A Comparative Study of Computational Intelligence Techniques Applied to $PM_{2.5}$ Air Pollution Forecasting

Mihaela Oprea
Automatic Control, Computers and Electronics Department
Petroleum-Gas University of Ploiesti, Romania
mihaela@upg-ploiesti.ro

Sanda Florentina Mihalache
Automatic Control, Computers and Electronics Department
Petroleum-Gas University of Ploiesti, Romania

Marian Popescu
Automatic Control, Computers and Electronics Department
Petroleum-Gas University of Ploiesti, Romania

***Abstract*— The paper presents the results of a comparative study performed between two computational intelligence techniques, artificial neural networks (ANNs) and adaptive neuro-fuzzy inference systems (ANFIS) applied to particulate matter (fraction $PM_{2.5}$) air pollution forecasting. The experiments were realized on datasets from the Airbase databases with $PM_{2.5}$ hourly measurements. The main statistical parameters that were computed are root mean square error (RMSE) and mean absolute error (MAE).**

Index Terms-computational intelligence; ANN; ANFIS; particulate matter; air pollution forecasting

I. INTRODUCTION

Computational intelligence provides several techniques for solving time series forecasting problems, neural ones being the most used and successful. However, depending on the application domain knowledge and the specific requirements related to the parameter that needs to be predicted, a more informed (accurate) forecasting solution has to be given. Thus, the neural method has to be properly tailored to the application. The research work reported in the literature presents such particular solutions, which combines neural methods with other techniques.

A. *The forecasting problem domain*

In this paper, the focus is on an environmental pollution forecasting problem, air pollution with particulate matter (PM) having the diameter less than 2.5 μm ($PM_{2.5}$, i.e. fine particles), which are of special interest in urban areas, due to their important potential effects on the health of humans and, particularly, on children's health. At present, $PM_{2.5}$ air pollutant, as well as the other fractions of PM (coarse one such as PM_{10} and more fine ones such as $PM_{1.0}$) are worldwide intensively studied at both levels, monitoring and analysis with early warning of potential PM air pollution episodes.

As the aerodynamic diameter of PM is less, much higher is the dangerous for vulnerable people (children, elderly people, and people with health problems) as they can penetrate the lungs and can affect the respiratory system and the cardiovascular system. Therefore, apart from the regulatory norms that were adopted by the Environmental Protection Agencies (e.g. US EPA, EEA in Europe), the air quality real-time monitoring networks developed at the level of each country or at local level, in a certain region, started to include a real-time forecasting system that provides the prediction of specific air pollutants and some warning messages (shown on their websites) to prevent or, at least, to minimize the effects of air pollution episodes on people health. Usually, two types of the air quality index (AQI) are computed, a specific one, for each air pollutant that is monitored, and a global one.

B. *Forecasting approaches*

Two approaches are currently applied to air pollution forecasting: deterministic and non-deterministic. The first category provides better solutions but in a longer period of time which cannot be used in real-time forecasting systems. From the second category, the approaches based on computational intelligence are the most promising to solve real-time forecasting problems. In particular, artificial neural networks approaches were applied to atmospheric forecasting. One of the first reviews on this topic was described in [1].

In this paper, we present a research study on the application of two computational intelligence techniques: artificial neural networks and adaptive neuro-fuzzy inference systems to $PM_{2.5}$ air pollution forecasting. A comparison between the results obtained with these two techniques was performed in order to learn more about their efficient use in real-time forecasting. Moreover, our research purpose was to provide a $PM_{2.5}$ forecasting model that is simple enough, provides a good prediction accuracy and is based solely on time series with $PM_{2.5}$ measurements. This last characteristic is very good when meteorological data or other air pollutants measurements are not available.

II. Neural methods for PM forecasting

The neural methods applicable to solve forecasting problems are: artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). They can perform air pollution forecasting more efficiently than the deterministic methods by capturing the knowledge accumulated in the historical data sets (time series) which is learned via a training algorithm and is used to accurately predict specific air pollution parameters (e.g. air pollutants concentrations).

A. Artificial neural networks

Artificial neural networks are universal approximators of non-linear functions [2]. They are composed by a number of non-linear processing units named artificial neurons which are structured in layers. An artificial neuron (represented in Fig. 1) has a number of inputs (an input vector, *x*), each one with a specific weight (*w*), an activation function (*f*), a bias (*b*) and one output (*y*), computed with relation

$$y_i = f_i\left(\sum_{j=1}^{n} w_{ij} x_j + b_i\right) \quad (1)$$

where, x_j is the j^{th} input, y_i is the output of the i^{th} neuron, w_{ij} is the weight of the connection between the j^{th} input and the i^{th} neuron, b_i is the bias of the i^{th} neuron and f_i is the activation function of the i^{th} neuron.

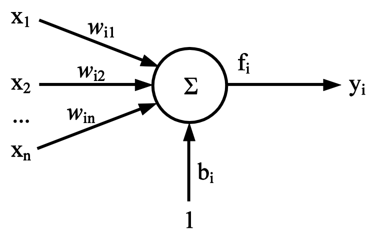

Fig. 1. The representation of an artificial neuron.

Forecasting problems are solved mainly by feed-forward ANNs (multi-layer perceptron - MLP and radial basis function - RBF) and recurrent ANNs. Fig. 2 shows the general architecture of a feed-forward ANN, which has an input layer, some hidden layers and an output layer.

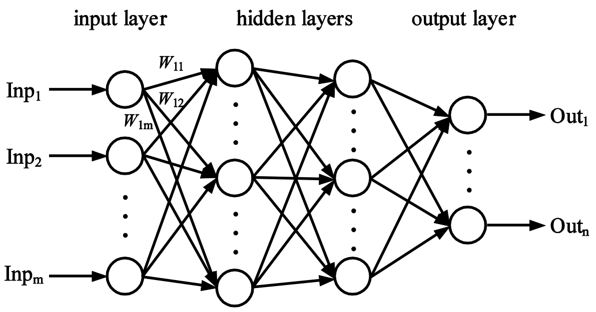

Fig. 2. The general architecture of a feed-forward ANN.

A recurrent neural network is a type of artificial neural network that has a directed cycle made by the connections between the artificial neurons, allowing it to have an internal state. Thus, it exhibits a dynamic behaviour which provides the ability to process and predict chaotic time series for long-terms [3].

The best structure of an ANN is experimentally determined. Usually, a single hidden layer is enough to capture the nonlinearity of any function. The number of hidden nodes is chosen by experiments, while the number of input and output nodes is set according to the forecasting problem that has to be solved. The number of input nodes represents the input window (in the case of time series, the number of past hours measurements) and the number of output nodes represents the forecast horizon (number of future time steps, in hours, days etc., for which the prediction is determined). The ANN is trained with a training set (which is extracted from a data set) by using a specific training algorithm (the most used being backpropagation and the Levenberg-Marquardt algorithm), after that following the validation and testing steps which are executed on the validation set and testing set, respectively. Details on the ANN computational algorithms are given in the literature (see e.g. [4]).

Several PM neural forecasting systems were reported in the recent literature. Most of them use meteorological parameters and air pollutants concentrations measurements. A brief review is presented as follows.

A feed forward ANN with backpropagation training algorithm is described in [5], for 3 days in advance forecasting of PM_{10}, SO_2 and CO air pollutants (AP) levels in the Besiktas district in Istanbul, Turkey. The ANN is integrated in the AirPol system (http://airpol.fatih.edu.tr). The ANN inputs are daily meteorological forecasts and the AP indicator values. The authors applied some geographical models, the most complex one being based on the distance between two sites in the case of using three selected neighborhood districts.

In [6] it is demonstrated the efficacy of using EnviNNet, a prototype stochastic ANN model for air quality forecasting in cities from Italy (Rome, Milan and Napoli) to predict PM_{10} in Phoenix, Arizona in comparison with the use of CMAQ system. The ANN is a MLP that uses the conjugate-gradient method for training.

Another research work that reports the successful use of ANNs and principal component analysis for PM_{10} and $PM_{2.5}$ forecasting in two cities, Thessaloniki (Greece) and Helsinki (Finland) is described in [7]. The authors used a MLP and the meteorological and AQ pollutants to predict the next day mean concentration of PM_{10} and $PM_{2.5}$.

A genetically optimized ANN and k-means clustering was applied in [8] to predict PM_{10} and $PM_{2.5}$ in a coastal location of New Zealand.

One of the recent work on $PM_{2.5}$ neural forecasting is described in [9]. The authors propose a novel hybrid model that combines air mass trajectory analysis with wavelet transformation in order to improve the accuracy of the average $PM_{2.5}$ concentration two days in advance neural forecasting. The ANN is a MLP trained with a backpropagation algorithm and the Levenberg-Marquardt (LM) algorithm. Also, early stopping was used to avoid overfitting, and some meteorological parameters were used.

B. Adaptive neuro-fuzzy inference systems

The ANFIS method applied to prediction uses a hybrid architecture composed by a fuzzy inference system FIS enhanced with ANN features proposed by Yang [10]. The advantages of FIS are mainly its design that emulates human thinking and the simple interpretation of the results. Integrating the ANN part into a fuzzy inference system enhanced the FIS part with learning/adapting capabilities. The prediction model does not use a mathematical model as well as the case of ANN. The ANFIS architecture has the structure given in Fig. 3.

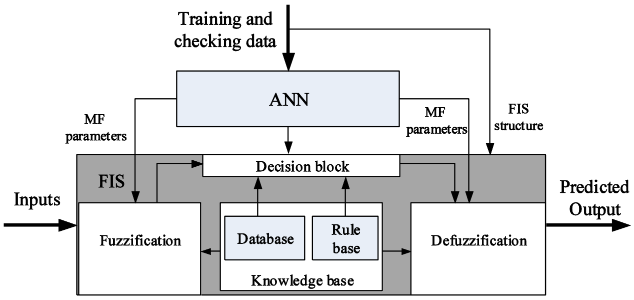

Fig. 3. ANFIS structure.

The FIS part is formed by five functional units: a fuzzification unit (from crisp value to fuzzy set), a defuzzification unit (from fuzzy set to a crisp value), the database unit (containing the description of membership functions for input/output variables), a rule base unit (all the rules defined for FIS), and the decision unit (performing the inference operations on the fuzzy rules) [11]. The neuro-fuzzy architecture is capable to learn new rules or membership functions, to optimize the existing ones (Fig. 3). The training data determine restrictions on the design methods for the rule base and membership functions. Usually the particular type of datasets for $PM_{2.5}$ eliminates the subclustering method in generating the FIS structure, a good choice being the grid partition method.

The ANFIS architecture (Fig. 3) has five layers, with Takagi-Sugeno rules. The first layer (adaptive) forms the premise parameters (the IF part with inputs and their membership functions). The second layer computes a product of the involved membership functions. The third layer normalizes the sum of inputs. In layer 4, the adaptive i-node computes the contribution of i-th rule to ANFIS output, forming the consequence parameters (the THEN part with output and its membership function). The fifth layer makes the summation of all inputs. The ANN part can improve the membership functions associated with FIS structure. Usually these membership functions are the tuning parameters of the FIS. Their initial values are chosen from experience or trial and error methods. In the training mode the ANN finds the most suited membership functions for the input-output relation described by FIS, according to training and checking dataset.

ANFIS applies a hybrid learning algorithm (H) or backpropagation (BP) algorithm. The hybrid learning algorithm identifies premise parameters with gradient method and consequence parameters with least square method. At feedforward propagation step from H, the system output reaches layer 4, and the consequence parameters are formed with least square method. With backpropagation (BP) optimization method, the error signal is fed back and the new premise parameters are computed through gradient method. The prediction method is tested with datasets that respect the main features of training dataset. The prediction precision decreases with enlarging the prediction window from one hour in advance to six hours in advance. The use of pure FIS for specific air pollutants is reported in literature [12, 13]. Examples of ANFIS based systems for the prediction of air pollutants concentrations are presented in [14, 15, 16, 17].

III. Comparative Study

A. Data sets

For this study were chosen two urban traffic $PM_{2.5}$ monitoring stations. Taking into consideration the lack of data for the Romanian stations, the data used for training and validation were from a München (Germany) station where sufficient data are available (4 years) in the Airbase database. In addition, the testing was done using also data (from 2015) from Ploiești (Romania), a city which presents interest as part of a larger project (ROKIDAIR project – www.rokidair.ro). The two data sets have the characteristics presented in Table I.

TABLE I. Data sets Characteristics

$PM_{2.5}$ properties / Station	No of samples	Range [μg/m^3]	Mean [μg/m^3]	Standard deviation [μg/m^3]
München (DEBY115)	34009	0.52-168.5	20.41	13.53
Ploiești (PH-2)	4073	3.24-36.45	14.80	5.67

In Fig. 4 is presented the hourly evolution of data from München for a period of four years (2009-2012).

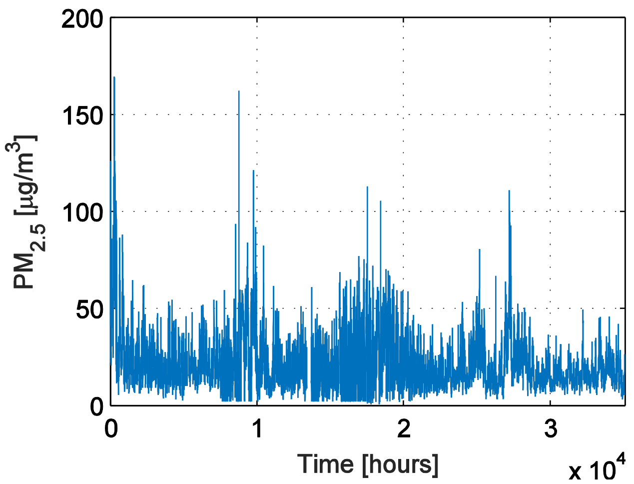

Fig. 4. Hourly $PM_{2.5}$ concentration from the München station (for 4 years).

In the time series there are missing data and in this case the valid data between the gaps were processed separately to meet the model requirements and then concatenated. The data were normalized and then divided (randomly) as follows: 70% for training, 15% for validation and 15% for testing the model.

As the most air pollution critical hours in urban regions during weekdays are 8, 12, 16 and 20 (mainly, due to higher traffic), it was chosen 4 as forecasting model's input window. This number of inputs was set by experiment and by taking into account that for any 4 consecutive values of the $PM_{2.5}$ measurements, the measurement of one critical hour is kept, when we analyse the air pollution episodes that arise during the day. In particular, the critical hours are also important for children who travel to kindergartens and schools or back to their homes.

Taking into consideration the above statements the proposed architecture (Fig. 5) for the forecasting model (ANFIS or ANN) has as inputs the values of $PM_{2.5}$ concentrations from current hour to three hours ago and as output the prediction for the next hour.

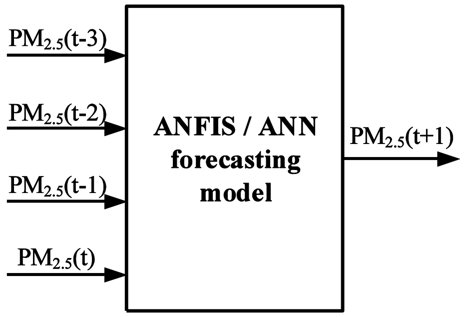

Fig. 5. Structure of the proposed architecture.

In order to determine the suitability of the models the following statistical parameters were calculated:

- MAE – mean absolute error;
- RMSE – root mean square error;
- IA – index of agreement;
- R^2 – coefficient of determination;
- R – correlation coefficient.

The two errors measure how close the predicted data are to the true values and have to be as small as possible, and the last three indices are numbers that indicate how well the data fit the prediction model and they should be close to 1.

The experiments presented in the following sections were performed using MATLAB® environment.

B. *Experimental results for ANFIS model*

In this study, for the generation of the fuzzy inference system was used the *grid partition* method, imposed by the specific of data. The membership functions were *triangular* and *Gaussian* and the optimization algorithms of the ANN were *backpropagation* and *hybrid*.

The ANFIS model was tested for all combinations between the modifiable parameters, namely the membership functions types and the optimization methods.

Table II presents the experimental results of the ANFIS model tested with data from München.

TABLE II. STATISTICAL PARAMETERS FOR ANFIS MODEL (MÜNCHEN)

ANFIS structure	MAE [µg/m³]	RMSE [µg/m³]	IA	R^2	R
Trimf/Hybrid	1.9614	3.2564	0.9796	0.9223	0.9604
Gauss/Hybrid	**1.9419**	**3.2089**	**0.9801**	**0.9246**	**0.9616**
Trimf/Backpr.	2.0494	3.3933	0.9778	0.9157	0.9573
Gauss/Backpr.	2.2177	3.4471	0.9770	0.9130	0.9556

Analyzing Table II it can be seen that the best configuration is when a Gaussian function is used for the membership functions associated to inputs, and for the training of the neural network a hybrid algorithm is used. In this case, the RMSE is the smallest, and the IA, R^2, and R have the biggest values. The smallest training and validation errors are around 0.02.

The variation of the testing error and a partial view of the testing vs. predicted data are illustrated in Fig. 6 and Fig. 7.

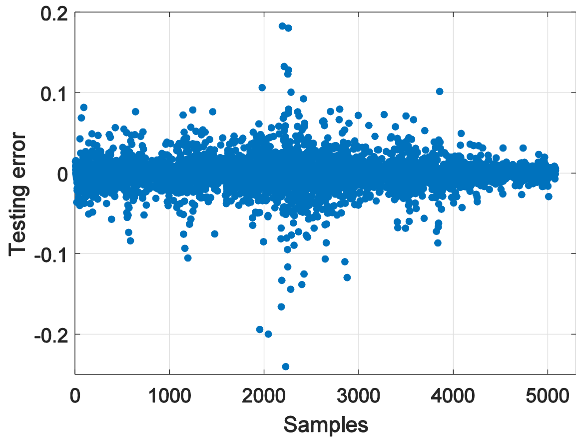

Fig. 6. Testing error for München data.

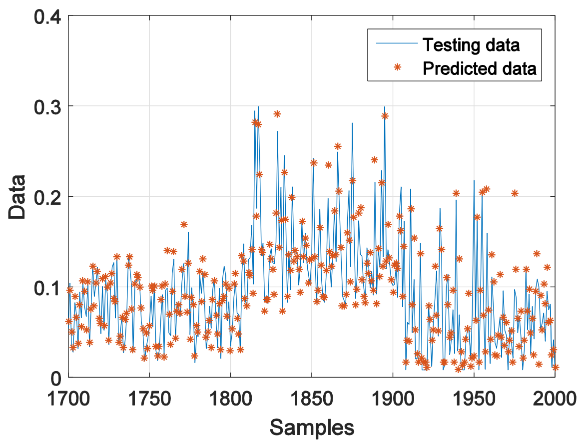

Fig. 7. Partial view of the comparison between testing and predicted data.

The best configuration of the ANFIS structure was also tested with data from the Ploiești station. The values of the statistical parameters are presented as follows: MAE [µg/m^3]: 1.0166; RMSE [µg/m^3]: 1.9160; IA: 0.9711; R^2: 0.8859; R: 0.9447.

C. Experimental results for ANN model

The structure of the neural network contains four neurons in the input layer, one hidden layer and one neuron in the output layer.

There were used two types of neural networks, namely *feed forward backpropagation (FF)* and *layer recurrent (LR)*, with *Levenberg-Marquardt* as training algorithm, and the adaptive learning functions were gradient descent with momentum weight and bias (*learngdm - LGDM*) and gradient descent weight and bias (*learngd - LGD*). The simulations were performed modifying also the number of neurons in the hidden layer (from three to six).

The training and validation errors are with two orders of magnitude smaller than the ones obtained in the case of ANFIS model, with values around 0.0004.

In Table III are presented the values of the statistical parameters for the ANN models.

TABLE III. STATISTICAL PARAMETERS FOR ANN MODELS (MÜNCHEN)

ANN structure		MAE [µg/m^3]	RMSE [µg/m^3]	IA	R^2	R
4x3x1/ LGDM	FF	1.9713	3.2430	0.9796	0.9230	0.9607
	LR	1.9613	3.2345	0.9798	0.9234	0.9609
4x3x1/ LGD	FF	1.9802	3.2378	0.9797	0.9232	0.9608
	LR	1.9605	3.2377	0.9798	0.9232	0.9608
4x4x1/ LGDM	FF	1.9421	3.2086	0.9802	0.9246	0.9616
	LR	1.9683	3.2359	0.9798	0.9233	0.9609
4x4x1/ LGD	FF	1.9421	3.2086	0.9802	0.9246	0.9616
	LR	**1.9278**	**3.1931**	**0.9804**	**0.9253**	**0.9619**
4x5x1/ LGDM	FF	1.9340	3.1966	0.9804	0.9252	0.9619
	LR	1.9836	3.2360	0.9797	0.9233	0.9609
4x5x1/ LGD	FF	1.9609	3.2152	0.9800	0.9243	0.9614
	LR	1.9486	3.2138	0.9801	0.9243	0.9614
4x6x1/ LGDM	FF	1.9415	3.2149	0.9801	0.9243	0.9614
	LR	1.9676	3.2207	0.9800	0.9240	0.9613
4x6x1/ LGD	FF	1.9471	3.2292	0.9799	0.9236	0.9611
	LR	1.9519	3.2182	0.9801	0.9241	0.9613

In Table III the best configuration of the ANN models is highlighted being associated to the layer recurrent structure with 4 neurons in the hidden layer and the *learngd* adaptation learning function. In this case the absolute mean error and the root mean squared error have the smallest values, and IA, R^2 and R indices have the biggest values.

Fig. 8 presents the ANN training performance, and Fig. 9 and 10 present the testing error evolution and a partial view of the comparison between testing and predicted data for München data set.

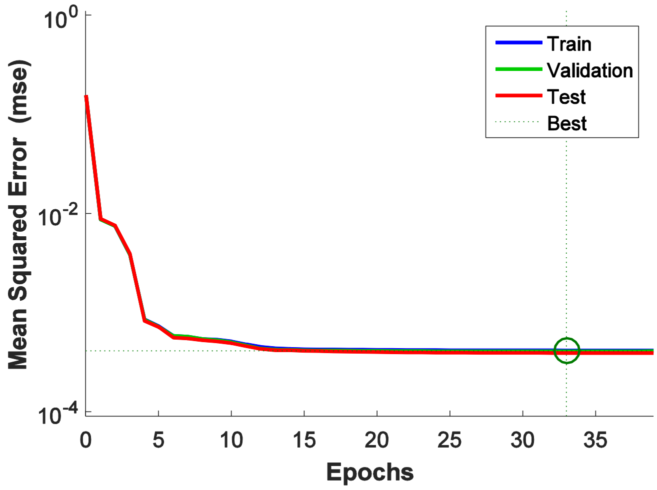

Fig. 8. ANN training performance.

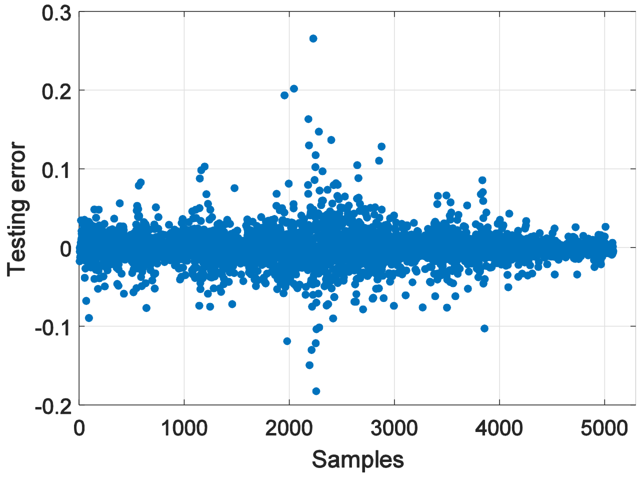

Fig. 9. Testing error.

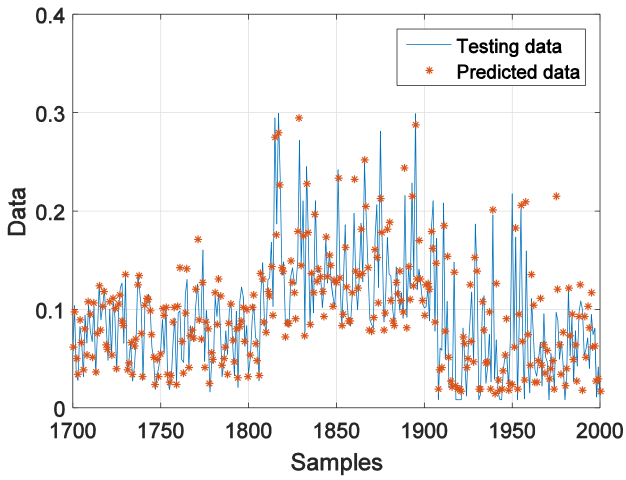

Fig. 10. Partial view of testing vs. predicted data.

In addition, as in the ANFIS case, the best configuration of the ANN structure was tested with data from the Ploiești station and the following values of the statistical parameters were obtained:

- MAE [μg/m^3]: 0.9672;
- RMSE [μg/m^3]: 1.3713;
- IA: 0.9852;
- R^2: 0.9416;
- R: 0.9714.

D. Discussion

Table IV presents a comparison between the best values of the statistical parameters obtained for the ANFIS model and the ANN model, respectively.

TABLE IV. COMPARISON BETWEEN BEST RESULTS

	MAE [μg/m^3]	RMSE [μg/m^3]	IA	R^2	R
Best ANFIS	1.9419	3.2089	0.9801	0.9246	0.9616
Best ANN	1.9278	3.1931	0.9804	0.9253	0.9619

From Table IV it can be observed that the results obtained with ANN are slightly better than the ANFIS results, and corroborated with the fact that for Ploiești data the statistical indices are much better in the case of ANN (see sections III-B and III-C) and the time for training is much smaller (with at least 10 times) for ANN, it can be concluded that the ANN forecasting model is most suitable for the prediction of $PM_{2.5}$ concentrations.

IV. CONCLUSIONS

This paper presented a comparative study between two computational intelligence techniques, ANN and ANFIS, applied to $PM_{2.5}$ short-term forecasting. The proposed neural forecasting model is simple, and thus, useful for real time systems, having a minimum number of past hours measurements that can capture the potential air pollution in urban regions at critical hours from weekdays. Moreover, as in real cases, for certain sites there are no meteorological recorded data, our study showed that neural models can provide accurate $PM_{2.5}$ forecasts, based solely, on $PM_{2.5}$ measurements time series.

ACKNOWLEDGMENT

The research leading to these results has received funding from EEA Financial Mechanism 2009-2014 under the project ROKIDAIR "*Towards a better protection of children against air pollution threats in the urban areas of Romania*" contract no. 20SEE/30.06.2014.

REFERENCES

[1] M.W. Gardner and S.R. Dorling, "Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences", Atmospheric Environment, 32, pp. 2627-2636, 1998.

[2] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators", Neural Networks, 2, pp. 356-366, 1989.

[3] D. Mandic and J. Chambers, Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability, Wiley, 2001.

[4] J. Hertz, A. Krogh, and R.G. Palmer, Introduction to the theory of neural computation, Addison Wesley, 1991.

[5] A. Kurt and A.B Oktay, "Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks", Expert Systems with Applications, 37, pp. 7986-7992, 2010.

[6] H.J.S. Fernando et al., "Forecasting PM10 in metropolitan areas: Efficacy of neural network", Environmental Pollution, 163, pp. 62-67, 2012.

[7] D. Voukantsis et al., "Intercomparison of air quality data using principal component analysis and forecasting of PM10 and PM2.5 concentrations using artificial neural networks, in Thessaloniki and Helsinki", Science of the Total Environment, 409, pp. 1266-1276, 2011.

[8] M.A. Elangasinghe, N. Singhal, K.N. Dirks, J.A. Salmond, and S. Samarasinghe, "Complex time series analysis of PM10 and PM2.5 for a coastal site using artificial neural network modeling and k-means clustering", Atmospheric Environment, 94, pp. 106-116, 2014.

[9] X. Feng et al., "Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation", Atmospheric Environment, 107, pp. 118-128, 2015.

[10] R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System", IEEE Transactions on Systems, Man and Cybernetics, 23 (3), pp. 665-685, 1993.

[11] M. Oprea, E.G. Dragomir, S.F. Mihalache, and M. Popescu, "Prediction methods and techniques for PM2.5 concentration in urban environment" (in Romanian), S. Iordache and D. Dunea (eds.), Methods to assess the effects of air pollution with particulate matter on children's health (in Romanian), MatrixRom, pp. 387-428, 2014.

[12] J.J. Carbajal-Hernández, L.P. Sánchez-Fernández, J.A. Carrasco-Ochoa, and J.F. Martínez-Trinidad, "Assessment and prediction of air quality using fuzzy logic and autoregressive models", Atmospheric Environment, 60, pp. 37-50, 2012.

[13] D. Domańska and M. Wojtylak, "Application of fuzzy time series models for forecasting pollution concentrations", Expert Systems with Applications, 39, pp. 7673-7679, 2012.

[14] A. Nebot and F. Mugica, "Small-Particle Pollution Modeling Using Fuzzy Approaches", M.S. Obaidat et al. (eds.), Simulation and Modeling Methodologies, Technologies and Applications, Advances in Intelligent Systems and Computing, Springer International Publ. pp. 239-252, 2014.

[15] M. Savić, I. Mihajlović, M. Arsić, and Z. Živković, "Adaptive-networkbased fuzzy inference system (ANFIS) model-based prediction of the surface ozone concentration", Journal of the Serbian Chemical Society, 79 (10), pp. 1323-1334, 2014.

[16] K. Polat and S.S. Durduran, "Usage of output-dependent data scaling in modeling and prediction of air pollution daily concentration values (PM10) in the city of Konya", Neural Computing and Applications, 21, pp. 2153-2162, 2012.

[17] S.F. Mihalache, M. Popescu, and M. Oprea, "Particulate Matter Prediction using ANFIS Modelling Techniques", Proc. of 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, Cheile Gradistei, Romania, pp. 895-900, 2015.