Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ecological Modelling 190 (2006) 99–115
ECOLOGICAL MODELLING
www.elsevier.com/locate/ecolmodel

Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions

S.M. Shiva Nagendra, Mukesh Khare *

Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

Received 17 October 2003; received in revised form 21 December 2004; accepted 3 January 2005
Available online 11 July 2005

Abstract

Artificial neural networks (ANNs) are useful alternative techniques in modelling the complex vehicular exhaust emission (VEE) dispersion phenomena. This paper describes a step-by-step procedure to model the nitrogen dioxide (NO_2) dispersion phenomena using the ANN technique. The ANN-based NO_2 models are developed at two air-quality-control regions (AQCRs), one, representing, a traffic intersection (AQCR1) and the other, an arterial road (AQCR2) in the Delhi city. The models are unique in the sense that they are developed for 'heterogeneous[1]' traffic conditions and tropical meteorology. The inputs to the model consist of 10 meteorological and 6 traffic characteristic variables. Two-year data, from 1 January 1997 to 31 December 1998 has been used for model training and data from 1 January to 31 December 1999, for model testing and evaluation purposes. The results show satisfactory performance of the ANN-based NO_2 models on the evaluation data set at both the AQCRs ($d = 0.76$ for AQCR1, and $d = 0.59$ for AQCR2).

Keywords: Air-quality management; Vehicular pollution; Meteorology; Traffic characteristic; Back-propagation training

1. Introduction

Air-quality models play a significant role in all aspects of air pollution control and planning, where prediction is a major component (Longhurst et al., 1996). The current deterministic and stochastic-based vehicular exhaust emission (VEE) models do not take into account some basic meteorological and traffic parameters, i.e. variations in the wind speed and direction, temperature, mixing height, pressure, atmospheric stability and heterogeneity in traffic compositions. The deterministic Gaussian models are based on steady state assumptions, and therefore, do not explain the non-linearity present in the VEEs dispersion (Esplin, 1995). The stochastic approach explains the non-linearity in the data and requires prior assumptions concerning the data distribution (Milionis and Davis, 1994). Further, these approaches are neither sufficiently comprehen-

* Corresponding author. Tel.: +91 11 26591212; fax: +91 11 26862037.
E-mail address: mukeshk@civil.iitd.ernet.in (M. Khare).

[1] It consists of light, heavy vehicles, three-wheelers: auto rickshaws and two-wheelers: scooter and motorcycles.

0304-3800/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2005.01.062

sive nor computationally tractable for predicting the vehicular pollutant concentrations to manage urban air quality during poor meteorological conditions (Chen and March, 1971). Unlike other modelling techniques, the artificial neural networks (ANN) makes no prior assumptions concerning the data distribution. ANNs are capable of modelling highly non-linear relationships and can be trained to accurately generalize when presented with a new data set. These features of ANN make it an efficient and accurate alternative tool for modelling the VEE dispersion (Gardner and Dorling, 1998; Viotti et al., 2002; Nagendra and Khare, 2004).

ANNs are parallel computational models, comprised of densely interconnected adaptive processing units. The important characteristic of neural networks is their adaptive nature, where 'learning by example replaces programming' (Bose and Liang, 1998). This feature makes the ANN techniques very appealing in application domains for solving highly non-linear phenomena (Zurada, 1997). In general, neural networks can be trained to approximate, virtually, any smooth measurable function (Simpson, 1989; Hornik et al., 1989).

ANNs have recently become important alternative tool to conventional methods in modelling complex non-linear relationships. In the recent past, the ANN has been applied to model large dimensionality of ecological data (Gevrey et al., 2003). Chon et al. (1996) used Kohnen network-type of ANN, to classify the benthic macro-invertebrate community data and to assign new data to previously created clusters. Levine et al. (1996) used ANN to classify the soil structure. Lek et al. (1996) compared multiple regression and ANN models in predicting density, biomass, reproduction potential, growth in brown trout management. In another study, Paruelo and Tomasel (1997) compared the performance of ANN models with regression models in predicting functional attributes of ecosystem. Both the studies indicated better performance of ANN models. Nunnari et al. (1998) used ANN technique to model the pollutants produced by alteration of photolytic cycle of NO_2, due to the presence of hydrocarbons released into the atmosphere. Lae et al. (1999) analyzed the relationships existing between environmental variables (catchment area, fishing, conductivity, depth, altitude and latitude) and fish yield. Scardi and Harding (1999) developed a neural network model for estimating primary production of phytoplankton in Chesapeake bay in the Eastern United States. Barciela et al. (1999) developed ANN models for predicting short-term temporal behaviour of phytoplankton biomass in a western Spanish Cost-Ria de Arousa. Dimopoulos et al. (1999) developed a neural network model to estimate the lead concentration in grasses using urban descriptors as model inputs in the Athens city, Greece. Manel et al. (1999) compared the performance of multiple discriminant analysis, logistic regression and ANNs in predicting the river bird's presence or absence from 32 variables consisting stream altitude, slope, habitat structure, chemistry and invertebrate abundance. The study indicated out-performance of the ANN model when compared with other traditional ecological modelling methods. Brosse et al. (1999) demonstrated the superiority of ANN models over the regression models in predicting spatial occupancy and abundance of fish in a mesotrophic lake. Ozesmi and Ozesmi (1999) developed the ANN model to select the habitat in marsh-breeding bird species in southwestern Lake Erie, USA. Lek and Guegan (1999) described the applications of back-propagation algorithm in ecological systems modelling. Karul et al. (2000) used a three-layer leven berg-Marquardt feedforward learning algorithm to model the eutrophication process in water bodies in Turkey. Heymans and Baird (2000) used ANN technique to analyze the carbon flow in the northern Benguela upwelling ecosystem of Namiba. Antonic et al. (2001) forecasted the forest survival after building the hydroelectric power plant on the Drava river, Croatia using the ANN. Olden and Jackson (2002) described randomization approach for statistically assessing the importance of network connection weights and the contribution of input variables in the neural network. Aitkenhead et al. (2003) developed a local interaction method for training neural networks and found that it was comparable with traditional back-propagation training method in predicting the complex behavior of environmental time series data. Park et al. (2003) used 'unsupervised' and 'supervised' network training algorithms to classify the sampling sites and to predict the aquatic insect species richness in running waters in France. Gevrey et al. (2003) evaluated partial derivatives, weights, perturb, profile, classical stepwise and improved stepwise methods to identify the environmental factors (model inputs) affecting trout abundance (model output) using the ANN. Ryan et al. (2004) used ANN technique to simulate nitrous oxide (N_2O) emissions from temperate grassland in New Zealand. The

results showed that ANNs were viable tool for simulating complex and highly variable biological systems.

In the recent past, ANN technique has become increasingly popular in modelling air-quality data (Nagendra and Khare, 2004). The multilayer neural network technique has been used to forecast the ozone (Comrie, 1997; Gardner and Dorling, 1996, 2000), the sulfur dioxide (Boznar et al., 1993), the NO_2 (Gardner and Dorling, 1999) and the particulate matter (Perez and Trier, 2001) in the ambient environment. However, this technique has very few applications in modeling the VEE dispersions (Perez and Trier, 2001; Viotti et al., 2002; Nagendra and Khare, 2002).

Moseholm et al. (1996) studied the usefulness of neural network in understanding the relationship between traffic parameters and carbon monoxide (CO) concentration measured near an intersection, which was sheltered from wind by multistoried buildings. In another work, Dorzdowicz et al. (1997) developed a line source neural network model for estimating hourly mean concentrations of CO in the urban area of Rosario, Italy. Eleven inputs, viz., vehicular flux in terms of vehicles per hour of cars, taxis, median vehicles, trucks and buses, wind speed and direction, solar radiation, humidity, pressure, rain intensity and temperature were used for developing three ANN-based models. The first, with 11 input variables, the second, with seven (excluding humidity, pressure, rain intensity and temperature) and the third with six input variables (excluding solar radiation, humidity, pressure, rain intensity and temperature). These models were validated for each type of network using approximately a set of 100 patterns. The results showed that model predictions were comparable. The elimination of variables from the input data set did not have any significant influence on predicted CO concentration. Gardner and Dorling (1999) developed multilayer perceptron (MLP) neural network models in the Central London using hourly NO_x, NO_2 and the meteorological data. The results showed better performance of the MLP models when compared to previously developed regression models (Shi and Harrison, 1997) for the same location. Perez and Trier (2001) developed ANN-based model to predict NO and NO_2 concentrations at a traffic junction in Santiago, Chile. The results showed that the model performed better than the persistence and regression models developed by them at the same location. Viotti et al. (2002) formulated ANN-based short- and long-term air-quality models for forecasting vehicular air pollutant concentrations in the city of Perugia, Italy. The models showed reasonable accuracy in predicting short- and long-term air pollutant concentrations. Kukkonen et al. (2003) evaluated five neural networks, a linear statistical and deterministic model in predicting the NO_2 and particulate matter concentrations in the central Helsinki, Finland. Recently, Nagendra and Khare (2004) developed ANN-based line source models for predicting CO concentrations on an urban roadway. Ten meteorological and six traffic characteristic variables were used for developing these models. The results showed that the neural network models are able to capture traffic 'wake' effects on the CO dispersion in the near field regions of the roadway.

Most of the ANN studies addressed the problem associated with pattern recognition, forecasting and comparison of the neural network with other traditional approaches in ecological and atmospheric sciences. However, the step-by-step procedure involved in development of ANN-based models are less discussed. This paper describes a methodology consisting of step-by-step approach in developing the ANN-based VEE models at urban roadways for heterogeneous traffic conditions and tropical meteorology. Further, the models have been used to predict 24 h average NO_2 concentrations at two AQCRs in the Delhi city, one, traffic intersection (AQCR1) and the other, an arterial road (AQCR2). The inputs to these models are 10 meteorological and six traffic characteristic variables.

2. Materials and methods

The ANN approach has several advantages over traditional phenomenological or semi-empirical models, since they require known input data set without any assumptions (Gardner and Dorling, 1998). It exhibits rapid information processing and is able to develop a mapping of the input and output variables. Such a mapping can subsequently be used to predict desired outputs as a function of suitable inputs (Schalkoff, 1992). A multilayer neural network can approximate any smooth, measurable function between input and output vectors by selecting a suitable set of connecting weights and transfer functions (Hornik et al., 1989; Gardner and Dorling, 1998). It consists of a system of layered interconnected 'neurons' or 'nodes' as illustrated in Fig. 1.

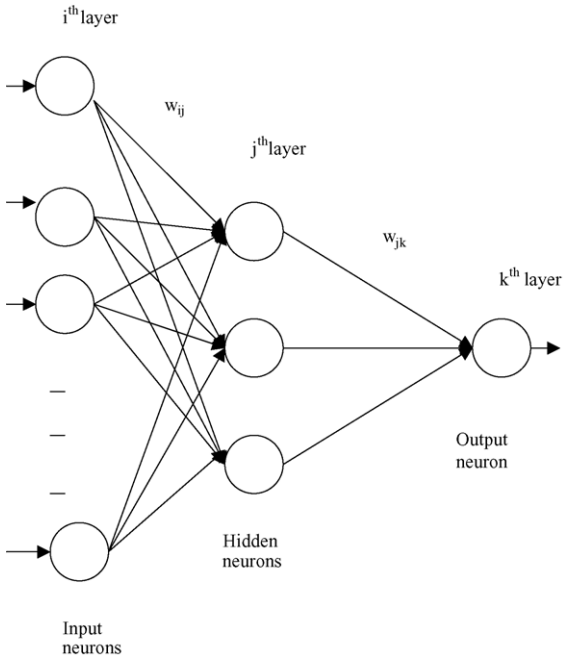

Fig. 1. General feed-forward neural network architecture.

2.1. ANN-based VEE modelling approach

The model building process consists of six sequential steps:

(i) selection of the optimal ANN-based VEE model architecture;
(ii) selection of the best activation function;
(iii) selection of the optimum learning parameters: 'n' (learning rate) and 'μ' (momentum rate);
(iv) initialization of the network weights and bias;
(v) training and testing of the model;
(vi) evaluation of the model.

2.1.1. Selection of the optimal ANN-based VEE model architecture

The number of neurons in the input layer equals the number of input variables (i.e. in the present work, meteorological and traffic characteristic variables). The output layer consists of one neuron, i.e. the pollutant concentration. The number of neurons in the hidden layer depends upon the number of training patterns, the number of input/output neurons, the amount of noise in the data, the network architecture, the type of activation function used in the hidden layer and the training algorithm (Alsugair and Al-Qudrah, 1998; Sarle, 1997). One hidden layer is sufficient to approximate any non-linear function in addition to input and output layers (Hornik et al., 1989).

The number of neurons in the hidden layer is obtained by training several networks and estimating the corresponding errors on the test data set. A few neurons in the hidden layer produce high training and testing errors due to under-fitting and statistical bias. On the contrary, too many hidden layer neurons lead to low training error, but high testing error, due to over-fitting and high variance (Sarle, 1997). In the past, researchers used 'rule of thumb' to find the number of neurons (H) in the hidden layer, as described below:

(i) H = number of input neurons + number of output neurons.
(ii) The maximum number of neurons in the hidden layer (H_{max}) is given by Swingler (1996) and Berry and Linoff (1997); H_{max} is two times the number of input layer neurons.
(iii) H = the number of the training patterns divided by five times of the number of input and output neurons.

The 'rule of thumb' failed to provide 'optimal' number of hidden layer neurons that subsequently affected the model prediction accuracy. However, this study has demonstrated that iterative approach is more efficient and accurate in determining the optimal number of hidden layer neurons, yielding minimum model prediction error on the 'test data set' (Haykin, 2001).

2.1.2. Activation function

The non-linear relationship between input and output parameters in any network requires a function, which can appropriately connect and/or relate the corresponding parameters (Sarle, 1997). Past air pollution-related studies by Gardner and Dorling (1999, 2000), demonstrated that the hyperbolic sigmoid activation function is faster and efficient in mapping the non-linearity among the hidden layer neurons than the logistic sigmoid activation function (Comrie, 1997; Rege and Tock, 1996). Hence, in the present study the hyperbolic tangent function has been used for hidden layer neurons. Further, the input and output layer neurons use the 'identity function' for their respective target values (Gardner and Dorling, 2000).

2.1.3. Learning parameter

Multilayer neural network has the ability to learn through training. Training requires a set of data consisting of a series of input and associated output vectors. A supervised back-propagation algorithm is most commonly employed in training the multilayer neural network (Haykin, 2001). In the back-propagation training, η and μ are used to 'speed up' or 'slow down' the convergence of error (Rumelhart et al., 1986). The back-propagation training algorithm gives an "approximation" to the trajectory in weight space, computed by the Gradient Descent Method (Battiti, 1992). The decrease in value of 'η' results in smaller changes in the synaptic weight from one iteration to the next and reduces the training speed. But, the increase in value of 'η' helps in faster training of the network due to the large changes in the synaptic weight and thus making the network as unstable (i.e. oscillatory). The term 'μ' has been used to avoid the network oscillation in back-propagation training algorithm. The values of 'η' and 'μ' are set between 0 and 1 (Rege and Tock, 1996; Rumelhart and McClelland, 1995). The following guidelines exist in evaluating the optimal value of 'η' and 'μ' (Haykin, 2001):

(i) The 'η' and 'μ' converge to a local minimum in the error surface of the network with the least number of epochs.
(ii) The 'η' and 'μ' converge to a least global minimum in the error surface with the least number of epochs.
(iii) The 'η' and 'μ' converge to the network configuration that has been best generalized with the least number of epochs.

In the present work, the optimal values of learning parameters have been evaluated using guideline (iii).

2.1.4. Initial network weights

Before starting the training, initialization of neural network weights and bias (free parameters) are required. The initial values of the synaptic weights and bias of the network help in fast convergence of the training processes. In the present study, all the free parameters of the network are set to the random numbers uniformly distributed within the range from $-2.4/F_i$ to $+2.4/F_i$, where F_i is the total number of inputs. The smaller distribution range reduces the probability of the saturation of the neurons in the network and so avoids the occurrence of the error gradients (Wasserman, 1989).

2.1.5. Training and testing

The neural networks are mostly trained using the 'supervised' learning algorithm. It is accomplished by providing known input and output data in an ordered manner to the network (Rumelhart and McClelland, 1995). Training involves finding the set of network weights thus enabling the model to represent the underlying patterns in the training data. It is achieved by minimizing the model error for all the input and associated output patterns (Gardner and Dorling, 1998). The 'under-training' of the network 'traps' the training algorithm in 'local' minima and 'over-training' results in high model prediction errors (Gardner and Dorling, 1998, 1999; Comrie, 1997). The over training can be avoided by training the network on a subset of inputs and outputs to determine weights and subsequently tested on the remaining (quasi-independent) test data to assess accuracy of the model predictions (Comrie, 1997). Therefore, the number of training epochs is decided avoiding under- and/or over-training of the network. The back-propagation learning algorithm is most suitable for air-quality modelling studies (Gardner and Dorling, 1998; Comrie, 1997). This algorithm divides the data into three partitions namely, the 'training data set', the 'test data set' and the 'evaluation data set'. The 'training data set' forms the bulk of the data used for the training purposes; the 'test data set' is used to check the generalization performance of the trained neural network model. The training is stopped when the performance on the 'test data set' results into minimum model error. Finally, the 'evaluation data set' is used to validate the model (Gardner and Dorling, 1998). The step-by-step procedure of the back-propagation training algorithm is given below.

(i) Multiply all the input by an initial random weight and sum the result as:

$$P_j = \sum_{j=1}^{H} w_{ij}\,(x_i + b_j);$$

$$i = 1, 2, \ldots, n;$$

$$j = 1, 2, \ldots, H \quad (1)$$

where P_j is the input to the 'j' hidden layer neuron, x_i the numerical value of the ith input layer neuron, w_{ij}

the weight of the ith input layer neuron to jth hidden layer neuron, n the number of the input layer neurons, H the number of the hidden layer neurons and b_j is the bias value for the jth hidden layer neuron.

(ii) Transform the hidden layer output by a sigmoid transfer function $f(P_j)$.

(a) Logistic function:

$$Q_j = \frac{1}{1+e^{-P_j}}; \quad j = 1, 2, \ldots, H \tag{2a}$$

(b) Hyperbolic tangent:

$$Q_j = \frac{2}{1+e^{-P_j}} - 1; \quad j = 1, 2, \ldots, H \tag{2b}$$

where Q_j is the output of the hidden layer neuron 'j'.

(iii) Multiply the hidden layer outputs by the hidden-output layer weight and sum as:

$$R_k = \sum_{j=1}^{H} w_{jk} Q_j + b_k; \quad k = 1, 2, \ldots, m \tag{3}$$

where R_k is the input to the kth output layer neuron, w_{jk} the weight of the jth hidden layer neuron to the kth output layer neuron, m the number of the output layer neuron and b_k is the bias value for the kth output layer neuron.

(iv) Transform the output, R_k by the transfer function to obtain the network outputs Y_k. The network outputs are then compared with observed values, and an error at the kth output neuron is computed:

$$E_k = T_k - Y_k \tag{4}$$

where T_k is the training (actual) value.

The general principle used in the back-propagation learning method is the 'delta rule', which is based on the minimization of the sum of squares of the error obtained in Eq. (4). The reduction in the sum of squares of the error is performed by iteratively modifying the numerical values of the weights in the direction of the steepest descent with respect to the error (Battiti, 1992). The weights in the hidden-output layer are adjusted first, followed by the adjustments of the weights in the input-hidden layer. The weights in the successive iterations are modified according to the following equation for layer 1 and tth iteration:

$$w_{jk}(t+1) = w_{jk}(t) + \eta\delta_j(t) Q_k(t) + \mu[w_{jk}(t) - w_{jk}(t-1)];$$
$$j = 1, 2, \ldots, H; \quad k = 1, 2, \ldots, m \tag{5}$$

where δ_j is the local gradient of the network.

The local gradient for the hidden-output layer δ_k is computed as follows:

$$\delta_k(t) = E_k(t) Y_k(t)[1 - Y_k(t)]; \quad k = 1, 2, \ldots, m. \tag{6a}$$

The local gradient for the input-hidden layer:

$$\delta_j(t) = Q_j(t)[1 - Q_j(t)] \sum \delta_k(t) w_{jk} \tag{6b}$$

2.1.6. Stopping criteria

The stopping criteria for the back-propagation algorithm are listed below.

(i) The back-propagation algorithm is considered to have converged when the absolute rate of change in the mean squared error (MSE) per epoch is very small.
(ii) After an each training iteration, the network is tested for its generalization performance. The training process stops when the generalization performance reaches the maximum on the test data set (Haykin, 2001).

The first criteria of minimizing the MSE over a training data set does not necessarily imply good generalization (Haykin, 2001). The second criteria trains the network iteratively based on number of training epochs. Each training epoch decides the value of synaptic weight and bias of the network. Thereafter, the trained network is tested on the 'test data' set which gives the prediction error. If the prediction error exceeds the statistical standards (here, the statistical standards is 'd' value), the network is again trained with increased number of epochs and so the process is repeated as described above. In the present study, the second stopping criteria has been adopted due to its superior learning efficiency.

2.2. Statistics for model testing and evaluation

The statistical indicators for testing and evaluating the model are systematic and unsystematic root mean

square error ($RMSE_S$ and $RMSE_U$), mean bias error (MBE), mean square error, coefficient of determination (r^2), linear best fit constant (a) and gradient (b), mean of the observed and predicted concentration (O and $\bar{P}$, respectively) and their standard deviations (σ_o and σ_p, respectively) and 'd' values (Willmott, 1982). It is not uncommon to find models that have been evaluated by the correlation coefficient 'r^2' (Gardner and Dorling, 2000). However, the 'r^2' statistic may not be appropriate in assessing the accuracy of air-quality model predictions (Comrie, 1997). The model evaluation based on 'r^2' statistics mostly fails due to the presence of 'lag' between source emission quantity and the ambient pollutant concentration. The 'lag' is due to adverse meteorological conditions (inversion) which implies the accumulation of pollutants in the ambient environment during 'odd' hours of the day when there are no source emissions (Khare and Sharma, 1999). The 'd' is a descriptive statistics. It reflects the degree to which the observed variate is accurately estimated by the simulated variate. The 'd' is not a measure of correlation or association in the formal sense, but rather a measure of the degree (based on ensemble average) to which the model predictions are error free. At the same time, 'd' is a standardized measure in order that it may be easily interpreted and cross-comparisons of its magnitudes for a variety of models, regardless of units, can readily be made. It varies between 0 and 1. A computed value of 1 indicates perfect agreement between the observed and predicted observations, while 0 connotes complete disagreement (Willmott, 1982). The value of 'd' is expressed as:

$$d = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}[|P_i - \bar{O}| + |O_i - \bar{O}|]^2} \quad (7)$$

where $\bar{O}$ is the average of the observed data, and p are the predicted data.

2.3. Development of ANN-based NO_2 models

Near the traffic intersections and busy roads, the VEE dispersion is influenced by two factors: first, the natural turbulence and second, the traffic generated turbulence (traffic wake). The natural turbulence is represented by meteorological variables and the 'traffic wake' relates to the traffic characteristic variables (Gardner and Dorling, 1999; Eskridge and Hunt, 1979). In the present work, ANN-based NO_2 models are developed at a traffic intersection and at an arterial road in the Delhi city, considering both the meteorological and the traffic characteristic variables. The criteria used in developing the ANN-based NO_2 models are listed in Table 1. Ten meteorological and six traffic characteristic variables have been used for the model formulation. The models have been formulated using three choices of input data sets. Firstly, considering both meteorological and traffic characteristics input data ($ANNNO2_A$); the second, considering only meteorological input data ($ANNNO2_B$); the third, considering only traffic input data ($ANNNO2_C$). The output corresponding to these inputs is the 24 h average NO_2 concentrations. The choice of inputs to model are directly connected to the quantity of information given to the neural network and is generally constituted from the meteorological and traffic characteristic data.

The 24 h NO_2 concentration data were collected from Central Pollution Control Board (CPCB), New Delhi for a period of 3 years from January 1997 to December 1999, for both the AQCRs. The meteorological data, including 24 h average observations of cloud cover, pressure, mixing height, sunshine hours, visibility, temperature, wind speed, wind direction and humidity were collected from Indian Meteorological Department, New Delhi. The 24 h average traffic characteristics data were collected from Central Road Research Institute (CRRI), New Delhi, for the respective AQCRs. The vehicles were classified into four groups, viz., two wheelers, three wheelers, four wheelers gasoline powered and four wheeler diesel powered, for which the emission factors (developed by the Indian Institute of Petroleum), were used for estimating CO and NO_2 source strengths (Pundir et al., 1994). Table 2 provides the details of input data used for developing the 24 h average ANN-based NO_2 models.

Two-year data from 1 January 1997 to 31 December 1998 was used for the model training and the data from 1 January to 31 December 1999 was used for model testing and evaluation purposes. At AQCR1, the total data set included, 783 values. About 67% (522) of the total data were used for the model training, 16% (128) for the model testing and 17% (133) for the model evaluation. About 51% of the total data values were missing at AQCR2. Therefore, 'training', 'testing' and 'evaluation data set' were selected ran-

Table 1
Criterion used in the ANN-based VEE modelling

Serial no.	Item	Criterion used in the present study	Similar criterion used in the previous studies
1	Criteria for selection of neural network architecture	Input neurons = number of input variables Output neurons = number of output variable Hidden neurons = smallest number of neurons that yields a minimum prediction error on the validation data set (Haykin, 2001)	Gardner and Dorling (1998, 1999), Comrie (1997), Perez and Trier (2001), Viotti et al. (2002)
2	Criteria for selection of neuron activation functions	Input neurons = identity function Output neurons = identity function Hidden neurons = hyperbolic tangent function (Waserman, 1989)	Gardner and Dorling (1999, 2000), Viotti et al. (2002)
3	Criteria for selection of learning parameters	The learning parameters converge to the network configuration and give best performance on the validation data with least number of epochs/iterations (Haykin, 2001)	Viotti et al. (2002)
4	Criteria for initialization of network weights	Network weights are uniformly distributed inside in the range of $[(-2.4/F_i)$ to $(+2.4/F_i)]$, where F_i is the total number of inputs (Haykin, 2001)	Gardner and Dorling (1998)
5	Training algorithm	Back-propagation (Rumelheart and McClelland, 1995)	Gardner and Dorling (1999), Comrie (1997), Rege and Tock (1996), Perez and Trier (2001), Viotti et al. (2002)
6	Stopping criteria for neural network training	*Stopping criteria*: after each training iterations/epochs the network is tested for its performance on validation data set. The training process is stopped when the performance reach the maximum on validation data set (Haykin, 2001; Sarle, 1997)	Gardner and Dorling (1998, 1999, 2000), Viotti et al. (2002)
7	Statistics for model validation	RMSE and '*d*' (Willmott, 1982)	Gardner and Dorling (1999, 2000), Comrie (1997), Viotti et al. (2002)
8	ANN modelling data set	*Training data set*: for training neural network *Test data set*: for testing of neural network during training *Evaluation data set*: for performance evaluation of trained neural network model	Gardner and Dorling (1999), Comrie (1997), Viotti et al. (2002)

domly for the development of the ANN-based NO_2 models. Out of 49% of available data, 33% (262) data were used for training, 10% (74) used for testing and 6% (47), for the model evaluation. The random selections of the data for training, testing and for the evaluation purposes are based on seasonal variations in meteorology and pollutant concentrations in the AQCRs.

2.3.1. Meteorological and traffic characteristic variables as model input

The $ANNNO2_{A1}$ and $ANNNO2_{A2}$ models were developed for AQCR1 and AQCR2, respectively, using daily average meteorological and traffic characteristics as predictor variables (17). Several hundred experiments were performed to determine the best combination of 'η', 'μ', the number of hidden layers, H, the learning algorithm and the transfer function. The guidelines (discussed under Section 2) were considered for choosing the optimum 'η', 'μ', the number of hidden layers, H, the learning algorithm and the activation function. The computational runs were conducted using the Stuttgart neural network simulator (SNNS) software (ftp://ftp.informatik.uni-stuttgart.de) to develop the optimum ANN-based NO_2 model. The inputs to these runs were the meteorological and traf-

Table 2
Input variable for the ANN-based NO_2 model

Model ID	Architecture	Input variables	
		Meteorological	Traffic characteristics
$ANNNO2_{A1}$, $ANNNO2_{A2}$	17:5:1	Cloud cover, humidity, mixing height, pressure, Pasquill stability[a], sun shine hour, temperature, visibility, sin (wind direction)[b], cos (wind direction)[b], wind speed	Two wheeler, three wheeler, four wheeler (gasoline), four wheeler (diesel), source strength (CO) and source strength (NO_2)
$ANNNO2_{B1}$, $ANNNO2_{B2}$	10:5:1	Cloud cover, humidity, mixing height, pressure, sun shine hour, temperature, visibility, sin (wind direction), cos (wind direction), wind speed	
$ANNNO2_{C1}$, $ANNNO2_{C2}$	5:5:1		Two wheeler, three wheeler, four wheeler (gasoline), four wheeler (diesel) and source strength of NO_2

[a] Estimated using Pasquill–Gifford stability scheme (Hanna et al., 1982).
[b] Wind direction data has been dichotomized using sine and cosine function.

fic characteristic variables in the input layer (17), the output was in terms of only pollutant concentration, i.e. NO_2. The number of neurons in the hidden layer were varied from 2 to 34. The descriptive statistics test, i.e. '*d*' value and RMSE (Willmott, 1982) were used to arrive at optimum number of neurons in the hidden layer. As a result, a fully connected feed-forward neural network with 17 neurons in the input layer, a single hidden layer, with five hidden neurons and a single neuron in the output layer shows best prediction on the 'test data set'. Table 3 shows the statistics of 24 h average ANN-based NO_2 models with the number of neurons in the hidden layer. Tables 4 and 5 lists the performance of the $ANNNO2_{A1}$ and $ANNNO2_{A2}$ models during generalization on 'test data set' at AQCR1 and AQCR2, respectively. After repeated experiments, the best model prediction on the test data set was achieved at 150 training epochs with 'η' = 0.01 and 'μ' = 0.7 at AQCR1; at AQCR2, the best $ANNNO2_{A2}$ model prediction was achieved after 250 training epochs with 'η' = 0.001 and 'μ' = 0.3. Fig. 2 shows the architecture of the models with 17 predictor variables (17:5:1).

2.3.2. *Meteorological variables as model input*

The $ANNNO2_{B1}$ and $ANNNO2_{B2}$ models were developed for AQCR1 and AQCR2, respectively, using daily average meteorological data as predictor variables. The purpose of formulating this model is two-fold. First, to develop ANN-based NO_2 models ($ANNNO2_B$) to forecast 24 h average NO_2 concentration using routinely monitored meteorological variables, second, to study the sensitivity of the traffic characteristic variables. The number of training and validation patterns remains same as that of the $ANNNO2_A$ model. The network architecture of 10:5:1 was used for the development of the $ANNNO2_B$ model (Fig. 3). At AQCR1, the $ANNNO2_{B1}$ model predictions were accurate after 500 training epoch having 'η' = 0.001 and 'μ' = 0.9 (Table 6). At AQCR2, the model predictions were accurate ('η' = 0.001 and 'μ' = 0.5) after 400 training epoch (Table 7).

2.3.3. *Traffic characteristic variables as model input*

The $ANNNO2_{C1}$ and $ANNNO2_{C2}$ models were developed for AQCR1 and AQCR2, respectively, using daily average traffic characteristics data as predictor variables. These models were developed with five traffic characteristic variables as input to the model, i.e. two-wheeler, three-wheeler, four-wheeler gasoline-powered, four-wheeler diesel-powered and the source strength of NO_2. The 5:5:1 network architecture was used for the development of the models (Fig. 4). At AQCR1, the best $ANNNO2_{C1}$ model prediction was obtained at 'η' = 0.001 and 'μ'=0.3, with 40 training epochs (Table 8). At AQCR2, the best $ANNNO2_{C2}$ model prediction was obtained on the test data set at 160 training epochs with 'η' = 0.001 and 'μ'=0.5 (Table 9).

Table 3
Experimental simulation results for optimization of the hidden layer neurons for the 24 h ANN-based NO_2 model

Number of hidden neurons	Mean square error after network stabilization	Statistical parameter	
		d	RMSE
2	0.04657	0.58	7.96
3	0.04046	0.44	10.29
4	0.03803	0.47	10.45
5	0.03777	0.63	7.18
6	0.03774	0.44	10.69
7	0.03802	0.48	10.18
8	0.03657	0.45	9.89
9	0.03647	0.44	11.13
10	0.03758	0.45	11.11
11	0.03703	0.40	11.98
12	0.03984	0.40	11.13
13	0.03994	0.42	10.49
14	0.04030	0.44	10.18
15	0.03982	0.42	10.52
16	0.03731	0.39	12.35
17	0.04014	0.40	11.08
18	0.03815	0.39	11.93
19	0.03798	0.39	11.99
20	0.04151	0.40	10.80
21	0.03896	0.39	12.00
22	0.04000	0.47	9.81
23	0.03888	0.40	11.83
24	0.04538	0.56	7.91
25	0.03896	0.40	11.87
26	0.03843	0.44	11.13
27	0.04216	0.56	9.07
28	0.04604	0.58	7.70
29	0.04464	0.52	8.34
30	0.04576	0.57	7.80
31	0.04606	0.58	7.67
32	0.03871	0.40	11.95
33	0.04299	0.43	9.74
34	0.04614	0.58	7.69

Table 4
Estimates of the statistics during generalization of the $ANNNO2_{A1}$ model

Epoch	*d*	RMSE
50	0.597	7.58
100	0.625	7.20
150	0.627	7.18
200	0.625	7.29
250	0.622	7.32
300	0.467	9.0
350	0.47	9.95
400	0.44	10.69
450	0.44	10.48
500	0.44	10.78

Table 5
Estimates of the statistics during generalization of the $ANNNO2_{A2}$ model

Epoch	*d*	RMSE
50	0.51	20.2
100	0.54	19.3
150	0.539	18.85
200	0.545	18.66
250	0.546	18.65
300	0.543	18.79
350	0.535	19.06
400	0.529	19.38
450	0.523	19.67

Table 6
Estimates of the statistics during generalization of the $ANNNO2_{B1}$ model

Epoch	*d*	RMSE
100	0.58	8.23
200	0.59	8.06
300	0.59	7.93
400	0.598	7.86
500	0.598	7.82
600	0.595	7.82
700	0.594	7.85
800	0.594	7.87
900	0.593	7.89
1000	0.592	7.90

Table 7
Estimates of the statistics during generalization of the $ANNNO2_{B2}$ model

Epoch	*d*	RMSE
50	0.524	19.4
100	0.528	19.16
150	0.528	19.16
200	0.528	19.15
250	0.529	19.13
300	0.529	19.11
350	0.53	19.11
400	0.53	19.10
450	0.539	19.11
500	0.539	19.11
550	0.533	19.12
600	0.533	19.13
650	0.529	19.16
700	0.529	19.18
750	0.528	19.23

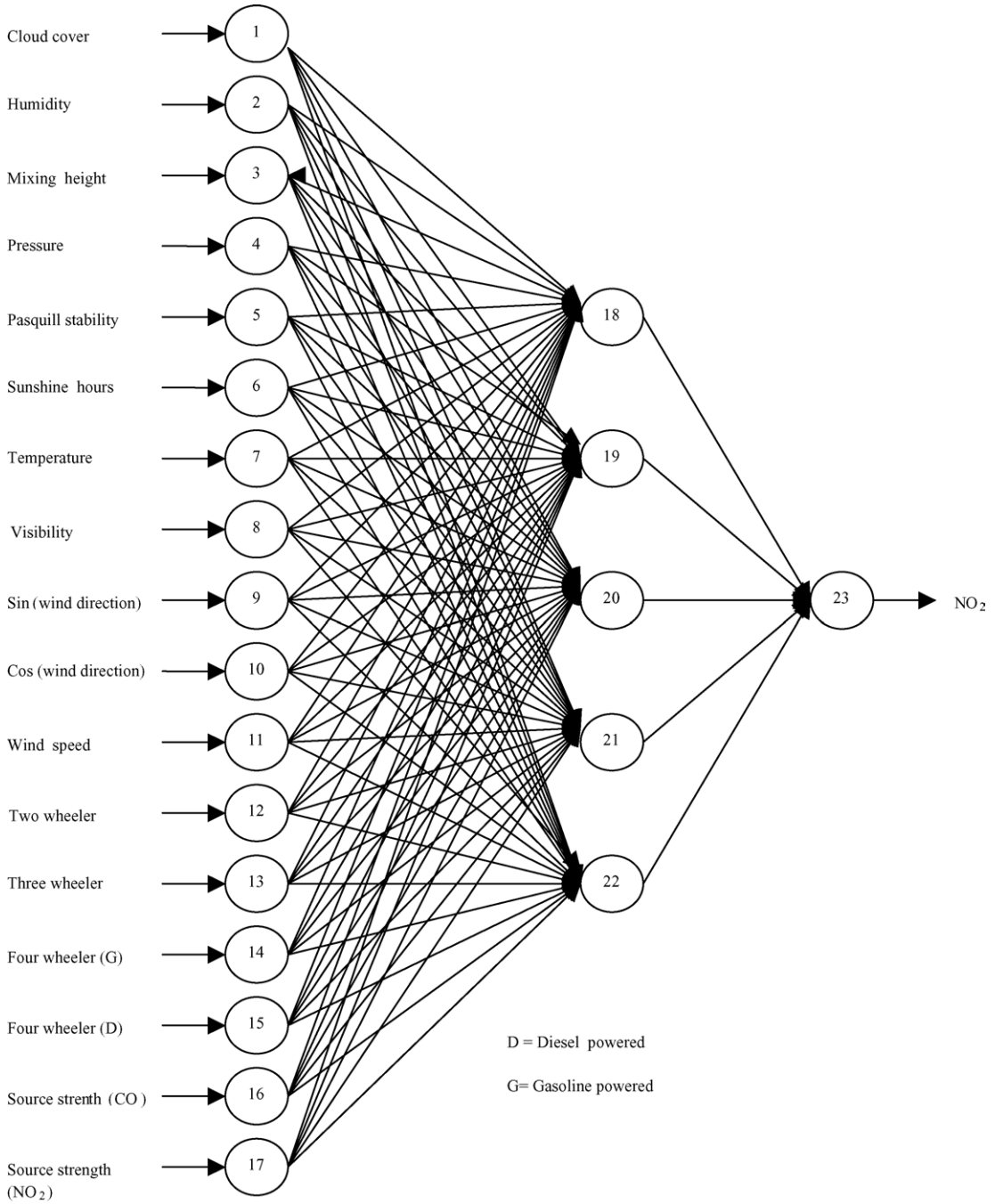

Fig. 2. Structure of 17:5:1 ANN-based NO_2 model.

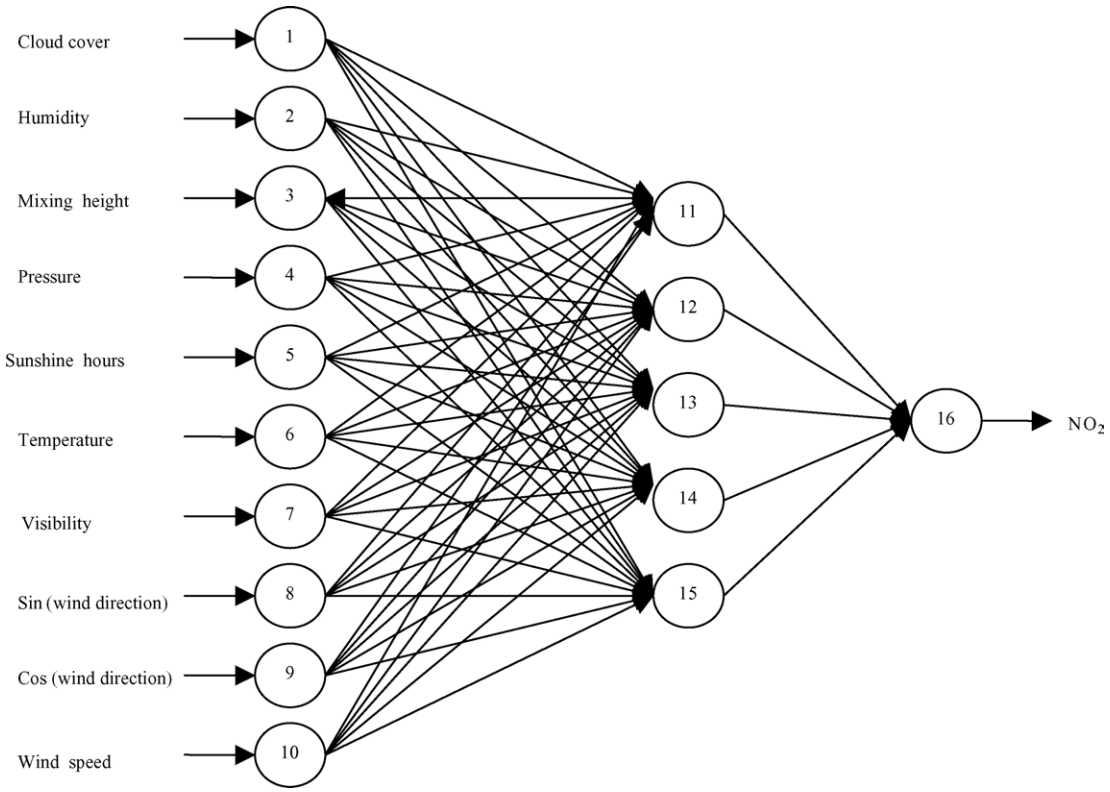

Fig. 3. Structure of 10:5:1 ANN-based NO_2 model.

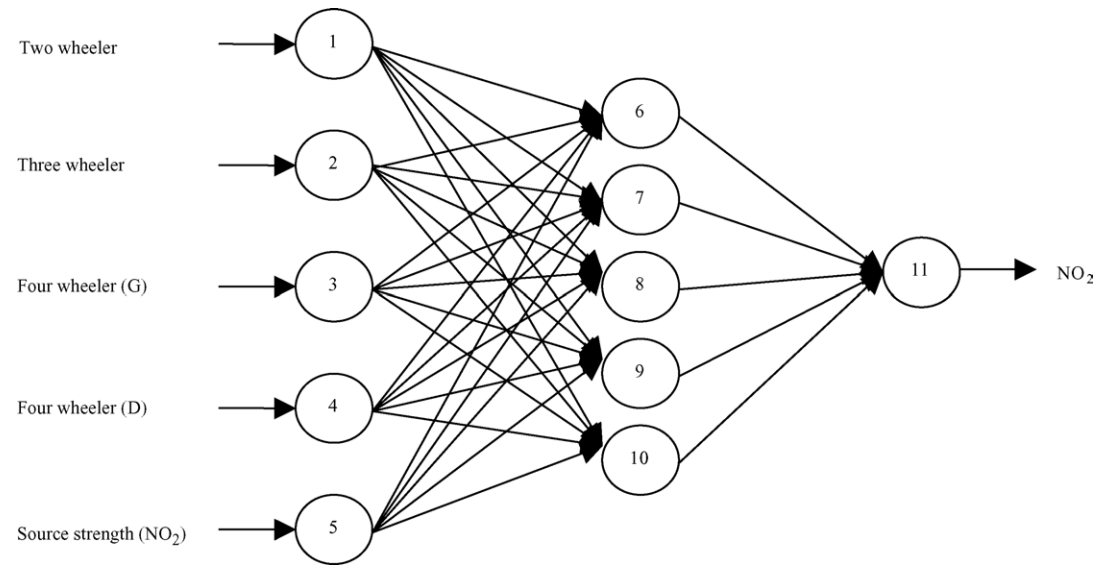

Fig. 4. Structure of 5:5:1 ANN-based NO_2 model.

Table 8
Estimates of the statistics during generalization of the $ANNNO2_{C1}$ model

Epoch	*d*	RMSE
5	0.265	29.21
10	0.34	23.07
15	0.43	12.59
20	0.44	9.76
25	0.44	9.03
30	0.44	8.84
35	0.44	8.77
40	0.44	8.75
45	0.43	8.75
50	0.40	8.75
55	0.40	8.75
60	0.40	8.75
65	0.40	8.75
70	0.40	8.75
75	0.40	8.75
80	0.40	8.75
85	0.40	8.75
90	0.40	8.75
95	0.40	8.75
100	0.40	8.75

Table 9
Estimates of the statistics during generalization of the $ANNNO2_{C2}$ model

Epoch	*d*	RMSE
20	0.41	19.55
40	0.41	19.55
60	0.41	19.55
80	0.41	19.54
100	0.41	19.52
120	0.41	19.52
140	0.41	19.51
160	0.41	19.48
180	0.41	19.50
200	0.41	19.52

3. Results and discussion

Table 10 gives the performance statistics of the trained ANN-based NO_2 models prediction on the evaluation data set at both the AQCRs. The mean values of $ANNNO2_{A1}$ and $ANNNO2_{A2}$ model predictions are slightly lower than the observed mean values. The MBE values at AQCR1 and AQCR2 are negative indicating the tendency of the models to under predict. The standard deviations (σ_P) of the $ANNNO2_{A1}$ and $ANNNO2_{A2}$ model predictions are 6.9 and 4.87 ppb, respectively. At AQCR1, σ_P is close to the standard deviation of the observed data. At AQCR2, the difference between the standard deviations of the observed and predicted data is quite high. This explains that the $ANNNO2_{A1}$ model is reproducing the variations in the evaluation data set at AQCR1 with better accuracy than the $ANNNO2_{A2}$ model. A low $RMSE_S$ value at AQCR1 indicates that the $ANNNO2_{A1}$ model predictions are closely matching with actual observations when compared with $ANNNO2_{A2}$ model predictions at AQCR2. Further, the '*d*' values for $ANNNO2_{A1}$ and $ANNNO2_{A2}$ models are 0.76 and 0.59, respectively. This explains that 76% of the model predictions are error free at AQCR1 and 59% at AQCR2. It shows that the $ANNNO2_{A1}$ model is more accurate than the $ANNNO2_{A2}$.

Table 10 summarises the performance statistics of the $ANNNO2_{B1}$ and $ANNNO2_{B2}$ model predictions on the evaluation data set at AQCR1 and AQCR2, respectively. The mean values of the predicted NO_2 concentration at both the AQCRs are lower than the observed mean values. The MBE values at AQCR1 and AQCR2 are -4.29 and -3.5 ppb, respectively, indicating the tendency of the models to under predict. The difference between the standard deviation of the observed and the predicted data at AQCR2 is higher than at AQCR1. It explains that the $ANNNO2_{B1}$ model predictions are closer to observed values when compared with the $ANNNO2_{B2}$ model. Further, a low $RMSE_S$ value at AQCR1 also indicates that the $ANNNO2_{B1}$ model predictions are closely matching with actual observations when compared to the $ANNNO2_{B2}$ model predictions. The '*d*' values for the $ANNNO2_{B1}$ and $ANNNO2_{B2}$ models explains that at AQCR1, 73% of the model predictions are error free and at AQCR2, it is 55%. It shows that the $ANNNO2_{B1}$ model at AQCR1 is more accurate than the $ANNNO2_{B2}$.

Table 10 provides the performance statistics of the $ANNNO2_{C1}$ and $ANNNO2_{C2}$ model predictions on evaluation data set at AQCR1 and AQCR2, respectively. At both the AQCRs, the mean values of model predictions are slightly lower than the observed mean values. The MBE values at AQCR1 and AQCR2 are negative, indicating the tendency of the models to under predict. The standard deviations (σ_P) values of the $ANNNO2_{C1}$ and the $ANNNO2_{C2}$ model predictions are very low when compared with the observed standard deviations (σ_O). It explains that both the models

Table 10
Performance statistics the ANN-based NO2 models

Site	Model ID	Statistic											
		$\overline{O}$ (ppb)	$\overline{P}$ (ppb)	σ_O (ppb)	σ_P (ppb)	MBE (ppb)	MSE (ppb)	RMSE		r^2	d	a (ppb)	b
								$RMSE_S$ (ppb)	$RMSE_U$ (ppb)				
AQCR1	$ANNNO2_{A1}$	35.1	31.7	10.4	6.9	−3.4	69.06	6.5	5.05	0.47	0.76	15.7	0.46
	$ANNNO2_{B1}$	35.1	30.8	10.4	6.56	−4.29	78.32	7.41	4.86	0.45	0.73	16.04	0.42
	$ANNNO2_{C1}$	35.1	33.1	10.4	0.74	−1.94	110.9	10.49	0.75	0.1	0.25	32.89	0.007
AQCR2	$ANNNO2_{A2}$	30.2	27.8	9.4	4.87	−2.34	77.44	8.01	4.41	0.18	0.59	21.2	0.2
	$ANNNO2_{B2}$	30.2	26.7	9.4	4.68	−3.5	90.25	8.53	4.35	0.12	0.55	21.46	0.17
	$ANNNO2_{C2}$	30.2	27.3	9.4	0.44	−2.85	96.43	9.86	0.44	0.03	0.3	27.57	−0.01

are inadequate to reproduce the variations in the evaluation data set. Further, the high $RMSE_S$ values also indicate that both the models perform poorly on the evaluation data set. The '*d*' values for the $ANNNO2_{C1}$ and $ANNNO2_{C2}$ models explain that at AQCR1, 25% of the model predictions are errors free and at AQCR2, it is 30%. It shows that both the models perform poorly on the evaluation data set.

3.1. Comparative performance of the models

For short-term average data (1 h), it is evident that the relationship between NO_2 with meteorological and traffic characteristic variables is complex and highly non-linear (Gardner and Dorling, 1999, 2000). Comrie (1997) and Gardner and Dorling (1998, 2000) observed the out-performance of the neural network model at sub daily time scale when the non-linearity of the system was more apparent and only small to marginal gains in model performance at the daily time scale. The present study supports the effect of the increase in averaging time period on the prediction performance of the models. As a result, marginal difference in the model performance has been observed after elimination of the traffic characteristic variables from the model input. It may be due to the increase in time averaging interval (1–24 h), which smoothens out the temporal variations of the pollutant concentration with meteorological and traffic characteristic variables, which, in turn, implies that the real non-linear pollutant–meteorology and pollutant–traffic relationships, gradually approach to the linear form (Gardner and Dorling, 2000). The facts are explained by the test statistics. The $RMSE_S$ values increase by 0.91 ppb for the $ANNNO2_{B1}$ model when compared to the $ANNNO2_{A1}$. However, the $RMSE_S$ value for the $ANNNO2_{B2}$ model is showing 0.52 ppb increase when compared to the $ANNNO2_{A2}$. Further, the '*d*' values for AQCR1 indicates that, the $ANNNO2_{B1}$ ($d = 0.73$) model performance decreases marginally when compared to the $ANNNO2_{A1}$ ($d = 0.76$). Similarly, at AQCR2, the $ANNNO2_{B2}$ model performance ($d = 0.55$) also shows marginal decrease when compared to the $ANNNO2_{A2}$ model ($d = 0.59$). The '*d*' value indicates that the $ANNNO2_{C1}$ model ($d = 0.25$) performs poorly at AQCR1, when compared with the $ANNNO2_{A1}$ ($d = 0.76$) and the $ANNNO2_{B1}$ ($d = 0.73$) models. At AQCR2, the $ANNNO2_{C2}$ model ($d = 0.3$) also shows poor performance when compared with the $ANNNO2_{B2}$ ($d = 0.55$) and $ANNNO2_{A2}$ models ($d = 0.59$). The poor performance of the models can be explained by the following facts. Firstly, these models are developed considering only traffic characteristic variables as their inputs. As a result, the models explain the NO_2 dispersion only due to the 'traffic wake' effects. Secondly, due to the absence of meteorological input variables, these models fail to take into account the 'lag effect' (Khare and Sharma, 1999). This phenomena frequently occurs during critical winter periods (November–March), when inversion conditions prevail during night time, particularly 4–6 h after 6:00 p.m. (10:00 p.m.–6:00 a.m.). As result, the models fail to explain the seasonal variations present in the NO_2 dispersion characteristics.

The predictive capability of ANN models depends on number of modelling parameters, i.e. selection of model inputs, number of hidden layers and its neurons, learning algorithm and learning parameters and stop-

ping criteria (Gardner and Dorling, 1998; Dimopoulos et al., 1999). Nevertheless, in most studies, the problem associated in developing optimal ANN models is reported (Gevrey et al., 2003; Aitkenhead et al., 2003; Olden and Jackson, 2002; Ozesmi and Ozesmi, 1999). Over-training is one of the main concerns in developing ANN models. It occurs when network learns the noisy details in training the data, which results in poor generalization capabilities. Paruelo and Tomasel (1997) used the two parameters of the network, i.e. selection optimum number of hidden neurons and error goal to reduce over-training problem. Lek et al. (1996) and Lae et al. (1999) used fixed number of iterations to avoid over-training of networks. However, the validation of these methods indicated the development of over trained ANN models. The step-by-step procedure discussed in the present work addresses the problems encountered during development of optimum ANN-based models using back-propagation training algorithm.

4. Conclusions

The present work provides vital statistics and guidelines on the choice of the ANN-based VEE modelling parameters, e.g. 'when to stop' the training process and determination of the learning parameters in back-propagation learning algorithm. Multilayer neural network technique has been used to develop short-term ANN-based NO_2 model for the air-quality prediction purposes at a traffic intersection and arterial road in the Delhi city. The daily time series of NO_2 concentration, meteorological and traffic characteristic variables, collected for the years from 1997 to 1999, have been used for training, testing and evaluation of the ANN-based VEE models. The models have been formulated following three choices of input data sets. Firstly, with both meteorological and traffic input data. Secondly, with only meteorological input data and lastly with only traffic input data. The results show that the ANN-based NO_2 models (with meteorological and traffic characteristic inputs) perform satisfactorily at both the AQCRs ($d = 0.76$, for AQCR1 and $d = 0.59$, for AQCR2). There is a marginal decrease in model performance, when only meteorological inputs have been used ($d = 0.73$, for AQCR1 and $d = 0.55$, for AQCR2). The models perform poorly when only traffic characteristic variables are used as inputs ($d = 0.25$, for AQCR1 and $d = 0.3$, for AQCR2).

Acknowledgement

We are thankful to Central Pollution Control Board, Indian Meteorological Department and Central Road Research Institute, New Delhi for providing necessary data.

References

Alsugair, A.M., Al-Qudrah, A.A., 1998. Artificial neural network approach for pavement maintenance. J. Comput. Civil Eng. ASCE 2 (4), 249–255.

Antonic, O., Hatic, D., Krian, J., Bukocev, D., 2001. Modelling groundwater regime acceptable for the forest survival after the building of the hydro-electric power plant. Ecol. Model. 138 (1–3), 277–288.

Aitkenhead, M.J., McDonald, A.J.S., Dawson, J.J., Couper, G., Smart, R.P., Billett, M., Hope, D., Palmer, S., 2003. A novel method for training neural networks for time-series prediction in environmental systems. Ecol. Model. 162 (1–2), 87–95.

Barciela, R.M., Garcia, E., Fernandez, E., 1999. Modelling primary production in a costal embayment affected by upwelling using dynamic ecosystem models and artificial neural networks. Ecol. Model. 120 (2–3), 199–211.

Battiti, R., 1992. First and second order methods for learning between steepest descent and Newton's method. Neural Comput. 4, 141–166.

Berry, T., Linoff, J., 1997. Data Mining Techniques. John Wiley and Sons, NY.

Bose, N.K., Liang, P., 1998. Neural Network Fundamentals with Graphs, Algorithms and Applications. Tata McGraw Hill Publishing Company Limited, New Delhi, India.

Boznar, M., Lesjak, M., Malker, P., 1993. A neural network based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. Atmos. Environ. 27B (2), 221–230.

Brosse, S., Guegan, J.F., Tourenq, J.N., Lek, S., 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. Ecol. Model. 120 (2–3), 299–311.

Chen, T., March, C.F., 1971. Effect of highway configurations on environmental problems dynamics of highway associated air pollution. In: Englund, H.M., Berry, T. (Eds.), Second International Clean Air Congress. Academic Press, NY, pp. 35–40.

Chon, T.S., Park, Y.S., Moon, K.H., Cha, E.Y., 1996. Patternizing communities by using an artificial neural network. Ecol. Model. 90 (1), 69–78.

Comrie, A.C., 1997. Comparing neural networks and regression model for ozone forecasting. J. Air Waste Manage. Assoc. 47, 653–663.

Dimopoulos, I., Chronopoulos, J., Sereli, A.C., Lek, S., 1999. Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). Ecol. Model. 120 (2–3), 157–165.

Dorzdowicz, B., Benz, S.J., Sonta, A.S.M., Scenna, N.J., 1997. A neural network based model for the analysis of carbon monoxide concentration in the urban area of Rosario. In: Power, H., Tirabassis, T., Brebbia, C.A. (Eds.), Air Pollution, vol. V. Computational Mechanics Inc., Southampton, Boston, pp. 677–685.

Eskridge, R.E., Hunt, J.C.R., 1979. Highway modelling. Part I: prediction of velocity and turbulence fields in the wake of vehicles. J. Appl. Meteorol. 18 (4), 387–400.

Esplin, G.L., 1995. Approximate explicit solution to the general line source problem. Atmos. Environ. 29 (12), 1459–1463.

Gardner, M.W., Dorling, S.R., 1996. Neural network modelling of the influence of local meteorology on surface ozone concentrations. In: Proceedings of International Conference on GeoComputation, University of Leeds, UK, pp. 359–370.

Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks: the multilayer perceptron: a review of applications in atmospheric sciences. Atmos. Environ. 32 (14–15), 2627–2636.

Gardner, M.W., Dorling, S.R., 1999. Neural network modelling and prediction of hourly NO_x and NO_2 concentrations in urban air in London. Atmos. Environ. 33 (5), 709–719.

Gardner, M.W., Dorling, S.R., 2000. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmos. Environ. 34 (1), 21–34.

Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 160 (3), 249–264.

Hanna, S.R., Briggs, G.A., Hosker, Jr. R.P., 1982. Handbook on Atmospheric Diffusion. Technical Information Center, U.S. Department of Energy, USA.

Haykin, S., 2001. Neural Networks: A Comprehensive Foundation, second ed. Pearson Education Inc., New Delhi, India.

Heymans, J.J., Baird, D., 2000. A carbon flow model and network analysis of the northern Benguela upwelling system. Namibia Ecol. Model. 126 (1), 9–32.

Hornik, K., Stinchcombe, M., White, H., 1989. Multi layer feed forward networks are universal approximators. Neural Netw. 2, 359–366.

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 (2–3), 145–452.

Khare, M., Sharma, P., 1999. Performance evaluation of general finite line source model for Delhi traffic conditions. Transport. Res. D4, 65–70.

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G., 2003. Extensive evaluation of neural network models for the prediction of NO_2 and PM10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki. Atmos. Environ. 37 (32), 4539–4550.

Lae, R., Lek, S., Moreau, J., 1999. Predicting fish yield of African lakes using neural networks. Ecol. Model. 120 (2–3), 325–335.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., Aulagnier, S., 1996. Application of neural networks to modelling nonlinear relationships in ecology. Ecol. Model. 90 (1), 39–52.

Lek, S., Guegan, J.F., 1999. Artificial neural networks as a tool in ecoligical modelling: an introduction. Ecol. Model. 120 (2–3), 65–73.

Levine, E.R., Kimes, D.S., Sigillito, V.G., 1996. Classifying soil structure using neural networks. Ecol. Model. 92 (1), 101–108.

Longhurst, J.W.S., Lindley, S.J., Watson, A.F.R., Conlan, D.E., 1996. The introduction of local air quality management in the United Kingdom: a review and theoretical framework. Atmos. Environ. 30 (23), 3975–3985.

Manel, S., Dias, J.M., Ormerod, S.J., 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. Ecol. Model. 120 (2–3), 337–347.

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.

Moseholm, L., Silva, J., Larson, T.C., 1996. Forecasting carbon monoxide concentration near a sheltered intersections using video traffic surveillance and neural networks. Transport. Res. D1, 15–28.

Nagendra, S.M.S., Khare, M., 2002. Line source emission modelling: review. Atmos. Environ. 36 (13), 2083–2098.

Nagendra, S.M.S., Khare, M., 2004. Artificial neural network based line source models for vehicular exhaust emission predictions of an urban roadway. J. Transport. Res. D Transport Environ. 9 (3), 199–208.

Nunnari, G., Nucifora, A.F.M., Randieri, C., 1998. The application of neural techniques to the modelling of time-series of atmospheric pollution data. Ecol. Model. 111 (2–3), 187–205.

Olden, J.D., Jackson, D.A., 2002. Illuminating the black box: a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Model. 154 (1–2), 135–150.

Ozesmi, S.L., Ozesmi, U., 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. Ecol. Model. 116 (1), 15–31.

Park, Y.S., Cereghino, R., Compin, A., Lek, S., 2003. Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. Ecol. Model. 160 (3), 265–300.

Paruelo, J.M., Tomasel, F., 1997. Prediction of functional characteristics of ecosystems: a comparison of artificial neural networks and regression models. Ecol. Model. 98 (2–3), 173–186.

Perez, P., Trier, A., 2001. Prediction of NO and NO_2 concentrations near a street with heavy traffic in Santiago. Chile. Atmos. Environ. 35 (10), 1783–1789.

Pundir, P.P., Jain, A.K., Gogia, D.K., 1994. Vehicle Emissions and Control Perspectives in India. Indian Institute of Petroleum, Dehradun, India.

Rege, M.A., Tock, R.W., 1996. A simple neural network for estimating emission rates of hydrogen sulphide and ammonia from single point source. J. Air Waste Manage. Assoc. 46, 953–962.

Rumelhart, D.E., McClelland, J.L., 1995. Parallel Distributed Processing: Explorations in the Microstructure of Cognitions, vol. 1. MIT Press, Cambridge, England.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning internal representations by error propagation. In: Rumelhart, D.E., McClelland, J.L. (Eds.), Parallel Distributed Processing, eighth ed. MIT Press, Cambridge, England, pp. 45–76.

Ryan, M., Muller, C., Di, H.J., Cameron, K.C., 2004. The use of artificial neural networks (ANNs) to simulate NO_2 emissions from temperate grassland ecosystem. Ecol. Model. 175 (2), 189–194.

Sarle, W., 1997. Neural network frequently asked questions, ftp://ftp.sas.com/pub/neural/FAQ.html.

Scardi, M., Harding Jr., L.W., 1999. Developing an empirical model of phytoplankton primary production: a neural network case study. Ecol. Model. 120 (2–3), 213–223.

Schalkoff, R., 1992. Pattern Recognition: Statistical, Structural and Neural Approaches. Wiley, NY.

Simpson, P.K., 1989. Foundations of neural networks. In: Sinencio, E.S., Lau, C. (Eds.), Artificial Neural Networks: Paradigms, Application and Hardware Implementations. IEEE Press, NY, pp. 3–24.

Shi, J.P., Harrison, R.M., 1997. Regression modelling of hourly NO_x and NO_2 concentration in urban air in London. Atmos. Environ. 31 (24), 4081–4094.

Swingler, 1996. Applying Neural Networks: A Practical Guide. Academic Press, London.

Viotti, P., Liuti, G., Genova, P.D., 2002. Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia. Ecol. Model. 148 (1), 27–46.

Wasserman, P.D., 1989. Neural Computing, Theory and Practice. Van Nostrand Reinhold, NY.

Willmott, C.J., 1982. Some comments on the evaluation of model performance. Bull. Am. Meteorol. Soc. 63, 1309–1313.

Zurada, J.M., 1997. Introduction to Artificial Neural Systems. West Publishing Company, Mumbai, India.