Indian Journal of Geo Marine Sciences
Vol. 47 (02), February 2018, pp. 365-377

Soft computing and statistical technique - Application to eutrophication potential modelling of Mumbai coastal area

V. S. Bharti[*1,] A. B. Inamdar [2], C. S. Purushothaman[3] & V.K.Yadav[4]

[1,4]Central Institute of Fisheries Education (CIFE), Versova, Mumbai-400061, India
[2] Centre of Studies in Resource Engineering (CSRE), IIT Bombay, Powai, Mumbai - 400076, India.
[3] Central Marine Fisheries Research Institute (CMFRI), Kochi-682 018, India

*[E.Mail: vidyabharti2003@yahoo.co.in]

Received 11 January 2016 ; revised 17 November 2017

Three water quality parameters – dissolved oxygen (DO), coloured dissolved organic matter (CDOM) and Chlorophyll - a loaded on the first principal component under the dimensional reduction method were used for deriving the Eutrophication Index (EI). Fuzzy logic (Mamdani) method of EI estimation is smoother than the Principal Component Analysis (PCA) method. Eutrophication potential obtained from the rule-based fuzzy approach and the multiple regressions derived from the first principal component were selected as the target variables for the artificial neural network (ANN) model in training and prediction. The performance of the ANN models with PCA-derived target and fuzzy-derived target was compared through the mean absolute error, root mean square error, and correlation coefficient computed from the measured and model-simulated EI values. EI predictions of this model has positive, high correlation (r = 0.968) with the measured EI values derived from the fuzzy approach as compared to the PCA-derived EI (r = 0.851) implying that the model predictions explain around 93.7% of the variation in the measured EI values derived by fuzzy approach as compared to 72.4% in the case of PCA-derived measured value.

[**Keywords**: Fuzzy logic, Principal component analysis, Eutrophication modeling, Mumbai coastal water, artificial neural network.]

Introduction

The eutrophication of aquatic ecosystems is usually accelerated by excessive nutrient inputs, thereby causing the deterioration of water quality and impairing the intended use of the water body for sustaining aquatic life and fisheries. Water quality indicators are widely used to predict the eutrophication levels of waters. However, the inter-dependence and inter-relationship of indicators increase the complexity in prediction as also the spatial and temporal distributions of indicators are affected by various climatic, geographical and ecological factors[1]. Statistically derived water quality models, assume normal distribution and linear relationship between response variables and prediction variables whereas artificial neural networks (ANN) are able to map the non-linear relationships among the variables that are characteristic of ecosystems[2].

During the recent years, the Mamdani FIS, using fuzzy set mathematical methodology, has been easily accepted by both researchers and decision makers due to its ability to handle the uncertainties in geoscience and water resources. In this study, the aim was also to develop the eutrophication index (EI) based on fuzzy logic instead of the conventional crisp classification method to remove the ambiguities in water quality

variables. In this method, the membership functions of the quality parameters and fuzzy rule bases were defined and then fuzzy logic toolbox of MATLAB R2012a package was used.

ANN is used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system. The ANN approach has several advantages over traditional phenomenological or semi-empirical models since they require known input data sets without any assumptions[3]. ANN develops a mapping of the input and output variables, which can subsequently be used to predict the desired output as a function of suitable inputs. A multi-layer neural network can approximate any smooth, measurable function between input and output vectors by selecting a suitable set of connecting weights and transfer functions[3]. ANN models have been widely applied to water quality problems[4,5,6,7,8]. The disadvantages of the ANN approach include its black-box nature, proneness to over fitting and the empirical nature of model development.

This paper demonstrates the application of ANN to model the eutrophication status of sea water, having the dynamic and complex processes hidden in the monitored data itself. The ANN model can reveal the hidden relationships in the sampling data, thus facilitating the prediction and forecasting of seawater eutrophication status.

The steps followed in the development of such models include the choice of performance criteria, division and pre-processing of the available data, determination of the appropriate model inputs and network architecture, optimization of the connection weights (training), and model validation. In this paper, a study of ANN modelling to predict and forecast eutrophication status based on the measurement of water quality parameters, e.g., CDOM concentration, dissolved oxygen (DO) content, and Chlorophyll – a (Chl-a) in Mumbai coastal waters is presented. These water quality parameters were measured monthly at various locations. These models could be used as a prediction tool, which complements the process-based model and field monitoring programme.

The main aim of the present work was to construct a principal component analysis (PCA) model and a Mamdani fuzzy model, and to compare their outputs with two ANN models to predict the eutrophication status in Mumbai coastal waters and demonstrate its application/ utility in improving the interpretation of the results and in identifying complex nonlinear relationships between input and output while dealing with the complex water quality data. Here, we have investigated the possibility of training ANN models correlating the primary water quality variables (independent) with their secondary attributes (dependent variables). The EI of the water was taken as the dependent variable here and the set of other parameters constituted the independent variables.

In this study, ANN models have been identified for computing EI. EI obtained from the rule-based fuzzy approach (Mamdani method) and the multiple regressions derived from the first principal component were selected as target variables for the ANN model in training and prediction. The main objective of the study was to assess and quantify the eutrophication status in shallow and turbid waters of Mumbai coastal area using soft computing and statistical techniques. Average depth of the study area varies from 8 to 15 m.

Materials and Methods

Mumbai lies on the west coast of India, spread across 437.77 km^2. The population of Mumbai, as per Census 2011, was 18.4 million, projected to reach 28.5 million by 2020. Mumbai used to be a group of seven islands in the Arabian Sea which lies off the northern Konkan coast on the west of the state of Maharashtra. These seven islands, which were once separated by creeks and channels were filled and bridged over the years by the inhabitants. Most of the year, Mumbai's climate is warm and humid. Between November and February, the skies are clear and the temperature is cooler. From March, the temperature becomes warm and humid till mid-June which is the beginning of the monsoon. During monsoon, there are torrential rains which sometimes cause flooding of major roads and streets of Mumbai. The average annual rainfall which is brought by the south-west monsoon winds in Mumbai is 180 cm. Monsoon ends by the end of September. October is comparatively hot and humid. The average annual rainfall is 1917.3 mm. About 94% of annual rainfall of the Greater Mumbai region is received during the south-west monsoon months of June-September. The area is rich in natural resources and being a part of the Western Ghats of India, offers rich biological diversity. It is a part of the upwelling belt of the west coast which has high primary

productivity and immense potential of fisheries. It is also important from the environmental point of view as it supports a vast area of mangrove forest besides the terrestrial flora and fauna in certain stretches. The coastal area is highly turbid, and is important for fisheries and recreation.

The study area (Fig. 1) has major creeks and river systems. The study area extends from $72^0$45' E to $73^0$0' E and $18^0$50' N to $19^0$15' N. The total coastal length of Mumbai is 140 km and in-situ data were collected from eight fixed sampling stations with varied ecological characteristics proximally on a monthly basis. The statistical summary of water quality variables is shown in Table 1.

Estimation of coloured dissolved organic matter

The concentration of CDOM was determined by its effect on light absorption by water. Many previous studies have shown that this optical approach has distinct advantages over analytical chemical techniques[9].Using optical methods; the concentration of CDOM in water is expressed in terms of its attenuation or absorption coefficient at a given wavelength in the UV or visible region (380- 440 nm). Water samples were collected once per month at selected stations with *MV Narmada* (boat) owned by the Central Institute of Fisheries Education during the pre-monsoon and post-monsoon periods (from May 2011 to January 2012). Samples were collected in 200-ml amber glass bottles that had been rinsed three times with the sample water before filling.

Table 1 – Range of water variables in study area during the sampling periods

Variable	Minimum	Maximum	Mean	Std. Deviation
Transparency (cm)	20	167	69.46	32.6
Air Temp. (^{0}C)	23	34.0	29.18	2.70
Water Temp. (^{0}C)	22	33.20	28.12	2.63
Salinity (‰)	29	36	34.7	2.21
pH	7.20	8.79	7.8222	0.390
DO (mg/l)	0.80	6.88	4.7841	1.547
Conductivity (mS/cm)	8.4	17.40	1.30	3.19
Available P (mg/l)	0.002	1.52	0.1721	0.235
Total P (mg/l)	0.024	3.80	0.5049	0.644
Ammonia-Nitrogen (mg/l)	0.0019	0.08	0.0260	0.016
Nitrate- nitrogen (mg/l)	0.0025	3.60	1.5643	0.893
Chl - a (µg l^{-1})	0.10	28.07	5.1493	3.568
TSS (mg/l)	1.33	252.80	61.828	50.998
$a_{CDOM440}$ (m^{-1})	0.09	2.72	1.5280	0.444

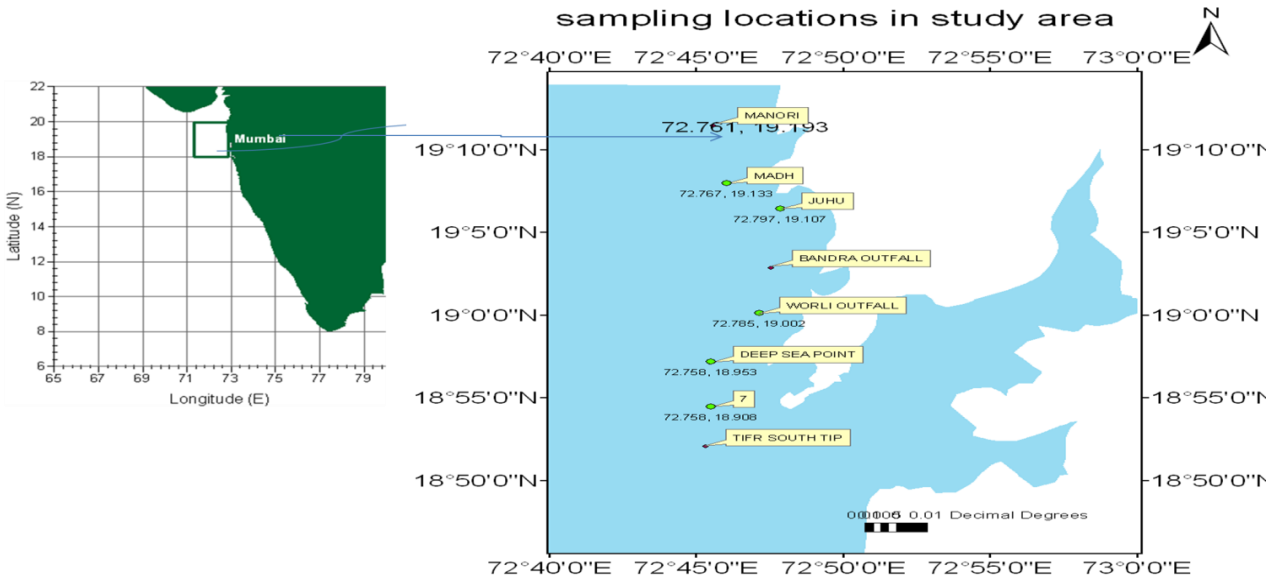


Fig. 1. Location map of study area

Water samples were filtered onboard through 0.2-µm cellulose acetate membrane filters; material remaining in the water was considered to be dissolved humic material and 0.4 ml of 0.5 M $HgCl_2$ was added to 200 ml of sample to avoid any bacterial degradation. The sample was then preserved at low temperature until analysis in the laboratory. The sample transparency was measured with a UV VIS spectrophotometer (Biochrom, Libra 32pc) over the spectral range 200 to 800 nm with an interval of 1 nm against Milli-Q water as blank. The spectral absorption coefficient was calculated by normalizing with respect to 440 nm[10].

$$C^*_f(\lambda) = C_f(\lambda) - C_R(\lambda) \qquad (1)$$

Where $C_f(\lambda)$ is the beam attenuation coefficient of the filtered water at wavelength (λ) and $C_R(\lambda)$ is the beam attenuation coefficient of distilled water at wavelength (λ) (equation 1) . The value of C*f(λ) is not identical to the absorption coefficient of CDOM (a_{CDOM}) because very small particles (colloids) may pass through the filter. To obtain the true values of the spectra of a_{CDOM} the following correction was made[9] shown in equation 2:

$$a_{CDOM}(\lambda) = C^*_f(\lambda) - C^*_f(\lambda_R)\,(\lambda_R/\lambda)^g \qquad (2)$$

Where $C^*_f(\lambda)$ and $C^*_f(\lambda_R)$ were obtained from spectrophotometer at wavelength at λ and reference wavelength (λ_R) respectively, and g is a parameter describing the contribution of scattering by colloids to $C^*_f(\lambda)$. Different publications [11] have used *g* values equal to 0, 1 and 2. We set λ_R to a value of 700 nm and *g* to a value of 1[(12)]. Most of the tested algorithms generated CDOM absorption coefficients aCDOM(λ) at 440 or 443 nm, which are widely accepted as the proxy of CDOM content . To describe the absorption coefficient of CDOM in study area, we set the wavelength λ to a value of 440 nm in the computation of a_{CDOM}.

Estimation of water quality parameters

The geographic locations of the sampling points were recorded using GPS. The parameters analyzed included secchi disk depth (SDD)/ transparency, pH, electrical conductivity (EC), salinity, total suspended solids (TSS), temperature, DO, , total carbon, total nitrogen, total phosphorus, available phosphorus, nitrate-nitrogen, ammonia-nitrogen and pigment concentration in water samples, and total carbon, total nitrogen, pH and EC in sediment samples. The analyses of the samples were done as per standard procedures[13]. SDD was determined as per the procedure[14]. DO was measured immediately after collection following the Winkler procedure. CDOM index was obtained after processing the OCM-2 images with SeaDAS software (6.4). CDOM index was used for comparing the CDOM (qualitatively) in the study area during pre- monsoon and post- monsoon.

Statistical and Soft Computing Techniques

EI can be calculated from the variables of physical parameters namely, SDD, temperature (water and air); chemical parameters namely, nutrients (nitrate, nitrite, ammonia, total and available phosphorus), DO, salinity; biological parameters namely, Chl-a; and optical property namely, aCDOM (440 nm). These variables can be analyzed using multivariate techniques, e.g., PCA, to produce an indicator on the trophic state of the water body. Being multidimensional, multivariate technique is an appropriate approach for developing the index due to the fact that it incorporates several parameters which have influence on eutrophication, to classify a trophic state of the water body.

The method of PCA as proposed by Primpas et al. (2009)[15] was implemented in this study for the development of the multivariate index for assessing eutrophication. Eutrophication Analysis was performed using the software SPSS 16. 0. The principal components (PC) can be expressed in equation (3) as:

$$Z_{ij}=a_{i1}x_{1j} + a_{i2}x_{2j}+ a_{i3}x_{3j} + ... + a_{im}x_{mj} \quad (3)$$

Where Z is the component score, a is the component loading, x is the measured value of variable, i is the component number, j is the sample number and m is the total number of variables. The PCs generated by PCA are sometimes not readily interpreted; therefore, it is advisable to rotate the PCs by varimax rotation. Varimax rotation ensures that each variable is maximally correlated with only one PC and a near-zero association with the other components[16, 17]. Varimax rotations applied on the PCs with eigenvalues more than 1 are considered significant[18] where the typical criteria are 75-95% of total variance[19]. The rotations were carried out in order to obtain new groups of variables. The variables with communality greater than 0.7 are considered, having significant factor loadings[20].

Mamdani fuzzy model

The process of fuzzy inference consists of membership functions, logical operations, and if-then rules. Mamdani fuzzy model is based on the collections of if-then rules with both fuzzy antecedent and consequent parameters. It is also called a linguistic model because both the antecedent and the consequent are fuzzy propositions. Mamdani fuzzy model due to its popularity and easy application is the most commonly seen fuzzy methodology. Mamdani model can be built by using linguistic relationships and observed data. The Mamdani-based fuzzy models use excessive number of rules for system modelling. The first step is to consider the inputs and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions.

In the fuzzy logic toolbox software, the input is always a crisp numerical value limited to the universe of discourse of the input variable (in this case, the interval between 0 and 10) and the output is a fuzzy degree of membership in the qualifying linguistic set (always the interval between 0 and 1). Fuzzification of the input amounts to either a table lookup or a function evaluation. For eutrophication modelling, three input membership functions and one output membership function were selected. The inputs were CDOM, Chl-a and DO for coastal system. The output was EI. Each input and output was divided further into low, average and high based on the concentrations and the range of inputs. The membership functions selected for eutrophication modelling were Gaussian as value of variable was non-zero at all points. In the membership function editor of Matlab, "Range" displayed is the range of the input which is generally the range between minimum and maximum of the parameter value. The "Params" describes the standard deviation and mean of the membership function which can be calculated manually using the existing data.

After the inputs are fuzzified, the degree to which each part of the antecedent is satisfied for each rule is determined. If the antecedent of a given rule has more than one part, the fuzzy operator was applied to obtain one number that represents the result of the antecedent for that rule. This number was then applied to the output function. The input to the fuzzy operator was two or more membership values from fuzzified input variables. The output was a single truth value. The logical operators that we used for our analysis were Fuzzy And. For And operator, output map was controlled by the smallest fuzzy membership value. Hence, it was pessimistic in nature. The rules were applied for eutrophication potential modelling.

The input for the implication process was a single number given by the antecedent, and the output was a fuzzy set. Implication was implemented for each rule. Two built-in methods were supported, and they were the same functions

that are used by the And method: min (minimum), which truncates the output fuzzy, set and prod (product), which scales the output fuzzy set. Eutrophication modelling was done by the minimum implication function as to be more pessimistic in approach and to get the areas which are highly suitable.

Decisions were based on the testing of all of the rules in a FIS; the rules must be combined in some manner in order to make a decision. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set. Aggregation only occurs once for each output variable, just prior to the fifth and final step, defuzzification. The input of the aggregation process was the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process was one fuzzy set for each output variable. Three built-in methods are supported in Matlab. They are max (maximum), probor (probabilistic or), sum (simply the sum of each rule's output set). Potential modelling prefers max operator as it gives the output for the larger area that is defuzzified.

The input for the defuzzification process is a fuzzy set (the aggregate output fuzzy set) and the output is a single number. As much as fuzziness helps rule evaluation during the intermediate steps, the final desired output for each variable is generally a single number. However, the aggregate of a fuzzy set encompasses a range of output values and so must be defuzzified in order to resolve a single output value from the set. The most popular defuzzification method is the centroid calculation, which returns the centre of area under the curve and eutrophication potential modelling uses the same method for defuzzification.

Artificial neural network

A neural network modelling may be appropriate to simulate eutrophication in the Mumbai coastal water due to following reason: it has limited data on nutrient inflows. The primary source of inflow is precipitation in the monsoon season from June to September. Continuous sewage drainage and sewage outfall inside the coastal water contribute to significant nutrient loads. Considering a number of complicating factors such as limited inflow data, large weather uncertainty and excessive watershed development, a neural network model may be appropriate to simulate eutrophication.

To determine the non-linear relationships between the water quality factors and eutrophication, an ANN model based on back-propagation training algorithm was chosen for the investigation. The ANN was structured such that the data measurements fed to the input layer, an eutrophication indicator is represented in the output layers. The input and the output layers are interconnected via a hidden layer consisting of neurons. The back-propagation training algorithm is applied to the ANN model to establish the relationship between the input and output through the data collected over months in the coastal water. Once trained, the connection weights in the ANN were fixed and the model was validated by assessing its predictive performance on a testing set of data excluded from the training set. Back-propagation neural network is popular because of its broad applicability to many problem domains such as principal prediction, classification and modelling. According to a supervised learning technique in the process, this neural network requires a set of training data in order to learn the relationships among data.

The back-propagation neural network architecture consists of two or more layers of neurons connected by weights. The information is captured by the network when input data pass through the hidden layer of neurons to the output layer. The weights connecting from neuron i to neuron j are denoted as wji. Each neuron calculates its output based on the amount of stimulation it receives from the given input vector x_i; x_i is the input of neuron I (Equation 4). The net input of a neuron is calculated as the weighted sum of its inputs and the output of the neuron is based on some active function, which indicates the magnitude of this net input. So, the net output u_i from a neuron can be indicated as equation 4:

$$u_j = \sum_{i=1}^{p} Wji\, Xi \qquad (4)$$

In this model, sigmoid function is chosen as its active function.

The general form of sigmoid function is equation 5:

$$\varphi(u_j) = 1/(1 + \exp(-u_j)) \qquad (5)$$

$$y_j = \varphi(u_j) \qquad (6)$$

where, y_j is the output of the j^{th} neuron in any layer equation 6.

For a given input set the network produces an output, and this response is compared to the known desired response of each neuron. The weights of the network are then changed to correct or reduce the error between the output of the neuron and the desired response, and this process goes on. The weights are continually changed until the total error of all training sets is reduced below the acceptable error or other stop mechanism. The back-propagation neural network, with an algorithm for determining the optimal weights for a given training set of data, has the improved potential of the function approximation when it is learning highly complex and non-linear data because of the increased number of the hidden layer or the neurons in the hidden layers. New weights are calculated by adding a correction to the old weights.

The parameters of ANNs

There are a number of key parameters in back-propagation neural network. First, the number of input units should be determined. The initial values of the weights are assigned randomly based on an input random number seed. The initial values of the weights are determined by a random starting, and randomly set between −0.1 and +0.1.This range is chosen where all the models created are able to run. All neural network models used in this study were composed of three layers with nodes in adjacent layers fully connected. That is, only one hidden layer was employed. Since the focus was on single variable forecasting, one output node was exclusively used in the output layer. The number of input nodes and the number of hidden nodes were the two major experimental factors. To speed up the training, or learning, and convergence of the weight values, the learning rate was adjustably made. It was necessary to have a learning rate that is small enough to converge but large enough that the computing time is reasonable. And it is better not to use a constant learning rate because the error will not smoothly converge to a minimum. Therefore, we selected the globally adaptive learning rate, where the learning rate is incremented or decremented by a small amount with each weight updated, depending on whether the error decreased or increased. As training progressed, weights were modified to minimize the error in the output. At the end of training, the inputs that were most important in prediction had the largest weights while those that were less important had lower weights.

Prediction models for EI with PCA-EI and fuzzy EI as target were built individually. Studies have shown that the cause and effect of eutrophication, and potential mitigating actions are complex for coastal water. Using the data reduction technique, we pre-screened the potential input variables. The field data were divided into training and testing sets randomly. The training set utilised three-fourths of the field data and the testing set one-fourth. According to these processes, the EI model was built by the input of DO, Chl-a and CDOM absorption coefficient.

Results and Discussion

Distribution of CDOM

The distribution of CDOM index in the study area during pre-monsoon and post-monsoon (Fig. 2) indicates that during post-monsoon, CDOM concentration is higher as compared to that in pre-monsoon. While analyzing the available data from OCM-2 sensor, it was observed that along with TSM, CDOM index is influencing the euphotic depth in the study area in post-monsoon. Monsoon causes terrestrial drainage of organic matter and sediment material to coastal water. CDOM being colloidal in nature remains suspended and affects the euphotic depth. Dissolved organic matter contributes more nutrients to the system during post-monsoon.

Carbon and nitrogen ratio

Carbon and nitrogen ratio in the sediment of the study area was varying from 11.79 to 17.39 (Table 2.) In the coastal sediment, carbon content is in the higher range due to the higher content of inorganic carbonate. Nitrogen content is lower as it is of organic origin. The lower range of C:N ratio indicates the faster mineralisation of organic matter which results in nutrient release and increase in CDOM content in the study area. Thus, the higher content of CDOM indicates the eutrophic condition of the study area.

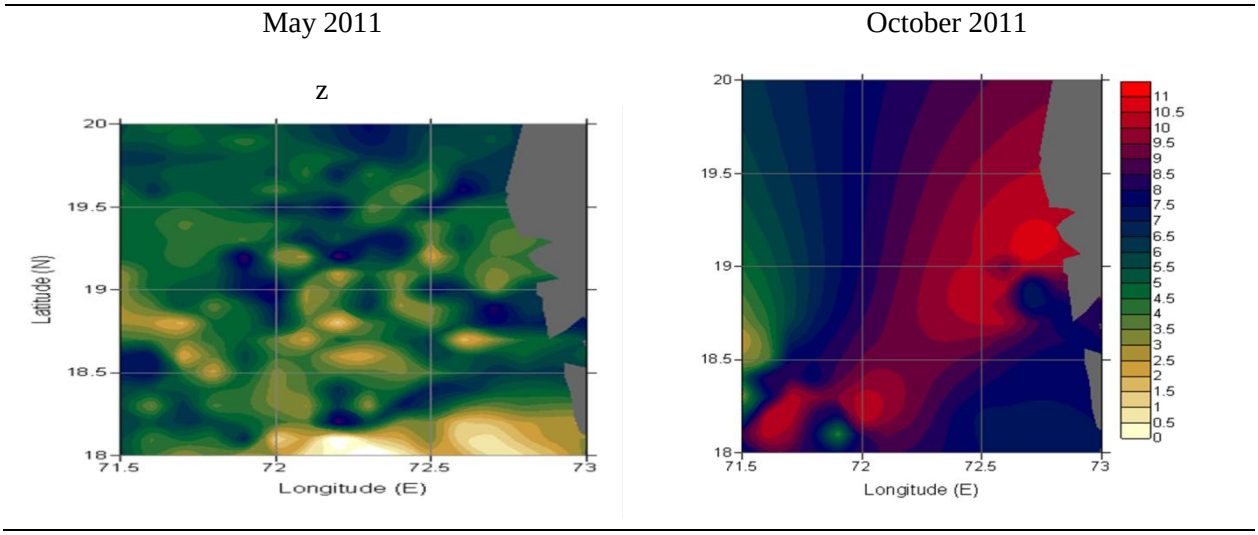


Fig. 2. Distribution of CDOM index in the study area

Table 2**.** Distribution (%) of carbon, nitrogen and sulphur in study area sediment

	C	N	S	C:N ratio
Station 1	2.2568	0.1387	1.0281	16.2679
Station 2	2.1422	0.1403	0.7442	15.2637
Station 3	2.1199	0.1628	0.5007	13.0182
Station 4	2.1794	0.1848	0.4453	11.7955
Station 5	3.5959	0.2931	1.1572	12.2690
Station 6	1.9862	0.1187	0.5627	16.7343
Station 7	2.0110	0.1241	0.4423	16.2014
Station 8	1.8535	0.1284	0.3820	14.4361

Eutrophication index

The totals of 240 out of 300 in-situ coastal water quality measurements were used to perform the PCA because after visual inspections, it was realized that some data were not given fixed values and they were unfit for analysis using PCA. The correlation matrix was further used to apply PCA. It was observed that the first component accounts for more than 20% of the whole variation (Table 3), whereas the other four components explain the parts of the remaining variation from 19 to 9 %. Therefore, a remarkable dimensional reduction was achieved, if the information from the first component was used. The correlation matrix of the three variables after standardization are given in (Table 4). It was observed that all coefficients are positive and three variables participate with weights varying from 64 to 81% to the formation of the first principal component and therefore, to the proposed EI having the formula (Equation 7). The coefficients of the first principal component for the three variables were extracted as shown in Table 5.

Table 3. Results of the PCA applied on variables from the study area

Principal component	Eigenvalue	% variance explained
1	2.234	20.313
2	2.118	19.252
3	1.723	15.663
4	1.138	10.344
5	1.020	9.272

Table 4. Lower triangular Pearson correlation matrix of the variables used in PCA analysis

Variable	DO	Chl-a	Acdom
DO	1		
Chl-a	0.419**	1	0.256
Acdom	0.430**	0.256	1

**. Correlation is statistically significant at the 0.01 level (2-tailed).
*. Correlation is statistically significant at the 0.05 level (2-tailed).

Table 5. Coefficients of the first principal component for the three variables in coastal system

Variable	Coefficient
DO	0.813
Chlorophyll - a	0.773
CDOM	0.640

Therefore, the EI derived from in-situ measurements as per the equation will be as follows:

$$\mathbf{EI_{Coastal} = 0.813C_{DO} + 0.640\ C_{CDOM} + 0.773\ C_{Chl\text{-}a}} \quad (7)$$

Using Equation 7, EIs were calculated with respect to their corresponding in-situ water quality parameters and are summarized in Table 6. EIs were varying from 3.7 to 28.7. The higher value of EI at Station 6 was supported by higher DO value and Chlorophyll - a content. In the study area, the growth rate of phytoplankton was influenced by sunlight, water temperature, salinity and nutrients. The eutrophication model is related to month, water temperature, salinity, pH, DO, SDD, TP, TSS, NO_3, NH_4, Avail.PO_4 and Chl-a. Month, SDD, CDOM and TSS would influence the sunlight intensity in water. Water temperature and DO may indicate how much oxygen is produced by eutrophic water. The pH variable may result from algal production because photosynthesis is a chemical reaction and pH is a major factor which controls the chemical reaction speed. After monsoon, the CDOM, SDD and nutrient contents increased in the study area which led to the phenomenon of eutrophication in the month of October.

Fuzzy logic based eutrophication index

In the Mamdani method of fuzzy logic output membership function is given by the modeler. The output membership function of EI is Gaussian curve as discussed in the methodology and plotted with a range (3.7 to 15.0) of EI derived by the principal component method. The standard deviation of the curve is assumed as 1.919 (Fig. 3). The mean of low EI membership function is 3.70, for medium, it is 9.35 and for high, it is 15.00. The outputs were analyzed using the rule viewer in the view menu (Fig. 4). The output of Juhu station is given in Fig. 5. The input given is in the form of [6.4 2.882 10] which are the concentrations of DO, CDOM absorption coefficient and Chl-a, respectively. The eutrophic Index of Juhu station was 13.3/15,00 (Fig. 4) which was the highest in the study supported by the high concentrations of Chl-a and DO which indicate presence of bloom. Similarly, the EI for each station was determined by this method. The results are presented in Table 7.

Table 6. Eutrophication Index of coastal water derived through PCA method

	9th May 2011	19th May 2011	10th 0ct. 2011	24th Nov. 2011	30th Jan. 2012
Station 1	3.7	6.0	6.9	9.2	7.3
Station 2	7.1	5.6	13.2	9.2	7.4
Station 3	9.7	6.2	13.3	7.2	9.6
Station 4	11.4	6.7	17.0	8.6	8.7
Station 5	8.1	4.5	15.5	11.2	10.6
Station 6	28.7	5.8	14.7	8.6	14.7
Station 7	7.5	6.8	14.3	8.4	11.8
Station 8	5.5	7.8	15.2	8.6	8.2

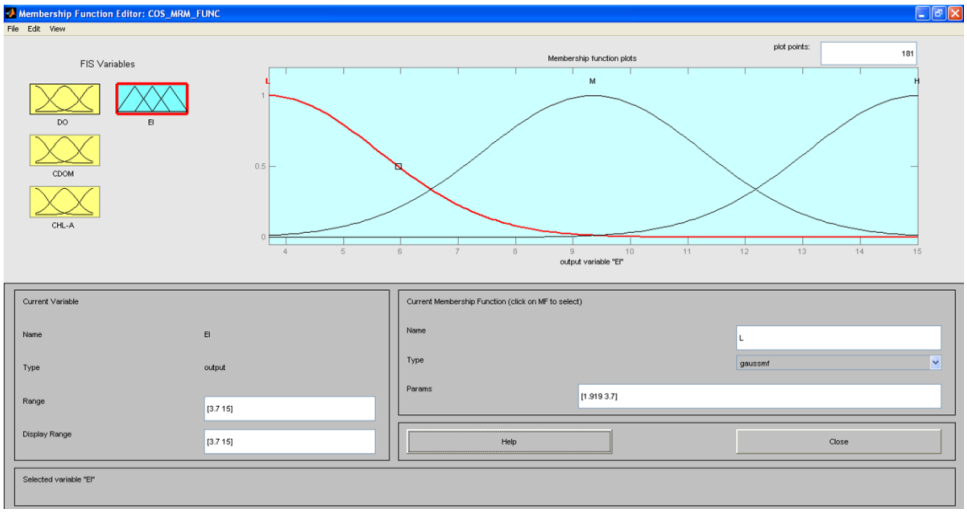


Fig. 3: Output membership function of low eutrophication potential

Juhu station had the highest EI in May (Table 7). January is the ideal month for bloom occurrence as reflected by the EI of coastal waters though it was neither visually observed nor remotely sensed as the CDOM content of the area was high. The result shows that the eutrophication problem is critical mostly during the month October, and that the method can efficiently capture the rapid changes in trophic states, i.e., critical periods for eutrophication in coastal water are autumn and spring as has already been reported[21].

The basic structure of ANN usually consists of three distinctive layers: the input layers, where the data are introduced to the ANN, the hidden layer, where data are processed, and the output layer, where the results of ANN are produced. The structure and operation of ANNs are discussed by a number of authors [22, 4, 23, 24]. The ANN is designed by putting weights between neurons, by using a transfer function that controls the generation of the output in a neuron and using adjustable laws that define the relative importance of weights for input to a neuron. In the training, the ANN defines the importance of the weights and adjusts them through an iterative process. The ANN models were developed to simulate EI. It uses the back propagation (BP) algorithm with two hidden layers and sigmoid activation functions. Three water quality parameters (DO, Chlorophyll - a and CDOM) loaded on the first principal component under the dimensional reduction method, were selected as input for ANN modelling. EI obtained from the rule-based fuzzy approach (Mamdani method) and multiple regressions derived from the first

principal component were selected as the target variables for the ANN model in training and prediction. The performance of the ANN model was compared in both the cases. The developed ANN models accurately simulated the EI. Typical ANN EI prediction model results are shown as a scatter diagram of computed and measured EI derived by the PCA and fuzzy approaches for training, validating and testing data sets in figures 5 and 6.The EI predictions of this model have positive, very high correlation (R=0.96) with the measured EI values derived from the fuzzy approach as compared to the PCA-derived EI (R=0.851) implying that the model predictions explain around 93.7% (R^2 = 0.937) of the variation in the measured EI values derived by fuzzy approach as compared to 72.4% (R^2 = 0.724) in the case of PCA-derived measured value.

Conclusions

The study area has significant spatial variability of CDOM and CN ratio between the stations with sewage outfall (stations 4 and 5) and chemical industry outfall (stations 1 and 2). CDOM is an indicator of eutrophication. The fuzzy logic method (Mamdani) of Eutrophication index (EI) estimation is smoother than the Principal Component Analysis (PCA) method. EI predicted through ANN taking PCA-derived EI as the target variable has R value of 0.851, whereas EI predicted through ANN taking the fuzzy approach derived EI as target variable has R value of 0.968. The fuzzy method of EI estimation is better as it includes the modeller view of eutrophication. It has already been found that the multiple regression principle of linear modelling often gives low performance when relationships between variables are nonlinear. On the other hand, netural networks are non-linear type models. These do not necessitate the transformation of variables and can give better results. We have also found that the PCA-derived EI model of ANN provides low accuracy. The ANN models can preserve the non-linear characteristics between the input and output variables, and are superior to the traditional statistical models. The ANN method restores the non-linear relationship among the water quality parameters. ANN models have been developed to predict the EI in Mumbai coastal waters, both spatially and temporally, using the measurements of water quality variables at different stations over time periods.

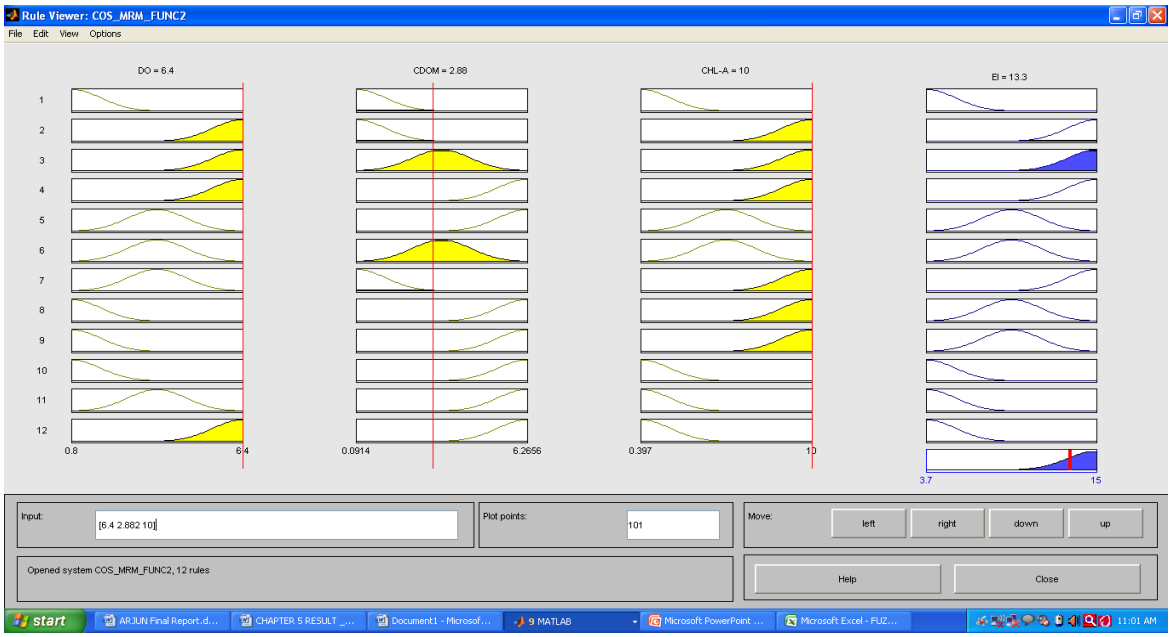
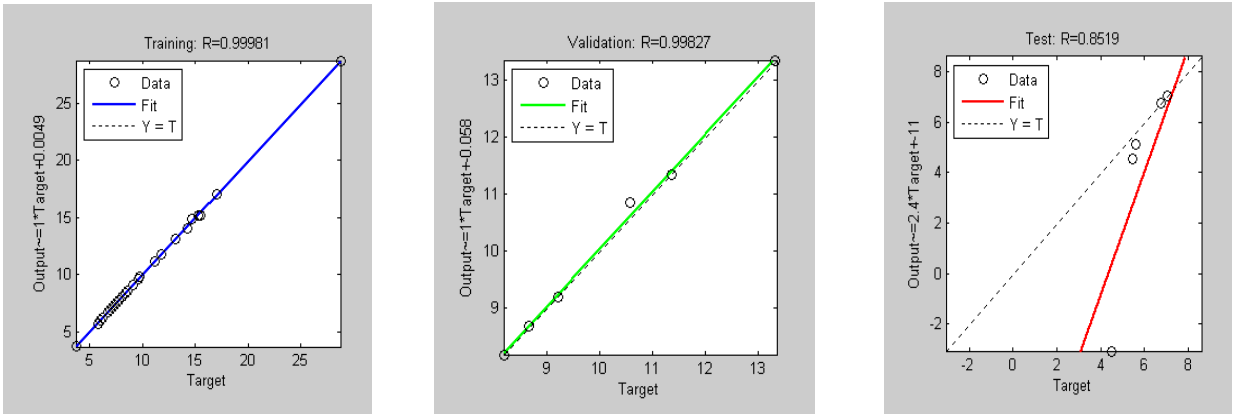


Fig. 4: Eutrophication index determination of Juhu station

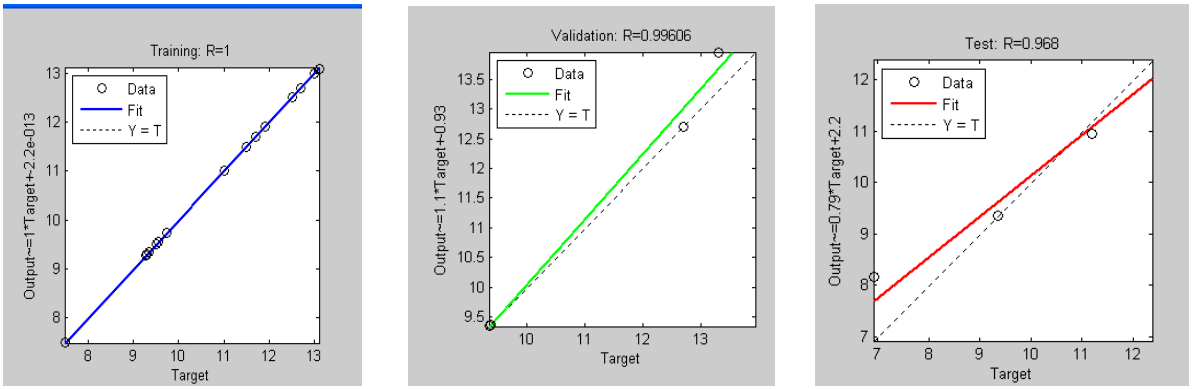
Table 7 : Eutrophication Index derived through the fuzzy method

	9th May 2011	19th May 2011	10th 0ct. 2011	24th Nov. 2011	30th Jan. 2012
Colaba	7.49	9.31	9.35	9.35	11
TIFR	9.35	9.27	9.56	9.35	9.35
Deep sea	11.2	9.34	9.74	9.34	13
Worli	9.35	9.35	12.7	9.35	12.2
Bandra	9.36	6.92	11.7	9.52	13
Juhu	13.3	9.27	11.5	9.35	13.1
Madh	9.35	9.35	11.9	9.35	12.7
Manori	9.35	9.35	12.5	9.35	9.35



MSE=0.009 and R=0.99 MSE=0.013 and R=0.98 MSE=0.976 and R=0.86

Fig. 5. Scatter diagram of predicted versus observed EI (derived from PCA) for training, validating and testing data sets



MSE=0.00001 and R=1 MSE=0.06 and R=0.99 MSE=0.27 and R=0.96

Fig. 6. Scatter diagram of predicted versus observed EI (derived by fuzzy approach) for training, validating and testing data sets

In spite of the largely unknown factors controlling seawater quality variation and the limited data set size, a relatively good correlation was observed between the measured and predicted values of EI. The limitations of this study include its limited data set. The lack of fit between the observed and estimated data indicates that new patterns must be incorporated into the model, and thus, the model should be recalibrated and revalidated as more data are collected. Even though the available data size was relatively small, reasonably good results were obtained for the water quality prediction of unseen validation or testing dataset from the training dataset stations. If more data become available, the proposed ANN approach should provide better predictions.

Acknowledgment

Authors are grateful to the Director, CIFE for support and facilities to carry out the work.

References

1. Kurunça, A., Yüreklia, K. and Çevikb, O., Performance of two stochastic approaches for forecasting water quality and stream flow data from Yesilurmak River, Turkey. *Environ Modell Softw*, 20 (2005) 1195-1200.
2. Lek, S. and Guegan, J.F., Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol Model*, 120 (1999) 65-73.
3. Gardner, M.W. and Dorling, S.R., Artificial neural networks (The multilayer perception) - A review of applications in atmospheric sciences. *Atmos Environ* , 32 (1998), 2627-2636
4. Dowla, U.F., Rogers, L.L., *Solving Problems in Environmental Engineering and Geosciences with Artificial Neural Networks*, (MIT press, MS, USA) 1995, pp. 249
5. Raman, H. and Chandramouli, V., Deriving a general operating policy for reservoirs using neural networks. *J. water resour. plan. manage*, 122 (1996), 342-347.
6. Wen, C.W. and Lee, C.S., A neural network approach to multiobjective optimization for water quality management in a river basin. *Water Rresour Res*, 34 (1998), 427-436.
7. Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. and Aulagnier, S., Application of neural networks to modelling nonlinear relationships in ecology. *Ecol Model*, 90 (1996), 39-52.
8. Bowers, J.A. and Shedrow, C.B., Predicting stream water quality using artificial neural networks.WSRC-MS-2000-00112. http:// www.osti.gov/bridge/.
9. Bricaud, A., Morel, A. and Prieur, L., Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnol Oceanogr* 26 (1981), 43-53.
10. Kowalczuk, P. and Kaczmark, S., Analysis of temporal and spatial variability of yellow substance absorption in the southern Baltic. *Oceanologia* , 38 (1996), 3–32.
11. Arst, H., *Optical Properties and Remote Sensing of Multicomponental Water Bodies*, (Springer Praxis Publishing, Chichester U.K) 2003, pp.231.
12. Davies-Colley, R.J. and Vant W.N., Absorption of light by yellow substance in freshwater lakes. *Limnol and Oceanogr* 32(1987), 416- 425.
13. APHA, *Standard methods for the examination of water and wastewater*, 21st Ed. American Public Health Association, Washington, DC (2005).
14. Saxena, M.M., *Environmental Analysis Water, Soil and Air*, (2nd Edn: Agro Botanica Bikaner) 1998, pp. 24-26.
15. Primpas, I. Primpas, I., Tsirtsis, G., Karydis, M. and , Giorgos, D.K., Principal component analysis: Development of a multivariate index for assessing eutrophication according to the European water framework directive. *Ecol Indic*, 10 (2009), 178-183
16. Abdul-Wahab, S.A., Bakheit, C.S. and Al-Alawi, S.M., Principal component and multiple regression analysis in modeling of ground-level ozone and factors affecting its concentrations. *Environ Modell Softw*, 20 (2005), 1263-1271.
17. Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M. and Pereira, M.C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, *Environ Modell Softw*, 22 (2007), 97-103.
18. Kim, J.-O. and Mueller, C.W., *Introduction to Factor Analysis: What it is and how to do it. Quantitative Applications in the Social Sciences Series*, (Sage University Press. Newbury Park) 1987, pp. 80.
19. Chen, Q. and Mynett, A.E., Integration of data mining techniques and heuristic knowledge in fuzzy logic modeling of eutrophication in Taihu Lake. *Ecol Model* , 162 (2003), 55-67.
20. Stevens, J., *Applied Multivariate Statistics for the Social Science*. (Hillsdale: Erlbaum) 1986, p.515.
21. Taheriyoun, M. Karamouz and M. Baghvand, A., Development of an entropy-based fuzzy eutrophication index for reservoir water quality evaluation. *Iranian J Environ Health Sci Eng*, 7 (2010), 1-14.
22. Fausett, L., Fundamentals of neural networks: architectures Algorithms and Applications, (Prentice Hall, Englewood Cliffs, N.J, USA), 1994.
23. Patterson, D., *Artificial Neural Networks*. (Singapore: Prentice Hall), 1996.
24. Gurney, K., *An Introduction to Neural Networks*, (Prentice Hall, UK) 1999.