Single hidden layer artificial neural network models versus multiple linear regression model in forecasting the time series of total ozone

G. Bandyopadhyay, *S. Chattopadhyay

Department of Information Technology of Pailan College of Management and Technology under West Bengal University of Technology, India

ABSTRACT: Present paper endeavors to develop predictive artificial neural network model for forecasting the mean monthly total ozone concentration over Arosa, Switzerland. Single hidden layer neural network models with variable number of nodes have been developed and their performances have been evaluated using the method of least squares and error estimation. Their performances have been compared with multiple linear regression model. Ultimately, single-hidden-layer model with 8 hidden nodes have been identified as the best predictive model.

Key words: *Arosa, total ozone, single-hidden-layer, artificial neural network, multiple linear regression, forecast*

INTRODUCTION

Processes involved in the formation of ozone (O_3) are highly multifaceted in nature. Ozone is a secondary pollutant and is not usually emanated straightforwardly from stacks. Instead is formed in the atmosphere as a result of reactions between other pollutants emitted mostly by industries and automobiles. The ozone precursors are generally divided into two groups, namely oxides of nitrogen (NOX) and volatile organic components (VOC) like evaporative solvents and other hydrocarbons. In suitable ambient meteorological condition (e.g. warm, sunny/clear day) ultraviolet radiation (UV) causes the precursors to interact photochemically in a set of reactions that result in the formation of ozone.

The process of ozone formation can be expressed as:

$$NO_2 + UV \rightarrow NO + O$$
$$O + O_2 + M \rightarrow O_3 + M$$
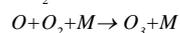
Where, M is a third body molecule that remains unchanged in the reaction.

Ozone produced this way gets simultaneously destroyed as:

$$O_3 + D \rightarrow DO + O_2$$
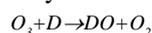
Where, D implies additional reactant that destroys the ozone via oxidation.

Because of its capability to absorb the incoming radiation, the stratospheric ozone is a major source of stratospheric heating, which further heats the troposphere. Again, because of radiation of IR the tropopause gets some cooling. Thus, stratospheric ozone exerts both heating and cooling effect on the land-troposphere system. Total ozone is a measure of the number of ozone molecules between the ground and the top of the atmosphere. In a more mathematical language, total ozone is simply the integral of the ozone concentration with respect to height. A literature survey by the authors of the present contribution has shown that statistical time series analysis approach in forecasting the atmospheric and environment pollution has been proved viable by a number of researchers (e.g. Milionis and Devis, 1994; Shi and Harrison, 1997 and many others). But, in recent times, artificial neural network (ANN) has been proposed by many scientists as a better alternative to the conventional regression approach in forecasting time series pertaining to complex atmospheric and environmental phenomena. Since formation of ozone is a highly intricate phenomenon, a number of researchers have concentrated on its prediction and consequently comparative studies have been carried out to discern the performance of ANN over conventional regression approach in predicting tropospheric ozone over different cities. Prybutok, *et al.,* (2000), Balaguer Ballester, *et al.,*(2002), Nunnari, *et al.,* (1998), Viotti, *et al.,* (2002) put ANN into practice in a number of case

*Corresponding author, Email: *surajit_2008@yahoo.co.in*
Tel./Fax: +9198 3073 6116

studies where they established supremacy of ANN over customary methodologies in ozone forecasting. Corani (2005) implemented ANN, pruned ANN, and lazy learning to predict ozone concentration over Milan. But, in all the aforesaid works, ANN models have been developed with other meteorological variables as predictors. In none of the ANN applications to ozone concentration, past values of the same data have been used as predictor. Since all other meteorological variables have their own chaotic characteristics and complexities, their inclusion to the input set would incorporate more complexity in the forecasting. Present approach viewed the prediction problem from a different point of view. Instead of tropospheric ozone, total ozone has been considered as the predictand and instead of incorporating other meteorological variables; past values of the total ozone time series have been explored in developing ANN predictive models. Performance of ANN models has been compared with conventional regression model. The research work explained in the present paper has been done in Kolkata, India, during the period April-June 2006.

MATERIALS AND METHODS
Artificial Neural Networks, an overview

Artificial Neural networks are mathematical analogue of biological neural systems, in the sense that they are made up of an interconnected system of nodes (neurons). Furthermore, a neural network can recognize patterns in numeric data in a similar fashion to the learning process in its biological counterpart. Neural networks are highly robust with respect to underlying data distributions (non-parametric), and no assumptions are made about relationships between variables (unlike the linear or pre-specified curvilinear relationships in regression). Since it does not depend upon any assumption regarding the data and it is robust to chaotic behavior of the data, ANN has opened up new avenues to pattern recognition and forecasting related to complex natural processes. According to Maqsood, *et al.,* (2002) ANNs are highly suitable to the cases where the underlying processes are less recognizable and are characterized by chaotic features. Though ANN can be used in a number of complex problems, the basic jobs that can be performed by ANN can be precisely enlisted as
• Pattern recognition
• Pattern classification
• Prediction

Since the present paper involves prediction problem, the features of ANN are discussed here from the point of view of prediction. The earliest trainable layered neural networks with multiple adaptive elements was the Madaline I structure of Widrow and Hoff (for details see Kartalopoulos, 2000). This ANN model had only two layers. The first one was the adaptive layer and the last layer was an output layer with fixed threshold function. Because of less flexibility, this Madaline I ANN of the early 60's could not prove itself suitable for practical purposes. Advent of the feed forward ANN or Multilayer Perceptron (MLP) with Backpropagation learning, an adaptation of the steepest descent method, opened up new avenues for the application of ANN for problems of practical interest (Kamarthi and Pittner, 1999; Gardner and Dorling, 1998; Hsieh and Tang, 1998). In MLP, each network consists of several simple processors called neurons, or cells, which are highly interconnected and are arranged in several layers. There are three basic types of layers: input layer, hidden layer(s), and output layer. The input and output layers are connected through hidden layer(s). There may be one to several hidden layers in between input and output layer. In mathematical form, the adaptive procedure of a feed forward MLP can be presented as (Kamarthi and Pittner, 1999):

$$w_{k+1} = w_k + \eta d_k \tag{1}$$

The above equation represents an iteration process that finds the optimal weight vector by adapting the initial weight vector w_0. This adaptation is performed by presenting to the network sequentially a set pairs of input and target vectors. The positive constant η is called the learning rate. The direction vector d_k is the negative gradient of the output error function E. Mathematically it is denoted as

$$d_k = -\nabla E(w_k) \tag{2}$$

The Backpropagation algorithm in which the weights of the network are updated immediately after the presentation each pair of input and target output is called the sequential learning. The other learning procedure in which the whole training set is considered as a batch is called the batch learning.

For sequential learning,

$$E(w_k) = E_p(w_k) \tag{3 a}$$

For batch-learning

$$E(w_k) = \Sigma_p E_p(w_k) \tag{3 b}$$

Where, *Ep*(*wk*) denotes the mean squared error (MSE)for a pair of input and output. The learning or training process may be supervised, unsupervised or competitive. The aim of supervised learning is to find out a set of weight vectors that minimizes the deviation between network output and the target output over the whole training. The optimal weight matrix obtained this way is applied to the test set to investigate the viability of the model.

Data and analysis

Present paper deals with mean monthly total ozone concentration in Arosa, Switzerland between 1932 and 1971. The measurements are taken in Dobson Units (DU) (300 DU=1layer of 3mm if the whole ozone column is taken at the sea level with standard conditions). First of all, an autocorrelation analysis is performed on the dataset. The autocorrelation coefficients (for lag k) are computed as (Wilks, 1995)

$$r_k = \frac{Cov\left[\left(\overline{x(n-k)}\right),\left(\underline{x}(n-k)\right)\right]}{\sqrt{Variance\left(\left(\overline{x(n-k)}\right)\right)}\sqrt{Variance\left(\left(\underline{x}(n-k)\right)\right)}} \quad (4)$$

Where, r_k denotes the autocorrelation of order *k*, $\overline{x}(n-k)$ denotes the first (*n-k*) data values, $\underline{x}(n-k)$ denotes the last (*n-k*) data values, and *k* varies from 1 to *n*. The autocorrelation coefficient (ACC) values are computed from the available data set up to 40 lags and their magnitudes are presented in Fig. 1. The Fig. makes it clear that the high ACC values are occurring at the lags 1,6,12,18 etc. That is, at lags separated by 5 time points. Thus, it can be inferred that mean monthly total ozone time series is completing a cyclic pattern in every 6 months. Thus, a predictive model can be developed with 5 months' data as predictor and the 6th month's data as predictand.

Multiple linear regression model

A multiple linear regression model is proposed as

$$\hat{y} = aX1 + bX2 + cX3 + dX4 + eX5 + \varepsilon \quad (5)$$

Where, the left hand side of equation (5) implies

the predicted value of mean monthly total ozone in the 6th month with X1, X2, X3, X4 and X5 as predictors i.e. the mean monthly total ozone in months 1,2,3,4 and 5. The constants a, b, c, d and e are the regression parameters computed by the method of least squares (Wilks, 1995). In the present study the regression equation comes out to be

$$\hat{y} = 0.30\ X1 + 0.18\ X2 + 0.10\ X3 + 0.20\ X4 + 0.54\ X5 + 282$$

From this linear multiple regression fit it comes out that the coefficient of determination is 0.7512, which is not very far from 1. Thus, this linear fit is not a very bad predictive model. The values of t-statistic are computed from this model and are compared to the critical t-statistic value. The comparison is presented in Fig. 2, which shows that only the value of the time series in month 3 falls just below the critical value. Thus, it can be said that mean monthly total ozone concentration in the month *t,* depend heavily upon the values in the months *t*-5, *t*-4, *t*-2 and *t*-1; but does not depend significantly upon the month *t*-3.

Single hidden layer ANN models

In the previous discussion it has been established the value corresponding to the month *t*-3 does not depend influence the month *t* very significantly. Since ANN models can work even with not-so-correlated predictors, all the values can be considered on the same foot while constructing single-hidden-layer ANN models. In this paper, seventeen single-hidden-layer ANN models have been developed with Backpropagation algorithm explained in equation (1) through (3b). The models have been denoted as H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, and H16and H17. Here Hi implies that the model contains 'i' number of nodes in the hidden layer. All the models have been trained separately with sigmoidal activation function up to 500 epochs sequentially with the aim that the mean squared error (equation (3b)) is to be minimized. From each input set, 75% data are considered as the training data and the remaining 255 data are considered as the test or validation data. To say more specifically, 120 data have been considered as validation set and 355 data are considered as training set. While training, learning rate has been taken as 0.9 and momentum rate has been taken as 0.2.
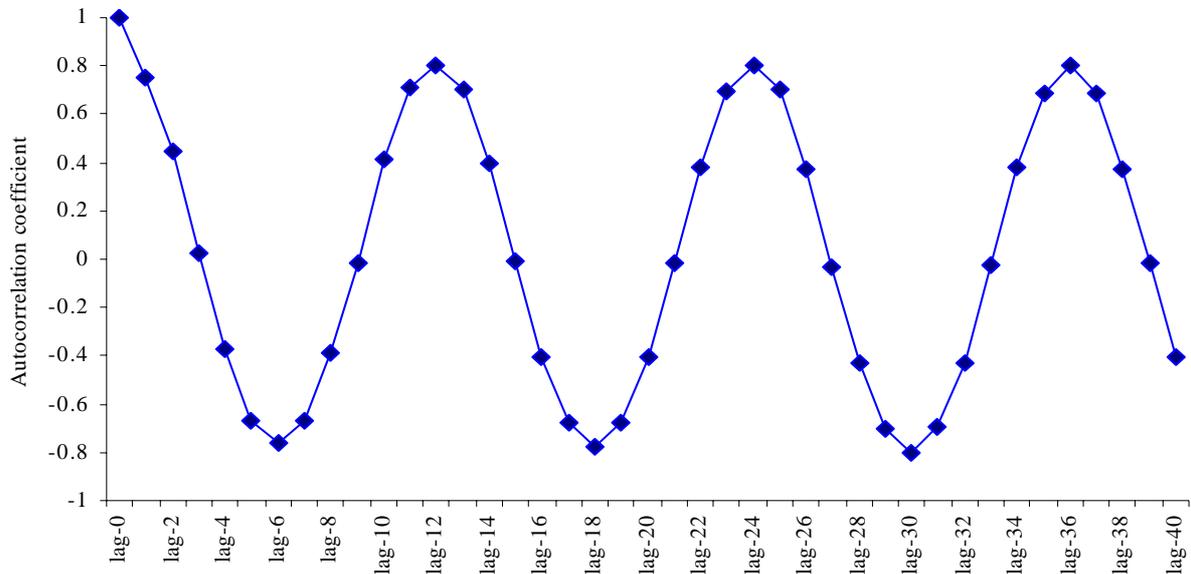
Fig. 1: Schematic showing the ACCs for different lags pertaining to mean monthly total ozone time series over Arosa, Switzerland between 1932 and 1971. The lags are taken in months
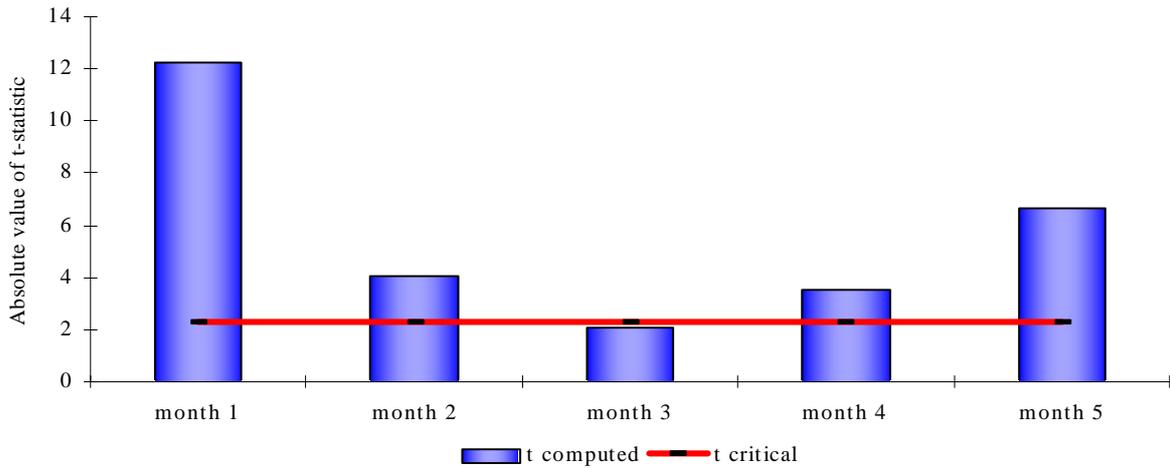

Fig. 2: Schematic showing the computed from multiple linear regression (MLR) model and critical t-statistic values

RESULTS

First of all, coefficients of determination conventionally denoted as R^2 (Comrie, 1997) are computed for all the predictive models. It has been mentioned earlier that R^2 for multiple linear regression (MLR) model is 0.7512. Now, R^2 are computed for all the 17 ANN models and are presented in tabular form in Table 1. To compare the performance of different ANN models between themselves and with MLR model, Fig. 3 is prepared. This Fig. shows that all the ANN models excepting H2 and H4 produce higher CD than MLR model. Thus, H2 and H4 are discarded from our discussion. From the rest of the models it is found that H1 produces almost the same CD as by MLR. Thus, a single hidden layer model with only one node does not produce any overwhelmingly better prediction than MLR.
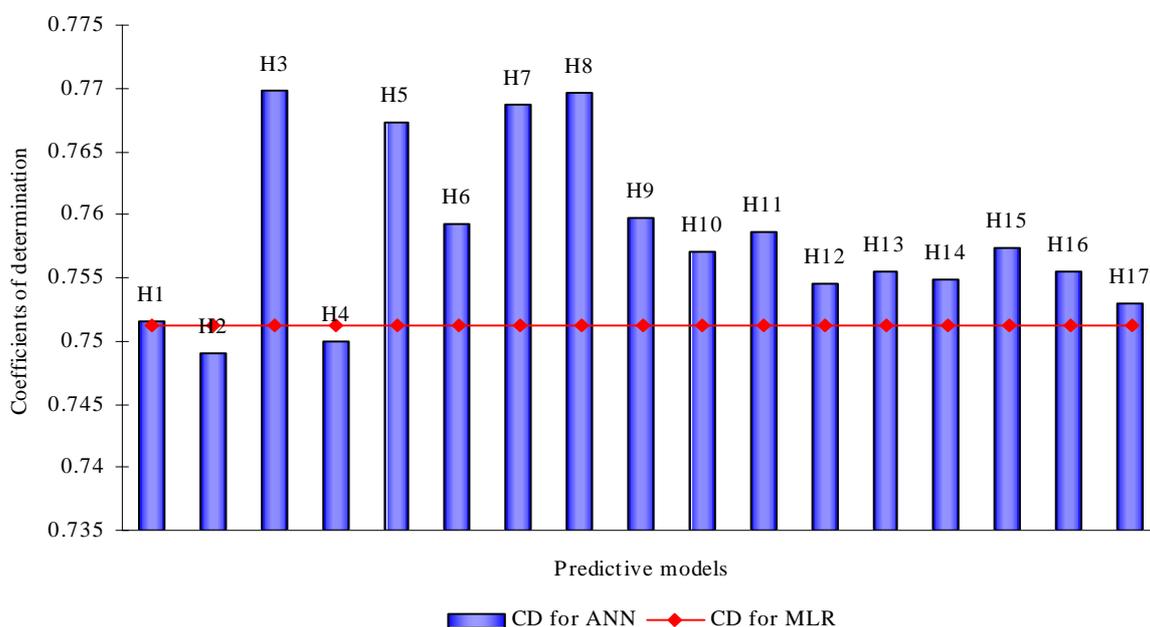
Fig. 3: Coefficients of determination for different ANN models and multiple linear regression model

Models H6, H9-H17 produce significantly better prediction than MLR but are distribute more or less regularly around an average value. It is very important to note that up to H8, addition of hidden nodes is producing significant changes in the CD, but after that the CDs did not vary drastically and all of them are significantly lower than that produced by H8. From here it can be inferred that after 8 hidden nodes, adding more hidden nodes does not create any room for further improvement in the prediction. The best CDs are produced by H3, H5, H7 and H8. The next task is to identify the best model among these four. Now, the prediction error (PE) (Perez and Reyes, 2001) over the whole test set are computed for each of the four models mentioned above and are schematically presented in Fig. 4. This Fig. clearly shows that the least PE is being produced by H8. From here it can be said that H8 is the best ANN model to forecast the mean monthly total ozone concentration over Arosa, Switzerland. But, to reach the final conclusion some qualitative tests are being executed. Scatter plots for the actual and predicted values corresponding to H3, H5, H7 and H8 are presented in Fig. 5a-d. The least square regression fits to the models are also presented in the Figs. It is apparent from those equations that the least y-intercept (79.021) by the trend equation occurs for H8. The other

y-intercept values are 80.023 (H3), 83.515 (H5) and 80.865 (H7). Since for highest correlation between actual and prediction the y-intercept becomes 0, thus, the lowest y-intercept producing model (H8) can be considered as the best predictive model. After all these tests, H8 is identified as the best predictive model and the prediction by this model with the actual values are presented in Fig. 6. A very close association between the actual and the prediction can be viewed from here. The structure of the final model (H8) is presented in Table 2.

DISCUSSION AND CONCLUSION

The rigorous study executed above leads us to conclude some important characteristics of the time series pertaining to total ozone time series over Arosa, Switzerland. The autocorrelation study reveals that mean monthly total ozone concentration over Arosa, Switzerland follows a cyclic pattern with 6 months cycle. Furthermore, the third month within the cycle has less significant impact upon the sixth month. Artificial Neural Network with Backpropagation learning has been established to be a potent predictive tool for the said time series. A comparative study with respect to prediction ability reveals that single hidden layer predicts the mean monthly total ozone concentration

over Arosa, Switzerland more efficiently than multiple linear regression model. A further comparative study among different neural net models proves that single-hidden-layer Artificial Neural Network with 8 hidden nodes is the best predictive model to forecast the mean monthly total ozone concentration over Arosa using the past data values.

Table 1: Tabular presentation of the CD (R2) for different single hidden layer ANN models

ANN model	R^2	ANN model	R^2
H1	0.751587223	H10	0.75711867
H2	0.749037151	H11	0.758648011
H3	0.769796595	H12	0.75457583
H4	0.749937679	H13	0.755545439
H5	0.767296424	H14	0.754808694
H6	0.759243692	H15	0.757303296
H7	0.768768298	H16	0.755478693
H8	0.769620271	H17	0.75292268
H9	0.759791222		

Table 2: The final weight matrix obtained through Backpropagation training of a single-hidden-layer ANN with 8 hidden nodes after 500

Hdn1_Nrn1	0.4733	-2.3168 -	0.6795	-1.0448	-0.7702	-2.9610
Hdn1_Nrn2	2.7856	-4.5471 -	3.6978	-0.5105	-0.5843	1.9933
Hdn1_Nrn3	-0.8968	0.1803	0.1131	-0.7821	-0.7530	-1.0905
Hdn1_Nrn4	3.1582	-0.8482 -	0.9615	-0.4335	-0.8715	-2.4837
Hdn1_Nrn5	-1.2774	-1.5098	-0.6341	-1.0410	-1.3477	-0.4665
Hdn1_Nrn6	-1.6250	1.5870	2.4146	-0.3212	-1.7871	-4.6208
Hdn1_Nrn7	-1.0507	-0.1162	-0.2167	-0.4801	-0.6661	-0.7367
Hdn1_Nrn8	-1.1073	-0.1166	-0.9859	0.1465	-0.7951	-0.5338
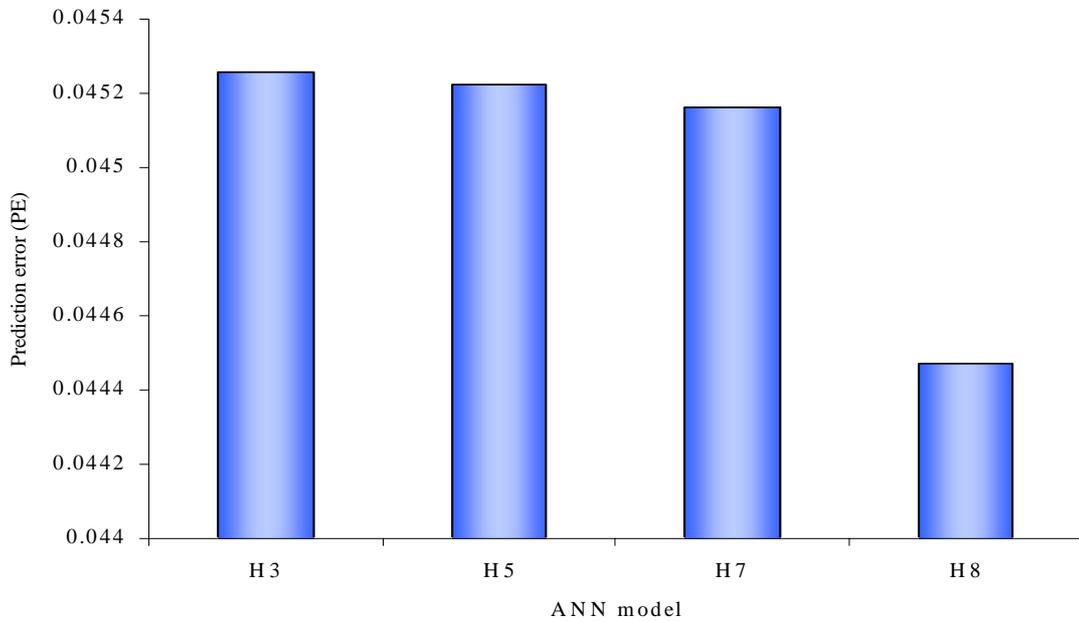
Fig. 4: Prediction errors (PE) for the ANN models with maximum coefficients of determination

(a)

$y = 0.7656x + 80.023$
$R^2 = 0.7698$

(b)

$y = 0.7542x + 83.515$
$R^2 = 0.7673$

(c)

$y = 0.764x + 80.865$
$R^2 = 0.7688$

(d)

$y = 0.7667x + 79.021$
$R^2 = 0.7696$
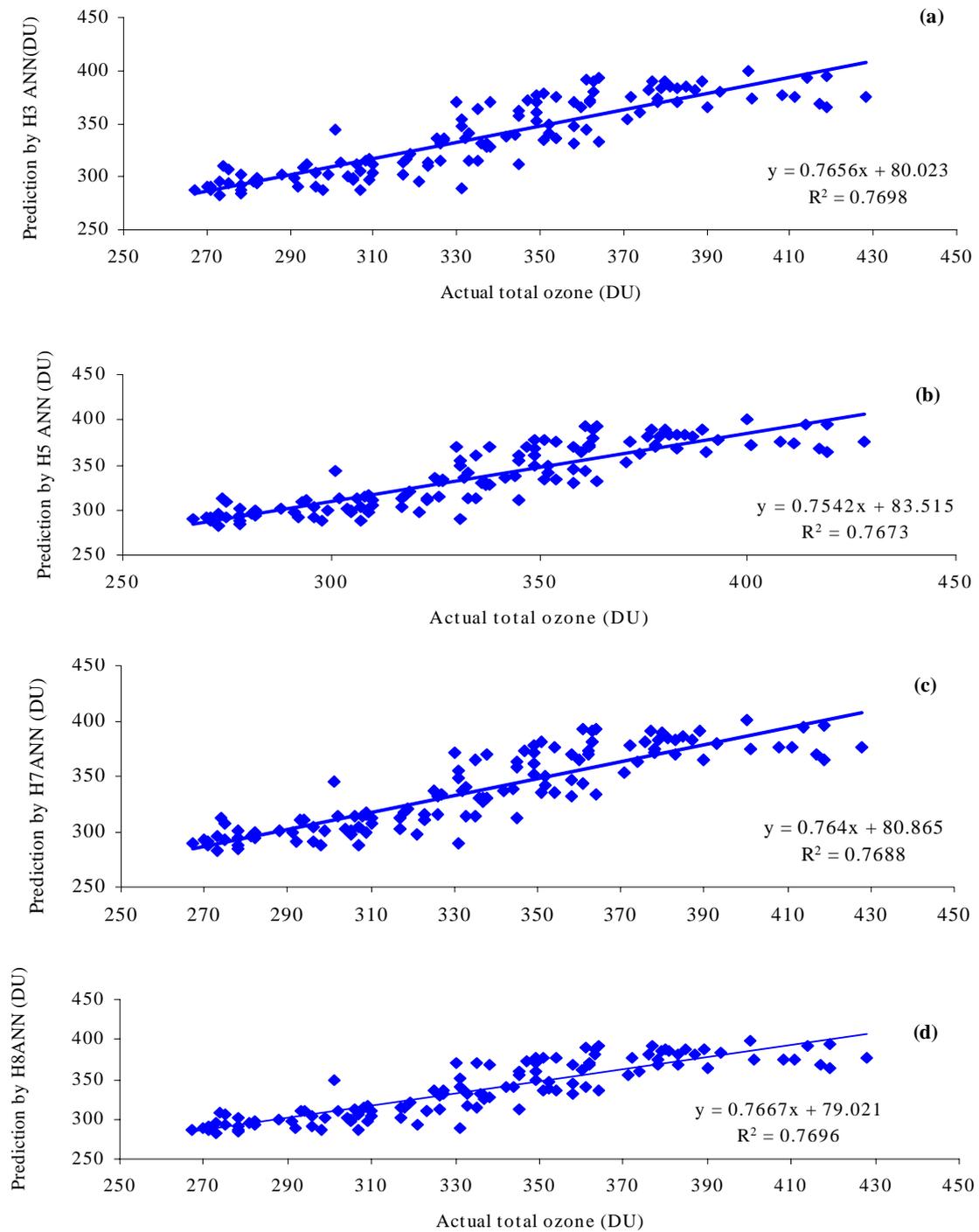
Figs. 5a-d: Scatterplots of actual versus prediction. Trend equations are presented in the corners with the coefficients of determination (R^2)
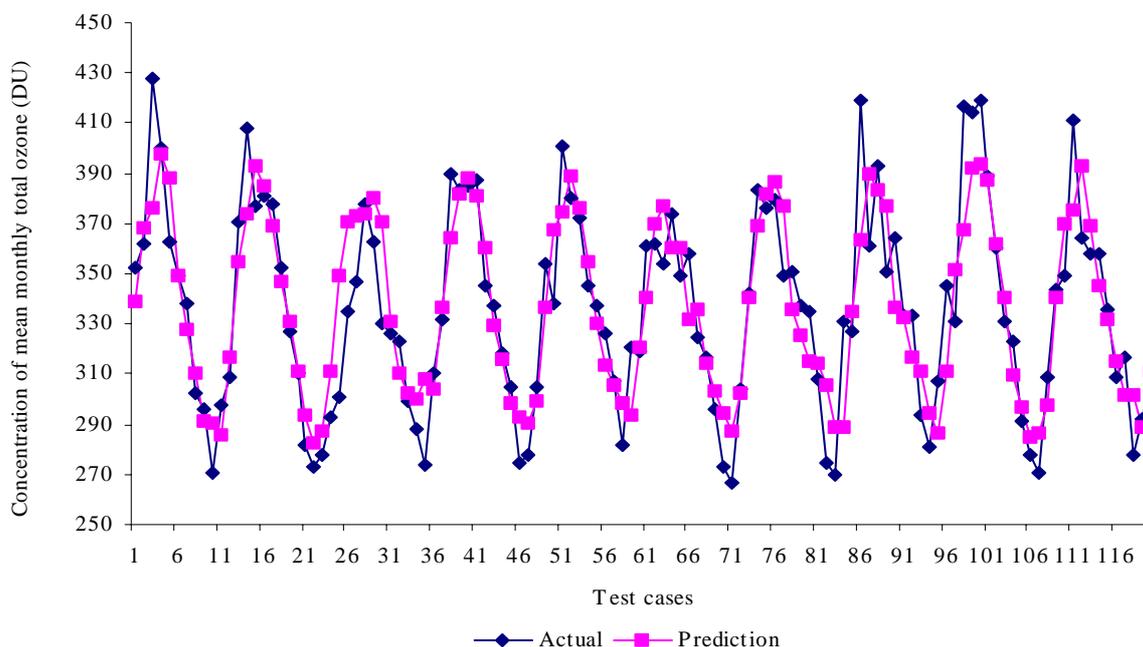
Fig. 6: Schematic showing the actual and predicted (by H8 ANN) mean monthly total ozone concentrations (the test cases)

REFERENCES

Balaguer Ballester, E., Valls, G., Carrasco-Rodriguez, J., Soria Oliva, E., Valle-Tascon, S., (2002). Effective 1-day ahead prediction of hourly surface ozone concentrations in eastern Spain using linear models and neural networks. Ecologic. Model., **156**, 27-41.

Corani, G., (2005). Air quality prediction in Milan: feed-forward neural networks, pruned neural networks and lazy learning, Ecologic. Model., **185**, 513-529.

Gardner, M.W., Dorling, S.R., (1998). Artificial Neural Network (Multilayer Perceptron)-a review of applications in atmospheric sciences. Atmos. Environ., **32**, 2627-2636.

Hsieh, W.W., Tang, T., (1998). Applying Neural Network Models to prediction and data analysis in meteorology and oceanography. Bulletin of the American Meteorological Society, **79**, 1855-1869.

Kartalopoulos, S.V., (2000). Understanding Neural Networks and Fuzzy logic-basic concepts and applications, Prentice Hall, New-Delhi

Kamarthi, S.V., Pittner, S., (1999). Accelerating neural network training using weight extrapolation. Neur. Net., **12**, 1285-1299.

Maqsood, I., Muhammad, R.K., Abraham, A., (2002). Neurocomputing based Canadian weather analysis. computational intelligence and applications. Dynamic Publishers Inc., USA, 39-44.

Milionis, A.E., Davis, T.D., (1994). Regression and stochastic models for air pollution, Part I: review comments and suggestions. Atmos. Environ., **28** (17), 2801-2810.

Nunnari, G., Nucifora, M., Randieri, C., (1998). The application of neural techniques to the modelling of time-series of atmospheric pollution data. Ecologic. Model., **111**, 187-205.

Prybutok, R., Junsub, Y., Mitchell, D., (2000). Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. Euro. J. Operat. Res., **122**, 31-40.

Shi, J.P., Harrison, R.M., (1997). Regression modeling of hourly NOx and NO2 concentration in urban air in London. Atmos. Environ., **31** (24), 4081-4094.

Viotti, P., Liuti, G., Di Genova, P., (2002). Atmospheric urban pollution: application of an artifical neural network to the city of Perugia. Ecologic. Model., **148**, 27-46.

Wilks, D. S., (1995). Statistical methods in atmospheric sciences, Academic Press, USA

AUTHOR (S) BIOSKETCHES

Bandyopadhyay, G., B.Sc. in chemistry, M.Sc., in Atmospheric Sciences from the University of Calcutta, India. Formerly, she was a junior research fellow of University of Calcutta. Ph.D. student in atmospheric sciences in the University of Calcutta. Email: *goutami15@yahoo.co.in*

Chattopadhyay, S., B.Sc. in Mathematics from Saint Xavier's College under the University of Calcutta, M.Sc. in applied mathematics from Jadavpur University, India. He is now as lecturer in the Department of Information Technology of Pailan College of Management and Technology under West Bengal University of Technology, India. Email: *surajit_2008@yahoo.co.in*