

Predicting PM_{2.5} Concentrations Using Artificial Neural Networks and Markov Chain, a Case Study Karaj City

Gholamreza Asadollahfardi*, Hossein Zangooei and Shiva Homayoun Aria

Civil Engineering Department, Kharazmi University, Iran

*Corresponding author. Tel: +982188734489, E-mail: asadollahfardi@yahoo.com

ABSTRACT

The forecasting of air pollution is an important and popular topic in environmental engineering. Due to health impacts caused by unacceptable particulate matter (PM) levels, it has become one of the greatest concerns in metropolitan cities like Karaj City in Iran. In this study, the concentration of PM_{2.5} was predicted by applying a multilayer percepteron (MLP) neural network, a radial basis function (RBF) neural network and a Markov chain model. Two months of hourly data including temperature, NO, NO₂, NO_x, CO, SO₂ and PM₁₀ were used as inputs to the artificial neural networks. From 1,488 data, 1,300 of data was used to train the models and the rest of the data were applied to test the models. The results of using artificial neural networks indicated that the models performed well in predicting PM_{2.5} concentrations. The application of a Markov chain described the probable occurrences of unhealthy hours. The MLP neural network with two hidden lavers including 19 neurons in the first layer and 16 neurons in the second layer provided the best results. The coefficient of determination (R^2) , Index of Agreement (IA) and Efficiency (E) between the observed and the predicted data using an MLP neural network were 0.92, 0.93 and 0.981, respectively. In the MLP neural network, the MBE was 0.0546 which indicates the adequacy of the model. In the RBF neural network, increasing the number of neurons to 1,488 caused the RMSE to decline from 7.88 to 0.00 and caused R² to reach 0.93. In the Markov chain model the absolute error was 0.014 which indicated an acceptable accuracy and precision. We concluded the probability of occurrence state duration and transition of PM_{2.5} pollution is predictable using a Markov chain method.

Key words: Air pollution, PM_{2.5} concentration prediction, Artificial neural network, Markov chain

1. INTRODUCTION

Prediction of particulate matter (PM) is one of the important issues in the control and management of pollutants in the air. Particulate matter is the term used for a mixture of solid particles and liquid droplets found in the air (Dong *et al.*, 2012). The health effects of exposure to fine particulate matter are increasing the risk of death from lung cancer, pulmonary illness (e.g., asthma), chronic bronchitis, heart attack and cardiovascular disease (Deng *et al.*, 2013; Goss *et al.*, 2004; Slaughter *et al.*, 2003).

The US Environmental Protection Agency (EPA) standards divide air quality into three categories of $PM_{2.5}$ pollution (0-12 µg/m³ hour concentrations as good quality, 12.1 to 55.4 μ g/m³ as sensitive quality, and 55.5 plus $\mu g/m^3$ unhealthy quality). By 2020, the benefits of reductions in fine particles and ozone are estimated to be \$113 billion annually (Dong et al., 2012). Availability of accurate and sufficient data for forecasting future emissions helps planning and control of air pollution in air quality management (AQM); therefore, forecasting air pollution for AQM in urban areas is essential. Several techniques have been developed for the prediction of particulate matter (PM) concentrations. Approaches for the PM prediction can be classified into five categories: (1) empirical models, (2) fuzzy logic-based systems, (3) simulation models, (4) data driven statistical models, and (5) model-driven statistical learning methods (Dong et al., 2012).

Using air pollution modeling software always has several limitations. In such models, several errors and inaccurate results may be caused because many factors are not considered, (Harsham *et al.*, 2008; Hanna *et al.*, 2007; Caputo *et al.*, 2003). The assessment of time series changes and their analysis using mathematical methods such as Markov chain model and artificial neural network (ANN) methods and using available data is appropriate and reliable methods, and usually have fewer errors.

Zickus et al. (2002), Owega et al. (2006), Kurt et al.



Fig. 1. The location of study area.

(2008), Kukkonen *et al.* (2003), Niska *et al.* (2005), Slini *et al.* (2006), Voukantsis *et al.* (2011), Feng and Moustris (2013) applied ANN models to predict air quality parameters. Li *et al.* (2015) applied ANN methods to simulate $PM_{2.5}$ and PM_{10} . Their results indicated that ANNs performed better than other methods and recommended this method as a reliable and accurate model.

The Markov chain model is a useful mathematical method in reliability research (Wang and Liu, 2012). Several studies have been conducted on the use of Markov models to predict air pollution in the world; Romanof (1982), Nicas (2000), Shamshad (2005). Chung and AitSahlia (2003), Sun *et al.* (2013) applied a Markov chain to determine the probability of various pollution scenarios of $PM_{2.5}$. Their results proved the workability of this method in modelling $PM_{2.5}$.

The main objective of this study was to predict $PM_{2.5}$ concentration and quality of air in Karaj City, Iran using past data on air pollution. We applied neural networks such as Multilayer Perceptron (MLP) and Radial Basis Function (RBF), and Markov Chain model. The MLP, RBF and Markov chain are independent models. The MLP and RBF belong to a family of artificial intelligence neural network, which they aim to predict future physical quantities of $PM_{2.5}$ in the city. The input parameters used in this study were temperature and hourly air concentrations of NO, NO₂, NO_x, CO, SO₂ and PM_{10} . Meanwhile, the Markov chain model was used to predict the probability of occurrence of $PM_{2.5}$ in different periods and to indicate air quality of the

city in three forms, including good quality, sensitive and unhealthy conditions. The models were developed and tested for hourly data for two months in the Karaj Metro area, and the feasibility was discussed in this paper.

1.1 The Study Area

Karaj City is the capital of Alborz Province, Iran. Its population is about 1.97 million, making it the fourthlargest city in Iran after Tehran, Mashhad and Esfahan. It is situated 20 kilometers west of Tehran, at the foothills of the Alborz Mountains. Its coordinates are 50 degrees, 55 minutes and 15 seconds east longitude and 35 degrees, 45 minutes and 50 seconds north latitude. Its area is about 858 km². The annual rainfall of the area is about 261 mm, and the mean annual temperature is between 5 and 13 degree centigrade (Ilanloo, 2011). A monitoring station was considered for this work which is located in the main metro station of the Karaj to Tehran metro (Fig. 1).

2. MATERIALS AND METHODS

Availability of accurate and sufficient data to train an ANN is very significant. The power of ANNs to respond to new problems depends on the primary data to some extent. In this study, air quality parameters were hourly temperature, SO₂, PM₁₀, PM_{2.5}, CO, NO, NO₂ and NO_x. About 1,488 data were available (62 days), of which 1,300 were applied to train the ANN

Parameters	Minimum	Maximum	Average	Standard deviation
			11,01480	
$PM_{2.5}(\mu g/m^3)$	1.00	91.00	26.50	18.02
$PM_{10} (\mu g/m^3)$	4.00	180.00	39.00	22.33
Temperature (°C)	-2.10	+25.80	+16.3	2.46
CO (ppm)	1.50	4.94	2.20	0.38
$SO_2(ppb)$	6	32	13	3.21
NO (ppb)	85	110	88	2.28
$NO_2(ppb)$	29	40	32	1.62
NO _x (ppb)	114	150	120	4.00

Table 1. The statistical summary of data on air quality in Karaj City

and Markov chain, and the rest of the data were used to compare the simulation data with observed data which was monitored by the Karaj department of environment. Table 1 presents the statistical summary of hourly air quality information from March 21, 2015 to May 20, 2015.

To allow better predictions, input and output data in ANNs were normalized in some iterations. Equation 1 was used to normalize data in this study. This function adjusts data in a range of 0 to 1 (Zurada, 1992).

$$N_i = \left| \frac{X_i - X_{min}}{X_{max} - X_{min}} \right| \tag{1}$$

Where N_i and X_i are scaled and the observed value of the parameters; X_{min} and X_{max} represent the lowest and highest amount of a series of the parameters.

2.1 Artificial Neural Networks (ANNs)

An artificial neural network is an idea to process information that is inspired by biological nervous systems and processes information like the human brain. The overall performance of ANN can be observed in Equations 2 and 3 (Hambli, 2011; Haykin, 1999).

$$y_i^m = f(v_i^m) \tag{2}$$

$$v_i^m = \sum_{j=1}^{L} w_{ji}^{m-1} y_j^{m-1} + b_i^m$$
(3)

Where y_i^m and v_i^m are the input and output of i-th neuron in m-th of the hidden layer, f is activation function. L is the number of connections to previously hidden layers and b_i^m represent the weight and bias.

Several different types of ANNs are available. We used Radial Basis function (RBF) and Multilayer Perceptron (MLP) neural networks in this study. We developed all the programs with MATLAB software (R2012a) produced by the Math Work Company.

2.2 MLP Neural Network Structure

MLP neural networks have the ability to determine the number of hidden layers, the number of neurons in each layer and transfer functions used in the layers. These functions can be log sigmoid functions, the one most usually applied in ANN (according to Eq. 4) or tan. Sigmoid functions (according to Eq. 5).

$$f(v_i^m) = \frac{2}{1 + exp(-2v_i^m)} - 1 \tag{4}$$

$$f(v_i^m) = \frac{1}{1 + exp(-\theta v_i^m)}$$
(5)

Where θ is the slope of the transfer function ($\theta = 0.9$).

Several algorithms for training MLP networks exist. In the simplest implementation of these networks, weight and bias are updated in the direction in which efficiency decreases (the opposite direction of the slope). Equation 6 illustrates a repeat of this algorithm (Rumelhart and McClelland, 1986).

$$x_{k+1} = x_k - \alpha_k g_k \tag{6}$$

Where x_k is weight and bias vector; g_k is the slope of the function and a_k is the learning rate. Fig. 2 indicates the MLP neural network used in this study.

Table 2 indicates the parameters of the MLP network design, where the parameter 'show' indicates the number of iterations after which the training status is displayed; α is the speed of learning; the goal is the target error rate. β is the coefficient of momentum and epochs are the frequency of training. Training stop when it reaches the number of levels determined in epochs, or when the amount of the performance function is less than the goal parameter. The learning rate is multiplied by the slope value and used to update weights and bias. If the value of this parameter is too large, the training process will not have enough stability and if it is too small, the algorithm will need a long time to converge. Momentum ratio (β) receives a value between 0 and 1. When the momentum ratio is zero, weight changes are only from the performance function slope and when it is one, weight changes are based on previous weight changes and the slope is ignored.

First of all, in the performance of MLP neural networks, the weights are selected randomly and then are



Fig. 2. Schematic of the MLP network in this study.

Table 2. Characters of MLP network training parameters.

Parameters	Value
Show	1000
α	0.4
Goal	$1e^{-5}$
β	0.9
Epochs	1000
Function	MLP(newff)

applied to the input of the neural network along with randomized bias. After that output prediction are compared with observed output data, and finally the mean square error (MSE) is calculated between the observed data and predicted data. If the error value is less than the desired error set of the network, training stops, otherwise, weights and bias will be changed to reduce errors.

2.3 RBF Neural Network

Radial functions are simply a class of functions. In principle they could be employed in any sort of model, linear or nonlinear. Fig. 3 presents a RBF network, each of n components of the input vector x feed forward to m basis functions whose outputs are linearly combined with weights $\{w_j\}_{j=1}^m$ into the network output f(x) (Orr, 1996).

Compared to MLP neural networks, RBF neural networks need less time to design and more neurons are necessary. When there are many training vectors, these networks have the best performance (Cohen and Intrator, 2002). The procedure in these networks is: the training process continues by increasing the number of hidden layer neurons until the performance function



Fig. 3. Diagram of RBF network.

reaches the target value or until it reaches the maximum number of neurons (the number of data).

The RBF neural networks have an easy architecture. Their structure includes an input layer, a single hidden layer, and an output layer, which at each output node makes available a linear combination of the outputs of the hidden-layer nodes. Training an RBF is comprised of two steps. First, the basic functions are established using an algorithm to cluster data in the training set. Kohohen self-organizing maps (SOMs) or a *k*-means clustering algorithm has been most typically used. Kohohen SOMs (Kohohen, 1984) are a form of 'selforganizing' neural network that learn to differentiate patterns within input data. A SOM will, consequently, cluster an input data according to perceived patterns without having to be given a corresponding output response. *K* means clustering and organizing all objects into a predefined number of groups by minimizing the total squared Euclidean distance for every object regarding its nearest cluster Centre. Nevertheless, other techniques, such as orthogonal least squares and maxi min algorithms, have also been applied (Song, 1996). Next, the weights linking the hidden and the output layer are calculated directly using simple matrix inversion and multiplication. The direct calculation of weights in an RBF makes it far quicker to train than an equivalent MLP (Dawson and Wibly, 2001).

2.4 Model Efficiency

To determine the amount of error in predicting $PM_{2.5}$ and to evaluate the performance of the models, we applied a Root Mean Squared Error (RMSE) and a Mean Bias Error (MBE) which are indicated in Equations 7 and 8. Also, we applied the Nash-Sutcliffe Efficiency Coefficient (E), coefficient of determination (R²) and the Index of Agreement (IA), between the observed and predicted data to illustrate the validity of the model (Feng *et al.*, 2015; Voukantsis *et al.*, 2011; Krause *et al.*, 2005).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P - M)^2}$$
(7)

$$MBE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{P - M}{M} \right| \tag{8}$$

$$E = 1 - \frac{\sum_{i=1}^{n} (\frac{P - M}{M})^{2}}{\sum_{i=1}^{n} (\frac{M - \overline{M}}{\overline{M}})^{2}}$$
(9)

$$R^{2} = \left(\frac{\sum(M - \overline{M})(P - \overline{P})}{\sqrt{\sum(M - \overline{M})^{2}\sum(P - \overline{P})^{2}}}\right)^{2}$$
(10)

$$IA = 1 - \frac{\sum (P - M)^2}{\sum (|P - \bar{P}| + |M - \bar{M}|)^2}$$
(11)

Where P and M are the predicted and the observed values of $PM_{2.5}$ at the time t, respectively, and \overline{M} and \overline{P} are the average of predicted and observed values, respectively and n is the number of data.

2.5 Markov Chain

Several mathematical methods have been used to measure the concentration of air pollutants such as Markov chain model, which was used in this study. Markov chain is a mathematical method for modeling of probabilistic processes. Two features characterize a Markov chain: (a) state space and (b) level. If we define the Karaj weather as a system, its state space (S) in a given hour will be one of three positions in Eq. 12 (Chung and AitSahlia, 2003).

$$\mathbf{S} = \{\mathbf{g}, \mathbf{s}, \mathbf{u}\} \tag{12}$$

Where g is the duration of the day with good air quality. s is sensitive hours and u represents the duration of the day with unhealthy air quality. A Markov chain level specifies that the current state of the system depends on how many of the previous states. To determine the most suitable level in a Markov chain several tests are available. We used the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) tests (Eq. 13 and 14). These tests were performed for different levels and the most suitable criteria were selected based on the lowest AIC or BIC. The AIC and the BIC tests are based on likelihood functions and the likelihood values of a Markov chain, which from zero likelihood (L_0) , the first (L_1) , second (L_2) and third (L_3) according to equations 15 to 18 (Taylor and Karlin, 1998).

$$AIC(m) = -2L^{m} + 2S^{m}(S-1)$$
(13)

$$BIC(m) = -2L^{m} + 2S^{m}Ln(S-1)$$
(14)

$$L_0 = \sum_{j=0}^{s-1} n_j \ln(\hat{P}_j)$$
(15)

$$L_{1} = \sum_{i=0}^{S-1} \sum_{j=0}^{S-1} n_{ij} \ln(\hat{P}_{ij})$$
(16)

$$L_2 = \sum_{h=0}^{S-1} \sum_{i=0}^{S-1} \sum_{j=0}^{S-1} n_{hij} \ln(\hat{P}_{hij})$$
(17)

$$L_2 = \sum_{g=0}^{S-1} \sum_{h=0}^{S-1} \sum_{i=0}^{S-1} \sum_{j=0}^{S-1} n_{ghij} \ln(\hat{P}_{ghij})$$
(18)

Where S is the number of states, m is the order of the Markov chain; n and \hat{P} are the transition count (number of data) and estimated transition probabilities, respectively. The n_{ij} is the observed transition count for a binary time series. For example, the transition count n_{00} specifies the number of consecutive pairs 0's in time series (Wilks, 2006). More information is available by Wilks, 2006.

In this study, according to the results of the AIC and BIC tests, a Markov chain was defined from the first level. Equation 19 indicates its mathematical expression (Logofet and Lensnaya, 2000).

$$\Pr\{X_t | X_{t-1}, X_{t-2}, \dots, X_1\} = \Pr\{X_t | X_{t-1}\}$$
(19)

According to equation 18, the state of a variable in time t, X_t , only depends on its state at the time t-1, X_{t-1} , and it does not depend on the path through which the system reaches its current state. The behavior of a Markov chain can be summarized in the form of a matrix of transition probabilities where each of its elements represents the probability of transition from one mode in the past to another mode later. Transition probability matrix is a k×k matrix, where k is the number of members of a state space. Eq. 20 expresses a case of a transfer matrix and Eq. 21 expresses the first level of a three-state Markov Chain transition probability matrix used in this study (Shamshad *et al.*, 2005).

$$P = (P_{i,j})_{(n \times n)} = \begin{bmatrix} P_{1,1} & \cdots & P_{1,n} \\ \vdots & \ddots & \vdots \\ P_{n,1} & \cdots & P_{n,n} \end{bmatrix}$$
(20)

$$P = (P)_{(3\times3)} = \begin{bmatrix} P_{gg} & P_{gs} & P_{gu} \\ P_{sg} & P_{ss} & P_{su} \\ P_{ug} & P_{us} & P_{uu} \end{bmatrix}$$
(21)

Where subtitles g, s and u as mentioned previously represent time with good, sensitive and unhealthy air quality, respectively, as the first subtitle(subscript) is related to time t–1 and the second subtitle (subscript) is related to time t (e.g. P_{uu} is the probability of occurrence of two consecutive unhealthy hours). Each element of this matrix is determined based on Eq. 22.

$$P_{ij} = \frac{n_{ij}}{\sum n_{ij}} \tag{22}$$

Where n denotes the hours, for example P_{gu} is the probability of occurrence of one hour unhealthy air quality after one hour healthy air quality.

By determining the transition probability matrix of the Markov chain, several analyses can be carried out with the most important of which is the continuing pollution of unhealthy $PM_{2.5}$. Considering that after the occurrence of a continuation of n hours of good air quality certainly an hour of sensitive quality or unhealthy certainly will occur. Eq. 23 is extracted to calculate the probability of n hours of continuing good air quality, and similarly Eq. 24 and 25 are extracted to calculate the probability of sustainability of n hours of unhealthy and sensitive air quality.

$$P_g(n) = P_{gg}^{n-1} \times P_{gs} + P_{gg}^{n-1} \times P_{gu}$$
(23)

$$P_s(n) = P_{ss}^{n-1} \times P_{sg} + P_{ss}^{n-1} \times P_{su}$$
(24)

$$P_u(n) = P_{uu}^{n-1} \times P_{ug} + P_{uu}^{n-1} \times P_{us}$$
⁽²⁵⁾

3. RESULTS AND DISCUSSION

Table 3 indicates the correlation between parameters

Table 3. The cross correlation coefficients between different air pollutant parameters.

	PM _{2.5}	PM_{10}	NO	NO_2	NO _x	SO_2	СО
PM_{25}	1						
PM_{10}	0.137	1					
NO	0.072	0.258	1				
NO_2	0.128	0.343	0.829	1			
NO _x	0.101	0.306	0.971	0.937	1		
SO_2	0.116	0.208	0.381	0.508	0.453	1	
CO	0.289	0.289	0.139	0.432	0.272	0.185	1

Table 4. Input parameters in different neural networks.

Models	NO	NO_2	NO_x	PM_{10}	SO_2	СО	Temperature
ANN1	*			*		*	
ANN2	*	*			*		*
ANN3			*	*	*		
ANN4		*	*			*	*
ANN5				*	*	*	
ANN6	*			*	*		*
ANN7			*	*		*	*

T: Tan-sigmoid, L: Log-Sigmoid

in our study. There should not be a high correlation between parameters in a stimulating process (Kuncheva, 2004), because then there is no need to utilize complex models such as neural network and the problem can be easily solved by regression methods. We used 1,300 data. Correlation was based on Pearson's. As indicated in Table 3, PM_{2.5} has significant correlations with PM₁₀, NO₂, NO_x, SO₂, CO at 1% significance level. However, according to the results of Table 3, very weak correlation exists between PM_{2.5} and each of other parameters including PM₁₀ (0.137), NO (0.072), NO₂ (0.128), NO_x (0.101), SO₂ (0.116) and CO (0.289) parameters. Because of very weak correlation exists between PM_{2.5} and each of the input parameters. Therefore, we used the parameters as inputs to the neural network.

3.1 Determining the Optimal Parameters in Predicting the Amount of PM_{2.5}

One of the major issues affecting the performance of a neural network is to select the input parameters to train the network. For this purpose, various algorithms have been used in previous studies (Niska *et al.*, 2006; Eleuteri *et al.*, 2005; Kohavi and John, 1997). The use of these methods has limitations and errors; therefore, in this study the decision on the choice of these parameters was based on their performance in the network training. Therefore, various models were applied and their performance were compared (Table 4). In all of these models, we applied two hidden layers with 15 neurons in the each layer (15 neurons in the first layer and 15 neurons in second layer) and type of transfer functions in the layers (First layer: tan-sigmoid transfer function; and second layer: log-sigmoid transfer function) were considered, so that the impact of changes in input parameters on the network was tangible. For example, our ANN6 model was made by using NO, PM_{10} , SO₂ and temperature as inputs with two hidden layers, 15 neurons in the first layer and 15 neurons in the second layer and transfer functions Tansig and Logsig.

3.2 MLP Neural Networks

We trained the network using 1,300 data, to predict hourly concentrations of $PM_{2.5}$ in Karaj City. To determine the most suitable network in predicting output, we changed the characteristics of the network such as, inputs, number of neurons in the hidden layers, type of transfer functions, learning rate and momentum factors. To choose the most accurate and reliable model, the amount of errors, R^2 , IA and E were computed. As a result, it was revealed that the network in which input parameters were CO, NO_x , PM_{10} and temperature, (ANN7), had a better performance than the other scenarios (Table 5).

After selecting the optimal input parameters and appropriate number of hidden layers (ANN7), the effects of the other main factors in ANN performance (normalizing data) were evaluated. Networks using input data CO, NO_x , PM_{10} and temperature were trained once by

Table 5. The results of the MLP model with different scenarios of inputs.

Models	R ²	IA	Е	RMSE	MBE
ANN1	0.77	0.78	0.68	5.72	0.331
ANN2	0.78	0.79	0.65	6.35	0.359
ANN3	0.74	0.77	0.61	7.04	0.386
ANN4	0.80	0.82	0.78	4.41	0.281
ANN5	0.79	0.81	0.75	4.89	0.316
ANN6	0.80	0.83	0.81	3.58	0.275
ANN7	0.83	0.84	0.88	2.76	0.063

Table 6. the characteristic of the model ANN7.

normalizing data and once without it. The number of neurons in the hidden layers in the MLP networks was changed using trial and error and were automatically changed in RBF neural networks from 0 to 1,488 (number of data). Table 6 indicates the characteristic of ANN7 applied in our study.

Fig. 4 indicates the performance of various networks. Normalization of the data improved network performance. Increasing the number of neurons in the hidden layer in the MLP and the RBF networks reduced forecast error and increased the coefficient of determination. A network made with two hidden layers including 19 neurons in the first layer and 16 neurons in the second layer, using the normalized data (FN19/16) had a coefficient of determination, efficiency (E), RMSE and MBE equal to 0.92, 0.981, 1.25 and 0.0545, respectively. The network had the best performance among the MLP networks which we developed (Fig. 4).

Fig. 5 illustrated the comparison of the observed data with the predicted data using MLP network with 188 data test for $PM_{2.5}$ parameter. The coefficient of determination was 0.92 and the RMSE was 1.25. The result indicates the good workability of the model.

3.3 The Effect of Learning Rate and Momentum Factor

Fig. 6 indicates the effect of the learning rate and momentum factor in the performance of MLP networks. First, we fixed the rate of learning (on each of the values: $\alpha = 0.05$, 0.2, 0.4, 0.65, 0.8, 0.95). We increased the momentum coefficient (β) from 0.05 to 0.95 in steps of 0.05 and trained and tested each step 20 times. Each of the network errors in predicting PM_{2.5} values were averaged and considered as the error value for a model with factor of momentum and learning rate. These tests were made on the MLP model with two hidden layers (FN19/16). The results indicated that increasing the learning rate caused weakens of the network performance. On the other hand, lower values in the learning rate made the learning process time-consuming.

Network type	MLP	MLP	RBF
Input data normalization	No	Yes	No
Train function	trainlm	trainlm	_
Number of layers	2	2	1
Properties for layer 1	0-20	0-20	Selected by network
number of neurons	(Trial and error)	(Trial and error)	(0-Number of data(1,488))
Transfer function	Tansig	Tansig	Radbas
Properties for layer 2	0-20	0-20	
number of neurons	(Trial and error)	(Trial and error)	—
Transfer function	Logsig	Logsig	-

Tansig: tan-sigmoid transfer function; Logsig: log-sigmoid transfer function



Fig. 4. The performance of MLP networks with various structures (N: normalized data, 0: without normalization).



Fig. 5. A comparison between observed and predicted $PM_{2.5}$ data using FN19/16.



Fig. 6. The effect of learning rate and momentum factors in the RMSE rates in the MLP neural network.

Increasing momentum from 0.05 to 0.95 first increased errors, then improved performance and reduced errors. Therefore, we concluded that determining changes of weight by using any of the slopes of efficiency function or previous weight changes alone improves network performance compared to when the both of these factors are involved in determining the new weight. To select the momentum factor in the range of 0 to 1 improved network performance in predicting $PM_{2.5}$ in our research.

Correct designing of the parameters of an MLP neural network such as input parameter, the number of layers, the number of hidden layer neurons, transfer functions, learning rate and momentum factors and normalizing data also increases the accuracy. We considered all mentioned parameters and determined the effect of learning rate and momentum factor in the amount of RMSE in the MLP neural network (Fig. 6) which causes a good performance of predicting PM_{2.5} in the Karaj City. Voukantsis et al. (2011) used a principal component analysis to select input parameters for MLP neural network and predicted PM_{10} and PM_{25} . They obtained IA = 0.8. We carried out other method for selecting input parameters and selected those input parameters which each of input parameter had a low correlation coefficient with PM2.5 parameter. Index of agreement (IA) in our study for ANN7 was 0.84. The advantage of their work was to select a few meteorological parameters as input to the ANN. Our input parameters for MLP and RBF were NO, NO₂, NO_x, PM₁₀, SO₂ and temperature. However, meteorological data was not available for authors. Bahari et al. (2014) predicted PM2.5 concentrations, in one station, in Tehran using an MLP neural network. Their input parameters were temperature, wind speed, wind direction, relative humidity, and cloud cover and inversion strength. They did not describe the method of selecting input parameters to the ANNs. The R^2 of the study were between 0.61 and 0.79. However, the R^2 in our study for MLP and RBF was 0.92 and 0.93, respectively.

We compared the results our MLP neural network with the results of Voukantsis *et al.* (2011) and Feng *et al.* (2015) and found that our model presented a suitable performance in predicting $PM_{2.5}$ concentrations in Karaj City. Feng *et al.* (2015) obtained RMSE rate between 28 to 36 for one day and two days $PM_{2.5}$ prediction using an MLP neural network. The amount of RSME in the MLP neural network was 1.25 for FN19/ 16 model. Their RSME results could be due to the smaller number of data used in their studies and may need a longer data collection. In fact, they changed in concentrations of suspended particles over the year lead to a reduction in network accuracy in predicting the amount of this parameter.

3.4 RBF Neural Network

In RBF networks that are formed from a hidden layer and Radial Basis transfer function, the number of neurons starts from zero and increases. At each stage, error calculated and reported. This process continues until the error decreases to zero, or the number of neurons is equal to the number of input data. Fig. 7 indicates the changes of prediction error for $PM_{2.5}$ due to the increase of the number of neurons of hidden layer using this method. The root mean square error of the neurons in this method decreased with increasing neurons from 7.88 to 2e-06 and the coefficient of determination reached 0.93. The results indicate the proper functioning of this network in predicting the concentration of $PM_{2.5}$, without requiring any design. The coefficient of determination between observed data and predicted data



Fig. 7. The changes of RMSE through increasing the number of neurons in RBF network.



Fig. 8. A comparison between observed and predicted $PM_{2.5}$ data using RBF network.

reached 0.93 which indicates the reliability of RBF in predicting of $PM_{2.5}$ (Fig. 8).

The training of neural network structure of artificial neural network which was used in our study was, according to design of continuous statistical model by using past data. This artificial neural network presents a numerical description of a mathematical structure which is able to predict the physical condition of air pollution for 24 or 48 hours in advance. By increasing the length of existing data, we can increase the possibility of predictions durations.

3.5 Markov Chain

Equation 26 presents the transition probability matrix and Fig. 9 indicates its graphical representation for 9 different transition mode. The results of this matrix state that, in the event of good or sensitive air quality, there is the possibility of repetition of the above condition (probability 71% and 89% respectively). However, in the event of pollution and poor quality in one of night and day hours, the possibility of repeating or transferring to sensitive state is somehow the same value (respectively 48 and 51 percent probability) and there is no high chance for the air to be healthy in the following hours. Generally, in the case of healthy, fresh air, the likelihood of its re-occurrence, unhealthy state, the possibility of switching to a sensitive air quality, and in the case of sensitive weather, the likelihood of its continuation get the highest value.

$$P_{(3\times3)} = \begin{cases} g & s & u \\ s & 0.7091 & 0.28 & 0.0109 \\ 0.0796 & 0.886 & 0.0346 \\ 0.0141 & 0.507 & 0.4788 \end{cases}$$
(26)

Table 7 presents the number of incidents and the probability of selection for each of the two data sets initial calculations (1,300 data) and test (data 188) with the error between them. By calculating the transition probability matrix for 188 test data, showed that the matrix with slight error is similar to the calculated matrix for the raw data. For example, the possibility of transfer of clean air quality to sensitive state (gs) was calculated as 0.28 that compared with the probability of the situation in test data that was equal to 0.2758, with absolute error of 0.014 has acceptable accuracy and precision. As a result, the probability of occurrence of PM_{2.5} pollution in different periods is predictable using this method.

Fig. 9(a-c) indicates the continuation of n hour for 2 hours to 1 day for each of air quality conditions. The vertical axis is the likelihood to maintain a state and the horizontal axis presents duration time. For example, in the first curve (Fig. 9(a)), P(n) is the probability that a good air quality status is maintained for n hours



Fig. 9. The probability of continuation of n hour (a) good, (b) sensitive, (c) unhealthy AQ statement.

and then the situation will change. Fig. 9(c) illustrates the high probability of remaining two-hour unhealthy state (25%) the poor status of $PM_{2.5}$ pollutant. However, the likelihood of continuation of this situation to fourth and fifth hours is reduced and there will be no substantial risk of the continuity of too much pollution. Also, due to the low likelihood of continuation of

*	•									
		gg	gs	gu	sg	SS	su	ug	us	uu
Calculating (1300 data)	Number (N) Probability (P)	195 0.709	77 0.28	3 0.011	76 0.0796	845 0.885	33 0.035	1 0.0141	36 0.507	34 0.479
Testing (188 data)	Number (N) Probability (P)	21 0.724	8 0.2758	0 0	9 0.0643	124 0.886	7 0.05	0 0	6 0.4	9 0.6
Error (%)		0.021	0.014	_	0.19	0.00003	0.31	_	0.21	0.2

Table 7. The transition probability of the various state.

healthy condition and opposite to that its relative high likelihood in some hours or daily continuation of sensitive quality, we will not experience favorable weather conditions in terms of particulate pollution, especially for sensitive groups.

The MLP, RBF and Markov chain are independent model. The MLP and RBF belong to a family of artificial intelligence neural network, which they can predict future physical quantities of $PM_{2.5}$ in a city. The RBF results may be accurate than MLP model. The Markov chain model can describe the probability of occurrence of $PM_{2.5}$ in different periods and also indicates the quality of air in a city in three forms, including good quality, sensitive and unhealthy conditions.

4. CONCLUSION

Considering the results and discussion of using the MLP neural network, RBF neural network and Marko chain to predict $PM_{2.5}$ pollutant in the Karaj city, Iran, we summarized the results as follows:

- 1. The MLP neural network needs a suitable design for optimal performance. We developed an MLP neural network containing two hidden layers with 19 neurons in the first layer and 16 neurons in the second layer. The MBE was 0.0545 which indicates the adequacy of the MLP neural network. The R² and Index of agreement (IA) between the observed data and the predicted data were 0.92 and 0.93, respectively.
- 2. Change of momentum and learning coefficients indicated that increasing the learning rate increases MLP network error, and thus choosing lower the learning rate improves network performance. On the other hand, low learning rate model reduces modeling speed. The increase momentum rate from 0 to 1 increases the error and then reduces it. This issue proves that the selection of new weight based on the performance function slope or previous weight will result in better network performance.
- 3. Selecting appropriate learning rate and momentum factors caused improving performance of artificial

neural network.

- 4. The RBF neural network using a hidden layer with transfer radial basis had an easy and good performance to predict hourly $PM_{2.5}$ pollutant. By increasing the number of neurons from zero to 1,488 (equal to the number of data) the errors of this network dropped from 7.88 to 2E-06 and coefficient of determination between observed data and predicted data reached 0.92.
- 5. The RBF prediction of hourly PM_{2.5} reaches to more accurate results if we compare to the MLP neural network.
- 6. Markov chain model results indicated that the air quality in the coming months of 2015, will continue in a sensitive state, which is dangerous for people with heart disease and respiratory problems.

REFERENCES

- Bahari, R.A., Ali Abssaspour, R., Pahlavi, P. (2014) Prediction of PM_{2.5} concentrations using temperature inversion effects based on an artificial neural network, The ISPRS international conference of Geospatial information research, 15-17 November, Tehran, Iran.
- Caputo, M., Gimenez, M., Schlamp, M. (2003) Intercomparison of atmospheric dispersion models. Atmospheric Environment 37, 2435-2449.
- Chung, K.L., Farid AitSahlia (2003) Elementary Probability Theory: With Stochastic Processes and an Introduction to Mathematical Finance, Springer Undergraduate Texts in Mathematics and Technology, ISSN 0172-6056.
- Cohen, S., Intrator, N. (2002) Automatic model selection in a hybrid perceptron/radial network; Information Fusion. Special Issue on Multiple Experts 3(4), 259-266.
- Deng, X., Zhang, F., Rui, W., long, F., Wang, L., Feng, Z., Chen, D., Ding, W. (2013) PM_{2.5}-induced oxidative stress triggers autophagy in human lung epithelial A549 cells. Toxicology in Vitro 27(6), 1762-1770.
- Dong, G.H., Zhang, P., Sun, B., Zhang, L., Chen, X., Ma, N. (2012) Long term exposure to ambient air pollution and respiratory disease mortality in Shenyang, China: a 12 year population - based retrospective cohort study.

78 Asian Journal of Atmospheric Environment, Vol. 10(2), 67-79, 2016

Respiration 84(5), 360-368.

- Eleuteri, A., Tagliaferri, R., Milano, L. (2005) A novel information geometric approach to variable selection in MLP networks. Neural Network 18(10), 1309-1318.
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J. (2015) Artificial neural network forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. Atmospheric Environment 107, 118-128.
- Goss, C.H., Newsom, S.A., Schildcrout, J.S., Sheppard, L., Kaufman, J.D. (2004) Effect of ambient air pollution on pulmonary exacerbations and lung function in cystic fibrosis. American Journal of Respiratory and Critical Care Medicine 169(7), 816-821.
- Hambli, R. (2011) Multiscale prediction of crack density and crack length accumulation in trabecular bone based on neural networks and finite element simulation. International Journal for Numerical Methods in Biomedical Engineering 27(4), 461-475.
- Hanna, S.R., Paine, R., Heinold, D., Kintigh, E., Baker, D. (2007) Uncertainties in air toxics calculated by the dispersion models AERMOD and ISCST 3 in the Houston ship channel area. Journal of Applied Meteorology and Climatology 46, 1372-1382.
- Harsham, D.K., Bennett, M. (2008) A sensitivity study of validation of three regulatory dispersion models. American Journal of Environmental Sciences 4(1), 63-76.
- Haykin, S. (1999) Neural networks: a comprehensive foundation. (2nd ed.) Upper Saddle River, New Jersey: Prentice Hal.
- Jones, R.M., Nicas, M. (2014) Benchmarking of a Markov multizone model of contaminant transport. Annals of Occupational Hygiene 58(8), 1018-1031.
- Kohavi, R., John, G.H. (1997) Wrappers for feature subset selection. Artificial Intelligence 97, 273-324.
- Kohohen, T. (1984) Self-organization and associative memory. New York: Springer-Verlag.
- Krause, P., Boyle, D.P., Bäse, F. (2005) Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences 5, 89-97.
- Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Li, P., Xin, J.Y., Wang, Y.S., Wang, S.G., Li, G.X., Pan, X.C., Liu, Z.R., Wang, L.L. (2015) Reinstate regional transport of PM_{2.5} as a major cause of severe haze in Beijing. Proceeding of the National Academy of Sciences of the United States of America 112, E2739-E2740.
- Kuncheva, L. (2004) Combining Pattern Classifiers: Methods and Algorithms. Wiley, New York, USA.
- Kurt, A., Gulbagci, B., Karaca, F., Alagha, O. (2008) An online air pollution forecasting system using neural networks. Environment International 34, 592-598.
- Logofet, D.O., Lensnaya, E.V. (2000) The mathematics of Markov models: what Markov chains can really predict in forest successions. Ecological Modelling 2(3), 285-298.
- Nicas, M. (2014) Markov modeling of contaminant con-

centrations in indoor air. American Journal of Environmental Sciences, 61(4), 484-491.

- Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G. (2003) Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modeling system and measurements in central Helsinki. Atmospheric Environment 37, 4539-4550.
- Niska, H., Heikkinen, M., Kolehmainen, M. (2006) Genetic algorithms and sensitivity analysis applied to select inputs of a multi-layer perceptron for the prediction of air pollutant time-series. Chapter Intelligent data engineering and automated learning-IDEAL2006 volume 4224 of the series lecture notes in computer science pp. 224-231 springer publisher.
- Niska, H., Rantamäki, M., Hiltunen, T., Karppinen, A., Kukkonen, J., Ruuskanen, J. (2005) Evaluation of an integrated modelling system containing a multi-layer perceptron model and the numerical weather prediction model HIRLAM for the forecasting of urban airborne pollutant concentrations. Atmospheric Environment 39(35), 6524-6536.
- Orr, M.J.L. (1996) Introduction to radial basis function networks, University of Edinbergh, EH89LW.
- Owega, S., Khan, B.U.Z., Evans, G.J., Jervis, R.E., Fila, M. (2006) Identification of long-range aerosol transport patterns to Toronto via classification of back trajectories by cluster analysis and neural network techniques. Chemo Metrics and Intelligent Laboratory Systems 83(1), 26-33.
- Romanof, N. (1982), A Markov chain model for the mean daily SO₂ concentrations. Atmospheric Environment 16(8), 1895-1897.
- Rumelhart, D.E., McClelland, J.L. (1986) Parallel distribution processing: Exploration in the microstructure of cognition, Cambridge, MA: MIT Press.
- Shamshad, A., Bawadi, M.A., Wan Hussin, W.M.A., Majid, T.A., Sanusi, S.A.M. (2005) First and second order Markov chain models for synthetic generation of wind speed time series. Energy 30, 693-708.
- Slaughter, J.C., Lumley, T., Sheppard, L., Koenig, J.Q., Shapiro, G.G. (2003) Effects of ambient air pollution on symptom severity and medication use in children with asthma. Annals of Allergy, Asthma and Immunology 91(4), 346-353.
- Slini, T., Kaprara, A., Karatzas, K., Moussiopoulos, N. (2006) PM₁₀ forecasting for Thessaloniki, Greece. Environ. Modell. Softw. 21, 559-565.
- Song, X.M. (1996) Radial basis function networks for empirical modeling of chemical process. MSc thesis, University of Helsinki.
- Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W., Liu, S. (2013) Prediction of 24-hour-average PM_{2.5} concentrations using a hidden Markov model with different emission distributions in Northern California. Science of the Total Environment 443, 93-103.
- Taylor, H., Karlin, S. (1998) An Introduction to Stochastic Modeling. Academic Press, San Diego, California.

- Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., Kolehmainen, M. (2011) Intercomparison of air quality data using principal component analysis, and forecasting of PM₁₀ and PM_{2.5} concentrations using artificial neural networks, in Thessaloniki and Helsinki. Science of the Total Environment 409, 1266-1276.
- Wang, X., Liu, W. (2012) Research on Air Traffic Control Automatic System Software Reliability Based on Markov Chain. Physics Procedia 24, 1601-1606.

Wilks, D.S. (2006) Statistical methods in the atmospheric

sciences. 2nd ed. Academic Press, xvii, 627 p.

- Zickus, M., Greig, A.J., Niranjan, M. (2002) Comparison of four machine learning methods for predicting PM₁₀ concentration in Helsinki, Finland. Water, Air and Soil Pollution 2(5), 717-729.
- Zurada, J.M. (1992) Introduction to Artificial Neural Systems, PWS; Singapore, 195-196.

(Received 30 November 2015, revised 7 February 2016, accepted 26 April 2016)