Available online at www.elixirpublishers.com (Elixir International Journal)

Chemical Engineering

Elixir Chem. Engg. 66 (2014) 20556-20560

ISSN: 2229-712X

Assessment on Air Quality Index in Kolkata and suburban area of West Bengal, India using cluster analysis and neural network Techniques

Keka Sinha[1] and Papita Das Saha[2]

[1]Biotechnology Department, National Institute of Technology, Durgapur.

[2]Department of Chemical Engineering, Jadavpur University. Kolkata, India

ARTICLE INFO

Article history:

Received: 4 November 2013;

Received in revised form:

23 December 2013;

Accepted: 1 January 2014;

Keywords

Air quality index,

Air pollution,

Cluster analysis,

ANN model.

ABSTRACT

Air quality index (AQI) is the single value numerical expression to evaluate the quality of a given air at a particular location. The present study classifies the locations with different air pollution characteristics in respect of AQI and to evaluate the parameters of the locations of the monitoring program. A selected cluster analysis is applied to the pollution monitoring dataset which including SPM, RPM, SO_2, NO_2. A computer-simulated artificial neural network (ANN) model is developed using Matlab to get a good interaction between the different components of air of different location in Kolkata responsible for Air Quality measurement and their relationship with AQI.

Introduction

The environmental pollution in India are growing rapidly. The rapidly growing population and increasing economic development are putting an effect on the environment and the natural resources. Different pollution like Industrial, soil erosion, rapid industrialization, urbanization, automobiles, domestic fuel burning, road side dusts, construction activities, and land degradation creates worsening problems. Air pollutants are associated with increased hospitalization [1] and mortality due to cardiovascular disease, [2-4] frequent arrhythmias, or both [5]. Ambient air quality monitoring under National Air Monitoring Programme (NAMP) provide air quality monitoring information that forms the basis for identifying areas with high air pollution levels and subsequently for planning the strategies for control the air pollution (National Summary Report, 2010; [6]). The Air Quality Index (AQI) is an indicator of daily air quality that based on air pollutants like ozone, fine particulate matter, nitrogen dioxide, carbon monoxide, sulphur dioxide and total reduced sulphur compounds which can affect on human health and the environment.

Largest numbers of people of Kolkata area are suffering from lung cancer [7] due to the pollution of the most polluted metropolitan cities in West Bengal [7]. There is also a very high impact of old vehicles on air quality. The present study of AQI is focused on the evaluation of pollution data of air collected from various locations in Kolkata of seven years.

Ambient Air Pollutants

Particulate Matter

There are many chemicals that can be detected in particulate matter (PM) which include nitrates, sulfates, elemental and organic carbon, biological compounds (eg, endotoxin, cell fragments), organic compounds, , and a variety of metals (eg, iron, copper, nickel, zinc, and vanadium) [9].

Nitrogen Oxides

Nitrogen oxides are commonly encompassed nitrogen dioxide (NO_2), nitric oxide (NO), nitrogen trioxide, nitrogen tetroxide (N_2O_4), and di-nitrogen pentoxide (N_2O_5)[10]. The main source of NO_X in the air is fossil fuel burning in motor vehicles and industrial processes, in generation of power.

Carbon Monoxide

Carbon monoxide (CO) is the product due to the incomplete combustion of carbonaceous fuels. The main sources are motor vehicles, engines on motorboats, chain saws, and from other devices that may require residential wood burning; improperly adjusted gas-burning; fossil fuel combustion; and coal combustion; and tobacco smoking [11-12].

Sulfur Dioxide

Sulfur dioxide (SO_2) is a colorless, highly irritating soluble gas that is released in various industrial processes. In contact with water, it will converted to sulfurous acid, which has a strong irritant effect on eyes, mucous membranes, and skin [13]

Objective of the study

This present work aims to identify the air pollution characteristic of all 15 stations covers northern, southern, central and peripheral part in Kolkata during seven years and to evaluate the correlation analysis and to design a artificial neural network approach between the pollutants in air and air quality index (AQI). A selected cluster analysis technique is applied to the pollution monitoring dataset which includes RPM, SPM, SO2, and NO2 of all stations. The obtained experimental results allowed to determine groups of stations with similar air quality index which can be determined from air pollutants dataset, and to compare air quality among these groups. Moreover an ANN model [14] has been developed to analyze the close interaction between experimental and simulated result of annual air quality index in Kolkata.

Materials and Methods

Dataset:

The dataset contains the values of selected pollution indicators for 15 monitoring locations in Kolkata. The population and human activities are mainly centered in those regions of Kolkata. The yearly average air quality of Kolkata is

Tele:

E-mail addresses: papitasaha@gmail.com

monitored through 15 fixed monitoring stations for the four selected air pollutants parameters respirable particulate matter (RPM), suspended particulate matter (SPM), sulphur dioxide (SO2), nitrogen dioxide (NO2) for the year 2002-2008. These air quality monitoring stations are evenly distributed throughout the state.

Several attempts have been found in the past to develop a suitable AQI method. Three basic approaches are followed by the researchers: Scope (F1), Frequency (F2) and Amplitude (F3). Various parameters like respirable particulate matter (RPM), suspended particulate matter (SPM), SO2, NO2 etc are considered for calculating the AQI.

Methods

The values of three measures of variances from selected objectives for quality are combined to create a vector in imaginary 'objective exceedance' space. Objectives may refer to the quality guidelines. The length of vector is then scaled to range between 0 to 100.and subtracted from 100 to produce an index which is 0 or close to closed to 0 for very poor quality, and close to 100 for excellent quality [15-17].

Calculation of AQI

FACTOR 1: (F_1) SCOPE

This factor is called scope because it assesses the extent of quality guideline non compliance over the time period of interest.

F_1= (Numbers of failed variables of air/Total no. of variables of air)*100 ---------- (1)

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

FACTOR 2: (F_2) FREQUENCY

F_2 = (Number of failed tests/Total no. of tests)*100 ---- (2)

FACTOR 3: (F_3) AMPLITUDE

F_3 represents the amount by which the failed test values do not meet their objectives, and it is calculated in three steps:

i) When the test value must not exceed the objective:

Excursion= (Failed test value/Objective) -1 ---------- (3)

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

Excursion= (Objective/Failed test value) -1 ---------- (4)

ii) nse= Summation of excursion values/Total no. of tests ---------- (5)

F3 is then calculated by a scales using the sum of excursions from objectives (nse) to yield a range between 0 and 100.

F_3= (nse/0.01*nse) +0.01 ---------- (6)

AQI is finally calculated from F_1 , F_2 , F_3 as

AQI = $100-\sqrt{(F_1^2+F_2^2+F_3^2)}/1.732$ ----------(7)

- If the AQI reading is within 0 to 44, the air quality is the very poor category, and this may have adverse effects on the environment. [Table 1].
- If the AQI reading is within 45 to 79, the quality is poor category, and this may have adverse effects for human and animal populations [Table 1].
- If the AQI reading is above 80, the air quality is in moderate category, and this may have some adverse effects for very sensitive people.

Cluster Analysis

Cluster analysis is a way to create groups of objects, or clusters, in such a way that the profiles of objects in the same cluster are very similar and the profiles of objects in different clusters are quite distinct. Cluster analysis can be performed on many different types of data sets.

The initial assumption is that the nearness of objects in the space defined by the variables reflects the similarity of their properties [9; 12-13]. In our study, a hierarchical clustering by applying complete linkage method and the squared Euclidean distance as matrix were performed by using the Matlab7 (The Mathworks, Inc. ver. 7.0.1).

Table 1: Grading scale

INDICATOR	CONDITION	GRADE
95-100	EXCELLENT	A
80-94	GOOD	B
65-79	FAIR	C
45-64	POOR	D
0-44	VERY POOR	F

ANN modeling

An artificial neural network (ANN) is a simplified computational mathematical model that is inspired by the structure of biological neural networks [18-19]. Artificial neural network model have been used successfully for monitoring air quality index for the four input variables in Kolkata. MATLAB 7 (The Mathworks, Inc. ver. 7.0.1) was chosen to develop the ANN model using neural network toolbox using the data.

In the present study, a three-layer feed forward back-propagation neural network with a linear transfer function was developed for yearly average air quality of Kolkata for the year 2002-2011.

Most ANN model consists of three layers: an input layer represents as independent variables; the output layer gives the dependent variables and one or more intermediate nodes acts as a collection of feature detectors. Respirable particulate matter (RPM), suspended particulate matter (SPM), sulphur dioxide (SO2), nitrogen dioxide (NO2) as three independent variables, and average air quality index (AQI) of Kolkata was chosen as dependent variables [20].

Table 2: Air Quality Index of different areas of Kolkata (2002-08)

Stations	Year	AQI range	Quality
DUNLOP	2002- 2008	62.77– 54.43	Marginal
TOLLYGUNGE	2002- 2008	63.35-54.25	Marginal
ULTADANGA	2002- 2008	63.58-54.48	Marginal-Fair
PICNIC GARDEN	2002- 2008	74.86-69.13	Marginal-Fair
MINTO PARK	2002-2008	63.22-54.93	Marginal
SHYAMBAZAR	2002-2008	63.01-54.64	Marginal-Fair
BELIGHATA	2002-2008	63.11-68.95	Marginal-Fair
MOMINPORE	2002-2008	63.39-54.36	Marginal
BAISHNABGHAT	2002-2008	75.58-55.47	Marginal-Fair
TAPSIA	2002-2008	62.66-55.51	Marginal
SALT LAKE	2002-2008	74.96-69.54	Marginal-Fair
MOULALI	2002-2008	62.38-52.48	Marginal
BEHLA CHAWRASTA	2002-2008	62.89-53.34	Marginal
HYDE ROAD	2002-2008	74.9-68.65	Marginal-Fair
GARIAHAT	2002-2008	62.42-70.01	Marginal-Fair

Results and discussion

The experimental results (Fig.1-2) showed that the yearly average values of RPM, SPM, SO2, and NO2 were different at different Stations in Kolkata. The figure 1-2 showed that the year of 2006-07 at Behala Chowrasta and the year 2003-04 at Tapsia, the respirable particulate matter (RPM) and suspended particulate matter (SPM) in air quality are higher than the other locations and year in Kolkata. In Table 2, the value of Air Quality Index from 2002-2008 and quality grade of all the 15 stations in Kolkata has been presented.

The AQI of these stations of Kolkata range between marginal to fair depicting not good quality of air. The annual AQI seems to decrease from 2002-2008 it has increased also for some stations.

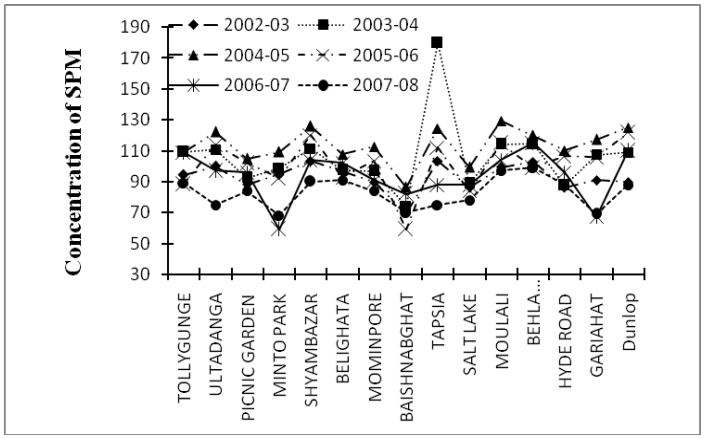

Fig. 1. SPM concentrations of air quality of Kolkata

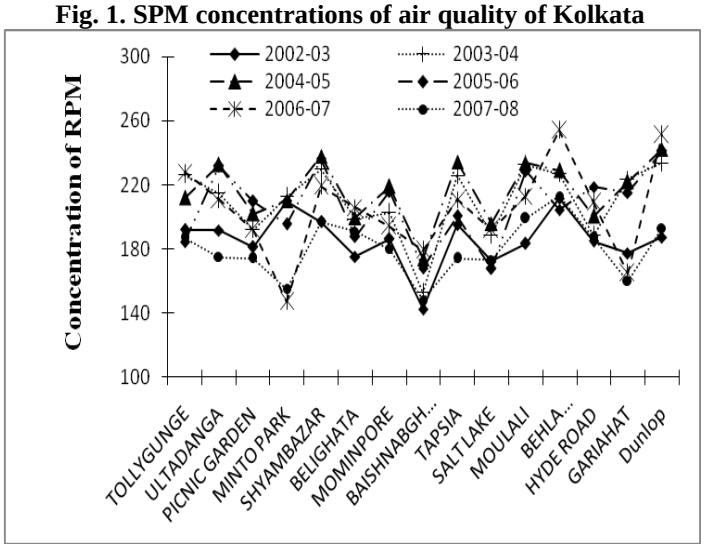

Fig. 2. RPM concentrations of air quality of Kolkata

Cluster analysis of Station's Data

The dendrogram, Fig 3, showed the results obtained from using hierarchical complete linkage clustering method and squared Euclidean distance. The all 15 stations in Kolkata can be classified into two major groups according to the average calculated air quality index [Table 2] from the year 2002-2008, cluster I: Dunlop, Behala Chowrasta, Tollygunge, Mominpore, Tapsia, Moulali, Minto park, Ultadanga, Shayambazar (the range of AQI is between 58.5 to 62) and cluster II: Picnic Garden, Baishnabghata, Belighata, Salt Lake, Hyde Road, Gariahat (the range of AQI is within 64 to 70) as presented in Fig 2. It was clearly observed that the former major cluster was characterized by the biggest Euclidean distance to another cluster (high significance of clustering). Three associations were proved the clustering. The association between Dunlop and Behala Chowrasta, Tollygunge and Mominpore and then Belighata and Salt Lake were most significant. The identified group including sampling stations at Dunlop and Behala Chowrasta, Tollygunge and Mominpore and then Belighata and Salt Lake were selected to analyse temporal variations and to evaluate whether the sampling frequency can be reduced. MATLAB 7 (The Mathworks, Inc. ver. 7) computing environment was used to design the neural network model from the experimental data. After testing, it was observed that 10 neurons produce minimum value of error of the training and validation sets. The yearly average (2002-2011) air quality monitoring result had been chosen the experimental response or output variable. The network model which gave a coefficient of correlation (*R*) between the model prediction and experimental results tends to 1 considered to be significant model and hence selected. Majority of ANN architectures were feed-forward networks which were mostly trained from the input data using back-propagation algorithm. The performance of the trained network was shown in Fig. 4

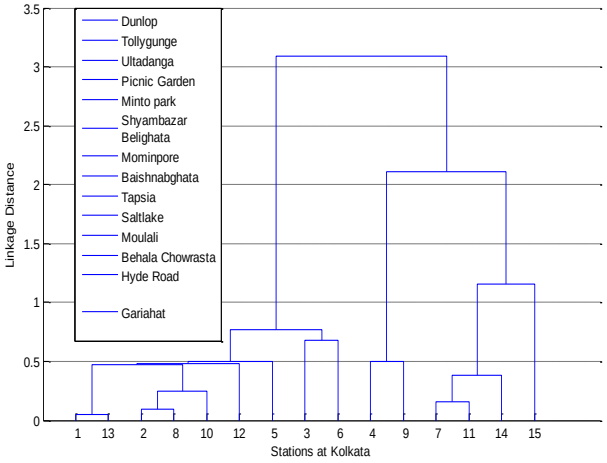

Figure 3: Dendrogram of the Cluster analysis based on regional air pollution characteristic

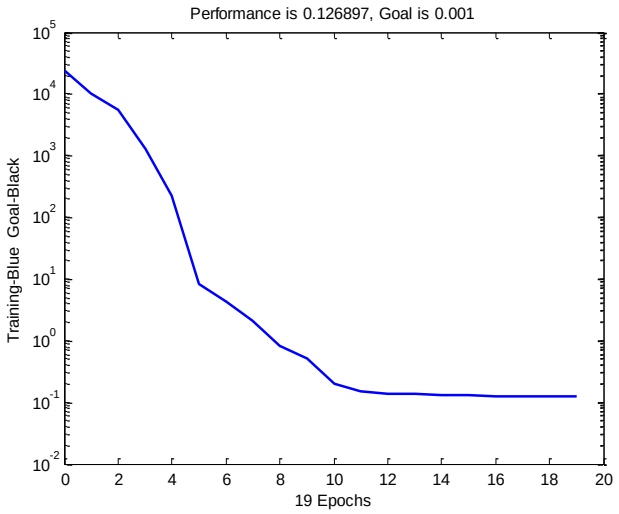

Fig. 4. Performance plot during the ANN training

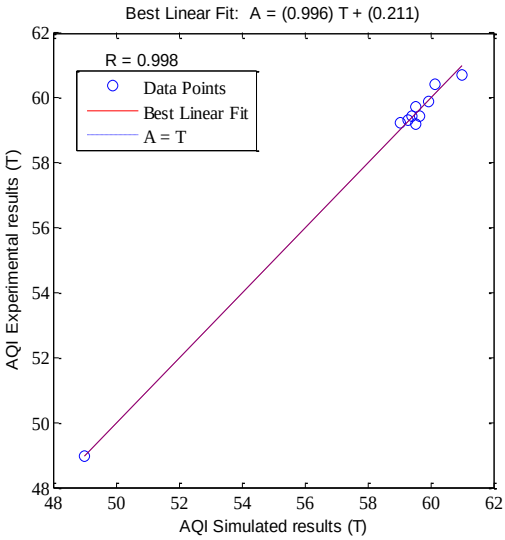

Fig. 5. Regression plot on AQI (Experimentally vs. Predicted) using four input variables, ten processing elements in hidden layer, and one output variable using ANN model

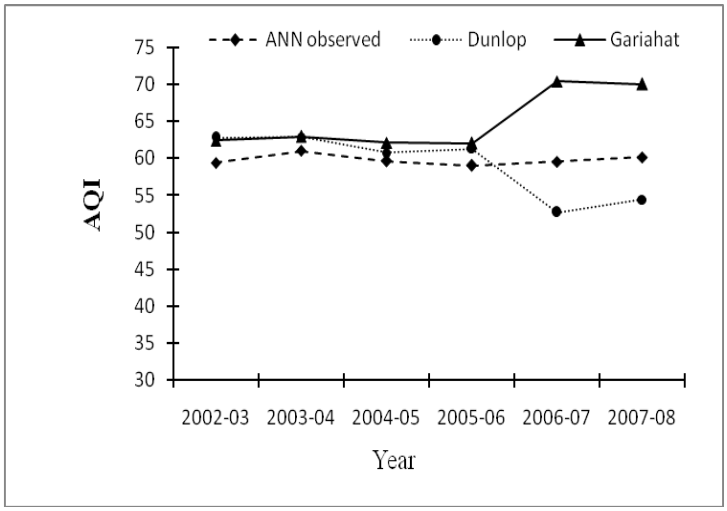

Fig. 6: Plot of AQI of ANN experimental result and AQI of two stations

Linear transfer function 'POSLIN' (being a positive linear transfer functions) was chosen for the input to hidden layer mapping while a 'PURELIN' (purely linear transfer function) was chosen for the hidden layer to the output layer mapping. The values of correlation coefficient (R) were calculated among the best of 10 repeated run [21]. The regression plot of the trained network was shown in Fig 5. A high correlation coefficient of this plot signified the reliability of the neural model with the experimental data. Fig 5 showed that the 'Levenberg-Marquardt backpropagation (trainlm)' algorithm tool with 'poslin' transfer function gave a correlation coefficient of 0.998 that suggests most satisfactory interaction.

Figure 6 compared the ANN experimental results in details with two stations among 15 stations. Above figure showed the plot of experimental ANN model relation for air quality index from Dunlop and Gariahat in Kolkata of the seven years dataset used for training. For all the training data sets, the data points of experimental ANN values and two stations AQI values more or less coincide with each other except 2006-07 and 2007-08. This indicates perfect training of the network.

Acknowledgement

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 air quality of Kolkata and West Bengal. I am also thankful to Mr. Souvik Ganguly, Environment Engineer, West Bengal Pollution Control Board for his valuable help and suggestions during analysis of the data of different locations.

Conclusion:

The present study showed to classify the different stations, based on annual average air quality index with different air pollution characteristics from 2002 to 2008 in Kolkata, and to evaluate the air quality of various clusters. The obtained results have proved that the stations with similar air quality monitoring index can be grouped in one using cluster analysis method. From this study, Kolkata's average SPM count is 211µg/m^3, PM count is 155µg/m^3, SO_2 count is 7.51 and NO_2 count is 76.68. All 15 stations in Kolkata are classified into two major groups based on the annual calculated air quality index. Cluster I is characterized as low AQI level (range of 58.5 to 62) than the cluster II with high AQI level (range within 64.5 to 70). The annual AQI level in Kolkata is successfully predicted by designing a three layered neural network model with 4 input and 10 neurons in the hidden layer and using linear transfer function with Levenberg-Marquardt backpropagation (trainlm) backpropagation algorithm from the experimental data.

References:

[1] Poloniecki, J.D., Atkinson, R.W., de. Leon, A.P., (1997). 'Daily time series for cardiovascular hospital admissions and previous day's air pollution in London, UK'. *J Ocup. Med* Vol.54, pp.535–540.

[2] Pope, C.A., Burnett, R.T., Thun, M.J., (2002). 'Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution'. *J Am. Med. Assoc* Vol. 287, pp. 1132–1141.

[3] Pope, C.A., Burnett, R.T., Thurston, G.D., (2004). 'Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease'. *Circulation* Vol. 109, pp.71–77.

[4] Samet, J.M., Dominici, F., Curriero, F.C., (2000). 'Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994'. *New Engl* J *Med* Vol. 343, pp.1742–1749.

[5] Mann, J.K., Tager, I.B., Lurmann, F., (2002). 'Air pollution and hospital admissions for ischemic heart disease in persons with congestive heart failure or arrhythmia'. *Environ. Health Perspect.* Vol.110, pp.1247–1252.

[6] Air quality monitoring, emission inventory and source apportionment study for Indian cities, National Summary Report, (2010). Central Pollution Control Board: www.cpcb.nic.in.

[7] Rashme,S., 2004. **'**Info Change urban India, News & Features' http://infochangeindia.org/urban-india/features/kolkata-has-the-highest-number-of-lung-cancer-cases-worldwide.html

[8] Seinfeld, J.H., Pandis, S.N., (1998). 'Atmospheric Chemistry and Physics: From Air Pollution to Climate Change New York', NY: *John Wiley & Sons*, pp1–1326.

[9] Hagberg, M., Kolmodin-Hedman, B., Lindahl, R., (1985). 'Irritative complaints, carboxyhemoglobin increase and minor ventilatory function changes due to exposure to chain-saw exhaust'. *Eur J Respir Dis.* Vol.66, pp.240–247.

[10] U.S. Environmental Protection Agency. (1993). 'Air Quality Criteria for Oxides of Nitrogen. Research Triangle Park, NC: US Environmental Protection Agency (1), 3–14.' Publication No. EPA/600/8-91/049aF.

[11] Hampson, N.B., Norkool, D.M., (1992). 'Carbon monoxide poisoning in children riding in the back of pickup trucks'. *J Am. Med. Assoc* Vol. 267, pp.538–540.

[12] Lipsett, M.J., Ozone-Sullivan, J.B., Krieger, G.R., (2001). 'Clinical Environmental Health and Toxic Exposures'. 2nd ed. Philadelphia, Pa: *Lippincott Williams & Wilkins*, pp. 806–818.

[13] Massart, D.L., Kaufman, L., (1983). 'The interpretation of analytical chemical data by the use of cluster analysis'. *Wiley*, New York.

[14] Sinha, K., Saha, D.P., Datta, S., (2012). 'Optimization and modeling of natural dye extraction assisted by microwave from pomegranate rind using Response Surface Methodology and Artificial Neural Network'. *Ind* Crop *Prod* Vol. 37, pp.408–414.

[15] Boyacioglu, H., (2007). 'Development of a water quality index based on a European classification scheme.' *Water SA*. Vol. 33(1), pp.101 – 106.

[16] Pesce, F., Wunderlin, D. A., (2000). 'Use of Water Quality Indices to verify the Impact of Cordoba City (Argentina) on Suquia River'. *Water Res* Vol. 34(11), pp. 915 – 926.

[17] Sargaonkar, A., Deshpande, V., (2003). Development of an overall index of pollution for Surface water based on a general classification scheme in Indian context'. *Environ Monit. Assess* Vol. 89**,** pp. 43–67.
[18] Lee, S., Cho, S., Wong, M., (1998). 'Rainfall prediction using artificial neural networks'. *Journal of Gender, Information and Development, Vol.* 2 (2), pp.233–242.
[19] Haykyn, S., (2003). 'Neural Networks, a Comprehensive Foundation'. *Prentice Hall, India.*
[20] Kumar, K.V., Porkodi, K., (2009). 'Modelling the solid-liquid adsorption process using Artificial Neural Networks trained by pseudo second order kinetics'. *Chem* Engg J Vol. 148(1), pp.20-25.
[21] Suman, D., Simon, A.P., Chiranjib, B., Sibdas, B., Siddhartha, D., (2010). 'Development of an artificial neural network model for adsorption and photocatalysis of reactive dye on TiO_2 surface'. *Expert Syst Appl* Vol. 37, pp. 8634-8638.