Performance Comparison between HMLP, MLP and Recurrent Networks with Applications to Carbon Monoxide Concentrations Forecasting

USHARANI RAJI[1], M.Y.MASHOR[2], A.H.ADOM, A.N.ALI and A.F. SADULLAH
Control and ELectronics Intelligent System (CELIS) Research Group,
School of Electrical & Electronic Engineering, Engineering Campus,
University Sains Malaysia,
14300 Nibong Tebal, Pulau Pinang,
MALAYSIA.
Email:vsharani@yahoo.com[1], yusof@eng.usm.my[2]

Abstract: - This paper compares the performance of Hybrid Multilayered Perceptron (HMLP) network, Multilayered Perceptron (MLP) network and Recurrent network. These networks are used to model and forecast carbon monoxide (CO) concentration. Two data sets are used for the comparison, one data set from simulated environment and one real data set obtained from Malaysian Environmental Department (ASMA). The forecasting performances of these models are evaluated using index of coefficient (R^2), one step ahead prediction (OSA) and multi step ahead prediction (MSA). The results obtained from both data sets indicate that HMLP network gives the best performance compared to MLP and Recurrent networks.

Key-Words:- carbon monoxide, forecasting, hybrid multilayered perceptron, multilayered perceptron recurrent network

1 Introduction

The impact of urban air pollution is broad especially towards human beings (WHO, 1987), since it can cause irritation, odour annoyance, acute and long term toxic effects [1]. Carbon Monoxide (CO) is a primary pollutant in urban area, due to the major emission from motor vehicles. CO is produced from incomplete burning of carbon contained fuels. According to the Journal of American Medical Association (JAMA), 1500 people die annually due to accidental CO poisoning and 10000 people seek medical attention [2]. Forecasting of CO or other gas pollutants concentration are very important since preventive action can be taken if the forecasted CO level exceeds certain value.

A lot of researches have been carried out using different methodology on CO concentrations forecasting. One of the methods was by using univariate linear stochastic models based on Box-Jenkins modelling technique [3]. This model sufficiently needs long historical data set for model formulation. Another approach was by using Box-Jenkins transfer function noise model (TFN) [4]. The forecasting performance was better compared to the first approach presented in [3]. Besides that, Gaussian and regression models were implemented for CO forecasting [5] [6]. In another study, performance comparison between the use of dispersion and stochastic models were carried out. It was reported that stochastic model performed better than dispersion model to predict the hourly mean value of CO concentrations [7].

Lately, the application of neural networks (NN) becoming very popular for forecasting air pollutants concentration. NN have been proved mathematically to be capable of representing nonlinear systems. A NN known as "Brainmaker" using back propagation algorithm was used to predict CO concentrations with an accuracy of R^2=0.69 [8]. Forecasting on other gases using NN were reviewed since not much of studies have been done specifically on implementation of NN on CO concentrations forecasting. The prediction of hourly time series of NO_2 was carried out using MLP network, the R^2 obtained was 0.96 [9]. In another study, AR model was used for prediction of NO_2 and NO_X concentrations with an accuracy of R^2=0.69 and 0.42, respectively [10]. The results obtained from [10] were compared with the implementation of MLP network by using the same data set. MLP network was found to perform better than AR model with an accuracy of R^2=0.86 and 0.88 [11]. In another study, prediction of $PM_{2.5}$ concentrations was carried out by using multilayer neural network, linear regression and persistence models. The predictions produced by these methods were compared and NN was found to give

the best result [12]. In another study, recurrent network with feedback in the hidden layer was used to predict SO_2 concentration[13]. The network was trained using Levernberg-Marquadt algorithm. The results obtained from recurrent network were compared with those obtained from multivariate regression model. The results indicated that neural network gave better prediction with less residual mean square error than those given by multivariate regression models.

In the present study, CO concentrations forecasting performance will be compared between HMLP, MLP and Recurrent networks. The HMLP network is trained using Modified Recursive Prediction Error (MRPE) algorithm. The MLP and Recurrents networks are trained using Levernberg-Marquadt algorithm. Their performances are evaluated using R^2 test, OSA and MSA test, respectively.

2 Neural Network Models

A hybrid multilayered perceptron with one hidden layer is shown in Figure 1. HMLP network with one hidden layer can be expressed by the following equation:

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 F\left(\sum_{i=1}^{n_i} w_{ij}^1 v_i^0(t) + b_j^1 \right) + \sum_{i=0}^{n_i} w_{ik}^{\ell} v_i^0(t); \quad for 1 \le k \le m \qquad (1)$$

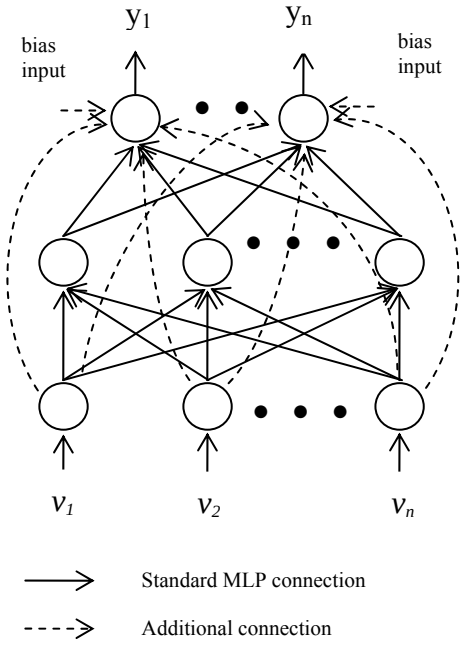

Fig.1 Hybrid Multilayered Perceptron

A multilayered perceptron with one hidden layer can be defined as shown in Equation (2).

$$\hat{y}_k(t) = \sum_{j=1}^{n_h} w_{jk}^2 F\left(\sum_{i=1}^{n_i} w_{ij}^1 v_i^0(t) + b_j^1 \right) \qquad (2)$$

where w_{ij}^1 denotes the weights that connect the input and the hidden layers; b_j^1 and v_i^0 represents the threshold in hidden nodes and input supplied to the network; w_{jk}^2 denotes the weights that connect the hidden and output layer; w_{ik}^{ℓ} are the weights connection between input and output layer; n_i and are the number of input nodes and hidden nodes; m represents the number of output nodes while $F(\bullet)$ is an activation function which is normally selected as sigmoidal function.

The weights w_{jk}^2, w_{ik}^{ℓ}, w_{ij}^1 and b_j^1 are unknown, and should be selected carefully in order to achieve minimum prediction error, defined as below:

$$\varepsilon_k(t) = y_k(t) - \hat{y}_k(t) \qquad (3)$$

where $y_k(t)$ and $\hat{y}_k(t)$ are the actual and predicted output.

In this study, recurrent network called as Elman network is applied for the comparison studies. Elman networks are commonly structured as two layer back propagation networks, with the additional feedback connection from the output of hidden layer to its input. The feedback connection allows the network to both recognize and generate time-varying patterns. An Elman network with one hidden layer is shown in Figure 2.

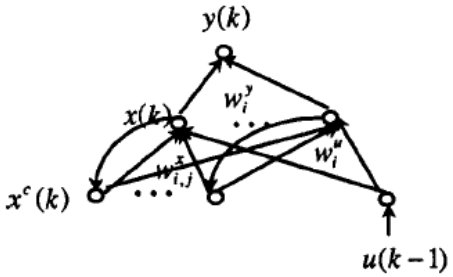

Fig.2 Elman network

The descriptive equations of Elman network can be written as shown below:

$$x^c(k) = x(k-1) \qquad (4)$$

$$x_i(k) = f\left(\sum_{j=1}^{N} w_{i,j}^x x_j(k-1) + w_i^u u(k-1) \right) \qquad (5)$$

$$y(k) = \sum_{i=1}^{N} w_i^y x_i(k) \qquad (6)$$

where $w_{i,j}^{x}$ is the weight that connects *i*-th hidden layer neuron and *j*-th context layer neuron; w_i^u is the weight linking the input neuron $u(k-1)$ and the *i*-th hidden layer neuron; w_i^y is the weight that connects output neuron $y(k)$ and the *i*-th hidden layer neuron; $f(\bullet)$ represents activation function in the hidden layer node and *N* is the number of hidden layer nodes. Usually, sigmoidal activation function is used for application to modelling non-linear systems.

In the literature, MLP and Recurrent networks were used to perform gasses forecasting. Besides that, both networks were trained using Levernberg-Marquadt algorithm. That is the main reason both networks were chosen for this comparison studies.

3 CO Concentrations Forecasting using HMLP network

In this section, the performance of HMLP network together with MRPE algorithm has been evaluated using one simulated environment data set and one real data set. The simulated environment data set contain 500 data samples which were sampled every 10 seconds. The real data set contain 1000 data samples consisting of hourly CO concentration measurements. In this study, the number of steps ahead to be forecasted has been limited to eight. Network input series are formed by lagged inputs of CO concentrations level.

3.1 Simulated Environment Data Set

The simulated environment data set plot is shown in Figure 3. The first 250 data are used to train the network, while the remaining 250 data are used to test the fitted model and to calculate index of coefficient (R^2). The network is trained by the following input configuration:

$$v(t)=[\ y(t-1)\ y(t-2)\ y(t-3)\ y(t-4)\ y(t-5)\ ];$$

For simulated environment data set, HMLP network only requires 5 past CO concentration values to achieve its best results. Number of hidden nodes used are 2, since it gave better results compared to others. The R^2 values achieved by HMLP network are shown in Table 1. From the results, it can be seen that HMLP network gives good results over the testing data set. The network gives good results even for higher number of steps ahead forecasting.

MSE calculated for the whole data set is shown in Figure 4, which indicates that the network parameters converge rapidly. The MSE converges to an acceptable value after 200 data samples, suggesting that HMLP network only requires about 200 data to be trained properly.

Number of Steps	R^2 Value
1	0.9807
2	0.9272
3	0.8521
4	0.7635
5	0.6702
6	0.5748
7	0.4832
8	0.3961

Table 1. R^2 Values Achieved for Simulated Environment Data Set

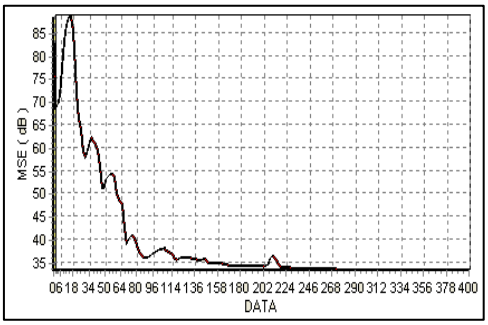

Fig.4 MSE for Simulated Environment Data Set

3.2 Real Data Set

The industrial data plot is shown in Figure 5. The first 600 data samples are used to train the HMLP network, while the remaining 400 data are used to test the network. The HMLP network is trained using the following input configuration:

$$v(t)=[\ y(t-1)\ y(t-2)\ldots\ldots..y(t-47)\ ];$$

From the configuration shown, it can be noted that HMLP network requires 47 past CO concentrations value to perform the task. For this data set, 2 hidden nodes are used since it gave the best results compared to others. The R^2 values achieved by HMLP network are shown in Table 2.

Number of steps	R^2 Value
1	0.7223
2	0.5303
3	0.4857
4	0.4581
5	0.4389
6	0.4265
7	0.4166
8	0.4107

Table 2. R^2 Values Achieved for Real Data Set

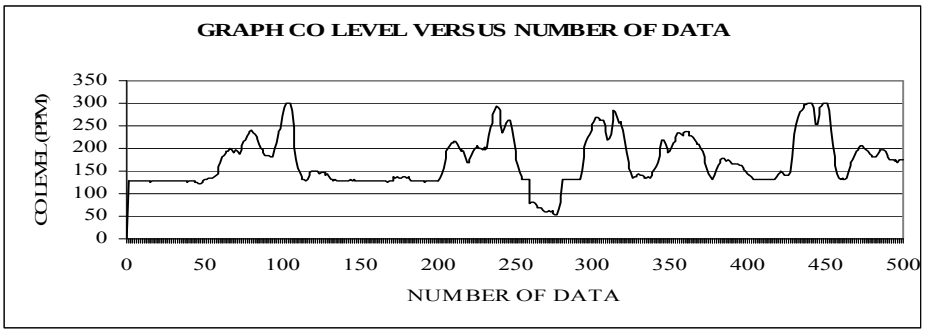

Fig.3 Simulated Environment Data Set

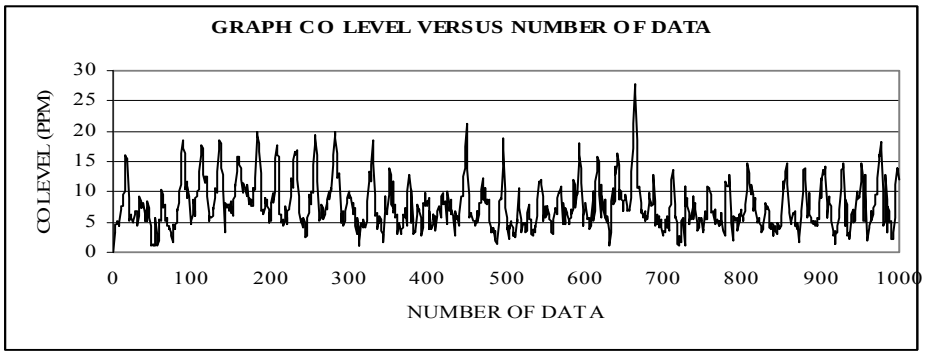

Fig.5 Industrial Data Set

From the results shown in Table 2, it can be noted that HMLP network produces good result for one step ahead only but manages to produce average results for higher number steps ahead of forecasting. The R^2 value drops drastically for two steps ahead, and decrease slowly for 3 steps ahead onward. The MSE calculated for the whole data set is shown in Figure 6. The plot shows that HMLP network parameters converge rapidly after 600 data samples. This means that HMLP requires about 600 data samples in order to be trained properly.

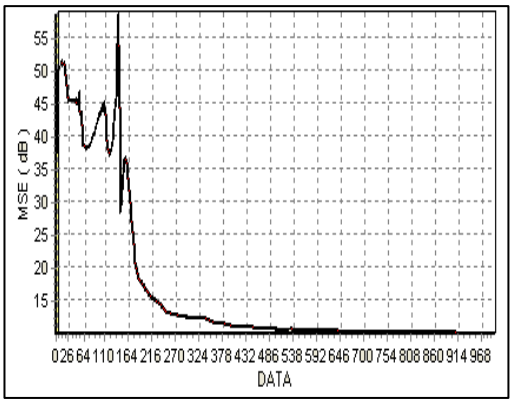

Fig.6 MSE for Real Data Set

For real data set, large numbers of input lags are required since the CO concentrations level fluctuates heavily. The number of input lags required to perform CO forecasting depends on the dynamic of data set. The network will not be able to represent nonlinear relationship between the input series if small numbers of input lags are used. Thus, more input lags are used to reach higher learning capability in order to achieve minimum prediction error.

4 Performance Comparisons

In this section, the effectiveness of the HMLP network is compared with MLP and Recurrent networks. The performance comparison will be divided into two sections, consisting of HMLP network versus MLP network and HMLP network versus Recurrent network. It is divided into two sections since the architecture of both the neural network are different. OSA, MSA and R^2 tests are used to evaluate the performance of these networks. In this comparison studies, OSA and MSA tests are described in terms of R^2 values. By using R^2 test, the comparison can be shown more accurately in form of quantitative analysis. The performance comparison for CO concentrations forecasting are carried out by using the same conditions mentioned in the previous section, such as number of training set and testing set, respectively. For fair comparison, analysis were carried out in order to choose the best input lags and hidden nodes in order to obtain the best forecasting performance produced by MLP and Recurrent network.

4.1 HMLP Network versus Standard MLP Network

In this section, performance of HMLP and MLP networks trained using MRPE and Levernberg-

Marquadt algorithm, respectively, are compared. For simulated environment data set, MLP network requires 5 past CO concentration values and 11 hidden nodes to achieve the results shown above. The R^2 values achieved by the standard MLP network are shown in Table 3. From the results, it can be noted that MLP network produces good overall results. The results indicated that HMLP network performs better compared to MLP network. The difference between both networks becomes noticeable with the R^2 values achieved for higher number of steps ahead. The maximum differences achieved by R^2 values between both networks are around 0.15.

Number of Steps	R^2 Values	
	HMLP	MLP
1	0.9807	0.9195
2	0.9272	0.8698
3	0.8521	0.7412
4	0.7635	0.6749
5	0.6702	0.5538
6	0.5748	0.4427
7	0.4832	0.3432
8	0.3961	0.2509

Table 3. R^2 values Achieved for Simulated Environment Data Set

For real data set, the R^2 values achieved by MLP network are shown in Table 4. MLP network requires 45 past CO concentration values to achieve the best results. For this data set, 5 hidden nodes need to be considered in order to obtain the best results from MLP network.

Number of Steps	R^2 Values	
	HMLP	MLP
1	0.7223	0.7070
2	0.5303	0.5610
3	0.4857	0.4622
4	0.4581	0.4368
5	0.4389	0.3991
6	0.4265	0.3595
7	0.4166	0.3226
8	0.4107	0.3106

Table 4. R^2 values Achieved for Real Data Set

From the results, it can be noted that MLP network only produces good result over 1 step ahead forecasting. The R^2 values decrease drastically from 2 steps ahead onward, and the values decrease slowly from 3 steps ahead onwards. It can be noted that MLP network gives higher R^2 value for 2 steps ahead forecasting compared to HMLP network, but it gives lower R^2 values for higher number of steps ahead. For real data set, HMLP network gave better results compared to MLP network. The maximum differences achieved by R^2 values between both networks are around 0.10. This means HMLP network performs 10% better compared to MLP network. Overall, HMLP network performs better compared to standard MLP network by using real and simulated environment data sets.

4.2 HMLP network versus Recurrent network

In this section, performance of HMLP network together with MRPE algorithm is compared to Recurrent network trained using Levernberg-Marquadt algorithm. The R^2 values achieved for simulated environment data set are shown in Table 5. For this data set, Recurrent network requires 24 past CO concentrations values and 15 hidden nodes to give the best results. From the results shown, it can be noted that Recurrent network does not produce good results over simulated environment data set. The R^2 values are very low from 1 step ahead and decrease slowly for higher number of steps ahead. Basically, the network fails to produce good results over simulated environment data set.

Number of Steps	R^2 Value	
	HMLP	Recurrent
1	0.9807	0.4315
2	0.9272	0.3809
3	0.8521	0.3301
4	0.7635	0.2716
5	0.6702	0.2612
6	0.5748	0.2543
7	0.4832	0.2400
8	0.3961	0.2211

Table 5. R^2 Values Achieved for Simulated Environment Data Set

In that case, HMLP network is found to perform better compared to Recurrent network. For 1 step ahead, the differences in R^2 values achieved by both the networks are around 0.5492.

For real data set, the R^2 values achieved by Recurrent network are shown in Table 6. Recurrent network requires 36 past CO concentrations value to achieve its best performance. For this data set, 5 hidden nodes need to be considered in order to obtain the best results from Recurrent network. From the results, it can be noted that Recurrent network gave good result for 1 and 2 step ahead forecasting. The R^2 values drop drastically after 2 steps ahead onwards. Generally, Recurrent network fails to provide multiple steps ahead forecasting for CO concentrations. From the Table 6, it can be seen that recurrent network provides higher R^2 values for 2 steps ahead forecasting compared to HMLP network. The network gives lower R^2 values for higher number of steps ahead

forecasting, where the maximum difference obtained are around 0.2. Overall, HMLP network is found to perform better than Recurrent network for both data sets. The comparison studies have proved that HMLP network achieved higher R^2 values compared to Recurrent network.

Number of Steps	R^2 Value	
	HMLP	Recurrent
1	0.7223	0.7157
2	0.5303	0.6993
3	0.4857	0.4947
4	0.4581	0.4001
5	0.4389	0.3375
6	0.4265	0.2893
7	0.4166	0.2575
8	0.4107	0.1809

Table 6. R^2 Values Achieved for Real Data Set

5 Conclusions

This study proves that HMLP network gives the best results compared to MLP and Recurrent networks for CO concentrations forecasting. The comparison between these networks becomes more noticeable with the number of steps ahead forecasting. In this study, dynamic of the data set and sampling time have significant contributions towards the performance of these networks. The forecasting performance of these networks can be improved if the data sets are sampled appropriately for the models to learn the trend of CO concentrations measurement properly.

Acknowledgement

The authors would like to acknowledge the support by Ministry of Science, Technology and Innovation, Malaysia for providing research grant. We would like to thank Malaysian Environmental Department (ASMA) for providing the industrial data set.

References:

[1] G. Maffeis, Prediction of Carbon Monoxide Acute Air Pollution Episodes. Model Formulation and First Application in Lombardy, *Atmospheric Environment*, Vol.33, 1999, pp. 3859-3872.

[2] Carbon Monoxide Alert, http:// www.homesafe.com/coalert, 2001.

[3] P. Sharma and M. Khare, Real- time Prediction of Extreme Ambient carbon Monoxide Concentrations due to Vehicular Exhaust Emissions using Univariate Linear Stochastic Models, *Transportation Research: Part D*, Vol.5, No. 1, 2000, pp. 59-69.

[4] P. Sharma and M. Khare, Short-term, Real-time Prediction of Extreme Ambient carbon Monoxide Concentrations due to Vehicular Exhaust Emissions using Transfer Function Noise Model, *Transportation Research: Part D*, Vol.6, No. 2, 2001, pp. 141-146.

[5] M. Zickus and K. Kvietkus, Urban Air Pollution Forecast based on the Gaussian and Regression Models, *Proceeding of 6th International Conference of Air Pollution*, 1998, pp. 515-523.

[6] A. C. Comrie and J. E. Diem, Climatology and Forecast Modelling in Phoenix, Arizona, *Atmospheric Environment*, Vol.33, No.30, 1999, pp. 5023-5036.

[7] G. N. Polydoras, J. S. Anagnostopoulos and G. Ch. Bergeles, Air Quality Predictions: Dispersion Model vs. Box-Jenkins Stochastic Models. An Implementation and Comparison for Athens, Greece, *Applied Thermal Engineering*, Vol.18, 1998, pp. 1037-1048.

[8] L. Moseholm, J. Silva and T. Larson, Forecasting Carbon Monoxide Carbon Monoxide Concentrations Near a Sheltered Intersection using Video Traffic surveillance and Neural Networks, *Transportation Research D*, Vol. 1, No. 1, 1996, pp. 15-28.

[9] M. Kolehmainen, H. Martikainen and J. Ruuskanen, Neural Networks and Periodic Components used in Air Quality Forecasting, *Atmospheric Environment*, Vol.35, 2001, pp. 815-825.

[10] J. P. Shi and R. M. Harrison, Regression Modelling of Hourly NO_X and NO_2 Concentrations in Urban Air in London, *Atmospheric Environment*, Vol.31, No.24, 1997, pp. 4081-4094.

[11] M. W. Gardner and S. R. Dorling, Neural Network Modelling and Prediction of Hourly NO_X and NO_2 Concentrations in Urban Air in London. *Atmospheric Environment*, Vol.33, 1999, pp.709-719.

[12] P. Patricio, A. Trier and J. Reyes, Prediction of $PM_{2.5}$ Concentrations Several Hours in Advance using Neural networks in Santiago, Chile. *Atmospheric Environment*, Vol.34, 2000, pp. 1189-1196.

[13] A. B. Chelani, C. V. Chalapathi Rao, K. M. Phadke and M. Z. Hasan, Prediction of Sulphur Dioxide concentration using Artificial Neural Networks, *Environmental Modelling & Software*, Vol. 17, 2002, pp. 161-168.