International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 2, February - 2014

A Novel Air Quality Prediction Model Using Artificial Neural Networks

Niharika 1st
Centre of Information & Technology Department,COES,
University Of Petroleum and Energy Studies Dehradun,
India

Venkatadri M.
Centr Of Information and Technology Department,COES,
University Of Petroleum and Energy Studies
Dehradun, India.

Abstract— **Due to industrialization and urbanization air pollution is becoming an environmental threat at a very rapid rate. Thus, the air quality prediction is becoming essential both for the environment as well as for the society. Air Quality prediction can be carried out using various traditional and soft computing techniques/methods. Soft computing techniques are considered to be an excellent predictive and data analysis technique for air quality prediction.**

Among various soft computing techniques Artificial Neural Networks (ANN) has proven to be the most optimized air quality data analysis tool. Hence this research carried out by the collection of air pollutants data for NAGPUR city of Maharashtra state and designed an optimized model for air quality prediction.

Keywords— Air Quality, Artificial Neural Networks, Industrialization.

I. INTRODUCTION

The demand of resource utilization increases due to the rapid exponential population growth and adverse increase in standard of living leads to limited availability of resources. In order to meet the increasing demand of different types of resources such as land, large scale deforestation is performed. This rapid unmanaged urbanization and industrialization has lead to the change of chemical composition of the atmosphere which is also associated with anthropogenic activities. The excessive use of industrial activities is a biggest cause of addition of air pollutants in the environment, resulting in the deterioration of air quality.

Air is one of the main elements that is responsible for existence of life on Earth. Thus damage to the air quality affects the life on Earth. Air pollution is a contamination of the atmosphere by gaseous, liquid, or solid wastes or by-products that have a serious effect on human health and the biosphere reduce visibility, and damage materials

The critical air pollution existence is observed in the areas where the geographical and meteorological conditions do not permit an easy circulation of air and a large part of the population moves frequently between distant places of a city. These events require drastic measures such as the closing of the schools and factories and the restriction of vehicular traffic. Indeed, many epidemiological studies have consistently shown an association between particulate air pollution and cardiovascular [41] and respiratory [42] diseases. The forecasting of such phenomena with up to two days in advance would allow taking more efficient countermeasures to safeguard citizens' health.

This Study is basically on the fact that the data related to various pollutants present in the air is available with different pollution control boards but there are no efficient predicting tools for utilizing this data to maximum. If the change in these parameters is predicted efficiently then effective policies can be in place to arrest this growth. This can be realized by developing an efficient algorithm which will exploit the predicting capacity of the Artificial Neural Network to give results with high reliability.

The following section II provides the geographical and information related to the sample data collections. Section III discusses the steps for proposed model; Section IV presents the data preparation for the proposed model; Section V discusses the results followed by conclusion.

II IDENTIFICATION AND COLLECTION OF THE DATA FROM THE MONITORING STATION:

The monitoring station for this project is NAGPUR NEERI and the site is allocated in NASHIK, for air quality prediction. Nashik is situated on the banks of Godavari River. It is 11 most populated districts in India. Nashik District is located between 18.33 degree and 20.53 degree North latitude and between 73.16 degree and 75.16 degree East Longitude at Northwest part of the Maharashtra state, at 565 meters above mean sea level. The overall area of Nashik is 360 Km^2 [40] and the density of the population is 6107187 [39]. Though average rainfall of the District is between 2600 and 3000 mm, there is wide variation in the rainfall received at various blocks. Most of the rainfall is received at various blocks. Most of the rainfall is received from June to September. The maximum temperature in summer is 42.5 degree centigrade and minimum temperature in winter is less than 5.0 degree centigrade. Relative humidity ranges from 43% to 62%. However in recent years it is noticed that the temperature is increasing and the rainfall is decreasing due to industrialization and fast deforestation.

The Nashik district is one the highly industrialized district of Maharashtra. All the different types of industries such as manufacturing of electrical goods, fabrication, consumer goods industries and also there are many more engineering industries. Also, Nashik is the religious city of Maharashtra. The high concentration of the industries, heavy traffic congestion, and human settlement are responsible for the degradation of the air quality.

For the study of the air quality the data is collected from the centre of National Environmental Engineering Research Institute (NEERI) Nagpur site. The three major air pollutants oxides of sulphur (SO_x), Oxides of Nitrogen (NO_x) and Respiratory Suspended Particulate Matter (RSPM) are selected for air quality prediction. Weather data for Nashik is collected from the website of the meteorological department of Maharashtra. The different meteorological parameters are ambient temperature, relative humidity, visibility, wind speed and solar radiation are selected for the model development.

The data provided by the pollution control board and the meteorological department is widely time varying thus need normalization of the data. The normalization of the data is done in the range [0,1]. The process of normalization is carried out in order to carry out training and testing of the data with less prone to error.
Mathematically, the normalization can be done by,
Y= (X-Xmin)/ (Xmax-Xmin).
In this formula, Y is the normalized value; X is the value of the variable. Xmax is the maximum value of the variable and Xmin is the minimum value of the variable.

IV Steps involved in development of the model

Air quality forecasting model is developed using Artificial Neural Network. Since the data of the parameters is time series data thus for developing the neural network for time series data. In this research, the input parameters are meteorological parameters and the output parameters are the air pollutants. Thus an external feature in the form of meteorological parameters is used in the network. For this reason, NARXNET (non-linear autoregressive network with exogenous input) is used for developing the model for predicting the air quality.

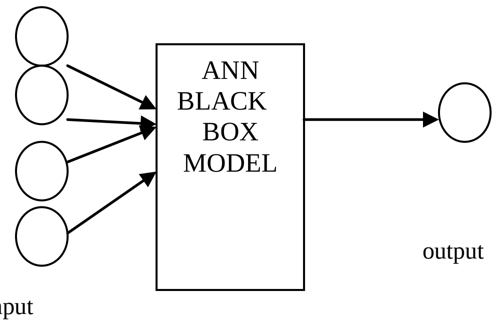

Fig1 ANN model

Setting the data for training, testing and validation:
In order to build the ANN model for air quality prediction separate data set is required for training, testing and validation of the model. Thus the three years data 2010-2012 is divided into three data sets.
It is important to note that training set should be chosen such that it can represent the entire data set including the testing and validation data set. The data set is divided into three categories randomly. During the development of the model the dataset is imported to the network and the network divides the dataset automatically. The percentage of divison of the dataset is as follows:
Training Dataset- 75%
Testing Dataset-15%
Validation Dataset-15%
Forward Selection Method is used for selecting the number and type of input variable in the model. According to the value of 'N' i.e. the number of steps ahead prediction required, computing criterion for the model may vary. When the value of N is small, computing criterion for the prediction model can be selected as RMSE, SSE, Cross Validation etc. whereas if the value of N increases, computational criterion increases. This is the reason to use Forward Selection procedure. This procedure is used by many researchers for building the prediction model (50)
Forward Selection procedure is based on Linear Regression model. The different steps in the model design is as follows:

- Ordering of the explanatory variables according to their correlation with the dependent variable (from the most to the least correlated variables).
- The explanatory variable, which is best correlated with the dependent variable, is selected as the first input.
- All the remained variables are then added one by one as the second input according to their correlation with the output and the variable which most significantly increases the correlation coefficient (R^2) is selected as the second input.
- Repeat the above step N-1 times for evaluating the effect of each variable on model output.
- Finally, among N obtained subsets, the subset with optimum R^2 is selected as the model input subset.

V RESULTS AND DISCUSSION

TRAINING FUNCTION	NUMBER OF HIDDEN LAYERS	NUMBER OF HIDDEN NEURONS	NUMBER OF EPOCHS	ERROR	R	mse
trainscg	1	18		±0.5	0.788	0.048
	2	1,8		±0.4	0.601	0.027
	3	2,6,10	8	±0.6	0.22201	0.016376
	4	2,6,8,10	9	±0.7	0.075739	0.032272
trains	1	18	1000	±0.07	0.992	0.409
	2	1,8	1000	±0.3	0.821	0.342
	3	2,6,10	1000	±0.5	0.89544	0.36633
	4	2,6,8,10	1000	±0.51	0.85904	1.1031
traingdx	1	18	128	±0.7	0.65589	0.082089
	2	1,8	67	±0.65	0.206	0.041
	3	2,6,10	62	±0.87	0.1891	0.047987
	4	2,6,8,10	52	±0.92	0.082697	0.045361
traingdm	1	18	16	±0.49	0.788	0.048
	2	1,8	22	±0.42	0.601	0.027
	3	2,6,10	8	±0.45	0.22201	0.016376
	4	2,6,8,10	9		0.0757	0.0322

ISSN: 2278-0181
Vol. 3 Issue 2, February - 2014

Air Quality prediction model for Sox, NOx and RSPM were developed. The training parameters were changed for the best and optimized results; they were changed according to hit and trial method. During the training of the neural network different training functions were used and also the number of hidden layers along with the number of neurons was varied in order to get back the optimized model for the air quality prediction.
The comparison of different models was carried out on the basis of different parameters such as error rate, mean square error (mse) and R value were selected as the standard parameter for the selection of the best and optimized model. Table illustrates the comparison between these models.
From the table above, it can be observed that that the best network output is given by the single hidden layer network, with 18 number of hidden neurons and by using “trains” as the training function. As the number of hidden layer increases the output of the network starts getting distorted. Thus, higher the number of hidden layers, more error prone network will be received from the above study and analysis of the output.
The best and the optimized model selected from the above table has the following parameters:

- Number of input neurons : 4
- Number of output neurons:3
- Number of hidden layers :1
- Number of hidden neurons:18
- Training Function: trains
- Activation Function: logsig and tansig

Using the above parameters, the training of the network is done using the neural network toolbox.
After training the network, the neural network obtained is as shown in the fig(2)

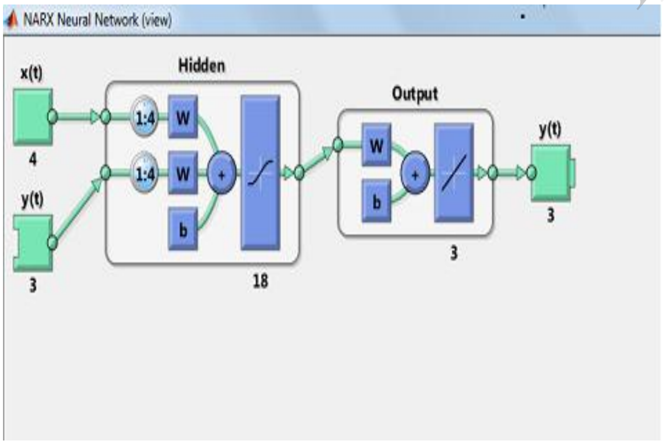

Figure No.2 Artificial neural network for air quality prediction.

As the training process is carried out, the open loop network gives the output which is basically the simple neural network. But this network alone does not allow to prediction for the air quality. Thus for the prediction close loop network is obtained which helps in prediction. The close-loop network for the above architecture is shown in fig-3

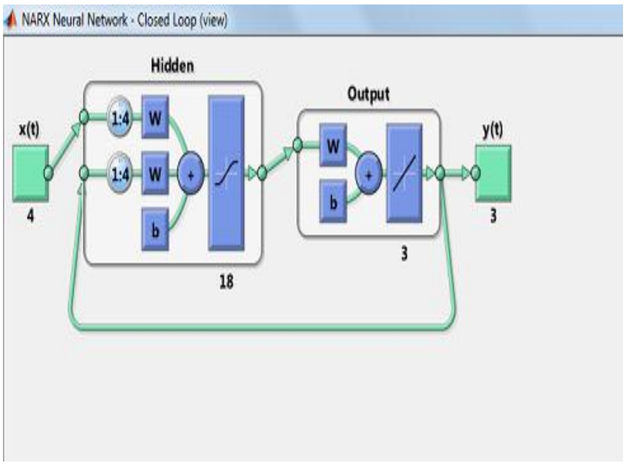

Figure No.3 Close loop network for air quality prediction.

Now using the close loop network, one step ahead prediction is performed. One-step ahead prediction is basically obtained from the output of the close-loop network. The feedback from the output is utilized for prediction. The network obtained for one-step ahead prediction is as shown in fig-3

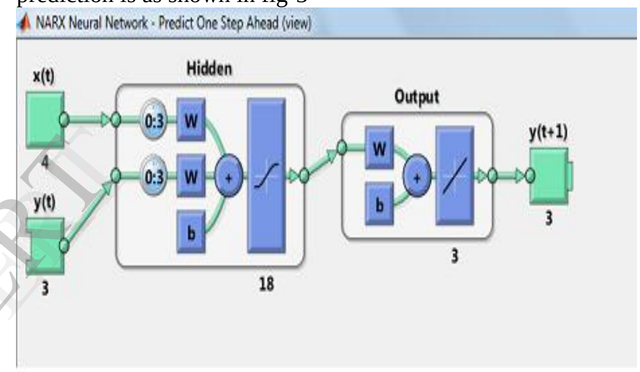

Figure No4 one-step ahead view of the air quality model.

After training the network the next step is the analysis of various results that helps in deciding the capability of the network best suitable for the network.
Fig- 5 (a-f) shows the different results for the above stated network.

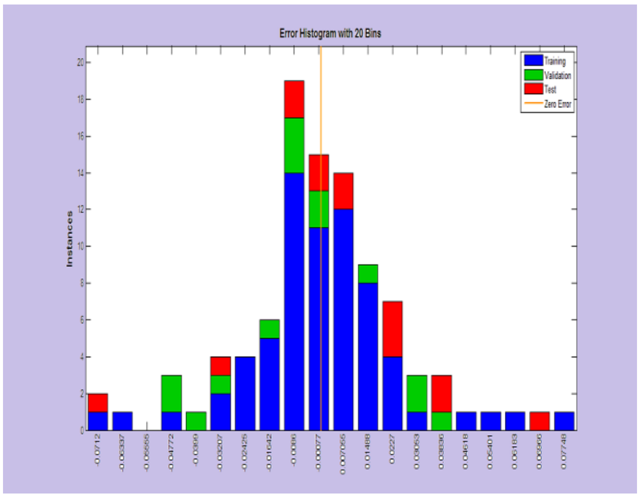

Figure No.5(a) Error Histogram for the network

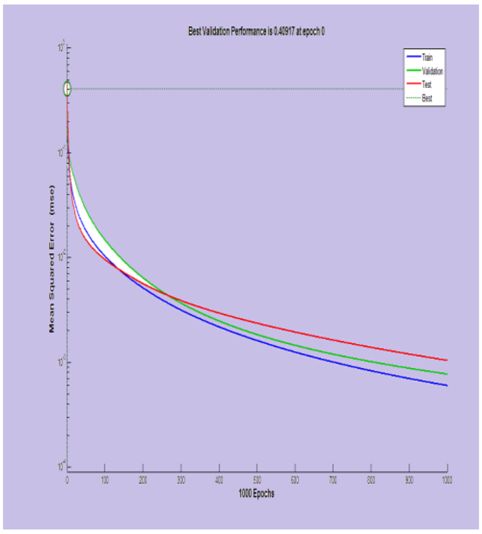

Figure No. 5(b) Performance graph for the network

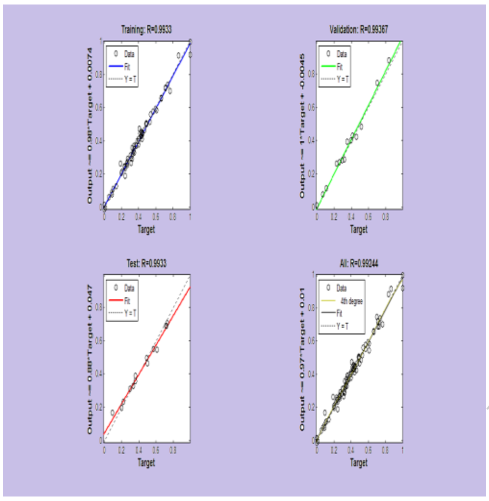

Figure No. 5(c) Regression plot for the network

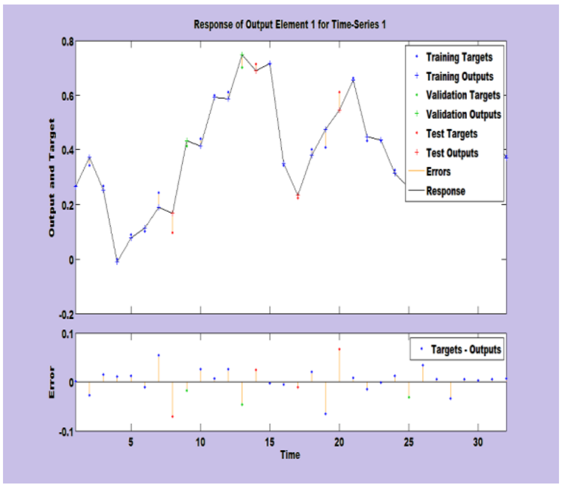

Figure No.5(d)Performance plot for the network

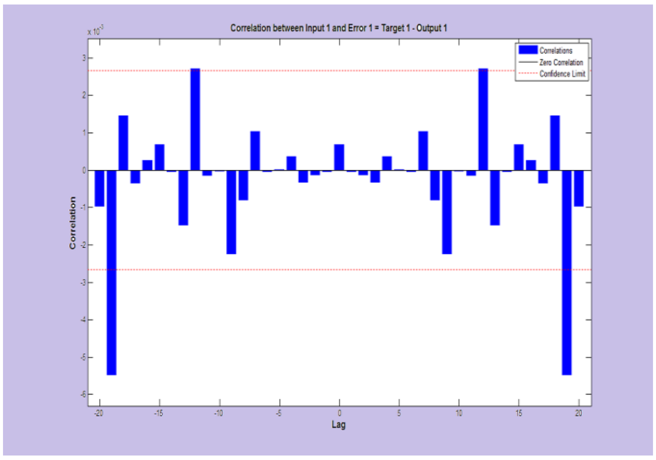

Figure No.5(e) Input-error correlation plot for the network

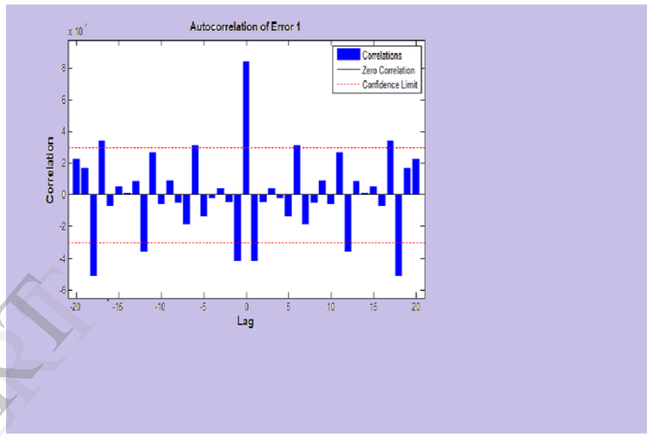

Figure No. 5(f) Error Autocorrelation of the network.

After the analysis of the network, the error between the output and the target in between which is under the confidence limits. Thus, the proposed model can be used be for the prediction of the air quality for the Nagpur city.

This model has the limitation that it cannot be used for the prediction of the air quality for other places since the input parameters used for this network are meteorological parameters whose values vary from place to place and also with as altitude.

IV Conclusion and Future Scope

Based on the results generated in, the air pollution depends on different metrological parameters and different air pollutants. Thus, in order to control the increasing concentration of air pollution it is necessary to take certain measures, necessary for the metrological parameters and automatically there will be a control on the concentration of different air pollutants. The model generated in this thesis gives the most optimized results by predicting the concentration of the different air pollutants in the atmosphere. On the basis of predicted output of our model, preventive measures can be taken by the government of the NAGPUR, Maharashtra state to maintain the human health and environmental balance.

Future Scope of this work can be summarized as follows:

- Present prediction model can be extended by the collection of more reliable and correlated data with more input parameters which can minimize forecasting error

- The proposed model constructed using of the technique comprising only ANN. Thus, due to certain limitations of this model, a hybrid model can be implemented for the air quality forecasting as an extension to this model.

- The model can be extended by taking other impact parameters of weather from the other cities of India and can be implemented a generalized model

REFERENCES

[1] Santosh Kumar Nanda, Debi Prasad Tripathi, S.S. Mahapatra, "Application of Legendre Neural Network for Air- Quality Prediction.", International Conference on Engineering and Technology,May 2011.

[2] N.Sharma, K.K. Chaudhary, CV ChalapatiRao, "Vehicular pollution modelling using artificial neural network technique-A review. " Journal of Scientific and Industrial Research, Vol 64, September 2005, pp. 637 647.

[3] SurendraRao, "Prediction OfParticualte Matter Concentrations Using Artificial Neural Network." Resources and Environment 2012,2(2):30-36 DOI: 10.5923/j.re.20120202.05.

[4] BindhuLal, Sanjaya Shankar Tripathi, "Prediction Of Dust concentration in open cast coal mine using artificial neural network." Atmospheric Pollution Research 3 (2012) 211-218.

[5] MahboudehAfzali, AfsahenAfzali, GholamrezaZahedi,"The Prediction Of Artificial Neural Network Technique in Daily and Monthly Ambient Air Temperature Prediction." International Journal Of Environmental Science and Development , Vol.3, No. 1, February 2012.

[6] MunirHussainNaveed, Sheikh Saeed Ahmad, Sobia Khalid and SehreshKhan,"DevelopmentOf Prediction Model For the Concentration Level Of Air Toxin in the city of Rawalpindi Using Artificial Neural Network. " World Applied Sciences Journal 10(1): 01-08,2010, ISSN 1818-4952.

[7] S. ShrutiTikhe, Dr. Mrs. K.C. Khare, Dr. K.N. Londhe, "Forecasting Criteria Air Pollutants Using Data Driven Approaches; An Indian Case Study. " IOSR Journal Of Environmental Science, Toxicology and Food Technology, Vol. 3, Mar-Apr 2013,PP.01-08, ISSN:2319-2399

[8] SurayaGhazali, Lokman Hakim Ismail, "Air Quality Prediction Using Artificial Neural Network .", IConCEES 2011,3-5 April 2012, Johor Bahru, Malaysia.

[9] RaihaneMechgoug,A.Taleb Ahmed, LakhmissiCherroun, "Optimization Of Neural Predictor For Air Pollution." Proceedings Of the World Congress on Engineering 2012, Vol 2, WCE 2012, July 4-6,2012, London, U.K.

[10] P.Zannetti, D.Al-Ajmi, S.Al- Rashied,"An Introduction to Air Pollution-Definitions, Classification and history." The Arab School for Science and Techonology(ASSST) and The EnviroComp Institute, 2007.

[11] W.Z. LU, H.Y.FAN, A.Y.T. Leung. J.C.K. Wong, "Analysis Of Pollutiant Levels In Central Hong Kong Applying Neural Network Method With Particle swarm Optimization.", Environmental Monitoring and Assessment 79:217-230.2002.

[12] BhupeshGour, AsifUllahkhan,"Atmospheric Condition based Clustering using ART Neural.", JICT, Vol2, February 2012, ISSN 2223-4985.

[13] Kadar Sheref, Dr.S.SanthoshBaboo,"A New Weather Forecasting Technique using BackPropagation Neural Network with Modified Levenberg-Marquardt Algorithm for learning." ,

[14] Ashish Mittal, BhardwajRashmi," Prediction of daily air pollution using wavelet decomposition and adaptive network-based fuzzy inference system. ", International Journal Of Environmental Sciences, Vol2, 2011, ISSN 0976-4402.

[15] S.V. Barai, AK Dikshit, Sameer Sharma,"Neural Network Models for Air Quality Prediction: A Comparitive Study."

[16]Eros Pasero, Luca Mesin,"Artificial Neural Network To Forecast Air Pollution."

[17]Maureen L.Cropper, Nathalie B.Simon, Anna Alberini, P.K. Sharma,"The Health Effects of Air Pollution in Delhi, India", Policy Research Working Paper.

[18]K.P. Moustris,G.T. Proias, I.K. Larissi, P.T. Nastos, K.V.Koukoulestos, A.G. Paliastsos,"Artificial Neural Networks For Surface Ozone Prediction In The Greater Athens Area, Greece.", Protection and Restoration of the Environment.

[19] Vidushi Sharma, SachinRai, AnuragDev,"A comprehensive Study Of Artificial Neural Network.",IJARCSSE, Vol2, Issue10,October 2012, ISSN:2277 128X.

[20] Min Li, Md. RafiulHassan,"Urban Air Pollution Forecasting using Artificial Intelligence-based tools."

[21] Aaron Daly, Paolo Zanneti,"An Introduction to Air Pollution-Definitions, classifications and history."

[22] Aceves-Fernandez Marco Antonio, Sotomayor-OlmedoArtemio, "Advances in Airborne pollution forecasting using soft computing techniques. "

[23] Wei-Zhen Lu, Wen-Jian, "Potential Assesment of the Support Vector Machine method in forecasting ambient air pollution trends.", 2005 ISSN-693–701

[24] Diana Domanska, "Model of air pollution forecasting."

[25] Ming Dong, Dong Yang, "PM2.5 concentration prediction using hidden semi-Markov model-based times series data mining", www.dl.acm.org,2009.

[26] Sun W, Zhang H, "Prediction of 24-hour-average PM(2.5) concentrations using a hidden Markov model with different emission distributions in Northern California." Sci Total Environment,2013.

[27] www.a2gov.org.

[28] www.princeton.edu/step/conference-reports/reports/ch5.pdf.

[29] www.worstpolluted.org/projects_reports/display/86.

[30] Cortes, Corinna; and Vapnik, Vladimir N.; "Support-Vector Networks", Machine Learning, 20, 1995.http://www.springerlink.com/content/k238jx04hm87j80g/

[31] SoawalakArampongsanuwat,Prediction of PM_{10}using Support Vector Regression ,2011 International Conference on Information and Electronics Engineering,IPCSIT vol.6 (2011) © (2011) IACSIT Press, Singapore

[32].Zadeh, L.A. (1965). "Fuzzy sets", Information and Control 8 (3): 338–353

[33].Davison Zimwara, LameckMugwagwa, Knowledge Nherera, "Soft Computing Methods for Predicting Environmental Quality:A Case Study of the Zimbabwe Sugar Processing Industry", Journal of US-China Public Administration, ISSN 1548-6591 April 2013, Vol. 10, No. 4, 345-357.

[34] Ming Dong, Dong Yang, Yan Kuang, David He , SerapErdal, Donna Kenski, "PM2.5 concentration prediction using hidden semi-Markov model-based times series data mining" Expert Systems with Applications 36 (2009) 9046–9055

[35].Shine –Wei Lin, Chih –Hong Sun, and Chin-Han Chen , Temporal data mining using Genetic Algorithm and Neural Network –A case study of Air pollutant forecasts.

[36]Elias Kalapanidas and Nikos Avouris, Feature Selection using a Genetic Algorithm applied on an Air Quality Forecasting Problem. Available www.hci.ece.upatras.gr/pubs.../C54_Kalapanidas_Avouris_BESAI_2002.pdf

[37]LiminWang,A PSO-based Model for Assessment of Atmospheric Quality,IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.3A,

[38] www.mathswork.com

[39]www.census2011.co.in

[40] Wikipedia.com

[41]Brook,2007

[42]pope,1991

[43]http://visual.merriam-webster.com/earth/environment/air-pollution.php

[44]Ulkusahin,2005

[45]Thomas,2012

[46]Perez,2012

[47]Kukkonen

[48]Asha B. Chelani, "Prediction of daily maximum ground ozone concentration using support vector machine" February 2009, springer science.

[49] Prediction of PM10 using Support Vector Regression SoawalakArampongsanuwat 1 and PhayungMeesad 2, 2011 International Conference on Information and Electronics Engineering IPCSIT vol.6 (2011) © (2011) IACSIT Press, Singapore.

[50] Chen et al., 2004; Eksioglu et al., 2005; Wang et al., 2006; Khan et al., 2007