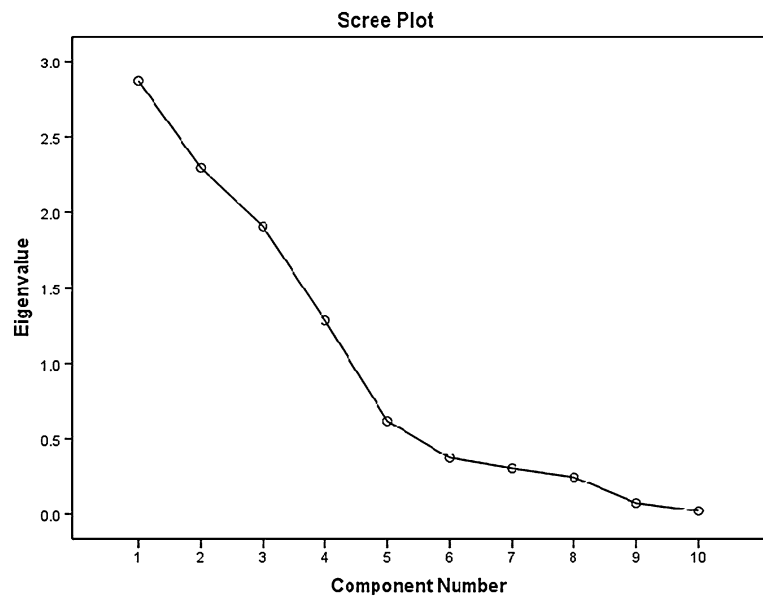
### Results and Discussion

Considering water sample parameters as shown in Tables 3a and 3b as variables, and applying the above methodology, the corresponding eigenvalues and eigenvectors for the Euclidean matrix were calculated using SPSS version 14.0 software. It is customary to use rotation of the components to obtain optimal distribution of variances in various components. Varimax rotation was applied to obtain optimal distribution of variances in various components. The number of factors (clusters) can now be selected based on principal components showing eigenvalues above one or number of principal components forming the cliff in scree plot or Akaike’s information criterion (AIC) (Basilevsky 1994; Kaiser 1960; Valarmathie et al. 2009). In this work, scree plot is used to select the number of clusters. It can be observed from the scree plots (Figs. 2a and 2b) for two seasons that only four clusters are needed to group the water samples. These four groups contribute 85.62 % in rainy and 88.22 % in winter seasons. Therefore, it is clear that the water sample data for two seasons can be clustered into four groups. In order to make a fair comparison with

**Fig. 2a** Scree plot for data of rainy season



**Fig. 2b** Scree plot for data of winter season



WQI values, it is decided that water quality can be categorized into four classifications depending on WQI values. Water quality can be treated as excellent, good, poor, and very poor if WQI lies in the ranges 0–25, 26–50, 51–75, and 76–100, respectively.

The next step is to assign the water samples into various clusters. In this study, absolute values of the elements of the eigenvectors (the factor loadings) are used to identify the clusters for water samples. The rotated factor loadings are shown in Table 7. For example, for water sample of  $W_{1.r.01}$ , the factor loading in cluster 1 is 0.93 which is higher as compared to loadings in other clusters, hence a stronger relationship with cluster 1 rather than with clusters 2, 3, and 4. It is to be noted that all the samples of well number 1 are clustered into group 1 irrespectively of seasons. From Table 6 it can be seen that the WQI of samples from well number 1 lying in the range 0–25 indicates that water samples belong to excellent category. Therefore, water samples belonging to cluster 1 (principal component 1 or PC1), cluster 2 (PC2), cluster 3 (PC3), and cluster 4 (PC4) are treated as of excellent, good, poor, and very poor quality, respectively. This procedure was repeated for all the samples to find out their respective clusters. In the same argument, sample  $W_{4.w.03}$  belongs to cluster 4 (very poor quality).

If a comparison is made on classification of water samples by Q-PCA mode (four clusters) and WQI method, it is observed that same classification has resulted in both the methods. It is found that nine samples belong to cluster 1, four samples to cluster 2, three samples to cluster 3, and three samples to cluster 4 for data from rainy season. Similarly, nine samples belong to cluster 1, three samples from cluster 2, two samples to cluster 3, and three samples to cluster 4 for data from winter season. The major advantage of clustering water quality data in an empirical manner lies in

the fact that it avoids subjectivity on weight assignment to parameters for WQI calculation. Furthermore, it provides a computationally elegant method with lesser dependence on choice of parameters. However, quality is a vague term that cannot be easily described using crisp data set, e.g. good quality water cannot simply be described as having a pH value of 7.0 or above. Instead, water quality can best be described based on its degree of potability and potential usages rather than expressing its constituents in numerical terms. Therefore, the non-parametric empirical method proposed here is efficient for such application. Any new water sample can be placed in any one of the above categories by knowing the constituents of physicochemical analysis, which is a routine analysis in the field and laboratory. The method is quite generic and can take care of any number of parameters. Although the method classifies water samples into proper groups, the performance of Q-mode clustering can be improved if size of data set is increased.

## Conclusions

In this work, PCA-based classification has been proposed for classification of water samples. It has been demonstrated that the methodology efficiently classifies into various clusters as far as the present data set is concerned. A similar classification can be obtained when WQI is calculated for the data set. The approach presented here is efficient and computationally elegant for classification of water samples. Importantly, it can be used in the field and laboratory due to easy accessibility and availability of statistical packages. The present approach has several advantages over other approaches:

**Table 7** Rotated factor loadings for water samples

Sample number	PC1	PC2	PC3	PC4	Sample number	PC1	PC2	PC3	PC4
W <sub>1.r.01</sub>	0.93	0.12	-0.06	0.15	W <sub>1.w.01</sub>	0.93	0.18	0.10	0.16
W <sub>1.r.02</sub>	0.92	0.19	-0.10	0.08	W <sub>1.w.02</sub>	0.94	0.12	0.06	0.16
W <sub>1.r.03</sub>	0.93	0.15	-0.04	0.17	W <sub>1.w.03</sub>	0.92	0.19	0.00	0.20
W <sub>2.r.01</sub>	0.21	0.91	0.27	0.00	W <sub>2.w.02</sub>	0.82	0.43	-0.16	-0.01
W <sub>2.r.02</sub>	0.93	0.25	0.04	-0.06	W <sub>2.w.03</sub>	0.82	0.42	-0.05	0.05
W <sub>2.r.03</sub>	0.92	0.25	-0.02	-0.03	W <sub>3.w.02</sub>	0.68	0.60	-0.07	0.18
W <sub>3.r.01</sub>	0.77	0.57	0.06	0.08	W <sub>3.w.03</sub>	0.84	0.30	-0.15	0.20
W <sub>3.r.02</sub>	0.87	0.39	0.17	0.10	W <sub>4.w.02</sub>	-0.33	-0.13	-0.32	-0.83
W <sub>3.r.03</sub>	0.92	0.31	0.14	-0.04	W <sub>4.w.03</sub>	-0.34	-0.13	-0.31	-0.81
W <sub>4.r.01</sub>	-0.10	0.12	0.95	-0.15	W <sub>7.w.01</sub>	0.32	-0.06	-0.08	0.90
W <sub>5.r.02</sub>	-0.15	0.94	0.12	-0.15	W <sub>8.w.01</sub>	0.28	0.89	-0.17	-0.15
W <sub>6.r.01</sub>	-0.08	0.04	0.95	-0.18	W <sub>8.w.02</sub>	0.25	0.91	0.01	-0.05
W <sub>6.r.03</sub>	0.52	0.08	0.79	0.16	W <sub>8.w.03</sub>	-0.23	0.81	0.17	-0.41
W <sub>7.r.03</sub>	0.60	0.74	0.20	0.00	W <sub>9.w.01</sub>	0.68	0.55	0.10	0.15
W <sub>9.r.01</sub>	0.51	0.79	0.13	0.11	W <sub>9.w.02</sub>	0.73	0.53	-0.02	-0.11
W <sub>9.r.03</sub>	0.72	-0.23	0.33	0.22	W <sub>10.w.01</sub>	0.40	-0.01	0.89	0.03
W <sub>10.r.01</sub>	0.21	-0.23	0.25	0.75	W <sub>10.w.03</sub>	-0.01	0.16	0.66	-0.30
W <sub>10.r.02</sub>	0.04	0.25	-0.11	-0.66					
W <sub>10.r.03</sub>	-0.37	-0.14	-0.14	-0.65					
Percent of total variance explained	32.81	23.32	16.72	12.76		34.29	28.46	14.40	11.06

- The physicochemical analysis of any water sample can be determined in a laboratory conveniently, since no sophisticated and costly experimental setup is required for the purpose. The classification system matches closely with the classification system based on calculation of WQI.
- It can be supported by available commercial software programs such as SPSS in order to facilitate industrial applications.
- It has the flexibility in allowing the user to identify the required number of clusters in advance, or consider it as a dependent variable.

The method of classification of water samples proposed in this work is quite generic and works well for present data set. Since such structured approach has already been applied in various fields of engineering due to its strong foundation, it is expected to work efficiently irrespectively of data sets. However, the efficiency of the method needs to be tested with water samples from other parts of the world.

**References**

Ahmed S, David KS, Gerald S (2004) Environmental assessment: an innovation index for evaluation water quality in streams. *Environ Manag* 34:406–414

Almasri ML, Kaluarachchi JJ (2004) Assessment and management of long-term water pollution of groundwater in agriculture-dominated watersheds. *J Hydrol* 295(1–4):225–245

Albadawi Z, Bashir HA, Chen M (2005) A mathematical approach for the formation of manufacturing cells. *Comput Ind Eng* 48:3–21

Ammann A, Eduard H, Sabine K (2003) Groundwater pollution by roof infiltration evidenced with multi-tracer experiments. *Water Res* 37(5):1143–1153

Brown RM, McClelland NI, Deiningner RA, Ronald GT (1970) A water quality index Do we dar? *Water Sew Works* 117(10):339–343

Basilevsky A (1994) *Statistical factor analysis and related methods*. Wiley, New York

Bhargava DS (1983) Use of a water quality index for river classification and zoning of the Ganga River. *Environ Pollut B* 6:51–67

Bhargava DS (1987) Nature and the Ganga. *Environ Conserv* 14:307–318

Chaudhari GR, Sohani D, Shrivastava VS (2004) Groundwater quality index near industrial area. *Indian J Environ Prot* 24(1):29–32

Couillard D, Lefebvre Y (1985) Analysis of water quality indices. *J Environ Manag* 21:161–179

Dahiya S, Singh B, Gaur S, Garg VK, Kushwaha HS (2007) Analysis of groundwater quality using fuzzy synthetic evaluation. *J Hazard Mater* 147(3):938–946

Dijksterhuis G (1998) European dimensions of coffee: rapid inspection of a data set using Q-PCA. *Food Qual Prefer* 9(3):95–98

Dinius SH (1972) Social accounting system for evaluating water resource. *Water Resour Res* 8(5):1159–1177

Dinius SH (1987) Design of an index of water quality. *Water Resour Bull* 23(5):833–843

- Faisal K, Tahir H, Ashok L (2003) Water quality evaluation and trend analysis in selected watersheds of the Atlantic region of Canada. *Environ Monit Assess* 88:221–248
- Hair JF Jr, Black WC, Babin BJ, Anderson RE (2009) *Multivariate data analysis*, 7th edn. Prentice Hall, New York
- Harrison TD, Cooper JAG, Ramm AEL (2000) Water quality and aesthetics of South African estuaries, Department of Environment Affairs and Tourism, South Africa. Available from [www.environment.gov.za/soer/reports/ehi/ehi\\_ch4.pdf](http://www.environment.gov.za/soer/reports/ehi/ehi_ch4.pdf)
- Horton RK (1965) An index number system for rating water quality. *J Water Pollut Control Fed* 37(3):300–306
- Iscen CF, Emiroglu O, Ilhan S, Arslan N, Yilmaz V, Ahiska S (2008) Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake. *Environ Monit Assess* 144(1–3):269–276
- Jha AN, Verma PK (2000) Physico-chemical properties of drinking water in town area of Godda district under Santal Pargana (Bihar), India. *Pollut Res* 19(2):245–247
- Kaiser HF (1960) The application of electronic computers to factor analysis. *Educ Psychol Meas* 20:141–151
- Karnchanawong S, Ikeguchi SKT (1993) Evaluation of shallow well water quality near a waste disposal site. *Environ Int* 19(6):579–587
- Landwehr JM (1979) A statistical view of a class of water quality indices. *Water Resour Res* 15(2):460–468
- Lind CJ, Carol L, Creasey CA (1998) In situ alteration of minerals by acidic groundwater resulting from mining activities: preliminary evaluation of method. *J Geochem Explor* 64(1–3):293–305
- Lohani BN, Todino G (1984) Water quality index for Chao Phraya River. *J Environ Eng* 110(6):1163–1176
- Lumb A, Sharma TC, Bibault JF (2011) A review of genesis and evolution of water quality index (WQI) and some future directions. *Water Qual Health Expos*: 11–24
- Lumb A, Halliwell D, Sharma T (2006) Application of CCME water quality index to monitor water quality: a case of the Mackenzie River Basin, Canada. *Environ Monit Assess* 113:411–429
- Maticic B (1999) The impact of agriculture on groundwater quality in Slovenia: standards and strategy. *Agric Water Manag* 40(2–3):235–247
- Montgomery DC, Runger GC (1999) *Applied statistics and probability for engineers*. Wiley, New York
- Nagarajan P, Priya GK (1999) Groundwater quality deterioration in Tiruchirapalli, Tamilnadu. *J Ecotoxicol Environ Monit* 9(2):155–159
- Nives SG (1999) Water quality evaluation by index in Dalmatia. *Water Resour* 33:3423–3440
- Park S, Choi D, Jun CH (2001) A clustering method for discovering patterns using gene regulatory processes. *Genome Inform* 12:249–251
- Parinet B, Lhote A, Legube B (2004) Principal component analysis: an appropriate tool for water quality evaluation and management-application to a tropical lake system. *Ecol Model* 178:295–311
- Sabal D, Khan TI (2008) Fluoride contamination status of groundwater in Phulera tehsil of Jaipur district, Rajasthan. *J Environ Biol* 29:871–876
- Sarkar C, Abbasi SA (2006) Qualidex—a new software for generating water quality indices. *Environ Monit Assess* 119:201–231
- Shaji C, Nimi H, Bindu L (2009) Water quality assessment of open wells in and around Chavara industrial area, Quilon, Kerala. *J Environ Biol* 30(5):701–704
- Shamruck M, Corapcioglu MY, Fayek AA, Hassona (2001) Modeling the effect of chemical fertilizers on groundwater quality in the Nile Valley aquifer, Egypt. *Groundwater* 39(1):59–67
- Sharma RK, Agarwal M (2005) Biological effects of heavy metals. *J Environ Biol* 26:301–313
- Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji River Basin, Japan. *Environ Model Softw* 22:464–475
- Singh KP, Parwana HK (1999) Groundwater pollution due to industrial wastewater in Punjab state and strategies for its control. *Indian J Environ Prot* 19(4):241–244
- Singh B, Sudhir D, Sandeep J, Garg VK, Kushwaha HS (2008) Use of fuzzy synthetic evaluation for assessment of groundwater quality for drinking usage: a case study of southern Haryana. *Indian Environ Geol* 54:249–255
- Singh UK, Kumar M, Chauhan R, Jha PK, Ramanathan AL, Subramanian V (2008) Assessment of the impact of landfill on groundwater quality: a case study of the Pirana site in western India. *Environ Monit Assess* 141(1–3):309–321
- Srinivas C, Shankar R, Venkateshwar C, Rao MSS, Reddy RR (2000) Studies on groundwater quality of Hyderabad. *Pollut Res* 19(2):285–289
- Swamee PK, Tyagi A (2000) Describing water quality with aggregate index. *J Environ Eng* 126(5):451–455
- Shiow-Mey L, Shang-Lien, Shan-Hsien W (2004) A generalized water quality index for Taiwan. *Environ Monit Assess* 96:35–52
- Tyagi P, Budhi D, Sawhney RL (2003) A correlation among physico-chemical parameters of groundwater in and around Pithampur industrial area. *Indian J Environ Prot* 23(11):1276–1282
- Tiwari TN, Mishra M (1985) A preliminary assignment of water quality index of major Indian Rivers. *Indian J Environ Prot* 5(4):276–279
- Ubala B, Farooqui M, Arif M, Zaheer A, Dhule DG (2001) Regression analysis of groundwater quality data of Chikalthana industrial area, Aurangabad (Maharashtra). *Orient J Chem* 17(2):347–348
- Valarmathie P, Srinath MV, Dinakaran K (2009) Increased performance of clustering high dimensional data through dimensionality reduction technique. *J Theor Appl Inf Technol* 5(6):731–733
- Walski TM, Parker FL (1974) Consumers water quality index. *J Environ Eng Div* 100(3):593–611
- WHO (2006) *Guidelines for drinking water quality first addendum to 3rd edn (I) recommendations*, Geneva, Switzerland
- Woocay A, Walton J (2008) Multivariate analyses of water chemistry: surface and ground water interactions. *Groundwater* 46(3):437–449
- Zhang WL, Tian X, Zhang N, Li XQ (1996) Water pollution of groundwater in northern China. *Ecosyst Environ* 59(3):223–231