



Hybrid intelligent system for air quality forecasting using phase adjustment



Paulo S.G. de Mattos Neto^{a,*}, Francisco Madeiro^b, Tiago A.E. Ferreira^c, George D. C. Cavalcanti^d

^a Department of Computing, University of Pernambuco, Brazil

^b Department of Informatics, Catholic University of Pernambuco, Brazil

^c Department of Statistics and Informatics, Federal Rural University of Pernambuco, Brazil

^d Centro de Informática, Universidade Federal de Pernambuco, Brazil

ARTICLE INFO

Article history:

Received 23 October 2013

Received in revised form

27 February 2014

Accepted 19 March 2014

Available online 12 April 2014

Keywords:

Air quality forecasting

Artificial neural network

Genetic algorithm

Time series

PM_{2.5}

PM₁₀

ABSTRACT

The pollution caused by particulate matter (PM) concentration has a negative impact on population health, due to its relationship with several diseases. In this sense, several intelligent systems have been proposed for forecasting the PM concentration. Although it is known in the literature that PM concentration time series behave like random walk, to the authors' knowledge there is no intelligent systems developed to forecast the PM concentration that consider this characteristic. In this paper, we present an architecture developed to forecast time series guided by random walk process. The architecture, called Time-delay Added Evolutionary Forecasting (TAEF), consists of two steps: parameters optimization and phase adjustment. In the first step, a genetic optimization procedure is employed to adjust the parameters of a Multilayer Perceptron neural network that is used as the prediction model. The genetic algorithm adjusts the following parameters of the prediction model: the number of input nodes (time lags), the number of neurons in the hidden layer and the training algorithm. The second step is performed aiming to reduce the difference between the forecasting and the actual concentration value of the time series, that occur in the forecasting of the time series with random walk behavior. The approach is data-driven and only uses the past values of the pollutant concentrations to predict the next day concentration; in other words, it does not require any exogenous information. The experimental study is performed using time series of concentration levels of particulate matter (PM_{2.5} and PM₁₀) from Helsinki and shows that the approach overcomes previous state-of-the-art methods by a large margin.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The study of the air conditions, as well as its prediction, is an important topic due to the relationship between high concentrations of pollutants, such as particulate matter (PM), and adverse effects on human health, that is an issue of increasing public concern. Pollutants concentration forecasting (Niska et al., 2004; Pisoni et al., 2009; Siwek et al., 2011) is a relevant task, since it enables governments to warn the population regarding high levels of pollution. Several epidemiological studies have associated the concentration of those pollutants with cardiovascular and respiratory diseases (Ebelt et al., 2005; Nel, 2005; Peng et al., 2008). The

Global Monitoring Report (Bank, 2008) affirms that the major urban air pollutant that affects the human health is PM.

Several types of Artificial Neural Networks (ANN) have been applied for forecasting of the PM concentration. Some works have compared the performance of different ANNs in the prediction of one step ahead. Sharma et al. (2003) used the concentration data of seven pollutants, among them PM_{2.5} and PM₁₀, of the California area to evaluate the performance of four ANN models: Recurrent Network Model (RNM), Change Point Detection Model with RNM, Sequential Network Construction Model and Self Organizing Feature Model. Ordieres et al. (2005) addressed the PM_{2.5} concentration in the cities of El Paso (Texas) and Ciudad Juárez (Chihuahua) to compare the performance of the Multilayer Perceptron (MLP) model, Radial Basis Function (RBF) and Square Multilayer Perceptron (SMLP). Other works proposed ANN based on MLP model. Perez and Reyes (2006) developed an integrated ANN to forecast the maximum average concentration for PM₁₀ per day for city of Santiago, Chile. Kukkonen et al. (2003) combined the MLP model with homoscedastic and heteroscedastic Gaussian

* Corresponding author.

E-mail addresses: paulo.salgado@upe.br (P.S.G. de Mattos Neto), madeiro@dei.unicap.br (F. Madeiro), tiago@deinfo.ufrpe.br (T.A.E. Ferreira), gdcc@cin.ufpe.br (G.D.C. Cavalcanti).

URLS: <http://www.ppgia.ufrpe.br/tiago> (T.A.E. Ferreira), <http://cin.ufpe.br/~viisar> (G.D.C. Cavalcanti).

noise (ANN-HeG) to forecast the PM₁₀ concentration in Helsinki. Although most of the works have evaluated the prediction of one step ahead, Kurt and Oktay (2010) used a MLP model of three layers to forecast sulfur dioxide (SO₂), carbon monoxide (CO) and PM₁₀ concentration levels for 3 days ahead for a Besiktas district.

Intelligent models based on artificial neural network (ANN) also have been widely used to predict the PM concentration aiming best results. Some works combined Principal Component Analysis (PCA) with ANN models, as Slini et al. (2006), that proposed a hybrid system (PCA-MLP) to forecast PM₁₀ concentration in Thessaloniki. Voukantsis et al. (2011) proposed a methodology that selects the input variables using PCA and the prediction is performed by combining the outputs of two methods: MLP and linear regression (LR). In that work, the time series concentration of PM_{2.5} and PM₁₀ in Thessaloniki and Helsinki were used. Other hybrid systems were also proposed. Niska et al. (2005) integrated a MLP model with a numerical weather prediction model HIRLAM (High Resolution Limited Area Model) to predict sequential hourly time series of concentrations of PM_{2.5} in Helsinki. In that work, an input selection method based on the use of a multi-objective genetic algorithm (MOGA) was applied to reduce the number of potential meteorological input variables. For PM₁₀ concentration management in Santiago, Chile, Perez (2012) combined a MLP model with a nearest neighbor method. Siwek and Osowski (2012) combined ANNs to forecast daily average concentration of PM₁₀ in Warsaw, Poland.

Despite many intelligent models have been applied for forecasting of the PM concentration, to the authors' knowledge, no one considered the fact that these time series consist of a random walk process (Sitte and Sitte, 2002), typical of a Brownian motion process (Grau-Bové and Strlic, 2013; Sahu and Nicolis, 2008; Cheng et al., in press). A random walk (Sitte and Sitte, 2002) is a stochastic process that consists of successive and connected steps in which each step is chosen by a random mechanism uninfluenced by any previous step. This phenomenon also is widely seen in the financial time series representation.

In this paper, we present an evolutionary hybrid system, called TAEF method (Ferreira et al., 2008), for time series prediction of particulate matter concentrations. The system was particularly developed to forecast time series from of the phenomenon guided by random walk (Sitte and Sitte, 2002). The approach is composed of two main parts: parameters optimization and phase adjustment. The parameters optimization uses a genetic algorithm to search for

the best parameters to train the predictor. A Multilayer Perceptron neural networks is used as predictor and the following parameters are adjusted: number of time lags to represent the series, number of hidden units and the algorithm to perform the training of the predictor. After, a phase adjustment procedure is performed to improve the accuracy of the predictor by automatically correcting time phase distortions. These distortions in the forecasting are common when the phenomena guided by a random walk process (Ferreira et al., 2008) are addressed, such as the PM concentration (Grau-Bové and Strlic, 2013; Sahu and Nicolis, 2008; Cheng et al., in press). Thus, in the case of PM concentration levels, the phase adjustment procedure could significantly improve the accuracy of the intelligent methods? This is a relevant aspect addressed in this paper.

This paper is organized as follows. Section 2 describes the method and the phase adjustment procedure. Section 3 presents a set of six evaluation measures used to analyze the prediction results of the architecture. Simulation results and concluding remarks are presented in Sections 4 and 5, respectively.

2. The architecture for PM concentration forecasting

Given an univariate time series database (I), the output of the architecture is a trained Artificial Neural Network (ANN) that is ready to predict the next day value of the time series. The first step of the architecture is the Normalization, where each time series is normalized to lie within the interval $[0,1]$. After normalization, the database (I) is divided into three disjoint parts: training (Ω), validation (Υ), and test (Δ).

Fig. 1 shows that the architecture is composed of two main modules: (i) Parameters Optimization (Section 2.1) and (ii) Phase Adjustment (Section 2.2).

2.1. Parameters optimization

The parameters optimization module is based on a modified Genetic Algorithm (GA) proposed by Leung et al. (2003), used also in other works (Moris et al., 2003; Xu et al., 2007), and is the basis for the GA used in the TAEF method. The GA searches for the best parameters of an ANN in order to improve its performance. In the search process, the GA combines exploration (global search) and exploitation (local search) strategies (Yannibelli and Amandi,

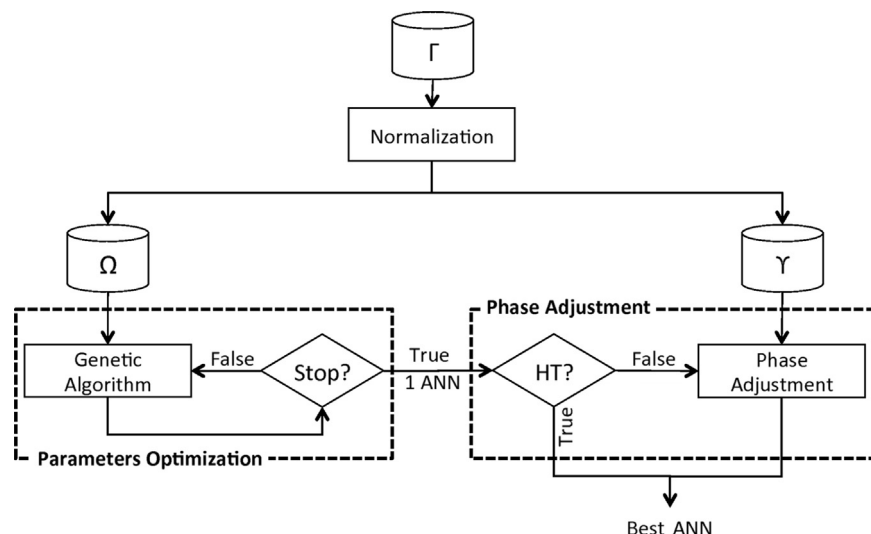


Fig. 1. Architecture of the system. "Stop?" corresponds to the stopping criteria of the Genetic Algorithm and "HT?" represents the output of a hypothesis test used to evaluate the forecasting of the Artificial Neural Network (ANN).

2013). The exploration process uses the genetic operators to search for new promising regions in the search space. The exploitation process is performed by the neural network training algorithm to fine-tune the solutions reached by the evolutionary search. The output of this module is the best ANN chosen from a set of candidates after the optimization procedure.

Inspired by Takens' (1980) theorem, the approach aims to search for the minimum dimensionality of time lags required to reproduce the generative phenomenon of the time series. In other words, the GA searches for the best and more optimized combination of time lags of the predictive model. On the design of a predictive model, the number of time lags is an important aspect because the larger the number of lags the higher the cost associated to represent the time series.

The initial population of the GA is a set of possible solutions randomly generated and each individual of the population is encoded by one chromosome. The chromosome is composed of the following parameters that defines an ANN (in our case, a multilayer perceptron—MLP): number of input nodes, number of units in the hidden layer, weights, training algorithm and an unit step function to each weight. This unit step acts like a switch to each weight of the ANN, since it enables that connections which are not important be eliminated in the evolution process.

After the initial population generation, each individual is trained using one of the four candidate algorithms: Levenberg–Marquardt (Mor, 1977), Scaled Conjugated Gradient (Moller, 1993), RPROP (Reidmiller and Braun, 1993) and One Step Secant Conjugate Gradient (Battiti, 1992). The stopping criteria of this training are the number of epochs, the increase in the validation error and the decrease in the training error.

After the training using a gradient descent algorithm, the GA is used to evolve the population towards a good fitness solution. So, each chromosome in the population is evaluated by a fitness function for the PM concentration forecasting proposed here. The fitness function is defined in Eq. (1) and it is composed of five well-known performance measures: Prediction Of Change In Direction (POCID), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), U of Theil Statistics (Theil) and Average Relative Variance (ARV). These measures are defined in Section 3.

$$\text{fitness} = \frac{\text{POCID}}{1 + \text{MSE} + \text{MAPE} + \text{Theil} + \text{ARV}} \quad (1)$$

This fitness function creates a global indicator of the forecasting performance (in the range 0–100). The higher the fitness value, the better is the quality of the prediction model.

After the evaluation phase, two chromosomes are selected as parents by the method of spinning the roulette wheel. The higher the fitness value of the chromosome, the higher is the chance of it being selected. So, it is expected that high potential parents will produce better offspring. The parents generate the offspring using the genetic operators proposed by Leung et al. (2003). For each new offspring created, the individuals are trained using a neural network training algorithm among four candidates and this algorithm is selected by the evolutionary process.

The stopping criteria of the GA (“Stop?”— Fig. 1) are (i) the number of epochs; (ii) an increase in the validation error; or (iii) a decrease in the training error.

When any GA stop criterion is reached, the algorithm compares the fitness of the best individual (f_{Best}) with the minimum acceptable fitness ($MinFit$) defined by the user. The best ANN (with the fitness value of f_{best}) of the population is the one with the higher fitness on the validation database (Y). If f_{Best} is smaller than $MinFit$, the variable that defines the maximum number of lags ($MaxLags$) in the search space of the GA is increased by one unit. After, the GA is reinitialized to search for a better solution. This update aims to increase the chance of selecting relevant lags. In contrast, if the fitness of the best individual is greater than $MinFit$, the variable $MaxLags$ is updated to the number of lags of this individual. After, the $MinFit$ is set as the fitness value reached by the best individual and the GA is started again to search for a better solution. In this case, we expect to find a solution that has a higher fitness than $MinFit$. This is possible because the vicinity of the best individual found, so far, can be a promising region for better individuals. This process is repeated until the maximum number of iterations of the approach be reached.

2.2. Phase adjustment

The best ANN obtained by the first module of the approach (Section 2.1) is evaluated by a statistical test (hypothesis test, “HT?”— Fig. 1). The t -test is used and it aims to verify if the obtained ANN has a predictive performance better than a random walk like model (null hypothesis) or not (alternative hypothesis). A random walk model (Sitte and Sitte, 2002) is the simplest forecast model, where the best prediction of one step ahead is given by the current value of the time series. Based on the result of the statistical test, the method selects one out of two operation modes:

- *In-phase*: The shapes of the actual and the predicted time series have a time matching. In other words, the trained ANN is not a random walk like model.
- *Out-of-phase*: A time delay mismatch of the predicted series is observed when compared with the actual time series. The trained ANN tends to behave like to a random walk like model.

When the operation In-phase mode is selected (null hypothesis is not rejected), this means that the trained ANN is ready for practical use. In contrast, when the operation Out-of-phase mode is selected, a phase adjustment procedure is performed to minimize the effects of the time delay mismatch. The phase adjustment procedure (Ferreira et al., 2008) is composed of two steps (Fig. 2): (1) given an input vector $[t_1, t_2, \dots, t_n] \in Y$, the output (y_1) of the ANN model is calculated; (2) the input vector is rearranged to include y_1 : $[y_1, t_1, t_2, \dots, t_{n-1}]$. This new vector is given as input to the same ANN model used in Step 1 and its output is the final forecasting.

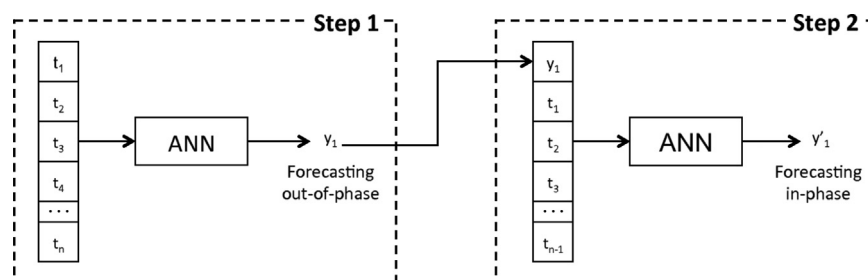


Fig. 2. Procedure to adjust the phase of the forecasting (adapted from Ferreira et al., 2008).

3. Evaluation measures

The evaluation of the method is performed by six measures: Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), *U* of Theil Statistics, Average Relative Variance (ARV), Prediction of Change in Direction (POCID) and Index of Agreement (IA). These measures are described below.

The MSE measure (Rodrigues et al., 2010; de Mattos Neto et al., 2010) is commonly used in the literature of time series forecasting and it is defined by the following equation:

$$MSE = \frac{1}{N} \sum_{j=1}^N (\text{target}_j - \text{output}_j)^2, \quad (2)$$

where *N* is the size of the series, target_j is the real value at period *j* and output_j is the predicted value at period *j*. It is worth mentioning that MSE cannot be considered a conclusive measure for comparison of different forecasting models (Clements and Hendry, 1993). Thus, other evaluation measures should be considered.

The MAPE measure (Rodrigues et al., 2010; de Mattos Neto et al., 2010) is given by the equation:

$$MAPE = \frac{100}{N} \sum_{j=1}^N \left| \frac{\text{target}_j - \text{output}_j}{\text{target}_j} \right|. \quad (3)$$

The *U* of Theil Statistics (Rodrigues et al., 2010; de Mattos Neto et al., 2010) can be defined as

$$\text{Theil} = \frac{\sum_{j=1}^N (\text{target}_j - \text{output}_j)^2}{\sum_{j=1}^N (\text{output}_j - \text{output}_{j+1})^2}. \quad (4)$$

The ARV measure (Rodrigues et al., 2010; de Mattos Neto et al., 2010) is defined as

$$ARV = \frac{\sum_{j=1}^N (\text{output}_j - \text{target}_j)^2}{\sum_{j=1}^N (\text{output}_j - \overline{\text{target}})^2} \quad (5)$$

where $\overline{\text{target}}$ is the mean of the series.

The POCID measure (Rodrigues et al., 2010; de Mattos Neto et al., 2010) is given by

$$POCID = 100 \frac{\sum_{j=1}^N D_j}{N}, \quad (6)$$

where

$$D_j = \begin{cases} 1, & \text{if } (\text{target}_j - \text{target}_{j-1})(\text{output}_j - \text{output}_{j-1}) > 0. \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

The IA measure (Voukantsis et al., 2011) is given by

$$IA = 1 - \frac{\sum_{j=1}^N |\text{output}_j - \text{target}_j|^2}{\sum_{j=1}^N (|\text{output}_j - \text{target}_j| + |\text{target}_j - \overline{\text{target}}|)^2}. \quad (8)$$

For MSE, MAPE, Theil and ARV, the lower the value of those measures, the better is the forecasting of the model. Theil is a measure used to compare the model performance with a random walk model performance. If the value of the measure is equal to 1, the model is equivalent to the random walk model. However, if its value is smaller or greater than 1, the model performance is better or worse than the performance of the random walk model, respectively. ARV is a measure that is used to compare the model forecasting with the forecasting of the mean of the series. If the ARV value is equal to 1, the forecasting of the predictor is equal to the mean of the series. However, if its value is less or greater than 1, the prediction of the model is better or worse than the mean, respectively. In case of POCID and IA, the higher the value the better is the performance of the model. The POCID can have values in the range [0, 100] and IA in the range [0,1].

4. Simulation and results

The method is evaluated using four time series that correspond to daily mean concentrations of PM_{2.5} and PM₁₀ from Helsinki. The data were measured in the time period of 2001–2003 from stations of Kallio and Vallila (Voukantsis et al., 2011). These stations are in locals with different characteristics of Helsinki. Vallila is more influenced by traffic in the center of the city, while the station of Kallio is an urban background. Generally, the population on urban background locals is less exposed to pollution, whereas the traffic locals represent urban environments more severely polluted. For each region, PM_{2.5} and PM₁₀ concentration time series are used in this work, composing a data set of four series. All series are normalized to lie within the interval [0,1] and divided in three sets: 80% for training, 10% for validation and 10% for test. For each time series, ten simulations were performed by the architecture and the best model based on the validation dataset (*Y*) is selected as the predictive model. All the results shown refer to one step ahead predictions generated using the test set (Δ). The parameters of the method are set up to (i) Initial acceptable fitness value (1% of error); (ii) Initial maximum number of time lags (10); (iii) Maximum number of hidden units (20); and (iv) Maximum number of iterations (10).

The parameters of the genetic algorithm used by the architecture are set up to (i) Mutation probability (10%); (ii) Population size (10); (iii) Maximum number of generations (1000); and (iv) Minimum fitness progress (10^{-4}). Three stopping conditions for the method are used: (i) Maximum number of iterations (1000); (ii) Generation loss (5%); and (iii) Progress training (10^{-6}).

4.1. Kallio station time series forecasting

Table 1 shows the forecasting of the PM_{2.5} and PM₁₀ concentration time series from the Kallio station using the architecture under consideration. Comparing the results before (In-phase) and after (Out-of-phase) the phase adjustment, we observe that the forecasting is consistently better when the phase of the time series is adjusted by the approach, independently of the time series analyzed and the measure used.

Figs. 3 and 4 show the real concentration series for the Kallio station (solid lines) and the forecasting generated by the method (dashed lines) for PM_{2.5} and PM₁₀, respectively. Figs. 3(a) and 4(a) show the results without phase adjustment and Figs. 3(b) and 4(b) show the results with phase adjustment. Observing these figures, we note that the best fit happens when the phases are adjusted.

After the optimization procedures, the method selected the following parameters for the best ANN when the PM_{2.5} concentration time series was used: (i) six lags: 1, 2, 4, 8, 9 and 10; (ii) eight

Table 1
Forecasting of the Kallio station time series using the architecture.

	Kallio station			
	PM _{2.5}		PM ₁₀	
	In-phase	Out-of-phase	In-phase	Out-of-phase
MSE	0.0047	0.0003	0.0053	0.0006
MAPE	127.32	27.41	105.93	35.93
Theil	0.7954	0.0596	0.7522	0.0969
POCID	41.66	97.19	45.28	97.16
ARV	1.8594	0.1349	2.5764	0.3077
IA	0.59	0.97	0.53	0.95
Fitness	0.32	3.40	0.41	2.60

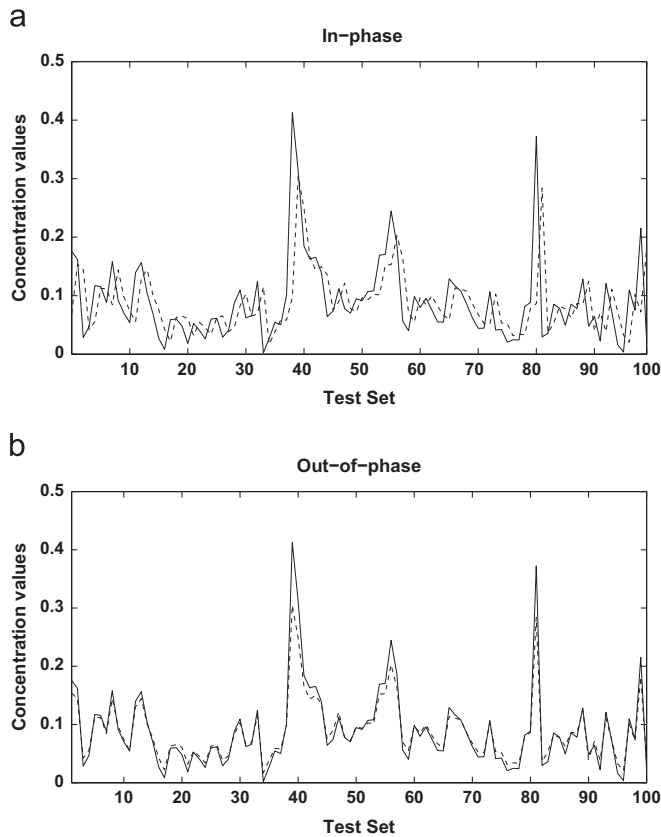


Fig. 3. Forecasting for the PM_{2.5} concentration time series for Kallio station (last 100 points). (a) Forecasting without the phase adjustment. (b) Forecasting with the phase adjustment.

neurons in hidden layer; and (iii) Levenberg–Marquardt training algorithm. For the PM₁₀ concentration time series, the parameters were (i) five lags: 2, 3, 4, 5 and 9; (ii) four neurons in hidden layer; and (iii) Levenberg–Marquardt training algorithm. For both cases, the method indicated the out-of-phase configuration for the predictive models.

4.2. Vallila station time series forecasting

Table 2 shows the forecasting of the PM_{2.5} and PM₁₀ concentration time series from the Vallila station using the architecture under consideration. Comparing the results before (In-phase) and after (Out-of-phase) the phase adjustment, we observe again that the forecasting is consistently better when the phase of the time series is adjusted by the approach, independently of the time series analyzed and the measure used.

Figs. 5 and 6 show the real concentration series for the Vallila station (solid lines) and the forecasting generated by the method (dashed lines) for PM_{2.5} and PM₁₀, respectively. Figs. 5(a) and 6 (a) show the results without phase adjustment and Figs. 5(b) and 6 (b) show the results with phase adjustment. We can note that again the best fit happens when the phases are adjusted.

After the optimization procedures, the architecture selected the following parameters for the best ANN when the PM_{2.5} concentration time series was used: (i) six lags: 1, 2, 3, 6, 7 and 8; (ii) six neurons in hidden layer; and (iii) Levenberg–Marquardt training algorithm. For the PM₁₀ concentration time series, the parameters were (i) four lags: 1, 4, 6 and 10; (ii) eight neurons in hidden layer; and (iii) Levenberg–Marquardt training algorithm. The method classified the predictive ANN with an out-of-phase configuration, for the both time series PM_{2.5} and PM₁₀.

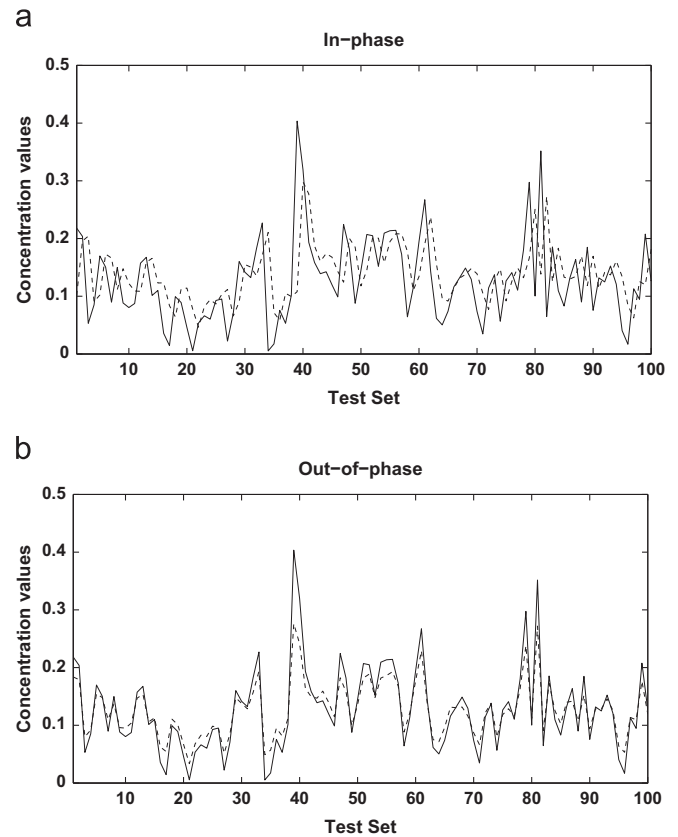


Fig. 4. Forecasting for the PM₁₀ concentration time series for Kallio station (last 100 points). (a) Forecasting without the phase adjustment. (b) Forecasting with the phase adjustment.

Table 2
Forecasting of the Vallila station time series using the architecture.

	Vallila station			
	PM _{2.5}		PM ₁₀	
	In-phase	Out-of-phase	In-phase	Out-of-phase
MSE	0.0028	0.0001	0.0056	0.0003
MAPE	77.56	21.66	71.04	29.66
Theil	0.7336	0.0313	1.2167	0.0727
POCID	43.00	98.11	44.68	97.82
ARV	2.1951	0.0630	2.7216	0.1910
IA	0.57	0.98	0.32	0.96
Fitness	0.53	4.31	0.59	3.16

4.3. Discussion

From the results in Tables 1 and 2, it is possible to observe that the method generated predictive models (ANNs) for PM concentration with a prediction performance better than the random walk model (based on the statistics *U* of Theil) and better than the forecasting using only the mean of series (ARV). The POCID measure shows that the prediction obtained by the predictive models follows the direction of the series in all cases. From the results shown in Tables 1 and 2, it is observed that the phase adjustment procedure improved significantly the performance of the hybrid system for all used measures and the fitness function. This result shows that this procedure can enhance the performance of the prediction when PM concentration time series are addressed.

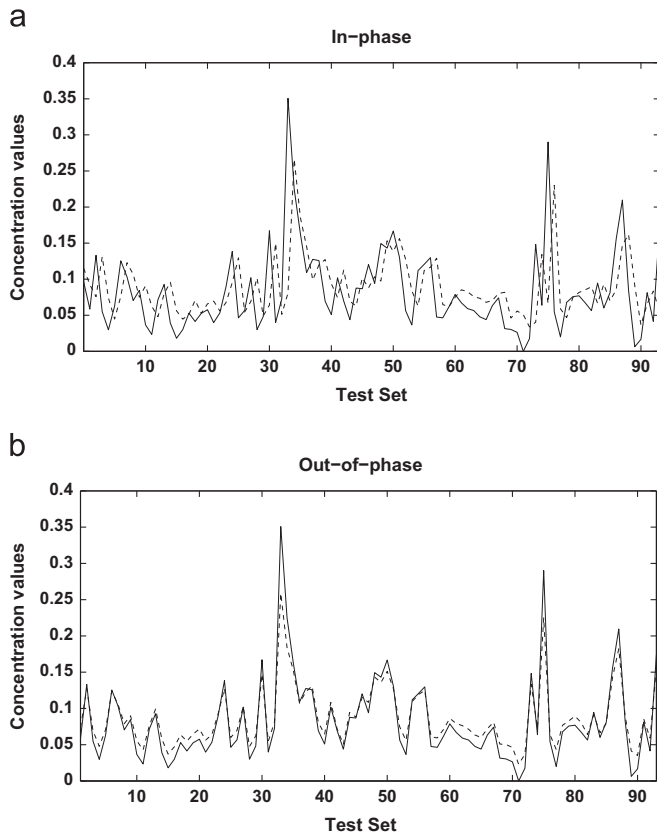


Fig. 5. Forecasting for the $PM_{2.5}$ concentration time series for Vallila station (last 95 points). (a) Forecasting without the phase adjustment. (b) Forecasting with the phase adjustment.

Table 3 shows the results, in terms of IA, obtained by the approach and by several techniques for different periods. The IA was considered because it is a performance measure widely used in the literature of the pollutant forecasting. The values used in the comparison are shown only as a reference. However, for all cases, the approach overcomes the results found in the literature by a large margin. Another alternative to measure the performance of the proposed method is to evaluate the mean IA using the k -fold cross-validation procedure. So, each time series was divided into three folds ($k=3$), where each fold corresponds to 1 year (2001, 2002 and 2003). The proposed approach obtained the following mean and standard deviation results in terms of the IA measure: (i) for the Kallio Station: 0.95 ± 0.03 for $PM_{2.5}$ and 0.94 ± 0.01 for PM_{10} ; and, (ii) for the Vallila Station: 0.96 ± 0.01 and 0.95 ± 0.01 for $PM_{2.5}$ and PM_{10} . These average results are similar to the ones shown in Table 3.

In contrast with previous models (Kukkonen et al., 2003; Voukantsis et al., 2011; Niska et al., 2005; Siwek and Osowski, 2012; Vlachogianni et al., 2011) that use several variables of meteorological data, the presented architecture does not require any exogenous information to perform the prediction.

5. Conclusions

In this paper, an architecture to forecast particulate matter concentration levels was presented. The architecture is composed of two main parts: parameters optimization and phase adjustment. In the parameters adjustment part, an optimization process is performed to find the best parameters of the predictor. After, a phase adjustment procedure is performed to minimize the effects of the time delay mismatch. This phase adjustment aims to

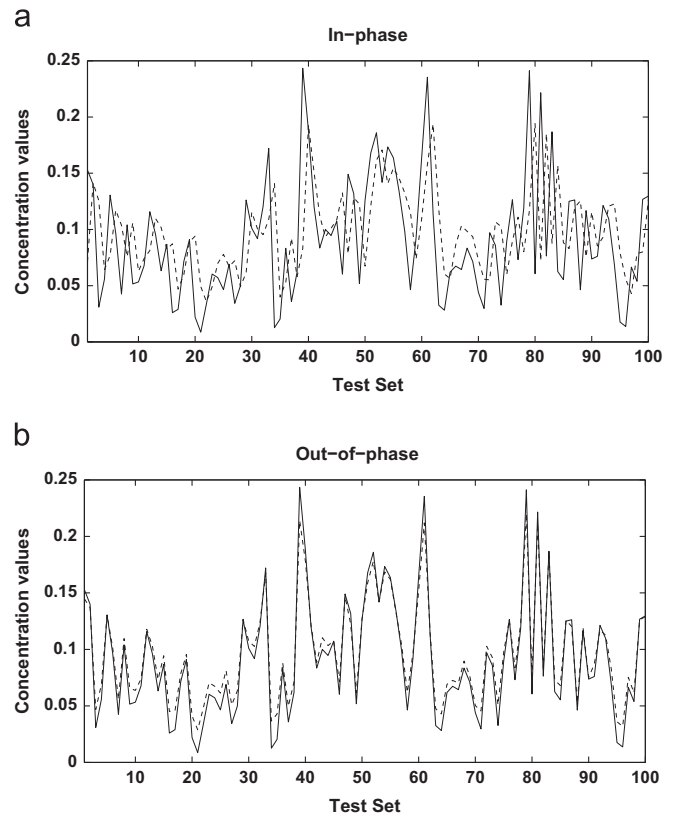


Fig. 6. Forecasting for the PM_{10} concentration time series for Vallila station (last 95 points). (a) Forecasting without the phase adjustment. (b) Forecasting with the phase adjustment.

Table 3

Results of the pollutant concentration prediction from Helsinki using the measure Index of Agreement. n.a., not available.

Reference	Method	Period	$PM_{2.5}$	PM_{10}
Kallio station				
Niska et al. (2005)	MLP-HIRLAM	May 2001–April 2002	0.85	n.a
Vlachogianni et al. (2011)	MLR	2005 (warm period)	n.a	0.90 ^a
Voukantsis et al. (2011)	PCA-MLP-LR	2001–2003	0.87	0.87
Method	TAEF	2001–2003	0.97	0.95
Vallila station				
Kukkonen et al. (2003)	ANN-HeG	1999	n.a	0.77
Vlachogianni et al. (2011)	MLR	2005 (warm period)	n.a	0.88 ^a
Voukantsis et al. (2011)	PCA-MLP-LR	2001–2003	0.87	0.87
Method	TAEF	2001–2003	0.98	0.96

^a In cold period, Vlachogianni et al. (2011) reached 0.85 and 0.82 for Kallio station and Vallila station, respectively.

improve the accuracy of the predictor by automatically correcting time phase distortions that can occur in the forecasting of time series, that has its generation process driven by Brownian motion process (also called “random walk”) (Grau-Bové and Strlic, 2013).

The performance of the method was assessed using six well-known performance measures and four time series. The time series consist of concentration levels of $PM_{2.5}$ and PM_{10} from the stations of Kallio and Vallila. The stations are located in urban background and urban traffic, each one containing different characteristics. The experimental results showed a consistent better performance of the approach when compared with other techniques (e.g. Voukantsis et al., 2011; Vlachogianni et al., 2011) for all investigated series. In contrast to other methods in the literature that use several exogenous time series to predict a time

series of interest, the approach uses only the time series under study to perform the prediction.

The performance of the method was reached mainly due to the use of the phase correction procedure. Thus, these results show that the phase adjustment procedure can be an interesting alternative to enhance the forecasting of the intelligent methods that are proposed to solve the forecasting problem of the PM concentration.

References

- Bank, W., 2008. Global Monitoring Report 2008: MDGs and the Environment: Agenda for Inclusive and Sustainable Development. Technical Report.
- Battiti, R., 1992. One step secant conjugate gradient. *Neural Comput.* 4, 141–166.
- Cheng, K.-C., Acevedo-Bolton, V., Jiang, R.-T., Klepeis, N.E., Ott, W.R., Kitanidis, P.K., Hildemann, L.M. Stochastic modeling of short-term exposure close to an air pollution source in a naturally ventilated room: an autocorrelated random walk method. *J. Expo. Sci. Environ. Epidemiol.* 1–8, <http://dx.doi.org/10.1038/jes.2013.63>, in press.
- Clements, M.P., Hendry, D.F., 1993. On the limitations of comparing mean square forecast errors. *J. Forecast.* 12 (8), 617–637.
- de Mattos Neto, P.S.G., Rodrigues, A.L.J., Ferreira, T.A.E., Cavalcanti, G.D., 2010. An intelligent perturbative approach for the time series forecasting problem. In: *IEEE World Congress on Computational Intelligence (WCCI 2010)*, IEEE, pp. 1–8.
- Ebel, S.T., Wilson, W.E., Brauer, M., 2005. Exposure to ambient and nonambient components of particulate matter: a comparison of health effects. *Epidemiology* 16 (3), 396–405.
- Ferreira, T.A.E., Vasconcelos, G.C., Adeodato, P.J.L., 2008. A new intelligent system methodology for time series forecasting with artificial neural networks. *Neural Process. Lett.* 28 (2), 113–129.
- Grau-Bové, J., Strlic, M., 2013. Fine particulate matter in indoor cultural heritage: a literature review. *Herit. Sci.* 1 (1), 1–17.
- Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., Cawley, G., 2003. Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmos. Environ.* 37 (32), 4539–4550.
- Kurt, A., Oktay, A.B., 2010. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Syst. Appl.* 37 (12), 7986–7992.
- Leung, F.H.F., Lam, H.K., Ling, S.H., Tam, P.K.S., 2003. Tuning of the structure and parameters of a neural network using an improved genetic algorithm. *IEEE Trans. Neural Netw.* 14 (1), 79–88.
- Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* 6, 525–533.
- Mor, J.J., 1977. The Levenberg–Marquardt algorithm: implementation and theory. In: *Watson, G.A. (Ed.), Numerical Analysis, Lecture Notes in Mathematics*, vol. 630. Springer-Verlag, Berlin, pp. 105–116.
- Moris, S., Dash, P., Basu, K., Sharaf, A.M., 2003. UPFC controller design for power system stabilization with improved genetic algorithm. In: *29th Annual Conference of the IEEE Industrial Electronics Society (IECON 2003)*, vol. 2. IEEE, pp. 1540–1545.
- Nel, A., 2005. Air pollution-related illness: effects of particles. *Science* 308 (5723), 804–806.
- Niska, H., Hiltunen, T., Karppinen, A., Ruuskanen, J., Kolehmainen, M., 2004. Evolving the neural network model for forecasting air pollution time series. *Eng. Appl. Artif. Intell.* 17 (2), 159–167.
- Niska, H., Rantamäki, M., Hiltunen, T., Karppinen, A., Kukkonen, J., Ruuskanen, J., Kolehmainen, M., 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. *Atmos. Environ.* 39 (35), 6524–6536.
- Ordieres, J., Vergara, E., Capuz, R., Salazar, R., 2005. Neural network prediction model for fine particulate matter (PM_{2.5}) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environ. Model. Softw.* 20 (5), 547–559.
- Peng, R.D., Chang, H.H., Bell, M.L., McDermott, A., Zeger, S.L., Samet, J.M., Dominici, F., 2008. Coarse particulate matter air pollution and hospital admissions for cardiovascular and respiratory diseases among medicare patients. *J. Am. Med. Assoc.* 299 (18), 2172–2179.
- Perez, P., 2012. Combined model for PM₁₀ forecasting in a large city. *Atmos. Environ.* 60, 271–276.
- Perez, P., Reyes, J., 2006. An integrated neural network model for PM₁₀ forecasting. *Atmos. Environ.* 40 (16), 2845–2851.
- Pisoni, E., Farina, M., Carnevale, C., Piroddi, L., 2009. Forecasting peak air pollution levels using NARX models. *Eng. Appl. Artif. Intell.* 22 (4–5), 593–602.
- Reidmiller, M., Braun, H., 1993. A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: *IEEE International Conference on Neural Networks (ICNN 1993)*, IEEE, San Francisco, pp. 586–591.
- Rodrigues, A.L.J., Silva, D.A., de Mattos Neto, P., Ferreira, T.A.E., 2010. An experimental study of fitness function and time series forecasting using artificial neural networks. In: *Genetic and Evolutionary Computation Conference (GECCO 2010)*, ACM, pp. 2015–2018.
- Sahu, S.K., Nicolis, O., 2008. An evaluation of European air pollution regulations for particulate matter monitored from a heterogeneous network. *Environmetrics* 20 (8), 943–961.
- Sharma, S., Barai, S.V., Dikshit, A.K., 2003. Studies of air quality predictors based on neural networks. *Int. J. Environ. Pollut.* 19 (5), 442–453.
- Sitte, R., Sitte, J., 2002. Neural networks approach to the random walk dilemma of financial time series. *Appl. Intell.* 16 (3), 163–171.
- Siwek, K., Osowski, S., 2012. Improving the accuracy of prediction of PM₁₀ pollution by the wavelet transformation and an ensemble of neural predictors. *Eng. Appl. Artif. Intell.* 25 (6), 1246–1258.
- Siwek, K., Osowski, S., Sowinski, M., 2011. Evolving the ensemble of predictors model for forecasting the daily average PM₁₀. *Int. J. Environ. Pollut.* 46, 199–215.
- Slini, T., Kaprara, A., Karatzas, K., Moussiopoulos, N., 2006. PM₁₀ forecasting for Thessaloniki, Greece. *Environ. Model. Softw.* 21 (4), 559–565.
- Takens, F., 1980. Detecting strange attractor in turbulence. In: *Dold, A., Eckmann, B. (Eds.), Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, vol. 898. Springer-Verlag, New York, pp. 366–381.
- Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., Kukkonen, J., 2011. Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM₁₀ in Athens and Helsinki. *Sci. Total Environ.* 409 (8), 1559–1571.
- 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. *Sci. Total Environ.* 409 (7), 1266–1276.
- Xu, J., Lu, Y., Ho, D.W.C., 2007. A combined genetic algorithm and orthogonal transformation for designing feedforward neural networks. In: *Third International Conference on Natural Computation (ICNC 2007)*, vol. 1. IEEE, pp. 10–14.
- Yannibelli, V., Amandi, A., 2013. Hybridizing a multi-objective simulated annealing algorithm with a multi-objective evolutionary algorithm to solve a multi-objective project scheduling problem. *Expert Syst. Appl.* 40 (7), 2421–2434.