We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 11

Analyzing the Impact of Airborne Particulate Matter on Urban Contamination with the Help of Hybrid Neural Networks

Daniel Dunea and Stefania Iordache

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/63109

Abstract

In this study, particulate matter (PM), total suspended particulate (TSP), PM_{10}, and $PM_{2.5}$ fractions) concentrations were recorded in various cities from south of Romania to build the corresponding time series for various intervals. First, the time series of each pollutant were used as inputs in various configurations of feed-forward neural networks (FANN) to find the most suitable network architecture to the PM specificity. The outputs were evaluated using mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and Pearson correlation coefficient (*r*) between observed series and output series. Second, each time series was decomposed using Daubechies wavelets of third order into its corresponding components. Each decomposed component of a PM time series was used as input in the optimal feed-forward neural networks (FANN) architecture established in the first step. The output of each component was re-included to form the modeled series of the original pollutant time series.

The final step was the comparison of FANN outputs with wavelet-FANN results to retrieve the wavelet utilization outcomes. The last section of the study describes the ROkidAIR cyberinfrastructure that integrates a decision support system (DSS). The DSS system uses artificial intelligence techniques and hybrid algorithms for assessing children's exposure to the pollution with particulate matter, in order to elaborate PM forecasted values and early warnings.

Keywords: air pollution, wavelet transformation, batch-learning algorithm, respiratory health, cyberinfrastructure

INTECH
open science | open minds

© 2016 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

1. Air pollution with particulate matter in urban areas

Quantifying the human exposure to air pollutants is a challenging task because air pollution is characterized by high spatial and temporal variability. The atmospheric physicochemical parameters of interest from the point of view of air pollution in urban areas are carbon monoxide (CO), sulfur dioxide (SO_2), nitric oxide (NO), nitrogen dioxide (NO_2), various fractions of particulate matter (PM_{10}, $PM_{2.5}$, PM_1, and UFPs or ultrafine particles), ozone (O_3), volatile organic compounds (VOCs), and polycyclic aromatic hydrocarbons (PAHs). The levels of these parameters are significantly influenced by meteorological factors (such as speed and direction of wind, precipitations, temperature, relative humidity, and solar radiation), seasonal and diurnal fluctuations, geographical factors (e.g., local topography, buildings), emission sources i.e., industrial activities and traffic in the area, as well as the air mass trajectories (e.g., long-range transport of pollutants).

Class	Description	Size (in diameter)
TSP	Airborne particles or aerosols that constantly enter the atmosphere from many sources having below 100 μm are collectively referred to as total suspended particles (TSP). TSP is assessed with high-volume samplers.	Below 100 microns (<100 μm)
Large particulates	Particles are retained by the nasopharynx area.	Over 10 microns (>10 μm)
PM_{10}	Particulates that can be inhaled below the nasopharynx area (nose and mouth) and are thus called inhalable particulates (coarse fraction).	Below 10 microns (0–10 μm)
$PM_{2.5}$	Fine particulates travel down below the tracheobronchial region, that is, into the lungs (fine fraction).	below 2.5 microns (0–2.5 μm)
UFP	Ultrafine particulates can penetrate into the deepest parts of lungs and can be dissolved into blood (ultrafine fraction).	below 0.1 microns (0–0.1 μm)

The most hazardous size classes to humans are $PM_{2.5}$ and UFP as they penetrate into the lungs and can even be dissolved into the blood.

Table 1. Airborne particulate matter classification depending on particle size [8].

In many urban agglomerations around Europe, the concentrations of airborne particles, NO_x, and O_3 exceed at least occasionally the limit or target values. Therefore, air pollution control focuses mostly on the surveillance of the above-mentioned pollutants [1]. Urban agglomerations are areas of increased emissions of anthropogenic pollutants into the atmosphere having adverse health effects on population.

Consequently, a major issue of environmental policy at regional level is the reduction of their concentrations in the ambient air [2].

Particle sizes range from a few nanometers up to more than 100 μm, and depending on particle size, there are several classes of particles (**Table 1**). However, epidemiological studies have

http://dx.doi.org/10.5772/63109

shown that the most hazardous size classes to human health are $PM_{2.5}$ and UFP, as they penetrate into the lungs and can even enter into the blood following the gas exchange. Diseases caused by UFP exposure primarily relates to lung cancer and heart disorders. Since the measurement of UFP is a difficult task requiring sophisticated equipment, one can monitor the submicrometric fraction that includes UFP using a reliable optical system, for example, Dusttrak DRX 8533 [3].

In the recent years, the most common size fraction that is usually monitored in the national air quality infrastructures at large scales in urban areas is $PM_{2.5}$. Recent long-term studies show the associations between PM and mortality at levels significantly below the current annual WHO air quality guideline level for $PM_{2.5}$, that is, 10 $\mu g/m^3$ (WHO, 2013).

The issue of studying the fine particulate matter is very complex and has many unknown variables mainly due to the multitude of sources from which it directly originate, as well as due to the physicochemical transformations that occur in the atmosphere, resulting in the formation of secondary $PM_{2.5}$ particulates [4–6]. Other major setbacks are the difficulties of compliance assessment and the setup of measurement methods equivalence. Furthermore, the methods of $PM_{2.5}$ measurement are still in the development period and the reference method was recently revised in EN 12341: 2014 standard [7].

2. Forecasting of particulate matter using neural networks

The analysis of environmental processes involves highly complex phenomena, random variations of parameters, and difficulty to perform accurate measurements in certain situations. In these conditions, the available data are incomplete, imprecise, and current applied models require further improvements.

Measuring and forecasting of atmospheric conditions is important for understanding the processes of formation, transformation, dispersion, transport, and removal of the pollutants. Reliable overall estimates regarding the identification of sources, effects on mixing, transformation, and transportation support the control of air quality and the implementation of preventive actions to reduce the anthropogenic emissions [8].

The performance of environmental management can be improved using forecasting tools of the potential pollution episodes that can affect the population from inner and surrounding areas where the episode might occur. Prediction of the evolution of an atmospheric parameter can be done for short term (1 h, 1 day, 1 month) or long term (1 or more years).

The interest in improving the forecasting performances of time series algorithms and models in air pollution studies has considerably grown. The applied methods may vary from statistical methods, artificial intelligence (AI) techniques, and probabilistic approaches to hybrid algorithms and complex models. The final purpose is to supplement monitored data and/or to complete the missing values in the time series of air pollutants.

The field of statistics, which deals with the analysis of time dependent data, is called time series analysis (TSA). One of the most widespread types of processing is the *time series forecasting*.

Many of these techniques are used in practice. We can mention, for example, random walks, moving averages, trend models, seasonal exponential smoothing, autoregressive integrated moving average (ARIMA) parametric models, Boltzmann composite lattice, etc.

Some of the traditional statistical models such as the moving average, exponential smoothing, and ARIMA model are linear techniques, which have been in the past the main research and application tools in air pollution research. Predictions of future values are constrained to be linear functions of past observations, under the assumption that the data series is stationary [9]. The general model ARIMA introduced by Box and Jenkins [10] involves the autoregressive and moving averages parameters, and explicitly includes differentiations in the formulation of the model. Three types of parameters are required in the model as follows: autoregressive parameter; differentiation passes, and moving averages parameters [10]. The ARIMA model assumes that a parametric model relating the most recent data value to previous data values and previous noise gives the best forecast for future data. However, one weakness of the ARIMA model resides in the assumption that the examined time series is stationary and linear, and therefore has no structural changes [9].

Air pollutants have a random evolution, which requires non-deterministic approaches. Advantages of neural computing techniques over conventional statistical approaches rely on faster computation, learning ability, and noise rejection [11]. Artificial neural networks (ANN), for example, succeeded to give good results for time series processing when the data present noise and nonlinear components. Their capacity of learning and generalization recommend them as valuable tools in a wide area of applications. The most popular architecture used in practice is the multilayer feed-forward neural network. Their processing units (neurons) are organized in layers and there exist only forward connections (i.e., their orientation is from the input layer toward the output). This type of networks started to be extensively used in the late 1980s when the standard back-propagation algorithm was introduced. Since that time, the multilayer feed-forward ANNs had a large applicability in various domains, that is, financial, health, meteorology, environmental protection, etc.

The research has been oriented to find faster algorithms for training the network and to provide algorithms to automate the design of an optimal network topology for a specific problem. We can mention the standard back-propagation with momentum or with variable learning rate, the adaptive Rprop, or algorithms based on the standard numerical optimization techniques (Fletcher-Powel, conjugate gradient, quasi-Newton algorithm, Levenberg-Marquardt, etc.).

Rprop algorithm introduced by Riedmiller and Braun [12] is a supervised batch learning which accelerates the training process in the flat regions of the error function and when the iterations get nearby a local minimum. This algorithm allows different learning rates for each weight. These rates are changed adaptively with the change of sign in the corresponding partial derivative of the error function. They change progressively but without getting out of an initially prescribed interval. The algorithm is described by four parameters denoted by η^+, η^-, Δ_{max} and Δ_{min}. The first two parameters give the increasing and decreasing factor for adjusting the update size and they are chosen such that $0 < \eta^- < \eta^+ < 1$. The size step of the update is

http://dx.doi.org/10.5772/63109

bounded by Δ_{min} and Δ_{max} . The following values of the parameters were used in our tests: $\eta^{+} = 1.25, \eta^{-} = 0.5, \Delta_{max} = 50, \Delta_{min} = 0$ [13].

Quickprop is a batch training algorithm introduced by Fahlman [14], which takes in consideration the information about the second-order derivative of the performance error function. Literature showed that Quickprop is a particular case of the multivariate generalization of the secant method for nonlinear equation [15]. The local minimum of the batch error function reached a critical point that is a zero of the gradient [13]. In practice, Newton's iteration is replaced by a quasi-Newton iteration, which uses an approximate of the Jacobian and saves the involved amount of computation. The approximation of the Jacobian by a diagonal matrix with its entries computed with finite difference formulas proves that Quickprop belongs to this category of quasi-Newton iterations. Its convergence is not anymore quadratic, but it remains linear in the vicinity of the solution. We have used the same value (equal to 1.75) for the maximum growth factor denoted by μ in [14], in all our tests with Quickprop algorithm.

3. Experimental setup

We used the resources of an AI forecasting system called RNA-AER [13] for the domain of air pollution forecasts in urban regions. RNA-AER stands for the Romanian abbreviation of ANN for air pollution. This is a part of a complex system for $PM_{2.5}$ forecasting based on various techniques of artificial intelligence (multi-agents, knowledge base system, ANNs, and neuro-fuzzy) and that is designed to analyze the pollution level of air within ROkidAIR system (http://www.rokidair.ro/en) [16]. A feed-forward neural network with a single hidden layer was used to perform the tests presented in this work (**Figure 1**).

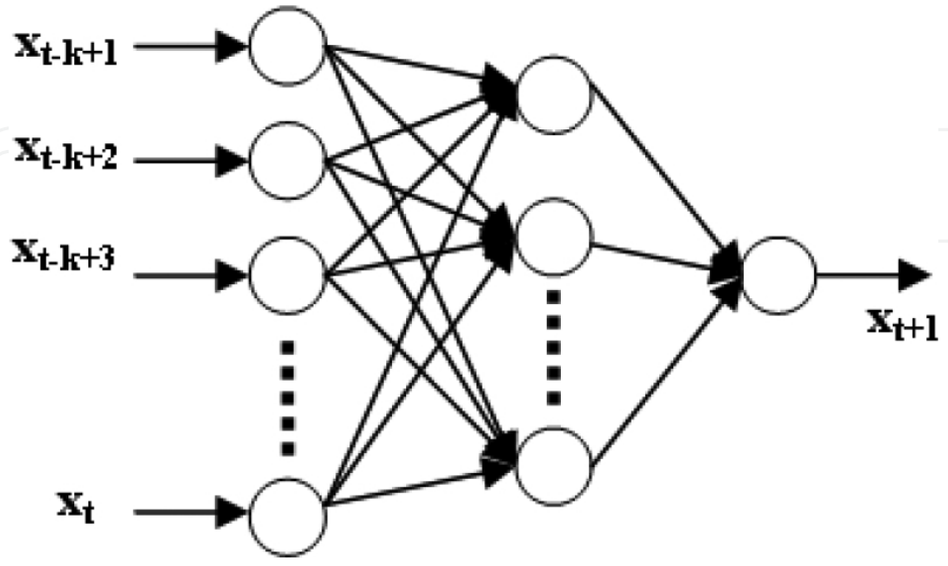


Figure 1. Example of feed-forward artificial neural network with one hidden layer.

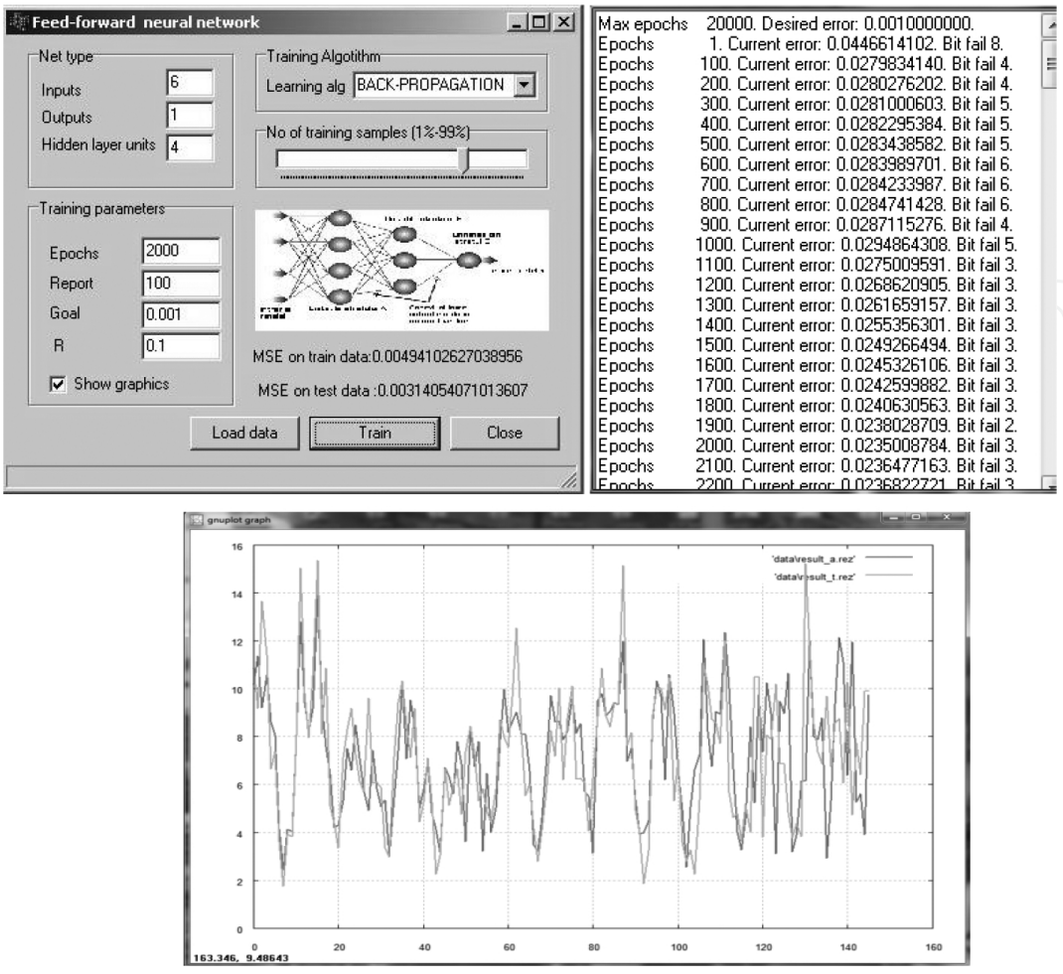


Figure 2. Front panel of the RNA-AER software with the feed-forward ANN configuration settings, error analysis, and observed and simulated time series.

The activation function used for the hidden and output layers was the symmetric sigmoid function (*tanh*). Since this function transforms the real axis into the range of (-1,1), the data were normalized before their use and transformed back in their real values after simulation.

In the training stage, we used various learning algorithms, but the most satisfactory results were obtained with Rprop and Quickprop. The application offers a friendly user interface from which one may choose various parameters that describe the network and the training algorithm. The program takes the raw data from one column text file and applies the necessary transformations in the preprocessing stage. After training, the application tests the network on the validation set of samples and shows the error. Then, the user is able to see the graphics for the evolution of the error in the training process and the observed and forecasted data (**Figure 2**). Best results were obtained with 4 or 6 units in the input layer, 6 neurons in the hidden layer and 1 output neuron. The output represents the one value ahead forecasted data. The network training comprised four learning algorithms. The first two were given by the batch and incremental implementations of the standard back-propagation learning [17]. These standard algorithms were tested with different values of the learning rate and momentum.

http://dx.doi.org/10.5772/63109

The other two algorithms were the resilient back-propagation Rprop and Quickprop. In this study, we present only the Rprop and Quickprop algorithms, which provided better results.

3.1. Steps for the development of a feed-forward neuronal model (FANN)

The use of raw data may rarely give satisfactory outputs. In this case, the training of the ANN will catch only general properties of the data series without being able to identify characteristics that are more refined. Therefore, a preprocessing step is often required in which the initial data are transformed such that the new data series eliminates some redundant characteristics from the analysis (e.g., interpolation, smoothing, wavelet decomposition, etc.).

The resulted series is then used to extract the required samples for training the network. Since our goal was to obtain one value ahead forecasting, each sample had the form $\left(x_{t-k+1}, x_{t-k+2}, \ldots, x_t, x_{t+1}\right)$, and the whole set of samples was obtained by moving window technique. Here, x_{t+1} represents the forecasting data while the other numbers are the corresponding inputs. Three-fourth of this set was used for training, while the rest was used in the validation process.

Inputs: particulate matter measurements of various PM fractions made in a certain default time window; outputs: one-step-ahead forecast of the PM pollutant.

Step 1. *Data preprocessing*. This stage involves the data processing, elimination of incomplete records, data interpolation to complete the missing values in the time series, data normalization validated by experts, and their redirection so that the database is compatible with the software used for forecasting.

Step 2. *Establishing the method of avoiding the overtraining of the ANN*. A common method is to divide the database into three sets of data: one for training (e.g., 75%), one for validation (e.g., 15%) and another one for testing (e.g., 15%). In some cases, the proportions that include the data in one of the datasets differ slightly around the value of 70–80% for the training set, 18–28% for the validation, and about 2% for the testing set. Alternatively, the cross-validation with 10 sets—9 sets used for training and the 10th for validation might be considered. This process is repeated until each of the 10 sets is used for validation.

Step 3. *Setting the ANN architecture*. This involves the establishing of the number of nodes in the input layer (optimal window time for the next value forecast of the pollutant), the number of nodes in the hidden layer, activation functions, etc.

Step 4. *Adjustment of training parameters*. The optimal number of epochs for network training, the learning rate, and momentum parameter are established experimentally, avoiding the overtraining of the network or an undertrained situation.

Step 5. *Network training* taking into account the parameters established in step 4 and step 5.

Step 6. *Validation of the resulted network architecture*.

Step 7. *Testing of the ANN*.

Step 8. *Analysis of the ANN performances*. At this stage, statistical parameters can be used such as the correlation coefficient between variables, mean absolute error (MAE), root mean square error (RMSE), the training error (MSE), and mean absolute percentage error (MAPE). The values of these parameters can be compared with the conventional limits established in the literature and those obtained with other models developed for forecasting the amount of particulate matter fractions.

The following tests present how different parameters, which describe the neural network model, affect the accuracy of the forecasted data. The topology of the neural network is denoted as $n_1 - n_2 - n_3$, where n_1 is the number of nodes in the input layer, n_2 is the number of nodes in the hidden layer, and n_3 is the number of nodes in the output layer. Since the training is sensitive to the initial values of the weights, 10 tests for each algorithm were performed and the mean of the resulted values was considered for all tables provided.

4. Results and discussion

4.1. Analysis of total suspended particulates time series

In the first test, we present the monthly average concentrations values of total suspended particulate (TSP) recorded between 1995 and 2006 in Targoviste, Romania. During that period, TSP often exceeded the limit value (75 μg/m^3) and the city was considered as a PM risk area at national level due to emissions from metallurgical industries. Later on, Romanian technical norms replaced the earlier TSP air quality standard with a PM_{10} standard.

We compared various (p, d, q) setups of ARIMA model [10] to identify the statistical model with the smallest magnitude of the errors during the estimation period. ARIMA (4,0,3) presented the smallest MAE and MAPE. A significant relationship ($p < 0.001$) with a correlation coefficient of 0.8 was noticed between the ARIMA (4,0,3) forecasted variables and observed data [9].

The tests performed with the feed-forward neural network using the TSP observed series provided good forecasting results with the Quickprop (4,6,1) algorithm. The correlation coefficient of ANN Quickprop (4,6,1) indicated a strong relationship between the forecasted variables and observed data (**Table 2**).

Indicator	ARIMA statistical model (4,0,3)	ANN model (4,6,1) Quickprop	ANN model (4,6,1) Incremental	ANN model (6,6,1) Rprop
r	0.801	0.946	0.779	0.652

Table 2. Correlation coefficients of forecasted/observed series of the ARIMA model and ANN algorithms using the time series of total suspended particulates (TSP) concentrations in Targoviste city.

http://dx.doi.org/10.5772/63109

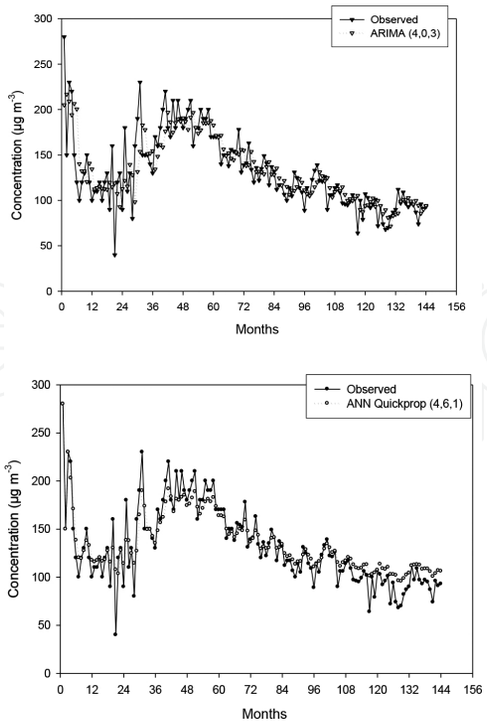


Figure 3. Comparison of monthly averages of total suspended particulates observations vs. the ARIMA (4,0,3) and Quickprop (4,6,1) simulations in Targoviste city (1995–2006) [9].

The ANN Quickprop (4,6,1) model presented a higher correlation coefficient (r = 0.94) than ARIMA (4,0,3) model. The neural network prediction algorithm provided a better fit to the TSP measured time series (**Figure 3**). Consequently, we observed that the use of a proper configuration of ANN could provide better results for TSP prediction than linear statistical models [9].

4.2. Analysis of PM_{10} time series

In the next test, we used daily time series of PM_{10} recorded by an optical analyzer in Targoviste city. We present a case with a time series of 101 values to test the influence of a short time series on the efficiency of the training.

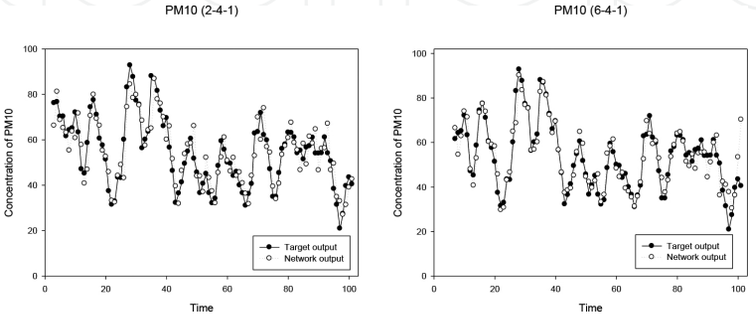


Figure 4. Observed and forecasted concentrations of PM_{10} (2-4-1) and PM_{10} (6-4-1) ANN configurations [13].

ANN	Rprop		Quickprop	
Network configuration	**MSE**	**MSE**	**MSE**	**MSE**
	Training data	**Validation data**	**Training data**	**Validation data**
2-4-1	0.00993	0.01095	0.01153	0.00866
4-4-1	0.00498	0.01659	0.00957	0.00950
6-4-1	0.00412	0.02069	0.00881	0.00949
8-4-1	0.00314	0.02285	0.00738	0.01368

Table 3. Dependence of training and validation errors with various topologies of feed-forward artificial neural network using short PM_{10} time series.

Figure 4 presents the graphics for the utilization of 2 and 6 neurons in the input layer.

We observed that the number of network inputs has a major influence over the forecasting performances. **Table 3** shows how the training error depends on the number of network inputs. For each case, we used the same number of values, that is, 80. Increasing the number of network inputs results in the decrease in the number of testing samples. Yet, the table shows an increase in the MSE of the validation data. This suggests that increasing the number of input neurons will improve the capability of the network to have a better response for the data close to ones used in the training process. On the other hand, the network loses its generalization abilities.

Table 4 shows how the network training and testing depend on the number of training samples. The selected network topology was 2-4-1.

The error of training data decreases with the increase in the number of samples, while the error of validation data increases for both tested algorithms.

ANN	Rprop		Quickprop	
No of training samples	**MSE**	**MSE**	**MSE**	**MSE**
	Training data	**Validation data**	**Training data**	**Validation data**
70	0.01060	0.00968	0.01242	0.00748
80	0.00991	0.01093	0.01151	0.00864
90	0.00962	0.01306	0.01075	0.01270

Table 4. Dependence of training and validation errors with the number of samples used in training a feed-forward artificial neural network (2-4-1).

http://dx.doi.org/10.5772/63109

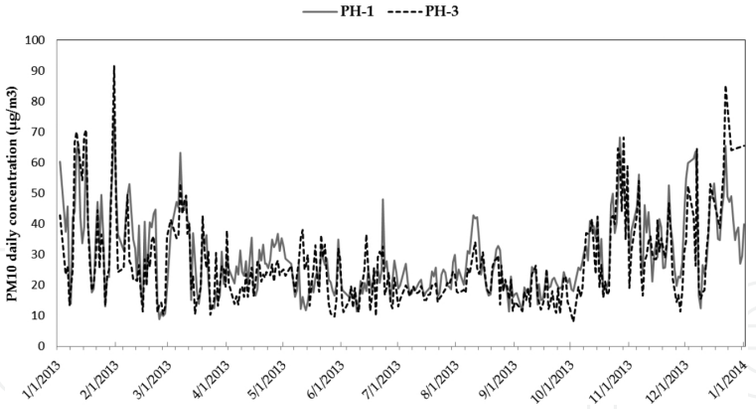


Figure 5. Plots of observed PM_{10} time series with daily averages from two automated stations i.e. PH-1 and PH-3 located in Ploieşti city in 2013.

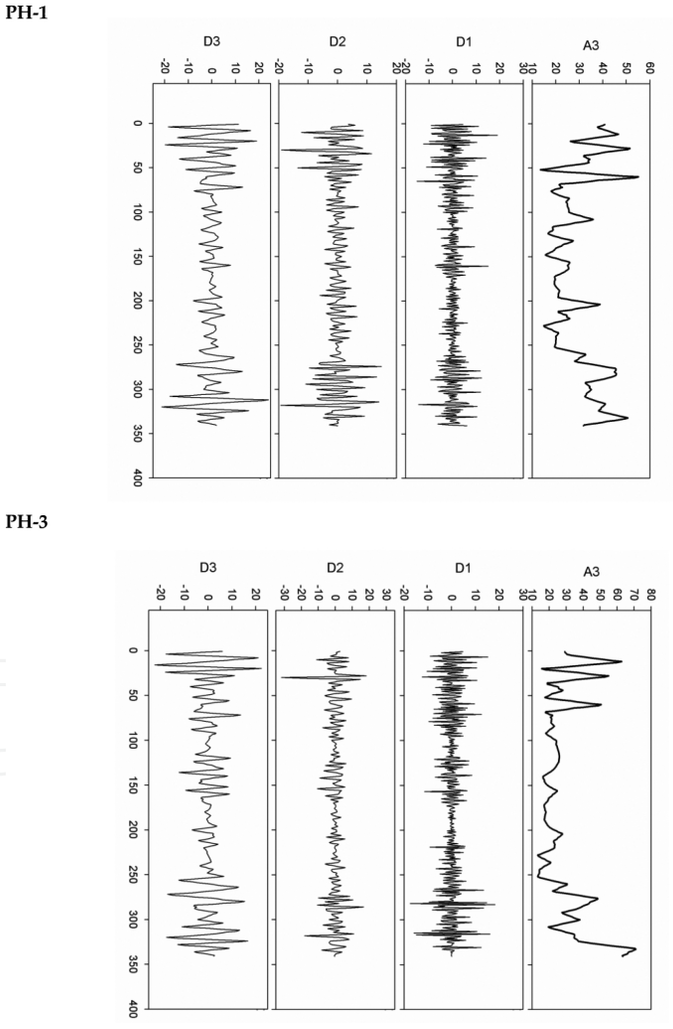


Figure 6. Decompositions of PM_{10} time series recorded at PH-1 and PH-3 automated stations in four components i.e. A3, D1, D2, and D3 using Daubechies wavelets of third order.

4.3. Analysis of PM_{10} time series

In this test, the daily averaged PM_{10} time series recorded at two automated stations located in Ploieşti city in 2013, that is, PH-1 and PH-3 were analyzed using the method of wavelet processing described in [11]. Data gaps (missing values 4 at PH-1 and 15 at PH-3) were interpolated based on existing measured values (**Figure 5**). Each air pollutant series (n = 365 values) was decomposed using the MATLAB Wavelet Toolbox in four components, that is, A3, D1, D2, and D3 using Daubechies wavelets of third order (**Figure 6**).

Automated station for monitoring air quality	**PH-1 (F)**	**PH-1 (WF)**	**PH-3 (F)**	**PH-3 (WF)**
Training data MSE	0.0098	A3: 0.00099	0.0067	A3: 0.00096
		D1: 0.00355		D1: 0.00289
		D2: 0.00099		D2: 0.00104
		D3: 0.00099		D3: 0.00099
Validation data MSE	0.0312	A3: 0.00053	0.0498	A3: 0.00263
		D1: 0.00677		D1: 0.01225
		D2: 0.00335		D2: 0.00238
		D3: 0.00274		D3: 0.00214
RMSE	7.7	3.4	9.9	4.4
MAE	5.5	2.5	6.8	3.2
Pearson coefficient (r)	0.78	0.96	0.75	0.95
Forecasted value ($\mu g\ m^{-3}$)	33.2	36.7	56.8	61.5
Observed value ($\mu g\ m^{-3}$)	39.9		65.6	
Studentized residuals >3.0	6	4	10	5

Table 5. Averages of 10 validation tests resulted from the Rprop (6-4-1) application to PM_{10} time series recorded in Ploiesti vs. Daubechies db3 wavelet—Rprop (6-4-1) results after recomposing the series; F—Rprop FANN, WF—Daubechies db3 wavelet—Rprop FANN.

The components resulted from decomposition of time series (A3, D1, D2, and D3) was used as input in an optimal FANN architecture established prior to this analysis, that is, Rprop (6-4-1). The simulated FANN output of each component was recomposed to form the modeled series of the original pollutant time series and the network performance was analyzed using MSE, MAE, RMSE, and r. The comparison of outputs when FANN is solely used with wavelet-FANN results allowed the evaluation of wavelet contribution to the improvement of forecasting abilities [11].

The application of Daubechies db3 wavelet as a decomposing preprocessor of daily averages time series has significantly improved the out-of-sample forecasted values (**Table 5**). The results showed that the exclusive use of Rprop (6-4-1) configuration was less fitted to the observed data at both stations. Wavelet preprocessing followed by the individual training of resulted components has substantially increased the r coefficient from 0.7 to 0.9 and decreased

http://dx.doi.org/10.5772/63109

the error indicators for both time series as compared to the exclusive use of FANN. Furthermore, the forecasted values were closer to the corresponding real observations.

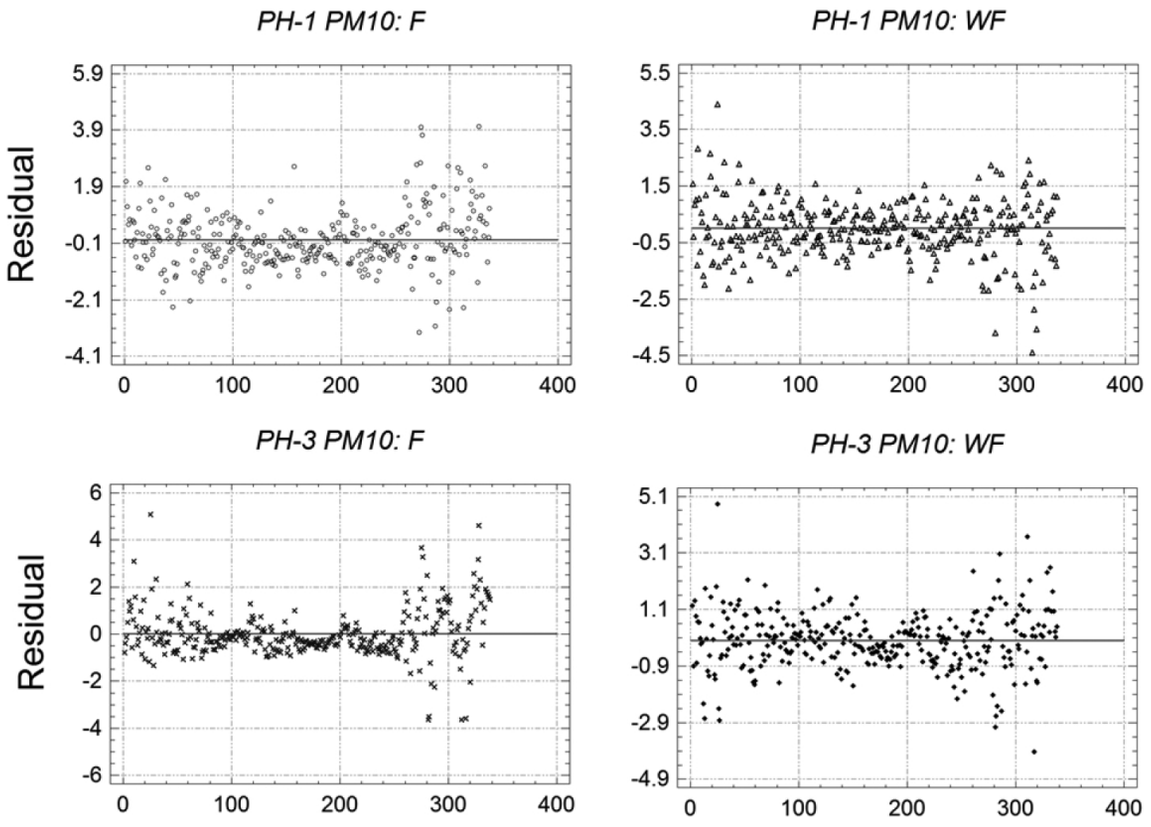


Figure 7. Plots of residuals resulted after correlating the daily averages of PM_{10} (μg m^{-3}) and Rprop FANN (6-4-1) modeled data, and Daubechies db3 wavelet—Rprop WFANN (6-4-1) data, respectively, recorded at two automated monitoring stations in Ploiesti.

A reduction of Studentized residuals number greater than 3.0 was observed using the wavelet processing of data from both stations compared to FANN (**Figure 7**), that is, from 6 Studentized residuals to 4 (PH-1), and from 10 to 5 (PH-3).

These aspects suggested that wavelet integration in processing of daily averages of PM_{10} series provided significant improvements of the forecasting ability recommending the use of the hybrid model. Compared to these results, the application of the hybrid model to hourly recorded PM_{10} time series at other Romanian stations showed also the improvements of correlation coefficient. However, the wavelet processing increased errors and provided more potential outliers [11]. Wavelet integration did not provide computational benefits taking into account the increase in time required for data processing. On the other hand, application of Rprop FANN to hourly recorded PM_{10} produced overfitting. For improved results when neural network is solely used, overfitting is required to be adjusted by using additional techniques, for example, early stopping [18], dropout [19], etc.

We observed in our study that wavelet integration diminished the overfitting tendencies.

4.4. Analysis of $PM_{2.5}$ time series

The section presents the results of Daubechies db3 wavelet—Rprop neural network (6-4-1) modeling using $PM_{2.5}$ time series of 24-h daily averaged concentrations recorded in Râmnicu Vâlcea city, south-west of Romania at VL-1 monitoring station. We selected this station for tests because VL-1 station was one of the two stations that recorded a substantial exceeding of the annual limit value (25 μg/m^3) at national level in 2012. The maximum value reached 149.13 μg/m^3 and the annual geometric mean was 23.8 μg/m^3.

Automated monitoring station in Ramnicu Valcea city (VL-1)	2012 (F)	2012 (WF)
RMSE	6.3	26.1
MAE	3.8	16.3
MAPE	31.6	50.4
Pearson coefficient (*r*)	0.86	0.93
Studentized residuals > 3.0	7	7

Table 6. Averages of 10 validation tests resulted from the Rprop (6-4-1) application to $PM_{2.5}$ time series recorded at VL-1 station vs. Daubechies db3 wavelet—Rprop (6-4-1) results after recomposing the series; F—Rprop FANN, WF—Daubechies db3 wavelet—Rprop FANN.

A significant increasing of the *r* coefficients was observed after the application of wavelet preprocessing. RMSE, MAE, and MAPE showed higher values compared to the exclusive use of Rprop configuration (**Table 6**). Both models overestimated the forecasted values in the last quarter of time series. However, the fluctuations observed in the original time series were simulated better by using Daubechies wavelets [11].

These results suggest that other models or algorithms with noise-filtering/smoothing properties may be applied in various stages of the simulation in conjunction with the Daubechies db3 wavelet—Rprop FANN utilization. The expected outcome would be a superior refining of the initial $PM_{2.5}$ forecasted values [11].

5. A cyberinfrastructure for the protection of children's respiratory health by integrating hybrid neural networks for PM forecasting—ROkidAIR

ROkidAIR cyberinfrastructure is currently developed in a European Economic Area (eeagrants.org) research project to facilitate the protection of children's respiratory health in two Romanian cities, that is, Targoviste and Ploiesti.

Recent developments in the management of urban atmospheric environment demonstrated the imperative need to ensure quick, efficient, and easy-to-understand information regarding the status of air quality. The negative impact of air pollution on human health requires improvements of contemporary systems for air quality management to reduce the human

http://dx.doi.org/10.5772/63109

exposure to various pollutants. Providing full and comprehensive information concerning the air quality is regarded as a mandatory service for citizens in the current air quality management systems. The authorities should establish an appropriate framework, especially in urban areas, where adverse health effects caused by poor air quality are more pronounced, to ensure the integration of relevant data regarding the maintaining of air quality at required standards. The systems for air quality management need to be adapted to decision makers' requirements (in order to reduce the ambient air quality issues through adequate policies) and citizens (for early warning and for providing useful recommendations). Aiming to reduce their exposure, citizens should receive adequate information on the spatiotemporal variation of air quality or the forecasts on short, medium, and long term. To achieve this goal, it is necessary to collect, integrate, and analyze data from multiple sources. Air quality forecast is one of the essential elements of modern air quality management in urban areas. However, the efficiency of the used forecasting methods is limited by the complex relationships between air quality, meteorological parameters, and specific characteristics of each study area. In addition, an important issue that needs to be considered in choosing the forecasting method is the variation of the input data quality. The methods to be used should be less sensitive to this factor [20]. Information related to air quality in urban areas is obtained by using specific methods and tools for processing the time series recorded by the monitoring stations. Mathematical methods and tools can provide air quality forecasting, so that decision makers can act with preventive measures that would "mitigate" or change the results of a foreseen critical pollution episode. There is an increasing demand regarding the development of cyber-platforms that may facilitate the air quality management providing real health benefits to the end user (e.g., ROkidAIR, *http://www.rokidair.ro*).

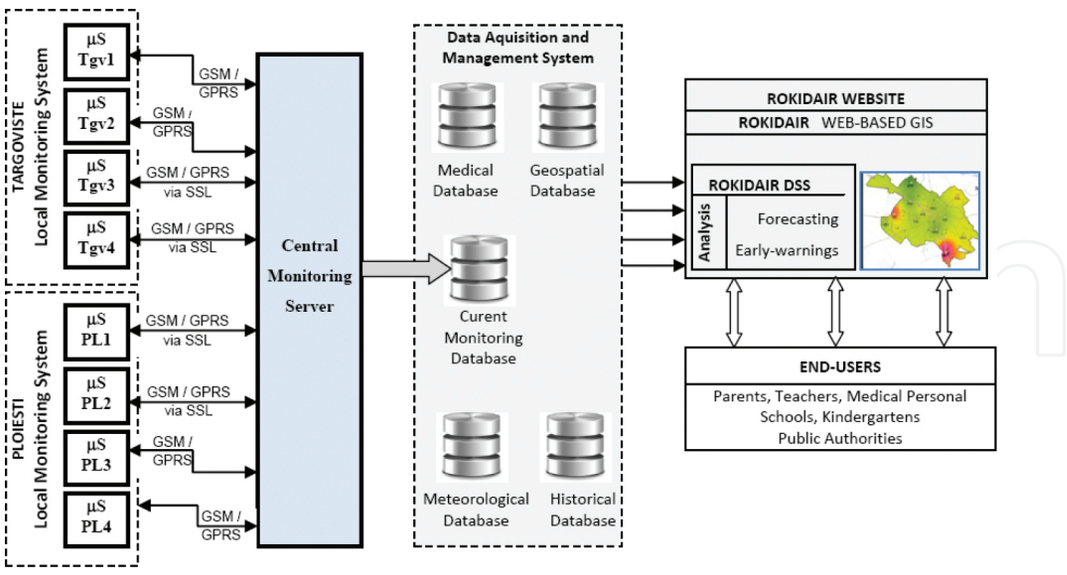


Figure 8. The architecture of the ROkidAIR system.

The main goal of the ROkidAIR project is to develop and deploy a monitoring network system and an adjacent early warning structure that provide synthesized data concerning the $PM_{2.5}$

levels obtained from simplified but reliable monitoring micro-stations and AI forecasting algorithms developed within the project . The architecture of the ROkidAIR system is presented in **Figure 8**. The ROkidAIR cyberinfrastructure is a pilot system, which is focused on fine particulate matter effects on children's health in two towns of Romania, that is, Targoviste and Ploiesti. It provides early warnings concerning the PM levels, tailored to the end-user requirements via several communication channels [3]. Collected time series obtained from the self-developed monitoring network system, based on PM micro-stations, are preprocessed and adapted to feed the forecasting module based on AI algorithms. All data are presented in a dedicated geo-portal adapted to be used by smartphones and other portable equipment. The main stream of information is transmitted both to the responsible authorities and to the sensitive persons, who are registered in the system. The expert advises and recommendations are transmitted via e-mails and SMSs to the registered users providing support for children's health management under the impact of air quality stressors and pressures. Early warnings are developed in cooperation with pediatric specialists, which synthesize the most relevant information concerning the protection of children's health against air pollution threats. The early warning data packages are also transmitted to the authorities (e.g., local EPAs—Environmental Protection Agencies and DPH—Public Health Protection Directions) for informational purposes. The monitoring network comprises eight PM micro-stations (four in each city), which are developed during the implementation of ROkidAIR project. These micro-stations provide continuous PM monitoring data that are processed to be used as inputs in forecasting algorithms based on AI. The raw data obtained from the eight micro-stations are also used in other modules of the cyber-platform: the ROkidAIR web-based geographic information systems (GIS) geoportal, and the decision support system (DSS) including the early warning module. The DSS system uses artificial intelligence techniques (ANNs and predictive data mining) and hybrid algorithms and models (Neuro-fuzzy ANFIS, and wavelet neural network, WNN) for assessing children's exposure to the pollution with particulate matter, in order to elaborate forecasted values and early warnings [16].

In ROkidAIR AI model, forecasting knowledge is extracted by using ANFIS (generating the fuzzy rules set), and other methods (e.g., a combination between some machine learning techniques) on the specific datasets (continuous monitoring data, historical data, meteorological data, and medical data). All the extracted forecasting rules and knowledge are included in a forecasting knowledge base that provide expert knowledge (heuristics) for a faster and optimal air pollution forecasting in a critical polluted area [21].

6. Conclusions

The contribution of artificial intelligence to the air quality monitoring systems under development relates to evolutionary computing, which provides stochastic search facilities that can efficiently assess complex spaces described by mathematical, statistical, neural network, or fuzzy inference models applied to assess the population exposure to air pollution in urban environments. Machine-learning techniques are currently contributing to the *online* air quality

monitoring and forecasting. Statistical and neural modeling techniques can also provide approximations to supplement results from computationally expensive analytic methods.

Significant results for PM data forecasting were obtained with Rprop (PM_{10}10 and $PM_{2.5}$), and Quickprop (TSP) algorithms. The exclusive use of the ANN algorithms showed difficulties in predicting pollutant peaks and limitations due to limited continuous observations and large local-scale variations of concentrations. WNNs is an alternative to overcome these drawbacks related to time series predictions by integrating a proper wavelet in the hidden nodes of WNNs or as a preprocessing step. The results of numerical tests provided that the application of wavelet transformation is a significant factor for improving the accuracy of forecasting. Further investigations are required using hourly, daily, and monthly air-quality data from other locations and regional level, by assessing and verifying the reliability, relevance, and adequacy of ANN data forecasting. An important step for reliable air quality forecasting is the optimal selection of ANN learning algorithm. The automation of this component is required to optimize the informational fluxes and to facilitate the decision-making process.

Acknowledgements

This study received funding from the European Economic Area 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

Author details

Daniel Dunea* and Stefania Iordache

*Address all correspondence to: dan.dunea@valahia.ro

Valahia University of Târgoviste, Aleea Sinaia, Târgoviste, Romania

References

[1] Ianache C., Dumitru D., Predescu L., Predescu M. (2015), Relationship between airborne particulate matter and weather conditions during cold months, In Proceedings of 15th International Multidisciplinary Scientific Geoconference SGEM 2015, June 18–24, 2015, Albena, Bulgaria, 1017–1024, DOI: 10.5593/SGEM2015/B41/S19.131

[2] Langner M., Draheim T., Endlicher W. (2011), Particulate matter in the urban atmosphere: concentration, distribution, reduction—results of studies in the Berlin metro-

politan area, In W. Endlicher et al. (eds.), Perspectives in Urban Ecology, Berlin, Heidelberg, Springer-Verlag, 15–41. DOI: 10.1007/978-3-642-17731-6_2

[3] Iordache St., Dunea D., Lungu E., Predescu L., Dumitru D., Ianache C., Ianache R. (2015), A cyberinfrastructure for air quality monitoring and early warnings to protect children with respiratory disorders, In Proceedings of the 20th International Conference on Control Systems and Computer Science (CSCS20-2015), Bucharest, 789–796.

[4] Dunea D., Iordache St., Alexandrescu D.C., Dincă N. (2014), Screening the weekdays/weekend patterns of air pollutant concentrations recorded in southeastern Romania, Environmental Engineering and Management Journal, 14(12), 3105–3115.

[5] WHO (2014), World Health Organization, Ambient (outdoor) air quality and health—Fact sheet N°313, Updated March 2014, http://www.who.int/mediacentre/factsheets/fs313/en, Accessed 10 February 2016.

[6] AQEG (2012), Fine Particulate Matter (PM2.5) in the United Kingdom, Defra, London.

[7] Dunea D., Iordache St. (2015), Time series analysis of air pollutants recorded from Romanian EMEP stations at mountain sites, Environmental Engineering and Management Journal, 14(11), 2725–2735.

[8] Vaisala (2016), Weather monitoring and urban air quality, Application note, http://www.vaisala.com/Vaisala%20Documents/Application%20notes/Urban%20air%20application%20note%20B210959EN-A.pdf, Accessed 10 February 2016.

[9] Dunea D., Oprea M., Lungu E. (2008), Comparing statistical and neural network approaches for urban air pollution time series analysis. In L. Bruzzone (ed.), Proceedings of the 27th IASTED International Conference on Modelling, Identification and Control, Acta Press, Innsbruck, Austria, February 11–13, 93–98.

[10] Box G.E.P., Jenkins G.M. (1970), Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, CA.

[11] Dunea D., Pohoaţă A., Iordache St. (2015), Using wavelet—feed forward neural networks to improve air pollution forecasting in urban environments, Environmental Monitoring and Assessment, 187, 477, 1–6.

[12] Riedmiller M., Braun H. (1993), A direct adaptive method for faster backpropagation learning: The RPROP algorithm, In H. Ruspini (ed.), Proceedings of the IEEE International Conference on Neural Networks, San Francisco, 586–591.

[13] Lungu E., Oprea M., Dunea D. (2008), An application of artificial neural networks in environmental pollution forecasting, In Proceedings of the 26th IASTED International Conference on Artificial Intelligence and Applications, Acta Press, Innsbruck, Austria, February 11–13, 187–193.

http://dx.doi.org/10.5772/63109

[14] Fahlman S.E. (1988), Faster learning variations on back-propagation: an empirical study, In D.S. Touretzky, G.E. Hinton, and T.J. Sejnowski (eds.), Proceedings of the 1988 Connectionist Models Summer School, Morgan Kaufmann, San Mateo, CA, 38–51.

[15] Vrahatis M.N., Magoulas G.D., Plagianakos V.P. (1999), Convergence analysis of the Quickprop method, In Proceedings of the International Joint Conference on Neural Networks (IJCNN'99), Washington DC, 848, Session: 5.3.

[16] Oprea M., Ianache C., Mihalache S., Dragomir E., Dunea D., Iordache Șt., Savu T. (2015), On the development of an intelligent system for particulate matter air pollution monitoring, analysis and forecasting in urban regions, In 19th International Conference on System Theory, Control and Computing (ICSTCC), 711–716.

[17] Oprea M. (2005), A case study of knowledge modelling in an air pollution control decision support system, AiCommunications, 18(4), 293–303.

[18] Guo X. (2010), Learning gradients via an early stopping gradient descent method, Journal of Approximation Theory, 162(11), 1919–1944.

[19] Srivastava N., Geoffrey H., Krizhevsky A., Sutskever I., Salakhutdinov R. (2014), Dropout: a simple way to prevent neural networks from overfitting, Journal of Machine Learning Research, 15, 1929–1958.

[20] Zhang Y., Bocquet M., Mallet V., Seigneur C., Baklanov A. (2012), Real-time air quality forecasting, part I: history, techniques, and current status, Atmospheric Environment, 60, 632–655.

[21] Mihalache S.F., Popescu M., Oprea M. (2015), Particulate matter prediction using ANFIS modelling techniques, In 19th International Conference on System Theory, Control and Computing (ICSTCC), 895–900.

IntechOpen

IntechOpen

IntechOpen