Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration

Rooollah Noori\textsuperscript{a,}*, Gholamali Hoshyaripour\textsuperscript{a}, Khosro Ashrafi\textsuperscript{a}, Babak Nadjar Araabi\textsuperscript{b}

\textsuperscript{a} Department of Environmental Engineering, Graduate Faculty of Environment, University of Tehran, Tehran, Iran
\textsuperscript{b} Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

**A R T I C L E   I N F O**

Article history:
Received 28 April 2009
Received in revised form 1 November 2009
Accepted 4 November 2009

Keywords:
Forward selection
Monte-Carlo

**A B S T R A C T**

This study aims to predict daily carbon monoxide (CO) concentration in the atmosphere of Tehran by means of developed artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) models. Forward selection (FS) and Gamma test (GT) methods are used for selecting input variables and developing hybrid models with ANN and ANFIS. From 12 input candidates, 7 and 9 variables are selected using FS and GT, respectively. Evaluation of developed hybrid models and its comparison with ANN and ANFIS models fed with all input variables shows that both FS and GT techniques reduce not only the output error, but also computational cost due to less inputs. FS--ANN and FS--ANFIS models are selected as the best models considering $R^2$, mean absolute error and also developed discrepancy ratio statistics. It is also shown that these two models are superior in predicting pollution episodes. Finally, uncertainty analysis based on Monte-Carlo simulation is carried out for FS--ANN and FS--ANFIS models which shows that FS--ANN model has less uncertainty; i.e. it is the best model which forecasts satisfactorily the trends in daily CO concentration levels.

\* Corresponding author. Tel.: +98 9374320526; fax: +98 2166407719. E-mail address: roohollahnoori@gmail.com (R. Noori).

\textsuperscript{1}352–2310/$ – see front matter \textcopyright 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.atmosenv.2009.11.005

1. Introduction

In recent years, artificial intelligence (AI) based methods have been proposed as alternatives to traditional statistical ones in many scientific disciplines. The literature demonstrates that AI models such as ANN and neuro-fuzzy techniques are successfully used for air pollution modeling (Nunnari et al., 2004; Perez-Roaee et al., 2006) and forecasting (Perez et al., 2000; Gautama et al., 2008). Moseholm et al. (1996) investigated the relationships between traffic and carbon monoxide (CO) concentrations using ANN model measured near an intersection which was sheltered from the wind by multi-story buildings. They compared ANN and MLR models and reported ANN as the superior model. Viotti et al. (2002) used an ANN model with a hidden layer to predict short-term and medium-term air pollutant concentrations (CO, ozone and benzene) in an urban area of Perugia city. Martin et al. (2008) used ANN and k-nearest neighbors classifiers as predictive tool in order to predicting future peaks of CO. Noori et al. (2008) compared use of ANN and PCA-MLR models in forecasting CO daily concentration in atmosphere of Tehran and reported ANN as the superior model. Modeling and controlling CO concentration using a neuro-fuzzy technique have been used by Tanaka et al. (1995) in a large city of Japan. Prediction results show that the fuzzy model is much better than the linear model. Yildirim and Bayramoglu (2006) proposed adaptive neuro-fuzzy inference system (ANFIS) to estimate the impact of meteorological factors on $\text{SO}_2$ and total suspended particular matter pollution levels over an urban area in Turkey. Carnevale et al. (2009) presented the application of neural network and neuro-fuzzy models to estimate nonlinear source–receptor relationships between precursor emission and pollutant concentrations (ozone and PM$_{10}$) in Northern Italy. The results show that, despite a large advantage in terms of computational costs, the selected source–receptor models are able to accurately reproduce the simulation of the 3D modeling system.

Input selection is a crucial step in ANN and ANFIS implementation. These techniques are not engineered to eliminate superfluous inputs. In the case of a high number of input variables, irrelevant, redundant, and noisy variables might be included in the data set, simultaneously; meaningful variables could be hidden (Seasholtz and Kowalski, 1993; Noori et al., 2009a). Therefore, reducing input variables is recommended. There are different methods for reducing the number of input variables such as forward selection (FS) (Chen et al., 1989; Wang et al., 2006) and Gamma test (GT) techniques (Corcoran et al., 2003; Moghaddamnia et al., 2009). In comparison with other statistical models, another important subject which rarely has been observed in ANN and ANFIS is uncertainty analysis...
of results. It is obvious that predictions are not certain; therefore, uncertainty analysis can be effective in application of results. Literature shows that just a few methods proposed for determination of uncertainty in ANN and ANFIS. Some of them are bootstrap and sandwich estimator (Tibshirani, 1994), maximum likelihood and Bayesian inference (Dybowski, 1997) and Mont-Carlo method proposed by Marce et al. (2004). In this research, Mont-Carlo simulation, which is based on locating the models in a Mont-Carlo random sampling process, is selected, because it has not only better performance but also more novelty. Agil et al. (2007) applied this uncertainty analysis method for evaluating outputs of ANFIS to predict weekly stream flow in the river and reported that it is appropriate for ANFIS model. Noori et al. (2009b) used Mont-Carlo method for uncertainty analysis of solid waste generation forecasting by means of wavelet transform-ANFIS and wavelet transform-ANN.

In this study, two techniques of input selection (FS and GT) have been applied in order to building hybrid models with ANN and ANFIS (FS–ANN, FS–ANFIS, GT–ANN, and GT–ANFIS), then have been compared with ANN and ANFIS fed with all input data. Finally, uncertainty analysis is done for two best models and the superior model is reported.

2. Material and methods

2.1. Case study and data

Tehran is the capital and the largest city of Iran which is located between 35° 34′–35° 50′N and 51° 02′–51° 36′E with the area about 570 km². It is surrounded by mountains to the north, west and east. It has current population of about 8,000,000 (Bayat, 2005). There are 11 air quality measurement stations in Tehran. The results of previous studies about air pollution of Tehran demonstrate that 90% by weight of total air pollutants are generated from traffic and only 10% from other sources (Bayat, 2005). In comparison with other air pollutants in the atmosphere of Tehran, CO is more significant because it has more than 75% by weight of air pollutants (Bayat, 2005). For this study, the pollution and meteorological data of 2004–2005 is obtained from Gholak station in north of Tehran at 35° 41′N and 51° 19′E with height of 1190.8 m above sea level. To predict CO concentration in the future, i.e. next 24 h, the daily arithmetic averages of six air pollutants: particulate matter less than 10 μm (PM10), total hydrocarbons (THC), nitrogen oxides (NOx), methane (CH4), sulfur dioxide (SO2) and ozone (O3) and also six meteorological variables: pressure (Press), temperature (Temp), wind direction (WD), wind speed (WS) and relative humidity (Hum) are used.

2.2. Artificial neural networks

ANN is a proper mathematical structure having an interconnected assembly of simple processing elements or nodes. ANN customary architecture is composed of three layers. Many theoretical and experimental works have shown that a single hidden layer is sufficient for ANNs to approximate any complex nonlinear function (Cybenko, 1989; Jalili-Ghazizadeh and Noori, 2008; Noori et al., 2009c). A major reason for this fact is that intermediate cells do not directly connect to output cells. Hence, they will have very small changes in their weight and learn very slowly (Gallant, 1993). Details for mastering the art of ANN model is published elsewhere (Gallant, 1993; Haykin, 1994). In this study, a model based on a feedforward neural network with a single hidden layer is used. The backpropagation algorithm is used to train the network. Also, the chosen activation functions are sigmoid and linear functions in the hidden and output layers, respectively.

2.3. Adaptive neuro-fuzzy inference system

A specific approach in neuro-fuzzy development is ANFIS, which has shown significant results in modeling nonlinear functions (Jang and Gulley, 1995; Jang et al., 1997). ANFIS uses a feedforward network to optimize parameters of a given fuzzy inference system (FIS) to perform well on a given task. The learning algorithm for ANFIS is a hybrid algorithm, which is combination of the gradient descent and least squares methods (Jang, 1991; Jang et al., 1997). The used FIS here is the Sugeno first-order fuzzy model with its equivalent ANFIS architecture (Jang et al., 1997; Noori et al., 2009d). Details for mastering the art of ANFIS and the hybrid learning algorithm are published elsewhere (Jang, 1993; Jang et al., 1997; Noori et al., 2009b), and thus, descriptions of these methods are not given here.

2.4. Forward selection

When the number of candidate covariates (N) is small, one can choose a prediction model by computing a reasonable criterion (e.g., RMSE, SSE, FPE or cross-validation error) for all possible subsets of the predictors. However, as N increases, the computational burden of this approach increases very quickly. This is one of the main reasons why step-by-step algorithms like FS are popular. FS has been successfully used by many researchers in order to build robust prediction models (Chen et al., 2004; Eksioglu et al., 2005; Wang et al., 2006; Khan et al., 2007). In this approach, which is based on linear regression model, first step is ordering of the explanatory variables according to their correlation with the dependent variable (from the most to the least correlated variable). Then, the explanatory variable, which is best correlated with the dependent variable, is selected as the first input. All remained variables are then added one by one as the second input according to their correlation with the output and the variable which most significantly increases the correlation coefficient (R²) is selected as the second input. This step is repeated N – 1 times for evaluating the effect of each variable on model output. Finally, among N obtained subsets, the subset with optimum R² is selected as the model input subset. The optimum R² is integral to a set of variables after which adding new variable dose not significantly increase the R² (Chen et al., 1989).

2.5. Gamma test

The GT estimates the minimum mean square error (MSE) that can be achieved during the modeling of unseen data using any continuous nonlinear models. The GT was firstly reported by Koncar (1997) and Agalbjörn et al. (1997) and later enhanced and discussed in detail by many researchers (Durrant, 2001; Tsui et al., 2002). Only a brief introduction on the GT is given here and the interested readers should refer to the aforementioned papers for further details. The basic idea is quite distinct from the earlier attempts with nonlinear analysis. Suppose that a set of data observations, \((x_i, y_i)\), \(1 \leq i \leq M\) exists, where the input vectors \(x_i \in R^m\) are vectors confined to some closed bounded set \(C \subseteq R^m\) and, without loss of generality, the corresponding outputs \(y_i \in R\) are scalars. The vectors \(x_i\) contain predicatively useful factors influencing the output \(y_i\). The only assumption is that the underlying relationship of the system is \(y = f(x_1 \cdots x_m) + \epsilon\), where \(f\) is a smooth function and \(\epsilon\) is a random variable which represents noise. Without loss of generality it can be assumed that the mean of the \(\epsilon\)'s distribution is zero (since any constant bias can be subsumed into the unknown function \(f\)) and that the variance of the noise Var(\(\epsilon\)) is bounded. The domain of a possible model is now restricted to the class of smooth functions which have bounded first partial derivatives. The GT is an estimate of the model’s output variance that cannot be accounted for a smooth data model. The GT is based on \(N[i, k]\), which are the \(k\)th \((1 \leq k \leq p)\)
nearest neighbors \(x_{N[i,k]}(1 \leq k \leq p)\) for each vector \(x_i(1 \leq i \leq M)\). Specifically, the GT is derived from the delta function of the input vectors:

\[
\delta_M(k) = \frac{1}{M} \sum_{i=1}^{M} |x_{N[i,k]} - x_i|^2 \quad (1 \leq k \leq p)
\]

where \(\cdot\cdot\cdot\) denotes Euclidean distance, and the corresponding Gamma function of the output values \(\gamma_M(k) = \frac{1}{M} \sum_{i=1}^{M} [y_{N[i,k]} - y_i]^2 , (1 \leq k \leq p)\), where \(y_{N[i,k]}\) is the corresponding y-value for the \(k\)th nearest neighbor of \(x_i\) in Eq. (1). In order to compute GT a least squares regression line is constructed for \(p\) points \((\delta_M(k), \gamma_M(k))\):

\[
y = A\delta + GT
\]

The intercept on the vertical axis \((\delta = 0)\) is the GT value, as can be shown \(\gamma_M(k) \rightarrow \text{Var}(r)\) in probability as \(\delta_M(k) \rightarrow 0\). Calculating the regression line gradient can also provide helpful information on the complexity of the system under investigation. A formal mathematical justification of this method can be found in Evans and Jones (2002). The graphical output of this regression line Eq. (2) provides very useful information. First, it is remarkable that the vertical intercept GT of \(y\) (or Gamma) axis offers an estimate of the best MSE achievable utilising a modeling technique for unknown smooth functions of continuous variables (Evans and Jones, 2002). Second, the gradient offers an indication of model’s complexity (a steeper gradient indicates a model of greater complexity). In practice, the GT can be achieved through WinGamma™ software implementation (Durrant, 2001). Noori et al. (2009e) applied the GT as a method for selecting the inputs of ANN in order to forecast the weekly solid waste generation in Tehran, Iran. This technique is very effective and could be potentially used for nonlinear modeling and forecasting of air pollution.

2.6. Determining the uncertainty in forecasting

In order to determine the uncertainty in the estimates of modeling procedures Monte Carlo simulation is used. In this method, the input parameter is described using a probability distribution and a single set of input data is randomly generated respecting this distribution. This single data set is run through the model and an output data set is obtained. The results of the run are stored and a new set of input data is generated. Multiple simulations, typically thousands, are carried out until the results of a new run do not affect the probability distribution of the output variable. Then, the resulting statistical performances (mean, median, variance, and percentiles) are collected, tabulated and their distributions are plotted. In present work, the above method is adapted by using random samples instead of random data which is relatively similar to Bootstrap Pairs resampling method (Tibshirani, 1994). In this method, the database is needed to be randomly resampled without replacement several times (here, 1000 times), maintaining the ratio between the training and validation sets constant. In this research, the 95 percent prediction uncertainties (95 PPU) are also calculated for forecasting models. This is calculated by the 2.5th \((X_L)\) and 97.5th \((X_U)\) percentiles of the cumulative distribution of every simulated point. The goodness of fit, therefore, is assessed by the uncertainty measures calculated from the percentage of measured data bracketed by the 95 PPU band, and the average distance \(\bar{d}_X\) between the upper and the lower 95 PPU (or the degree of uncertainty) determined from (Abbaspour et al., 2007):

\[
\bar{d}_X = \frac{1}{k} \sum_{i=1}^{k} (X_U - X_L)
\]

where \(k\) is the number of observed data points. The best outcome is that 100% of the measurements are bracketed by the 95 PPU, and \(\bar{d}_X\) is close to zero. However, because of model uncertainty, the ideal values generally will not be achievable. A reasonable measure for \(\bar{d}_X\), is calculated by the d-factor expressed as (Abbaspour et al., 2007):

\[
d\text{-factor} = \frac{\bar{d}_X}{\sigma_X}
\]

where \(\sigma_X\) is the standard deviation of the measured variable \(X\). The best value for d-factor is zero. A value less than 1 is a desirable measure for the d-factor (Abbaspour et al., 2007). Greater d-factor means higher uncertainty. The 95 percent prediction uncertainties (95 PPU) are calculated as:

\[
\text{Bracketed by 95 PPU} = \frac{1}{n}\text{count}(Q|X_L \leq Q \leq X_U) \times 100
\]

It should be noted that, since in all above mentioned methods the used data must be standard, in this study, all variables are standardized by mapping to \([-1, 1]\).

3. Result and discussion

3.1. Input selection

3.1.1. Forward selection

In this study, the FS method is used as a linear input selection technique in order to select the best subset of 12 input candidates. In other words, a linear model is developed using best correlated subset of inputs. First, correlation between each input variable and the desired output is evaluated. Second, the variable with highest correlation, i.e. Temp with \(R^2 = 0.26\), is selected as the first and the most important input. Then, remained candidates are implemented into the model one by one and the new variable which provides the best modeling result is selected as new input and added to the previously selected input (i.e. Temp). For evaluation of modeling goodness, correlation coefficient \(R^2\) is used. This step is repeated several times until that adding new variable to inputs dose not significantly improve the model output. In other words, if the increase in \(R^2\) is more than 5%, the new variable is selected. Finally, input variables with most significant effect on output are selected and other variables are removed. Results of FS are shown in Table 1 where seven candidates according to their importance are selected as input variables: Temp, \(\text{SO}_2\), THC, Press, \(\text{CH}_4\), \(\text{NO}_x\), and \(\text{O}_3\) and other candidates are eliminated.

3.1.2. Gamma test

The GT can greatly reduce the model development workload and provide input data guidance before developing a model (i.e. its result is independent from the models to be developed). In this study, different combinations of input data are explored to assess

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Results of forward selection procedure.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input subset</td>
<td>(R^2)</td>
</tr>
<tr>
<td>Temp</td>
<td>0.2699</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2)</td>
<td>0.3458</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC</td>
<td>0.4594</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC, Press</td>
<td>0.4615</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC, Press, (\text{CH}_4)</td>
<td>0.5193</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC, Press, (\text{CH}_4), (\text{NO}_x)</td>
<td>0.6055</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC, Press, (\text{CH}_4), (\text{NO}_x), (\text{O}_3)</td>
<td>0.6325</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC, Press, (\text{CH}_4), (\text{NO}_x), (\text{O}_3), Solar</td>
<td>0.6337</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC, Press, (\text{CH}_4), (\text{NO}_x), (\text{O}<em>3), Solar, (\text{PM}</em>{10})</td>
<td>0.6339</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC, Press, (\text{CH}_4), (\text{NO}_x), (\text{O}<em>3), Solar, (\text{PM}</em>{10}), Hum</td>
<td>0.6345</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC, Press, (\text{CH}_4), (\text{NO}_x), (\text{O}<em>3), Solar, (\text{PM}</em>{10}), Hum, WD</td>
<td>0.6347</td>
</tr>
<tr>
<td>Temp, (\text{SO}_2), THC, Press, (\text{CH}_4), (\text{NO}_x), (\text{O}<em>3), Solar, (\text{PM}</em>{10}), Hum, WD, WS</td>
<td>0.6349</td>
</tr>
</tbody>
</table>

* After this value, variations of \(R^2\) are negligible and thus, inputs related to this value are selected.
their influence on the CO prediction. There are $2^n - 1$ meaningful combinations of $n$ inputs from which the best one can be determined by observing the Gamma value, which indicates a measure of the best prediction attainable using any modeling methods for unseen smooth functions of continuous variables. For determining more important variables, first, the Gamma value must be calculated for a combination of all variables (12 input candidates). In the next step, one of the variables is omitted and Gamma value is calculated for the combination of remaining variables (11 variables). Then, the omitted variable in the previous stage is returned and another variable is omitted from the original combination and Gamma value is then calculated for the new combination which again contains 11 candidates. This process is repeated for each variable one by one and in each step the Gamma value is computed for the combination of remaining variables (11 variables). The results of different combinations are showed in Table 2. This table indicates that SO2 is the most important variable because of having the biggest Gamma value related to original combination; therefore, they are considered as ANN inputs, i.e. 7 inputs (FS−ANN model). Finally, the subset of inputs selected by GT is fed to ANN, i.e. 9 inputs (GT−ANN model). To improve the generalization of these models, the stop training algorithm (STA) is used (Coulibaly et al., 2000). For implementing STA in practice, the available data is divided into three parts: calibrating and testing step. In order to selecting the best architecture for mentioned models (ANN, FS−ANN, and GT−ANN model). In this research because of huge amount of input variables, this method cannot be used. For example by having 12 input variables and three MFs for each input variable, the rules will be $3^{12}$ rules (531,441 rules) that obstacle the calculation of parameters. Therefore, in this research a subtractive fuzzy clustering is used in order to establish the rule base relationship between the input and output variables. This method is based on a measure of the density of data points in the feature space that calculates based on the given search radius ($r_o$) (Chiu, 1994). The

### Table 2

<table>
<thead>
<tr>
<th>Input variables</th>
<th>Gamma value</th>
<th>Input variables</th>
<th>Gamma value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All inputs</td>
<td>0.017924</td>
<td>All inputs</td>
<td>0.000939</td>
</tr>
<tr>
<td>All inputs − PM</td>
<td>0.007779</td>
<td>All inputs−WD</td>
<td>0.018981</td>
</tr>
<tr>
<td>All inputs − Hum</td>
<td>0.020894</td>
<td>All inputs−WS</td>
<td>0.020894</td>
</tr>
<tr>
<td>All inputs − NO</td>
<td>0.005172</td>
<td>All inputs−CH</td>
<td>0.015712</td>
</tr>
<tr>
<td>All inputs − Press</td>
<td>0.007220</td>
<td>All inputs−O2</td>
<td>0.007251</td>
</tr>
<tr>
<td>All inputs − SO2</td>
<td>0.027841</td>
<td>All inputs−Solar</td>
<td>0.021780</td>
</tr>
<tr>
<td>All inputs − THC</td>
<td>0.019738</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: The Gamma test results applied on input data set.

### Table 3

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of input variable</th>
<th>Calibrating</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$R^2$</td>
<td>MAE</td>
</tr>
<tr>
<td>ANN</td>
<td>12</td>
<td>0.87</td>
<td>0.42</td>
</tr>
<tr>
<td>FS−ANN</td>
<td>7</td>
<td>0.9</td>
<td>0.36</td>
</tr>
<tr>
<td>GT−ANN</td>
<td>9</td>
<td>0.9</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Table 3: Results of calibrating and testing ANN, FS−ANN, and GT−ANN models.

### Table 4

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of rules</th>
<th>Calibrating</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANFIS</td>
<td>64</td>
<td>0.91</td>
<td>0.35</td>
</tr>
<tr>
<td>FS−ANFIS</td>
<td>51</td>
<td>0.9</td>
<td>0.36</td>
</tr>
<tr>
<td>GT−ANFIS</td>
<td>54</td>
<td>0.88</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Table 4: Results of calibration and testing of ANFIS, FS−ANFIS, and GT−ANFIS models.

3.3. ANFIS model development

In the ANFIS system, each input parameter might be clustered into several class values to build up fuzzy rules, and each fuzzy rule would be constructed using two or more membership functions (MFs). Several methods have been proposed to classify the input data and making the rules, that the most conventional of them are grid partition (Jang and Sun, 1995) and subtractive fuzzy clustering (Chiu, 1994). When there are a few input variables, grid partition is a suitable method for data classification. But in this research because of huge amount of input variables, this method cannot be used. For example by having 12 input variables and three MFs for each input variable, the rules will be $3^{12}$ rules (531,441 rules) that obstacle the calculation of parameters. Therefore, in this research a subtractive fuzzy clustering is used in order to establish the rule base relationship between the input and output variables. This method is based on a measure of the density of data points in the feature space that calculates based on the given search radius ($r_o$) (Chiu, 1994). The

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

where $\bar{O}$ is the average of the observed data, and $P$ are the predicted data. Results indicated that architecture with 8, 4, and 4 neurons in hidden layer for ANFIS, FS−ANN, and GT−ANN are the best model, respectively. Table 3 shows the results of calibration and testing of the models with the best structures. According to this table, although the accuracies of all models are relatively similar, the FS−ANN and GT−ANN models are superior because they have not only better accuracy, but also less number of inputs. Among these two models, FS−ANN model is selected as the best model because of having the least number of inputs.
optimal values for $r_a$ are identified through a trial and error procedure by varying the $r_a$ from 0.1 to 0.9 (in increments of 0.05). In this paper a Gaussian-type MFs is used. The optimum values for $r_a$ and also, the optimum number of rules for each model are shown in Table 4. After constructing the initial FIS structure, the training phase is done. In this step the number of iteration of hybrid algorithm for correction of model parameters and objective error are considered 30 and 10 respectively. The calibrating and testing results of ANFIS, FS–ANFIS, and GT–ANFIS models are presented in Table 4. According to this table, although the accuracies of all models are relatively similar, the FS–ANFIS model is superior because it has not only better accuracy than other models, but also the least number of inputs.

### 3.4. Developed discrepancy ratio statistic

According to Tables 3 and 4, the FS–ANN and FS–ANFIS models have better results than others and $R^2$, $d$, and MAE indexes are better for these models. In this paper, we developed a proper statistic; based on discrepancy ration statistic defined by White et al. (1973), to check the model robustness.

$$ DR = \log \left( \frac{\text{Predicted value}}{\text{Observed value}} \right) $$

(7)

DR is commonly used as an error measurement in the literature and is used widely by many researchers such as Seo and Cheong (1998), Deng et al. (2001), Kashefpour and Falconer (2002), Tayfur and Singh (2005). However, it is not utilizable for negative and zero values. For solving this problem, the developed discrepancy ratio (DDR) statistic is present by authors:

$$ DDR = \left( \frac{\text{Predicted value}}{\text{Observed value}} \right) - 1 $$

(8)

For better judgment and visualization, the Gaussian function of DDR values could be calculated and illustrated in a standard normal distribution format. Therefore, DDR values must be standardized and then, normalized value of DDR ($Q_{DDR}$) is calculated using Gaussian function. The $Q_{DDR}$ values for all models in testing stage are calculated and the standard normal distribution for each model is illustrated in Fig. 1. It should be noted that in error distribution graph, more tendencies to the centerline and also, larger value of the maximum $Q_{DDR}$ indicate more accuracy. The maximum $Q_{DDR}$ values for models are plotted in this figure. These values for ANN, FS–ANN, GT–ANN, ANFIS, FS–ANFIS, and GT–ANFIS models are 0.43, 0.92, 0.88, 0.84, 0.90, and 0.81, respectively. Considering Tables 3 and 4 and maximum $Q_{DDR}$ values, FS–ANN and FS–ANFIS models have better performance than other models.

### 3.5. Evaluation of models during episodes

It must be emphasized that by this modeling approach, it is intended to forecast the daily trend of pollutant concentration as correctly as possible, with a minimum number of false alerts. Thus, it might be possible, in the context of an environmental warning system, to reschedule urban activities in case of critical estimations above air quality standards. In other words, the accuracy of models in predicting air pollution episodes should be investigated. In this research, all models have been evaluated during episodes in the testing step. Daily CO concentration above 7 ppm is selected as a representative value for episodes and sum absolute errors (SAE) for each model are calculated (Table 5). This table indicates that FS–ANFIS and FS–ANN models have better performance for SAE than other models during episodes. SAE for these two models during episodes are 8.36 and 8.39, respectively.

### 3.6. Uncertainty analysis

For calculating uncertainty of FS–ANN and FS–ANFIS models during the network calibration, prediction models are generated

![Fig. 2. 95% confidence intervals for the estimates of daily CO concentration during the calibrating step using FS–ANN model.](image-url)
from historical data. Uncertainty analysis of the predicted daily CO concentration during the calibrating and testing steps has been quantified by estimating the confidence intervals of the simulation results. The 95% confidence intervals are determined by finding the 2.5th and 97.5th percentiles of the associated distribution of the simulation results. Plots of the range 95% confidence intervals for the estimates of daily CO concentration values for FS−ANN and FS−ANFIS during the calibrating step are shown in Figs. 2 and 3, respectively. Also, plots of the testing steps are shown in Figs. 4 and 5. According to Figs. 2 and 3, it is obvious that during the calibration stage, both models consistently predict the trend of decrease and increase in daily CO concentration. Moreover, FS−ANN model has wider 95% confidence bound than FS−ANFIS; thus, \( d \)-factor of FS−ANN is greater than FS−ANFIS. Besides, number of predictions bracketed by 95% confidence bound in FS−ANFIS is lower than FS−ANN. These results are also shown in Table 6. A similar trend is also found at the testing step (Figs. 4 and 5). Regarding the fact that \( d \)-factor lower than 1 is appropriate, both FS−ANN and FS−ANFIS are acceptable in prediction of daily CO concentration; although, FS−ANFIS is superior. It is also observable that predictions bracketed by 95% confidence bound in testing step of FS−ANFIS is 51% which is lower than FS−ANN (76%). Considering the fact that \( d \)-factor and \( R^2 \) values of both models are acceptable, amount of predictions bracketed by 95% confidence bound plays the most important role in determining the superior model. Therefore, in this research, the FS−ANN model is selected as the best model for prediction of daily CO concentration.

4. Conclusion

Considering the importance of daily CO concentration in the atmosphere of Tehran, this research aims to develop proper prediction models using ANN and ANFIS models. Since input selection is a significant step in modeling, FS and GT methods are used and six models are developed. The goodness of each model is evaluated using \( R^2 \), \( d \), and MAE statistics and also, DDR. Finally, uncertainty analysis of FS−ANN and FS−ANFIS, as superior models, is carried out. The following conclusions could be drawn from the present study:

1. Input selection improves prediction capability of both ANN and ANFIS models. It reduces not only the output error, but also time of calculation due to having less input variables.
2. Number of selected input variables using FS is seven out of twelve candidates while it is nine for GT. In general, FS based models are more proper than GT based models.

Table 6

<table>
<thead>
<tr>
<th>Index</th>
<th>FS−ANN Calibrating</th>
<th>FS−ANN Testing</th>
<th>FS−ANFIS Calibrating</th>
<th>FS−ANFIS Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d )-factor</td>
<td>0.74</td>
<td>0.77</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>Bracketed by 95 PPU (%)</td>
<td>75.5</td>
<td>76</td>
<td>50</td>
<td>51</td>
</tr>
</tbody>
</table>

Fig. 3. 95% confidence intervals for the estimates of daily CO concentration during the calibrating step using FS−ANFIS model.

Fig. 4. 95% confidence intervals for the estimates of daily CO concentration during the testing step using FS−ANN model.

Fig. 5. 95% confidence intervals for the estimates of daily CO concentration during the testing step using FS−ANFIS model.
3. Considering $R^2$, $d$, and MAE statistics, FS–ANN and FS–ANFIS are superior models. The same result is obtained using DDR.

4. Superiority of FS–ANN and FS–ANFIS models is also proved considering their prediction accuracy during air pollution episodes.

5. Uncertainty analysis of superior models showed that FS–ANN predicts daily CO concentration more appropriate than FS–ANFIS.

References


