Desalination and Water Treatment
www.deswater.com
doi: 10.1080/19443994.2014.880379

54 (2015) 28–36
April

Taylor & Francis
Taylor & Francis Group

Assessment of water quality index using cluster analysis and artificial neural network modeling: a case study of the Hooghly River basin, West Bengal, India

Keka Sinha[a], Papita Das (Saha)[b,*]

[a]*Department of Biotechnology, National Institute of Technology – Durgapur, Mahatma Gandhi Avenue, Durgapur, WB 713209, India*
[b]*Department of Chemical Engineering, Jadavpur University, Kolkata, India*
Tel. +91 9903739855; email: papitasaha@gmail.com

Received 23 August 2013; Accepted 21 December 2013

ABSTRACT

River Hooghly, considered as an important tributary of the River Ganga, has been affected by indiscriminate discharging of polluted and untreated sewage sludge and industrial waste into the waterways. The assessment of water quality for natural river waters was done using a water quality index (WQI), developed by DELPHI and the Council of Ministers of the Environment methods. These two methods reflect the quality of the water measured with respect to its pollution level. Multivariate statistical techniques, such as cluster analysis, were applied to the data-set on water quality of the Hooghly River (India) which was generated during the years 2002–2008 controlling at eight different sites for five parameters. The relationships among the stations are highlighted by cluster analysis to characterize the WQI. The study represents a computer-simulated artificial neural network model for the evaluation of the relationship between the different parameters of water bodies collected at different stations along Hooghly River responsible for water quality measurement. Finally, both the water quality methods (CCME and DELPHI) were statistically compared by the coefficient of determination (R^2), root mean square error, and absolute average deviation based on the validation data-set.

Keywords: ANN model; CCME method; Cluster analysis; DELPHI process; River Hooghly; Water quality index

1. Introduction

Pollution of surface water as well as ground water is a big problem due to rapid urbanization and unlimited discharge of sewage, agricultural, and industrial waste. The city of Kolkata and its metropolitan area is situated at the lower most stretch of the Gangetic delta and spread over almost the bank of River Hooghly. Howrah and Hooghly are mainly located on the western banks of Hooghly River. Due to increase in population and industrialization, large amounts of domestic and industrial effluent, waste, and waste water reached the River Hooghly through various canals.

The West Bengal Pollution Control Board has been controlling the water quality of the river for

*Corresponding author.

assessment of its water quality level since late 80s. Water pollution is the contamination of water bodies from chemical, particulate, or bacterial matter that affects the water quality level. The index of water quality [1,2] is a numerical indicator of physical, chemical, biological, or radiological condition of the water sources and determining their quality before use for various purposes such as drinking, agricultural, aquatic life, recreational, and industrial water [3]. Mixed domestic sewage and treated and untreated waste effluents from different industries such as tannery, steel plants, and thermal power plants are directly discharged into the River Hooghly along its length. The contaminated water bodies may have undesirable color, odor, taste, turbidity, organic matter contents, harmful chemical contents, toxic and heavy metals, pesticides, oily matters, industrial waste products, radioactivity, high total dissolved solids, acids, alkalies, domestic sewage content, virus, bacteria, protozoa, rotifers, worms, etc. The contaminated drinking water may also cause human health risks such as tumors, ulcers, and skin disorders [4].

Water quality index (WQI) has been built to assess the suitability of water for a variety of uses. A WQI is a single number quantitative expression that provides overall water quality at a certain location and time based on several water quality parameters [5–12]. Different techniques have been used in an attempt to change complex water quality data into simpler information that is easy to understand and available to the public [13,14]. This WQI is a dimensionless value that ranges between 0 and 100 [15]. A good water quality is represented by a higher index value of water quality [16,17]. However, a water index is based on some very important parameters that can provide a simple measurement of water quality. It gives the public a general idea of the possible problems with water in the region.

The study aims to:

(1) to collect the waste water samples from each sampling station at suitable intervals of each month along Hooghly River,
(2) to improve the water quality systems of the Hooghly River by different techniques to obtain WQI with respect to their physical, chemical, and biological characteristics, and
(3) to identify the natural clustering patterns on the basis of similarities between data measured from different sampling points by cluster analysis method.

2. Materials and methods

2.1. Study sites and collected data

In order to determine the WQI, water samples were collected each month across the river width at all the eight stations (Berhampore, Palta, Srirampore, Howrah (Shibpur), Garden Reach, Dakhshineswar, Uluberia, and Diamond Harbour) during the study period (Fig. 1). Five selected water quality parameters, their units, and methods of analysis are summarized in Table 1.

2.2. Experimental procedure

pH was measured using pH meter (Elico, India).

Dissolved oxygen and biochemical oxygen demand (BOD) was analyzed using Winkler azide method. In this method, 300 mL sample was taken in a BOD bottle. Then 2 mL manganous sulfate, and 1 mL alkaline iodide-azide was mixed with it. A separate pipette was used to add the reagent well below the liquid surface. The bottle should be closed without inclusion of any air bubble and was mixed well thoroughly and allowed for 2–3 min the precipitate to settle down. The stopper was removed and 2 mL sulfuric acid was mixed with it. Two hundred and three milliliter sample of this solution was taken and was titrated with sodium thiosulfate (0.025 N) with starch as indicator.

For determination of total coliform and fecal coliform, multiple tube method was used. Ten test tubes were taken and 10 mL of media was distributed in each tube. The tubes were dipped in an inverted way into each test tube. All the tubes were autoclaved and after that 10 mL aliquots of the sample was inoculated into five tubes. 1 mL aliquots of the sample was added into marked tubes and 0.1 mL aliquots of sample was transferred into the rest of the marked tubes. All the incubated tubes were kept for 48 h at 37˚C. The production of gas in Durham's tube indicates the presence of coliform bacteria in the sample being analyzed.

2.3. Water quality index

Several methods have been introduced in the past to generate a suitable WQI method. Each method had some advantages and some disadvantages also.

2.3.1. Calculation of CCME WQI [18]

Each WQI (indicator) was calculated using methods developed by Canadian Council of Ministers of the Environment (CCME) [18] based on three different measurements of water quality: scope (F_1), frequency (F_2), and amplitude (F_3) results as,

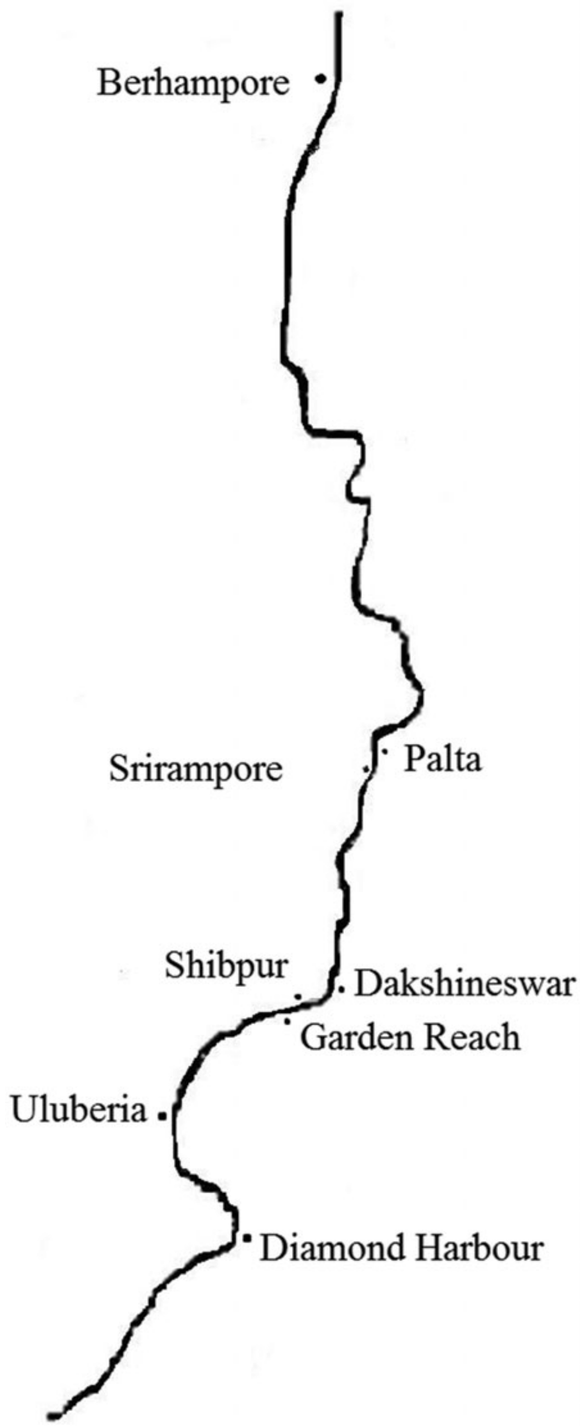

Fig. 1. Locations of the sample taken.

$$\text{WQI (indicator)} = 100 - \sqrt{\frac{(F_1^2 + F_2^2 + F_3^2)}{1.732}} \quad (1)$$

For each indicator, the grading scale following the "ranking" scale is used in five categories or levels that corresponded to specific levels of water quality which is shown in Table 2.

Where F_1 (scope) describes the extent of quality guideline non-compliance over a time period of interest and it was calculated as:

$$F_1 = \left(\frac{\text{Number of failed variables}}{\text{Total number variables}}\right) \times 100 \quad (2)$$

where the failed variables indicate the water quality variables with objectives which are tested during the time period for the index calculation.

F_2 (frequency) represents the percentage of individual tests that do not exceed failed tests.

$$F_2 = \left(\frac{\text{Number of failed tests}}{\text{Total number of tests}}\right) \times 100 \quad (3)$$

F_3 (amplitude) represents the value by which the failed test values do not meet their objectives, and it was calculated in three steps as:

(1) When the test value must not exceed the objective and the objective is termed an "excursion," then it expressed as follows:

$$\text{Excursion} = \left(\frac{\text{Failed test value}}{\text{Objective}}\right) - 1 \quad (4)$$

For cases where test value should not fall below the objective:

$$\text{Excursion} = \left(\frac{\text{Objective}}{\text{Failed test value}}\right) - 1 \quad (5)$$

(2) The collective amount by which individual tests are out of compliance is calculated by summing the excursions of individual tests from their objectives and dividing by the total number of tests. This variable referred to as the normalized sum of excursions (NSE) calculated as:

$$\text{NSE} = \frac{\sum_{i=0}^{n} \text{Excursion values}}{\text{Total number of tests in the results}} \quad (6)$$

Table 1
Water quality parameters, abbreviations, units, and analytical methods as measured during 2002–2008 for the Hooghly River

Parameters	Abbreviations	Units	Analytical methods
pH	pH	pH unit	pH-meter
Dissolved oxygen	DO	mg L^{-1}	Winkler azide method
Fecal coliform	F coli	MPN/100 mL	Multiple tube method
Biochemical oxygen demand	BOD	mg L^{-1}	Winkler azide method
Total coliform	TC	MPN/100 mL	Multiple tube method

Table 2
Grading scale used for the water quality indicator in the CCME process

WQI	Condition	Grade
95–100	Excellent	A
80–94	Good	B
65–79	Fair	C
45–64	Marginal	D
0–44	Poor	F

(3) F_3 is then calculated by an asymptotic function that ranges the NSE from objectives to yield a range between 0 and 100.

$$F_3 = \frac{\text{NSE}}{[0.01(\text{NSE}) + 0.01]} \tag{7}$$

Once the CCME WQI value has been determined, water quality can be categorized by corresponding it to one of the following level.

2.3.2. Calculation of WQI using DELPHI process

The systematic technique was attempted to incorporate the judgments of a large diverse system in water quality management process [19–21]. Two basic approaches are followed by the researchers: aggregative method and multiplicative method. In both cases, experts' ratings were taken and the scoring functions were done using regression analysis. An overall quality rating is derived by multiplying the final weights (w_i) of each individual parameter with the corresponding quality rating (q_i), the sum of which gives the required single number WQI. The quality rating is measured on a scale of 0–100 point (i.e. highest to lowest polluting).

Method 1: aggregative method [20].

The WQI considered is of the form:

$$\text{WQI}_a = \sum_{i=1}^{n} q_i w_i \tag{8}$$

where WQI_a is the aggregative WQI between 0 and 100, q_i the quality of ith parameter between 0 and 100, w_i the weight of ith parameter (between 0 and 1), and n is the total number of parameters.

In this type of index, if any significantly relevant parameter exceeds the permissible limit, the mean weighted indices do not consider sufficient lowering of the WQI. Table 3 is used to describe that the high indicator values corresponded to low levels of contamination (i.e. good water quality) and low values indicated high levels of contamination (i.e. poor water quality).

Method 2: multiplicative method.

Multiplicative form of index may be considered by:

$$\text{WQI}_m = \prod_{i=1}^{n} (q_i) \tag{9}$$

In this index, weights are calculated to the individual parameters based on a subjective opinion. The classification is shown in Table 4.

2.4. Cluster analysis

In order to avoid univariate statistical analysis problem, multivariate analysis such as cluster analysis is used in the study to describe the correlation amongst a large number of meaningful data without losing much information. Cluster analysis is a

Table 3
Classification of water quality based on WQI-DELPHI aggregative method

WQI	Class	Description
63–100	A	Good to excellent
50–63	B	Good to moderate
38–50	C	Bad
Below 38	D, E	Bad to very bad

technique to classify groups of objects, or clusters, in such a way that the resulting groups are similar to each other but distinct from other groups [22–27]. Cluster analysis can be performed on many different types of data-sets. Hierarchical clustering is a way to investigate grouping in data, simultaneously over a variety of scales, by creating a cluster tree. The tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next higher level. Hierarchical agglomerative cluster analysis was performed on the normalized data-set by means of the Ward's method for sample classification using squared Euclidean distances as a measure of close proximity [28,29]. A dendrogram has been developed using the MATLAB7 (The Mathworks, Inc. version 7.0.1) with WQI data-set of the Hooghly River to find out the similar sampling sites spread over the river stretch.

2.5. Artificial neural networks modeling

In this present study, different neural network models and algorithms were tested and optimized to obtain the best model structure for the prediction of WQI of sampling stations along the River Hooghly. Based on the principles of the feed-forward back-propagation algorithm, the modeling method has been developed [30]. The artificial neural networks (ANN) architecture typically comprises three types of neuron layers: an input layer (independent variables), one or more number of hidden layers, and an output layer (dependent variables). The input layer, which only connects one input value with its associated weighted values, receives information from external sources and transfers this information to the hidden layer for processing [31]. The net input for each neuron (a_j) is the sum of all input values x_i; each multiplied by its weight W_{ji}, and added a bias term Z_j which may be formulated as:

$$a_j = \sum w_{ij}x_i + z_j \tag{10}$$

Table 4
Classification of water quality based on WQI–DELPHI multiplicative method

WQI	Description
0–20	Bad
21–50	Medium
51–80	Good
81–100	Very good

The output value (t_j) can be generated by processing all the data of hidden layer and net input neuron into the linear transfer function (purelin) of the neuron:

$$t_j = f(a_j) \tag{11}$$

In this present study, two types of transfer function have been applied: a tan-sigmoid transfer function (tansig) at the hidden layer and a linear transfer function (purelin) at the output layer. The Levenberg–Marquardt back-propagation (BP) algorithm was used for network training. The input and output parameters to the ANN model were identical to the factors considered in the cluster analysis approach, namely pH, dissolved oxygen, BOD, total coliform, fecal coliform, and WQI, respectively. All neural network calculations were implemented using Neural Network Toolbox of MATLAB version 7.0.1.

3. Results and discussion

Table 5 represents the range of the parameters at different locations.

Cluster analysis was performed to identify the spatial similarity for clustering of WQI of sampling sites under the monitoring network. It represented a dendrogram (Figs. 2 and 3) by using two different methods, grouping all the eight sampling stations based on the WQI of the Hooghly River.

From the results it was observed that in the CCME method, the clustering procedure generated two statistically significant groups according to the average calculated WQI for six years. Cluster 1 (sites 2, 3, 4, 5, and 7) and another Cluster 2 (site 6 is distantly related to sites 1 and 8) can be classified corresponding to a relatively moderate pollution, low pollution, and very high pollution stations, respectively. From the DELPHI technique it was evident that all eight stations on the river can be grouped into three major significant clusters with similar characteristic features, Cluster 1: Srirampore, Howrah (Shibpur), Garden Reach, Dakhshineswar (the range of WQI is between 24 and 28.32) Cluster 2: Palta and Uluberia (WQI are 36.67 and 35.97), and Cluster 3: Berhampore and Diamond Harbour (WQI are 62.25 and 62.77) as presented in Fig. 2. It was clearly found that the third major clustering group (high significance of clustering) was characterized by the highest Euclidean linkage distance than the other two clustering group. CA technique is useful in reliable classification of WQI in the whole region across the river basin and will make it possible to design a future spatial sampling strategy in an optimal manner. Thus, the number of sampling sites in the evaluating network will be reduced.

Table 5
Water quality parameters at different seasons in range

Parameters	pH	Dissolved oxygen, mg/L	Fecal coliform, MPN/100 mL	Biochemical oxygen demand, mg/L	Total coliforms, MPN/100 mL
Berhampore	7.28	6.983–7.658	31.85–315	1.59–2.85	131.41–583.33
Palta	7.06	6.044–7.45	116.66–741.66	1.88–3.08	401.04–1,400
Srirampore	7.88	6.225–7.815	424.16–3,475	1.75–3.64	866.66–10545.83
Howrah (Shibpur)	7.49	5.65–6.926	59.16–2,550	2.073–3.516	176.66–5,350
Garden Reach	7.79	5.78–6.771	107.50–3762.50	2.918–4.087	453.33–9337.50
Dakhshineswar	8.48	5.919–6.373	238.95–4241.66	3.34–3.98	810–12485.42
Uluberia	8.28	5.733–6.285	53.58–1009.16	3.037–3.925	171.75–2342.08
Diamond Harbour	8.39	6.051–6.88	5.13–318.00	1.33–1.9625	34.26–569.91

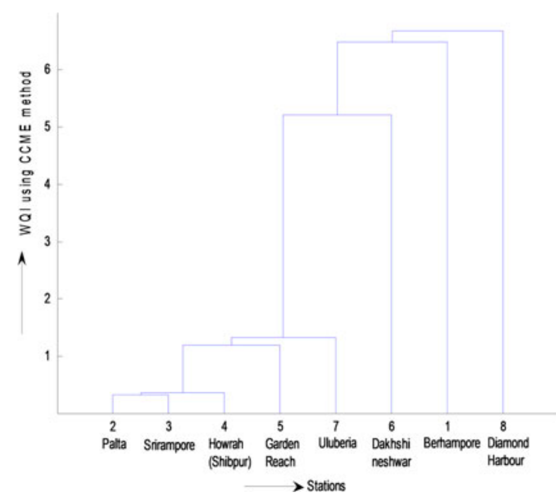

Fig. 2. Dendrogram based on agglomerative hierarchical clustering using the CCME method.

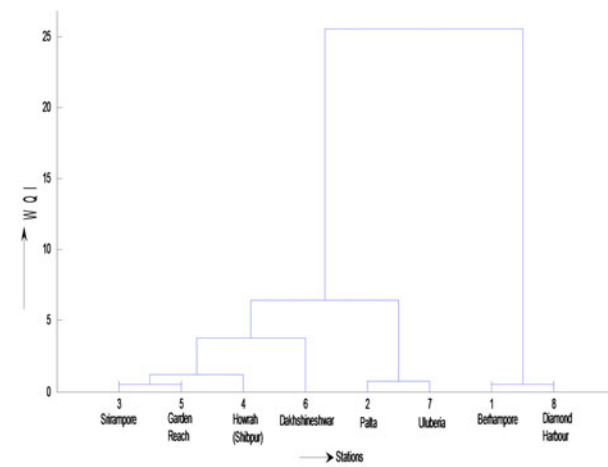

Fig. 3. Dendrogram based on agglomerative hierarchical clustering using the DELPHI method.

Machine learning techniques such as ANNs has increased recently as a powerful tool in simulation of data modeling and could be useful in ecological aspects [32–34]. In the present study, the data-set generated from the average calculated WQI of all the eight sampling stations along the river basin of Hooghly during the period 2002–2008 using both CCME and DELPHI methods. All analyses were based on the calculated data-set. The original training data-set comprises eight data points. The training procedure is iterated 10 times until the errors are minimized and the value of correlation coefficient (R) between the model prediction and experimental results reaches 1. The ANN model for the present problem involved a feed-forward neural network with five inputs, one hidden layer (one layer with 10 neurons) and one output layer (including one neuron). This feed-forward neural network is signified as multilayer perceptron (5:10:1), trained using BP method based on Levenberg–Marquardt algorithm. The goodness of fit of the trained network is shown in Figs. 4 and 5. Regression plot in Fig. 4 has a correlation coefficient of 0.987 using the DELPHI method and the correlation coefficient of Fig. 4 is 0.954 using the CCME technique.

The performance of the constructed DELPHI method and CCME method was also statistically analyzed by the root mean square error (RMSE), coefficient of determination (R^2), and absolute average deviation (AAD) as follows [35,36]:

$$\text{RMSE} = \left(\frac{1}{n} \sum_{i=0}^{n} (T_{\text{WQI,Pr}ed} - T_{\text{WQI,Exp}})^2 \right)^{\frac{1}{2}} \quad (12)$$

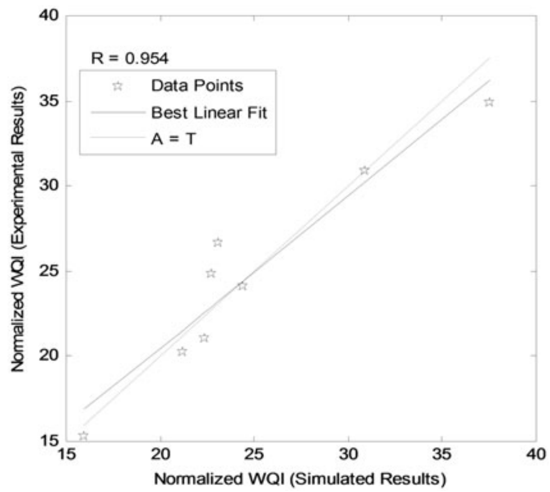

Fig. 4. Regression plot on WQI (experimental vs. predicted) using the CCME Method with five input variables, 10 processing elements in the hidden layer, and one output variable.

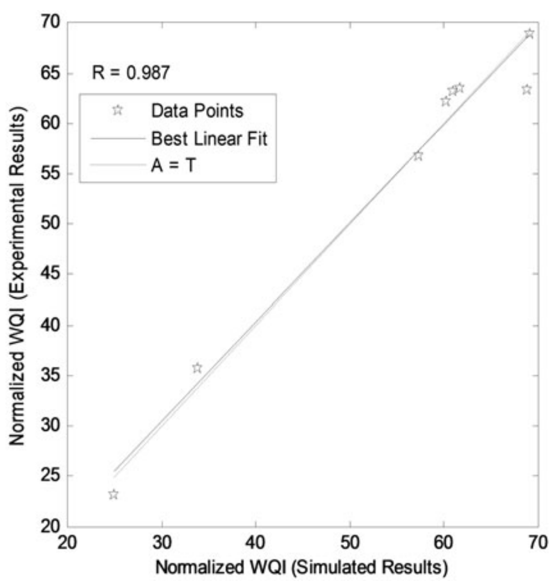

Fig. 5. Regression plot on WQI (Experimentally vs. Predicted) using the DELPHI method with five input variables, 10 processing elements in the hidden layer, and one output variable.

$$R^2 = \frac{\left(\sum_{i=1}^{n}\left(T_{\mathrm{WQI,Exp}} - \overline{T_{\mathrm{WQI,Exp}}}\right)\left(T_{\mathrm{WQI,Pr}ed} - \overline{T_{\mathrm{WQI,Pr}ed}}\right)\right)^2}{\sum_{i=1}^{n}\left(T_{\mathrm{WQI,Exp}} - \overline{T_{\mathrm{WQI,Exp}}}\right)^2\left(T_{\mathrm{WQI,Pr}ed} - \overline{T_{\mathrm{WQI,Pr}ed}}\right)^2} \quad (13)$$

Table 6
Comparison of DELPHI and CCME methods for determining WQI

Parameter	DELPHI method	CCME method
RMSE	1.687	2.16
R^2	0.92	0.805
AAD (%)	0.18	1.23

$$\mathrm{AAD} = \left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{T_{\mathrm{WQI,Pr}ed} - T_{\mathrm{WQI,Exp}}}{T_{\mathrm{WQI,Exp}}}\right)\right) \times 100 \quad (14)$$

where n is the number of data points, $T_{\mathrm{WQI,Pr}ed}$ is the predicted value from ANN results, $T_{\mathrm{WQI,Exp}}$ is the actual WQI value calculated by CCME and DELPHI methods, and the symbol "–" is the average of the related values. Table 6 represents the statistical comparison between CCME and DELPHI techniques. In this present study, both CCME and DELPHI methods provided good determination of water quality of the Hooghly River, yet the DELPHI method showed a clear superiority over the CCME method for both data fitting by ANN model development and estimation capabilities.

Thus it would be more rational and reliable to calculate the annual WQI using five different parameters such as pH, dissolved oxygen, biochemical oxygen demand, total coliform, and fecal coliform at eight different sampling stations through the DELPHI method.

4. Conclusion

From the present case study of water quality of various stations along the River Hooghly, it is noticed that water of the Hooghly River is somewhat polluted with the pollutants from various industries and domestic sources. Both DELPHI and CCME methods were applied to calculate the average data of every month of a year. The hierarchical cluster analysis and developed ANN model were applied to the Hooghly River basin to measure WQI, which has a very poor water quality. Overall water quality ranges from poor to marginal quality depending on the river reach and sample year. Hierarchical cluster analysis grouped eight sampling stations into three major clusters of similar characteristics reflecting the WQI calculated by the DELPHI method. The WQI was formulated by both DELPHI and CCME techniques and the RMSE, R^2, and ADD were used together to compare the water quality performance of the CCME and DELPHI methods. The DELPHI method was found to have higher predictive capability than the CCME method.

Acknowledgements

We sincerely acknowledge the co-operation extended by the West Bengal Pollution Control Board of Durgapur and Kolkata, India for providing the valuable data and suggestion regarding the water quality of river Hooghly at different locations of West Bengal. I am also thankful to Mr Souvik Ganguly, Environment Engineer, West Bengal Pollution Control Board for his valuable help and suggestions.

References

[1] S. Abbasi, A Water Quality: Sampling and Analysis, Discovery Publishing House, New Delhi, 1998, pp. 200–250.
[2] F. Coulston, E. Mrak, Water Quality, Proceedings of an international Forum, American Press, San Francisco, CA, 1977, pp. 51–56.
[3] A. Sargaonkar, V. Deshpande, Development of an overall index of pollution for surface water based on a general classification scheme in Indian context, Environ. Monit. Assess. 89 (2003) 43–67.
[4] C.M. Hogan, Water pollution, Encyclopedia of Earth, Topic ed. Mark McGinley; ed. in chief C. Cleveland, National Council on Science and the Environment, Washington, DC, 2010.
[5] A. Adriano, A. Bordalo, T. Rita, J. William, A water quality index applied to an international shared river basin: The case of the Douro river, Environ. Monit. Assess. 38(6) (2006) 910–920.
[6] H. Boyacioglu, Development of a water quality index based on a European classification scheme, Water S. Aust. 33 (2007) 101–106.
[7] R.M. Brown, N.I. Mcclelland, R.A. Deinnings, R.G. Tagore, A water quality index, do we dare? Water Sewage Works 11 (1970) 339–343.
[8] J.M. Landwehr, R.A. Deininger, A comparison of several water quality index, J. Water Pollut. Control Fed. 48(5) (1976) 954–958.
[9] S.F. Pesce, D.A. Wunderlin, Use of water quality indices to verify the impact of Córdoba city (Argentina) on Suquía river, Water Res. 34 (2000) 2915–2926.
[10] E. Sánchez, M.F. Colmenarejo, J. Vicente, A. Rubio, M.G. García, L. Travieso, R. Borja, Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution, Ecol. Indic. 7 (2007) 315–328.
[11] K.P. Singh, A. Malik, D. Mohan, S. Sinha, Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti river (India)—A case study, Water Res. 38 (2004) 3980–3992.
[12] R.K. Singh, H. Anandh, Water quality index of some Indian rivers, Indian J. Environ. Health 8 (1996) 21–34.
[13] D.A. Bennetts, J.A. Webb, D.J.M. Stone, D.M. Hill, Understanding the salinisation process for groundwater in an area of south-eastern Australia, using hydrochemical and isotopic evidence, J. Hydrol. 323 (2006) 178–192.
[14] P. Pulido-Leboeuf, A. Pulido-Bosch, M. Luisa Calvache, Á. Vallejos, J. Miguel Andreuc, Strontium, SO_4^{2-}/Cl^- and Mg^{2+}/Ca^{2+} ratios as tracers for the evolution of seawater into coastal aquifers: The example of Castell de Ferro aquifer (SE Spain), Geosci. 335 (2003) 1039–1048.
[15] K. Parmar, V. Parmar, Evaluation of water quality index for drinking purposes of river Subernarekha in Singhbhum district, Int. J. Environ. Sci. 1 (2010) 77–81.
[16] C. Cude, Oregon water quality index: A tool for evaluating water quality management effectiveness, J. Am. Water Resour. Assoc. 37 (2001) 125–137.
[17] M. Pandey, S.M. Sundaram, Trend of water quality of river Ganga at Varanasi using WQI approach, Int. J. Ecol. Environ. Sci. 28 (2002) 139–142.
[18] Canadian Council of Ministers of the Environment (CCME) (2001), Canadian water quality guidelines for the protection of aquatic life: CCME water quality index 1.0, technical report, in: Canadian Environmental Quality Guidelines, 1999. Winnipeg: Canadian Council of Ministers of the Environment. http://www.ccme.ca/assets/ pdf/ wqi_techrprtfctsht_e.pdf.
[19] S.A. Alexande, D.A. Jeyakar, M. Chellaraj, J. Princy, A. Rajendran, Formulation of a new water quality index: HWQI2, Ind. J. Environ. Prot. 19 (1999) 842–845.
[20] P. Saha, P.K. Banerjee, S. Datta, Assessment of water quality of Tolly's Nullah by water quality index (WQI), J. Inst. Eng. 88 (2007) 3–9.
[21] T.M. Walsk, F.L. Parker, Consumers water quality index, J. Environ. 5 (1974) 93–611.
[22] B. Helena, R. Pardo, M. Vega, E. Barrado, J.M. Fernandez, L. Fernandez, Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis, Water Res. 34 (2000) 807–816.
[23] R. Reghunath, T.R.S. Murthy, B.R. Raghavan, The utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka, India, Water Res. 36 (2002) 2437–2442.
[24] P. Simeonova, V. Simeonov, G. Andreev, Water quality study of the Struma river basin, Bulgaria, Cent. Eur. J. Chem. 1(2) (2003) 121–136.
[25] V. Simeonov, J.A. Stratis, C. Samara, G. Zachariadis, D. Voutsa, A. Anthemidis, M. Sofoniou, Th. Kouimtzis, Assessment of the surface water quality in Northern Greece, Water Res. 37 (2003b) 4119–4124.
[26] V. Simeonov, P. Simeonova, R. Tsitouridou, Chemometric quality assessment of surface waters two case studies, Chem. Eng. Ecol. 11 (2004) 449–469.
[27] M. Vega, R. Pardo, E. Barrado, L. Deban, Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis, Water Res. 32 (1998) 3581–3592.
[28] J.W. Einax, H.W. Zwanziger, S. Geiss, Chemometrics in Environmental Analysis, Wiley, Weinheim, 1997.
[29] R. Fovell, M.Y. Fovell, Climate zones of the conterminous United States defined using cluster analysis, J. Clim. 6 (1993) 2103–2135.
[30] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (1986) 533–536.

[31] U. Özdemir, B. Özbay, S. Veli, S. Zor, Modeling adsorption of sodium dodecyl benzene sulfonate (SDBS) onto polyaniline (PANI) by using multi linear regression and artificial neural networks, Chem. Eng. J. 178 (2011) 183–190.

[32] M.G. Moghaddam, M. Khajeh, Comparison of response surface methodology and artificial neural network in predicting the microwave-assisted extraction procedure to determine zinc in fish muscles, Food Nutr. 2 (2011) 803–808.

[33] F. Recknage, Applications of machine learning to ecological modeling, Ecol. Mod. 146 (2011) 303–310.

[34] K. Sinha, P.D. Saha, S. Datta, Response surface optimization and artificial neural network modeling of microwave assisted natural dye extraction from pomegranate rind, Ind. Crops Prod. 37 (2012) 408–414.

[35] A.S. Tokar, A.P. Johnson, Rainfall-runoff modeling using artificial neural networks, J. Hydrol. Eng. 399 (2011) 232–239.

[36] F. Geyikçi, E. Kılıç, S. Çoruh, S. Elevli, Modelling of lead adsorption from industrial sludge leachate on red mud by using RSM and ANN, Chem. Engg. J. 183 (2012) 53–59.