Research Paper
Volume : 2 | Issue : 6 | June 2013 • ISSN No 2277 - 8179

Modelling of Chlorophyll-A Concentrations in Pulicat Lagoon, Southeast Coast of India Using Artificial Neural Network

Engineering

KEYWORDS : ANN Model, Chlorophyll-a, Pulicat Lagoon, Primary Production.

Mageshkumar P.	Assistant Professor in Civil Engineering, K.S.Rangasamy College of Technology, Namakkal–637215, Tamilnadu, India
Harini Santhanam	Post Doctoral Fellow, Centre for Earth Sciences, Indian Institute of Science, Bangalore - 560012, India
Stalin John M.R.	Associate Professor, SRM University, Chennai 603203, India
Amal Raj S.	Associate Professor, Centre for Environmental Studies, Anna University Chennai 600025, India

ABSTRACT

Formations of algal blooms increasingly pollute both salt and fresh water ecosystems throughout the world. Owing to its negative impacts on human health and aquatic life, this widely reported phenomenon has become a serious environmental problem. While many process-based, statistical and empirical models exist for water quality prediction, Artificial Neural Network (ANN) models are increasingly being used for water related applications as they are often capable of modelling complex systems for which behavioural rules or underlying physical processes are either unknown or difficult to simulate. In the present study, a feed forward neural network is proposed to model the primary productivity of Pulicat lagoon. The commonly used back propagation learning algorithm has been employed for training the ANN. The model was constructed using five years of seasonal data set on the mouth part of Pulicat lagoon which is the most dynamic part of the lagoon. Despite the very complex and peculiar nature of this region of the lagoon, a very good correlation (R = 0.998) was observed between the measured and predicted values during model validation. The Mean Square Error between the measured and predicted values was found to be 0.018. Thus, the resulting prediction of Chlorophyll-a values clearly indicated that ANNs can fit the complexity and nonlinearity of ecological phenomena such as phytoplankton production to a high degree.

INTRODUCTION

Coastal zone is a dynamic area with many cyclic processes owing to a variety of resources and habitats. Coastal lagoons represent a tiny part (less than 1%) of the surface covered by oceans, but they are nonethe¬less characterised by high biodiversity and intense primary productions, that lead to both ecological and economical considerable importance. Since algae are at the bottom of the food chain in lake ecosystems, their responses can mostly be attributed to physical and chemical changes [5]. Estimates of phytoplankton primary production based on empirical models are increasingly used as an alternative to direct data acquisition. Conventional models that attempted to address this problem by means of multiple linear regression, or the use of semi-analytic formulations [1], did not perform significantly better than much simpler empirical models. An alternative approach, involving the use of neural networks has recently generated significant improvement in estimating production [12] or other complex non-linear ecological processes [7] where sufficient training data were available. In this paper, an attempt has been made to develop a reliable modelling tool for Pulicat lagoon using ANN to calculate the concentrations of chlorophyll-a; the primary photosynthetic pigment in the water column of Pulicat lagoon.

1.1 Artificial Neural Networks

The main processes that determine phytoplanktonic production can be approximated by linear or simple non-linear functions only to a limited extent [6]. Therefore, such models are not able to reproduce the behaviour of real systems when very low or high values of the independent variables are considered. ANNs imitate the learning process of the animal brain and can process problems involving very nonlinear and complex data even if the data are imprecise. Thus they are ideally suited for the modelling of ecological data which are known to be very complex and often non-linear [4, 13]. Several studies have highlighted the fact that, neural networks with at least one hidden layer can model non-linear systems independently of their complexity.

1.2 Study Area – Pulicat Lagoon

Pulicat lagoon is the second largest brackish water lagoon in India and is located between 13º 26' and 13º 43' N latitude and 80º 03' and 80º 18' E longitudes. Pulicat lagoon lies almost parallel to the Bay of Bengal and covers an area of about 461 km2 [5]. A study area map is shown in Fig. 1 with the sampling locations. The average depth of the lagoon is about 1.5 m and the minimum and maximum depth varies between 0.5 to 6.0 m respectively.

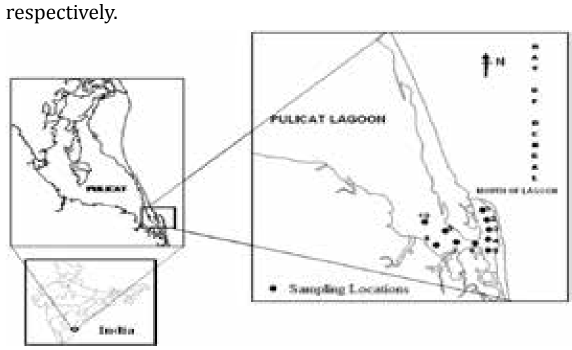


Fig. 1. Study area and sampling locations map

The lagoon-mouth is one of the most dynamic features which determines the mixing and circulation of waters, of not only widely varying salinities and dissolved-oxygen concentrations, but also of primary production, plankton, biodiversity and fisheries in this lagoon. Depending on the width and depth of the lagoon-mouth, during various seasons (summer, pre-monsoon, monsoon and post-monsoon), the salinity in the lagoon varies from 0-52 ppm from the monsoon to summer seasons [10]. Thus, due to the high spatial variation in lagoon characteristics, it is not possible to develop a unique model for the whole lagoon. So a model was developed exclusively of the mouth part of the lagoon, where the water exchange takes place between lagoon and sea has been presented in this study.

2. MATERIALS AND METHODS

2.1. Data and Selection of Input Variables

In order to develop the ANN, the available dataset on the water quality of Pulicat lagoon has been selected. The rationale behind the selection of the dataset is based on the fact that any lagoon ecosystem is characterized by a very broad range of environmental conditions [2] and estimating the Chlorophyll-a concentration in lagoons is therefore a challenging task for Neural Network based models. Data of past five years (1997 to 2001) were collected from the literatures [9, 15].

After the compilation of the data, correlation analysis was employed to select inputs for the ANN model [8]. This task was performed keeping in mind the fact that the correlation of concentration of Chlorophyll-a with the water quality parameters can be an indicator to the extent to which each of these parameters influenced its production by the phytoplankters in the lagoon. Model validation was performed using the values of the different parameters from laboratory analysis and a part of secondary data which were not included in the model training process.

2.2. Model Development and Training

In this study, a multilayer perceptron type network (Feed Forward) was developed to model the chlorophyll-a concentration. In line with the ANN architecture, the present model was also divided into three components or layers.

1. Input layer – Consisting of water quality parameters i.e., Temperature, pH, DO, NO3, PO4 and Salinity.
2. Hidden layer – Consisting of a series of nodes associated with transfer function.
3. Output Layer – Consisting of the desired output i.e., concentration of chlorophyll-a.

A three layer Feed Forward Neural Network (FFNN) has been developed using Neural Network Toolbox in the MATLAB 7.0 software. The Backpropagation algorithm was used to train the network. Fig. 2 illustrates the feed forward neural network model used for this study.

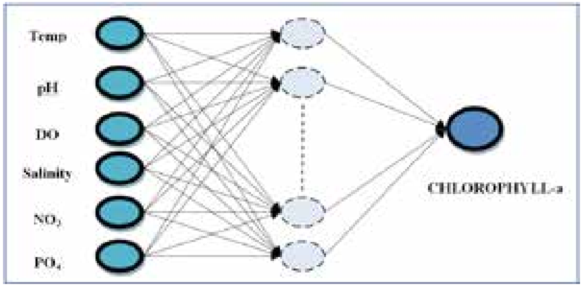


Fig.2 Structure of FFNN model used for the study

The number of hidden layer nodes was determined by optimizing the performance of the network. The lowest mean squared error was obtained by network with 20 nodes in the hidden layer. The activation function determines the relationship between inputs and outputs of a node and a network [14]. Logistic Transfer function or Sigmoid function was used for hidden nodes as well as output node in this model. The Mean Squared Error (MSE) was used as an objective function to calculate the error between predicted outputs and measured values. Initially the mean square error of 0.001 was set as a performance goal. The input data sets were presented at the input layer and the measured chlorophyll-a concentration was set at the output layer as a target for the training process. During the training process with Backpropagation algorithm the MSE starts decreasing for each input epochs. But after 37577 epochs the MSE reached 0.018 and tends to increase. So that point was considered as a optimum performance for the model and the training process was stopped.

2.3. Model Validation

The validation or testing process was used for measuring the generalization ability of the network and fine tuning the network. For validation, water samples were collected from ten locations in the mouth part of the lagoon (Fig. 1) and were analyzed for the concentrations of the following parameters in the laboratory using the procedures described by Grasshoff and Kremaling (1998) [3]: Temperature, pH, Dissolved Oxygen, Nitrate, Phosphate, Salinity and Chlorophyll-a. The results were used for the validation process.

2.4. Sensitivity Analysis

Sensitivity analysis was performed as a part of the present investigation to gain insights into internal weightings of the inputs in a trained neural network [11]. This analysis was carried out using the whole data set to assess the effect of small changes in each inputs on the output of the neural network model. The results of this analysis was also used to understand the underlying ecological processes, i.e. the relative importance of the predictive variables i.e. Temperature, pH, Dissolved Oxygen, Nitrate, Phosphate and Salinity to chlorophyll-a concentration in Pulicat lagoon [17].

3. RESULTS AND DISCUSSION

3.1. Correlation Analysis

In order to derive the strength of relationships between different water quality parameters, correlation analysis was preformed. The Karl Pearson's coefficients of correlation computed between the different water quality parameters are given in Table 1.

Temperature and pH are the two important physical parameters, having a major impact on the algal production (Chlorophyll-a concentration). Even though the Ammoniacal Nitrogen and Nitrite Nitrogen are also having higher correlation, Nitrate Nitrogen (NO3) only has been selected as one of the input parameters. Because there is a linear correlation between these three parameters and NO3 shows the highest correlation with Chlorophyll-a than other two. Hence NO3 has been selected as a Nitrogen source. In the same way Salinity also selected in comparison with the Electrical Conductivity. Based on the correlation analysis, Temperature, pH, Salinity, Dissolved Oxygen, Nitrate Nitrogen and Phosphate Phosphorous showing high values of pair wise correlation ($r > 0.44$) with Chlorophyll-a concentration were selected as the model inputs.

3.2 Model Training

Chlorophyll-a concentrations in the Pulicat lagoon was estimated by neural network as a function of temperature, pH, DO, salinity, PO4 and NO3. The model performance in predicting Chlorophyll-a concentration was obtained by comparing calculated output from the model and the actual measured output from the field (Table 2a). In the present study, a linear correlation coefficient of 0.98 was obtained between predicted and measured concentrations of chlorophyll-a. The Mean Squared Error (MSE) between predicted and measured output was found to be very low (MSE=0.018). A graph has been plotted for the predicted and the measured outputs in predicting the chlorophyll-a concentrations (Fig. 3a). It is observed that the two graphs show maximum convergence except some points which indicates the good prediction accuracy of the model. A scatter plot also has been drawn for the predicted output and the measured output values (Fig. 3b). An almost linear scatter of the points was obtained indicating that the model performance was good.

3.3 Model Validation

It is to be noted that the accuracy of the neural network model and its generalization capabilities were also tested on an independent data set, which was collected during 1997 and 2009. Results from the laboratory analysis of ten samples taken from mouth part of the lagoon (Fig.1) are shown in Table 3. After the validation process, it has been observed that the correlation coefficient between the predicted output and the measured value is 0.998 and the mean squared error is 0.018. Predicted output and the measured output are shown in Table 2b.

A graph was plotted between the predicted and the measured outputs in predicting the chlorophyll-a concentrations during the validation phase (Fig.3c). It is observed that the two graphs show maximum convergence except some points which indicates the good ability in prediction of the Chlorophyll-a concentration. A comparison between the calculated and measured output (Chlorophyll-a) values is presented in the scatter plot (Fig.3d). During the validation of the ANN, it has been observed that the prediction accuracy was far higher than the ordinary linear models.

3.4 Sensitivity Analysis

Results from the sensitivity analysis indicate that the temperature and Phosphate have major influence on the Chlorophyll-a concentration. DO result shows moderate sensitiveness on the output. Other three parameters, Nitrate, pH and Salinity have considerable influence on the Chlorophyll-a concentration. Fig.4 shows the resultant effect on chlorophyll-a concentration as percent average change in calculated values.

The results of the sensitivity analysis showed that increasing error additions caused increasing mean square errors in the output, even though this relationship was not absolutely monotonic, because less sensitive variables, that did not affect the neural network output very much, showed a few negative increments. However it was observed that, the relative sensitivity of the input variables did not vary significantly when very large amounts of error with such input variables were added. These results suggest that the chlorophyll-a model that was embedded in the neural network was probably consistent with the ecological processes.

Table 1. Pearson Correlation Matrix variables of water quality

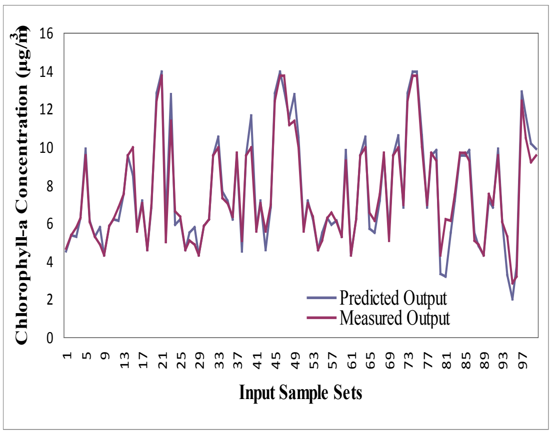
	Chl-a	Temp	pH	EC	HCO_3	Cl	SO_4	Na	K	Ca	Mg	TDS	DO	NO_3	NO_2	NH_4	PO_4	SAL
Chl-a	1																	
Temp	0.66	1																
pH	0.566	-0.285	1															
EC	-0.11	0.169	0.215	1														
HCO_3	-0.173	0.023	0.291	0.093	1													
Cl	-0.283	-0.099	-0.123	0.487	0.041	1												
SO_4	-0.253	-0.084	-0.104	0.462	0.089	0.847	1											
Na	-0.313	-0.159	0.167	0.428	0.007	0.884	0.77	1										
K	-0.135	-0.281	0.1	0.265	-0.001	0.488	0.555	0.55	1									
Ca	-0.018	-0.157	0.235	0.748	0.351	0.512	0.36	0.624	0.194	1								
Mg	0.222	-0.165	0.155	0.809	0.476	0.563	0.44	0.637	0.29	0.933	1							
TDS	-0.278	-0.134	-0.109	0.532	0.03	0.584	0.86	0.943	0.548	0.6	0.649	1						
DO	0.441	-0.698	0.451	0.328	-0.415	0.44	0.42	0.279	0.157	0.28	0.277	-0.406	1					
NO_3	0.866	0.06	-0.169	0.009	0.145	0.449	0.391	0.41	-0.014	0.225	0.109	0.426	0.427	1				
NO_2	-0.259	0.237	-0.164	0.053	0.081	0.368	0.291	0.247	-0.107	0.042	-0.036	0.31	0.373	0.822	1			
NH_4	0.051	0.183	0.065	0.087	-0.086	0.222	-0.085	-0.104	0.129	-0.131	-0.104	-0.053	0.102	0.078	0.177	1		
PO_4	0.725	-0.075	-0.056	0.06	0.161	0.409	0.352	0.154	0.12	-0.244	-0.147	0.313	0.302	0.17	0.269	0.471	1	
SAL	-0.468	0.084	0.475	0.19	0.243	0.612	0.199	0.095	0.204	0.245	0.193	0.154	0.065	0.525	0.489	0.022	0.328	1

Table 2. Predicted and Measured outputs of Chlorophyll-a concentrations (µg/m3) during Training and Validation processes

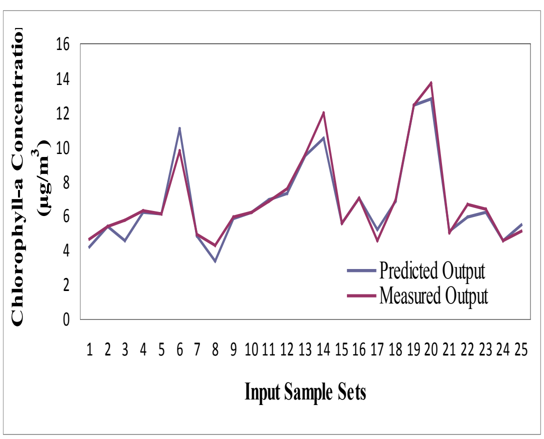
(a) During Training process				(b) During Validation process			
Predicted output	Measured output	Predicted output	Measured output	Predicted Output	Measured output	Predicted output	Measured output
6.91	6.87	13.37	12.47	4.71	4.7	12.45	12.47
4.69	4.70	13.88	13.73	5.38	5.4	13.78	13.73
22.84	20.33	5.12	5.03	5.58	5.8	5.08	5.03
4.34	4.26	18.51	16.37	6.26	6.3	16.32	16.37
4.25	4.13	6.78	6.64	6.12	6.1	5.97	6.64
3.75	4.68	5.97	6.37	5.10	5.3	6.25	6.37
6.09	6.00	6.35	6.35	4.89	4.9	4.59	4.60
6.73	6.68	10.64	10.55	4.36	4.3	5.53	5.12
5.78	5.70	6.25	6.00	5.88	5.9		
8.38	8.60	5.20	5.07	6.19	6.2		
6.96	6.87	4.58	4.54	6.96	6.87		
7.71	7.56	5.03	6.13	7.33	7.56		
10.55	9.60	3.65	3.60	9.55	9.60		
10.22	10.00	5.86	5.78	10.54	10.00		
5.60	5.60	4.63	4.56	5.61	5.60		
7.07	7.04	8.05	7.00	7.01	7.04		
3.68	4.60	4.60	4.60	4.18	4.60		
7.06	6.98	5.12	5.12	6.88	6.98		

Table 3. Sample analysis results for selected parameters from ten locations

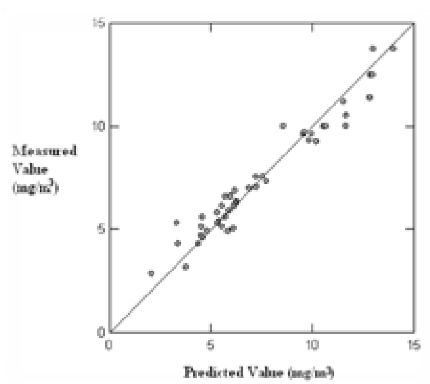
Location No.	Parameters						
	Temp (°C)	pH	DO (ml/L)	Sal (‰)	NO_3 (µg/L)	PO_4 (µg/L)	Chl-a (µg/m³)
1	29.5	7.7	2.0	25.5	25.2	19.3	4.7
2	29.2	7.6	2.2	23.8	29.7	21.7	5.4
3	28.8	7.8	2.4	20.3	19.4	35.8	5.8
4	29.3	7.8	2.1	18.8	31.8	26.8	6.3
5	29.7	7.9	2.4	19.5	22.3	18.5	6.1
6	28.9	8.0	2.3	21.6	30.6	25.4	5.3
7	29.0	7.8	2.1	20.5	18.1	23.7	4.9
8	28.6	7.6	2.2	17.8	27.4	32.8	4.3
9	30.0	7.7	2.2	18.6	23.6	29.1	5.9
10	29.4	7.9	2.2	17.9	15.8	30.8	6.2



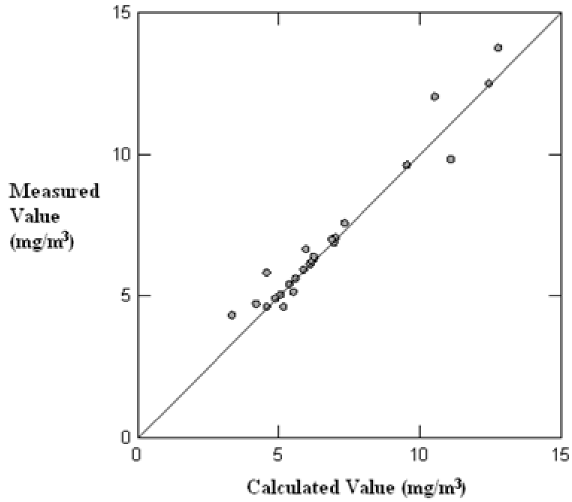
(a) Comparison during Training process



(c) Comparison during Validation process



(b) Scatter Plot after training process



(d) Scatter Plot after Validation process

Fig. 3. Comparison and Scatter Plot of measured and predicted output of Chlorophyll-a concentrations in Training and Validation processes

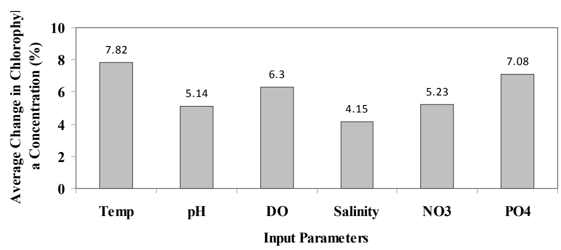


Fig.4. Average percent change in the calculated value of chlorophyll-a concentration

4. CONCLUSION

The variations in the results were mainly due to the inflow of seawater during premonsoon and the riverine input during postmonsoon. A seasonal cycle in the phytoplankton production was observed, especially during summer and postmonsoon seasons. This work presented a study on the application of artificial neural networks to the task of modelling of chlorophyll-a concentrations in the Pulicat lagoon. The six parameters have been selected from the correlation analysis for model input. The backpropagation training algorithm was used to train the model. Then model validation has been performed to test the predictability of the model. Finally a sensitivity anlysis was performed to investigate the effect of each input parameter on the output (Chlorophyll-a concentration).

For both, training and validation processes, the correlation coefficient between the calculated output and the measured output and the mean squared error were found to be 0.98 and 0.018 respectively. The validation of the neural network model using independent data not included in the training procedure showed good agreement for predictions of the chlorophyll-a concentrations. Thus the chlorophyll-a concentration was successfully predicted by the ANN.

In conclusion the work supports the concept that the neural network approach is a successful method of modelling such complex and nonlinear behaviour of Chlorophyll-a concentrations in the lagoon ecosystem with different environmental conditions. As a complement to traditional modelling techniques, the neural network approach was shown to be useful for identification of the primary driving mechanisms in the system dynamics.

REFERENCE

[1] Bowden, G.J., Dandy, G.C., Maier, H.R (2003), "Data Transformation for Neural Network Models in Water Resources Applications", J. of Hydro info, 5(4): 245-258. | [2] Duarte, P., Bernardo J. M, Costa A.M, Macedo F, Calado G., Cancela, L. da Fonseca (2002), "Analysis of coastal lagoon metabolism as a basis for management", Aqua. Ecol, 36: 3-19. | [3] Grasshoff, K., Kremaling, K (1998), "Method of sea water analyses", Third edition, Wiley VCH. | [4] Karul C, Soyupak S, Cilesiz A.F, Akbay N., Germen E (2000), Case studies on the use of neural networks in eutrophication modelling, Ecol. Model, 134:145–152. | [5] Lagoons of India: State-of-the-art report (2001), Environmental Information System Centre, Centre of Advanced Study in Marine Biology, Parangipettai, Tamilnadu, India. | [6] Lee, J.H.W., Huang, Y., Dickman, M., Jayawardena, A.W (2003), "Neural network modelling of coastal algal bloom", Ecol. Model, 159:179-201. | [7] Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. and Aulagnier, S (1996), "Application of neural networks to modelling nonlinear relationships in ecology", Ecol. Model, 90:39–52. | [8] Nitin Muttil, Kwok-Wing Chau (2007), "Machine-Learning paradigms for selecting ecologically significant input variables", Engg. App. Art. Intel, 20:735-744. | [9] Puranik Gayathri Ranganath (2000), "Environmental Geochemistry of the Pulicat Lake, South India", Ph.D Thesis, Institute for Ocean Management, Anna University, Chennai, India. | [10] Ramesh, R., Nammalwar, P., Gowri, V.S (2008), "Database on Coastal Information of Tamilnadu", Institute for Ocean Management, Anna University, Chennai, India. | [11] Recknagel, F., French, M., Harkonen, P., Yabunaka, K (1997), "Artificial neural network approach for modelling and prediction of algal blooms", Ecological Modelling, 96:11–28. | [12] Scardi, M (1996), "Artificial neural networks as empirical models of phytoplankton production", Mar. Ecol. Prog. Series, 139:289–299. | [13] Scardi, M (2001), "Advances in neural network modelling of phytoplankton primary production", Ecol. Model, 146:33-45. | [14] Scardi, M., Harding, L.W (1999), "Developing an empirical model of phytoplankton primary production: a neural network case study", Ecol. Model, 120:213–223. | [15] Shalini Akella (2002), "Methane Cycling in a Natural Brackish Water Lake: A Case Study on Pulicat Lake, South India", Ph.D Thesis, Institute for Ocean Management, Anna University, Chennai, India. | [16] State of Environment and related issues, Environmental Information System Centre, Department of Environment, Government of Tamil Nadu, India. | [17] Zhang, G., Patuwo, B.E., Hu, M.Y (1998), "Forecasting with artificial neural networks: The state of the art", International Journal of Forecasting, 14:35-62.