Scientia Iranica E (2021) 28(3), 1699–1710

SCIENTIA IRANICA

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
http://scientiairanica.sharif.edu

A hybrid model for online prediction of $PM_{2.5}$ concentration: A case study

Y.S. Sadabadi[a], M. Salari[a,*], and R. Esmaili[b]

a. *Department of Industrial Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.*
b. *Environmental Pollution Monitoring Center of Mashhad, Deputy of Services, and Urban Environment, Municipality of Mashhad, Iran.*

Received 4 September 2017; received in revised form 16 May 2019; accepted 2 September 2019

KEYWORDS
$PM_{2.5}$;
Prediction;
Neural networks;
Wavelet transformation;
Monte Carlo simulation;
WRF model.

Abstract. The present study aims to develop a model to predict the daily average concentration of particulate matter with a diameter of less than 2.5 micrometers ($PM_{2.5}$). This model employed Weather Research and Forecasting (WRF) meteorological model, Monte Carlo simulation, wavelet transform, and Multi-Layer Perceptron (MLP) neural networks. In particular, the MLP and wavelet transformation were combined for prediction purposes. The WRF meteorological model was employed to predict such input parameters of the model as wind speed, wind direction, temperature, rainfall, and temperature inversion. Finally, to achieve a more accurate prediction considering the uncertainty associated with the input data, Monte Carlo simulation was employed. Further, to assess the effectiveness of the proposed model in the real world, it was run in the online mode for 35 days. Numerical results revealed that the combined model yielded acceptable accuracy when using widely used measures. To be specific, according to R measurements, this accuracy was measured as 0.831 over a set of test instances.

1. Introduction

Several large cities in developing countries are increasingly facing high levels of air pollution that affect their life quality and overall health [1]. Located in the northeast of Iran, Mashhad is one of these cities dealing with air pollutions. The dominant pollutant in this city is particulate matter with a diameter of less than 2.5 micrometers ($PM_{2.5}$). This pollutant increases the risk of contracting some diseases that involve heart and lung. A reliable prediction of this pollutant can lead to efficient activities done to prevent air pollution crisis or, at least, reduce its adverse effects [2]. However, due to a large number of resources as well as the complexity and variety of physical and chemical processes involved in the formation and transformation of pollutants, it is quite difficult to ensure an accurate prediction. To this end, different types of approaches have been presented: empirical, deterministic, and statistical approaches [3].

Empirical approaches, which are based on simple principles, are classified into three categories of persistence, climatology, and empiricism [3]. In the persistence approach, it is assumed that today's observed pollutant level determines the predicted value of tomorrow [4]. In fact, in this technique, sudden changes in the air quality are not taken into account because considering such changes could attenuate the accuracy. The climatology approach uses the mean of pollutant values accumulated over 2–5 years for

*. *Corresponding author. Tel.: +98 5138805113*
E-mail address: msalari@um.ac.ir (M. Salari)

doi: 10.24200/sci.2019.5120.1107

prediction [3]. It is proved to be effective only when there are no changes in the data pattern in a specific time period. In case of changes in the pattern of data, this technique cannot yield an acceptable prediction. Finally, the empiricism approach (criteria or rules of thumb) is based on the assumption that the thresholds (i.e., criteria) of meteorological or air quality variables can lead to future pollutant concentrations [3,4].

In the deterministic approach, the physical and chemical relations associated with the production and dispersion of pollutants are employed to propose a model for simulation purposes. This approach does not require a large computation volume. However, it requires precise and complete information about pollution-producing sources, emission ways, and physical and chemical processes [5]. Although this type of approach can predict the concentration of pollutants in places using no measurement stations, the knowledge required for this class of methods is still limited. Therefore, in case of insufficient knowledge, some approximations and simplifications are employed, hence a number of errors [6]. In many cases, use of the mentioned approach is computationally costly [3,7].

Statistical approaches are formulated based on the existing relationship between the concentration of pollutants and effective parameters of air pollution. This type of approach often requires a large volume of data measured under different weather conditions. In this respect, the model can describe the relationship among variables using the available data [5]. To its disadvantage, this method is only applicable to a specific area from which the data are taken and thus, it cannot be generalized to other regions with different meteorological conditions [5]. However, to explore the complex relationship between the concentration of pollutants and possible parameters, using statistical approaches is preferred for a specific area over other approaches [6]. In general, statistical approaches provide more accurate results than deterministic ones in terms of reproducing local measured concentrations over short-term periods [8]. Furthermore, statistical approaches are easier, quicker, and more economical tools for predicting the concentration of pollutants than other approaches [9]. The most common statistical approaches are regression models [10,11], Artificial Neural Networks (ANN) [12,13], and Support Vector Machine (SVM) [14,15].

Recently, ANNs have been increasingly used as nonlinear tools for modeling air pollution with satisfactory results [5,16]. The first study on the prediction of pollution concentration using neural networks was conducted by Boznar et al. [17]. The main objective of this study was to provide an hourly prediction of sulfur dioxide (SO_2) concentration for a region around the biggest Slovenian thermal power plant at Sostanj. Perez and Reyes [18] proposed a multi-layer neural network model and a linear model to predict the average concentration of PM_{10} for 24 hours in advance. They showed that although ANN models could yield relatively more accurate results than linear models, choosing input parameters was of greater significance than choosing the nature of the models (i.e., linear or nonlinear). Multi-Layer Perceptron (MLP) models were used for predicting the daily average of air pollution index in [19]. They also suggested that implementing the MLP model and early stopping criterion could yield the desired results. A two-stage neural network for daily prediction of ozone concentration was introduced in [20]. In particular, at the first stage, the meteorological conditions were clustered into meteorological regimes using self-organizing map neural networks. At the second stage, the MLP model was utilized to approximate the nonlinear relation of ozone meteorology in each of the meteorological regimes. Hrust et al. [6] proposed an MLP neural network to predict the concentration of CO, O_3, NO_2, and PM_{10} on an hourly scale. In this paper, a new approach based on a group of univariate regression models was used to determine the averaging intervals for input variables. Paschalidou et al. [21] proposed several models for hourly prediction of PM_{10} concentration in Larnaca, Limassol, Nicosia, Paphos, and Cyprus using ANNs, including MLP and Radial Basis Function (RBF), as well as a model based on Principal Component Regression Analysis (PCRA). Their evaluation showed that MLP outperformed other models. They believed that the generally poor performance of the RBF was mainly because this model was specifically appropriate for dealing with clustering and pattern recognition problems. A hybrid model based on ANN and Taylor expansion was also proposed in [22] to predict the concentrations of PM_{10} and SO_2 on a daily scale. The authors claimed that their model could be useful when the existing input parameters were not complete. The results of their proposed model overcame the shortcomings of those achieved through the MLP and SVM. Perez and Gramsch [7] developed a model for hourly forecasting of $PM_{2.5}$ concentration via neural network in Santiago, Chile. The input data used in their model were collected by 7 PM on the day of prediction. Moreover, forecasting was done up to 15 hours in advance, starting at 8 PM of the present day. Gao et al. [23] investigated the feasibility of using the ANN model with seven input variables to predict the average ozone concentration in the daytime (9:00 am–6:00 pm) in the urban area of Jinan, a metropolis in Northern China. In this model, input variables were selected using the forward selection procedure. Finally, uncertainty and sensitivity analyses were performed based on Monte Carlo simulations. The results of uncertainty analysis demonstrated that the ANN model could properly predict ozone level, although while a number of extremely high values for ozone level fell outside of the 95% confidence interval. Furthermore, Monte Carlo simulation technique was employed to investigate the sensitivity of the output ozone concentration to the meteorological and temporal input variables. Maximum temperature, atmospheric pressure, sunshine duration, and maximum wind speed were identified as the input variables that could significantly affect the range of predicted ozone concentrations.

Wavelet transform is a tool integrated by some other methods to predict the pollutant concentration. In particular, Siwek and Osowski [24] employed the wavelet transformation in combination with a group of neural networks for daily prediction of PM_{10} concentration. In this approach, the outputs of seven types of neural networks were regarded as the inputs for another neural network to make a final prediction. The obtained results confirmed the effectiveness of the wavelet transform in combination with the neural network in predicting the air quality. Feng et al. [5] proposed a hybrid model composed of the analysis of air mass trajectory, wavelet transformation, and MLP neural network to predict $PM_{2.5}$ concentration for prediction two days in advance. Another hybrid model composed of Stationary Wavelet Transformation (SWT) and Back Propagation Neural Network (BPNN) was proposed in [25] to predict the concentrations of PM_{10}, SO_2, and NO_2 on a daily basis in Nan'an District of Chongqing, Chile. This model functions based on the following procedure. First, the SWT is applied to decompose the time series of the pollutant concentrations into different scales, in which the obtained information represents the wavelet coefficients of pollutant concentration. Next, the wavelet coefficients together with other inputs are used to train an ANN on different scales. Finally, the estimated coefficients using BPNN are employed to reconstruct the forecasting results through inverse SWT. The results of the proposed model overcame those obtained by the BPNN.

In this study, a hybrid model for daily prediction of $PM_{2.5}$ pollutant concentration was developed. The introduced model incorporates Weather Research and Forecasting (WRF) meteorological model, Monte Carlo simulation, wavelet transform, and MLP neural networks. In particular, the MLP and wavelet transformation were combined for prediction. WRF meteorological model was employed to predict a set of input parameters of the model including wind speed, wind direction, temperature, rainfall, and temperature inversion. Finally, the Monte Carlo simulation was utilized to achieve a more accurate prediction according to the available uncertainty in the input data.

The paper is organized as follows: Section 2 introduces a brief explanation of the area under study, existing data, and various methods used in this paper. Section 3 provides computational results. Finally, Section 4 reports the concluding remarks.

2. Data and methods

This section reports a description of the area under study, existing data, and the characteristics of the forecasting methods used to predict the $PM_{2.5}$ concentration.

2.1. Study area and available data

Mashhad, in the northeastern part of Iran, is geographically located in the range of longitude $59°15'$ to $60°36'$ and latitude $35°43'$ to $37°8'$. There are 11 monitoring stations in this city to measure the air pollutants including CO, PM_{10}, $PM_{2.5}$, SO_2, and NO_2 on an hourly basis.

The first category of effective parameters is meteorological data. The effects of climate and meteorological factors on urban problems such as air pollution have drawn much attention in recent years [26]. In particular, the meteorological parameters contain wind direction, wind speed, temperature, pressure, rainfall, and humidity, which are taken from the Meteorological Organization of Mashhad, Iran [27]. The meteorological standard for wind direction is as follows: 0°= wind from the north, 90°= wind from the east, 180°= wind from the south, and 270°= wind from the west. The required data are collected eight times per day (every three hours). The speed and direction of the wind are integrated using the following relations [5,13,24]. The obtained results offer exemplary outcomes.

$$\mathrm{wind}_x = w\cos(\varphi),$$

$$\mathrm{wind}_y = w\sin(\varphi), \tag{1}$$

where w and φ are wind speed and wind direction, respectively. Of note, for each involved parameter, the data were collected from March, 2014 to September, 2015.

Another category of parameters is traffic volume. Motor vehicles represent one of the most important sources of air pollution in several cities due to the high levels of pollution emission. The Sydney Coordinated Adaptive Traffic System (SCATS) data were employed to assess the traffic volume of the studied area. To be specific, SCATS uses sensors upon each traffic signal to detect the presence of vehicles on each line [28]. The objective of this study was to present an appropriate and applicable model in reality; to this end, it is recommended that predictable input parameters be used. In particular, the traffic data collected from 10% of the crossroads in Mashhad, Iran in the time span of March 2014 to September 2015 were processed. The results indicated that there was a significant difference in the traffic volume gleaned during the working days and the weekends. Therefore, in our developed model, the parameter corresponding to the traffic volume was considered as a binary parameter. Essentially, it takes 1 and 0 for weekends and working days, respectively.

In certain situations, a layer of cool air on the surface is overlain by a layer of warmer air, and this phenomenon is called temperature inversion. Temperature inversion is an important factor that plays a fundamental role in air pollution. In general, it is divided into two types: subsidence and radiation. Information about this parameter was taken from the University of Wyoming in Laramie, Wyoming [29] and was then analyzed using RAwinsonde OBservation (RAOB) software in terms of the existence and type of the inversion. RAOB Automatically decodes the data derived from more than 100 different formats and plots them on 12 interactive displays including skew-Ts, hodographs, and cross-sections. Then, it produces displays of over 200 atmospheric parameters including inversions, icing, turbulence, wind shear, clouds, and so on [30]. The airport station is located in Shahid Hashemi Nejad international airport, Mashhad, Iran. The mentioned input parameter is incorporated into the model such that the input value takes the value 1 if the inversion type of radiation or subsidence occurs at an altitude below 1200 meters and above the sea level; otherwise, it will be equal to 0.

Time-dependent variables were added to the input data set to reflect the relation between the pollutant concentration and seasonal cycles. To this end, Values 1–4 were attributed to the season from spring to winter, respectively. Finally, as shown in Figure 1, there exists a high correlation between the pollutant concentration on the day previous to prediction and the actual day. As a result, the pollutant concentration with lag 1 is considered as an input of the model. The data of $PM_{2.5}$ concentration during the required period was taken from Environment Pollution Monitoring Center (EPMC), Mashhad, Iran [31].

There could be a day when at least one of the input data is missing. In this case, the corresponding day will not be considered in the final set of data. As a result, the final set of data contains the information about 469 days for which the statistical description is given in Table 1.

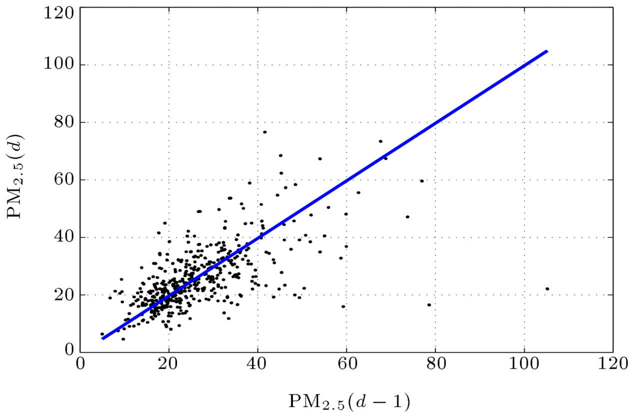

Figure 1. The relationship of $PM_{2.5}$ concentration on the previous $(d-1)$ and the actual (d) days.

Table 2 gives the correlation coefficients of the environmental parameters. The results indicated that there was a high correlation between each pair of parameters: pressure and humidity; temperature and pressure; and temperature and humidity. Figure 2 displays a distribution of available data corresponding to (a) the pressure and humidity, (b) temperature and pressure, and (c) temperature and humidity parameters.

In fact, in case of a correlation between two parameters, only one parameter can be considered as an input. Therefore, the following choices are possible: removal of pressure and humidity; temperature and pressure; temperature and humidity. To select the best scenario, three MLP neural networks were trained. In each network, one of the three mentioned combinations of parameters was removed from the inputs of the model. In addition, 20% of the data were employed

Table 1. Statistical description of the collected input data from March 2014 to September 2015.

Variable	Unit	Range	Mean	St. Dev.
$PM_{2.5}$	$\mu g/m^3$	[4.83,105.06]	26.38	11.61
$Wind_x$	m/s	[–8.16,4.65]	1.03	1.56
$Wind_y$	m/s	[–4.05,6.35]	0.03	1.58
Temperature	°C	[–2.19,34.56]	18.86	9.90
Pressure	hPa	[991.91,1039.28]	1012.83	8.45
Rainfall	mm	[0,4]	0.07	0.28
Humidity	%	[10.25,99.5]	41.34	26.06
Type of day	—	{0,1}	—	—
Season of year	—	{1,2,3,4}	—	—
Temperature inversion	—	{0,1}	—	—

Table 2. The correlation coefficients between different parameters used in prediction of $PM_{2.5}$.

	Wind$_x$	**Wind$_y$**	**Temperature**	**Pressure**	**Rainfall**	**Humidity**	***PM$_{2.5}$***
Wind$_x$	1	–0.398	0.086	0.046	–0.056	–0.022	–0.034
Wind$_y$	–0.398	1	0.401	–0.369	–0.022	–0.325	0.063
Temperature	0.086	0.401	1	–0.887	–0.25	–0.904	–0.003
Pressure	0.046	–0.369	–0.887	1	0.115	0.767	–0.038
Rainfall	–0.056	–0.022	–0.25	0.115	1	0.369	0.107
Humidity	–0.022	–0.325	–0.904	0.767	0.369	1	0.027
PM$_{2.5}$	–0.034	0.063	–0.003	–0.038	0.107	0.027	1

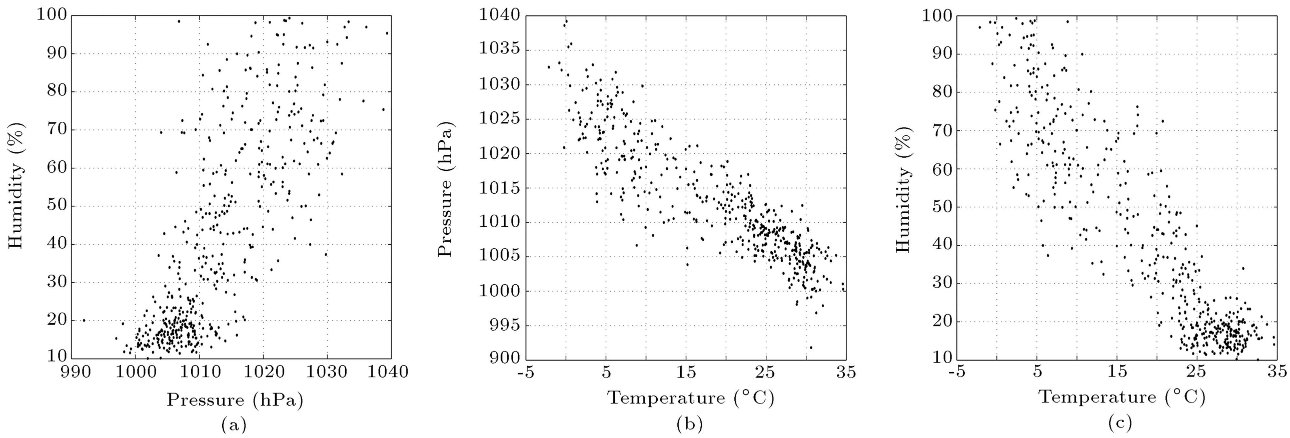

Figure 2. The relation between (a) the humidity and pressure, (b) the pressure and temperature, and (c) humidity and temperature.

to test the three neural networks. The results of the Mean Absolute Error (MAE) measure indicated that the removal of pressure and humidity was the best scenario in generating the minimum average MAE value. Hence, in the rest of the paper, pressure and humidity were removed from the input parameters.

2.2. Solving method

This study proposed two models, i.e., offline and online, to predict $PM_{2.5}$ concentration for one day ahead. While the exact values of $PM_{2.5}$ and meteorological data for prediction were utilized in the offline model, the predicted values for the uncertain data were employed in the online model. The details are provided in the following.

The main objective of the developed offline model was to provide a tool to predict the average value of $PM_{2.5}$ concentration for one day ahead. To this end, a data set including the $PM_{2.5}$ concentration, temporal variables, and meteorological data including temperature, rainfall, wind, and temperature inversion was collected on the day of prediction (see Section 2.1). In the developed model, first, a model was established based on a combination of the wavelet transform and MLP neural network. To this end, the time series of the pollutant concentration with high variability was decomposed into several sub-series with low variability using discrete wavelet transformation (see Section 2.2.2). Following this step, the MLP was applied at each decomposition level, and the results at each level were summed up to determine the final concentration of the pollutants in every time series (see Section 2.2.2).

In the online model, first, the WRF meteorological model was used to predict the meteorological input parameters (see Section 2.2.3). Uncertainty associated with the model could result from different corresponding input data in terms of its accuracy and quality. Among the eight input parameters, three of them are not subject to uncertainty including day, season, and pollutant concentration on the day previous to the prediction. However, other parameters, namely temperature, wind in two directions, rainfall, and temperature inversion, cannot be predicted with certainty. Therefore, the uncertainty attributed to the input data cannot be eliminated, but can only be reduced due to the predicted input parameters using the Monte Carlo simulation (see Section 2.2.4). This prediction can be obtained through the trained neural networks in each of the decomposition levels, and the

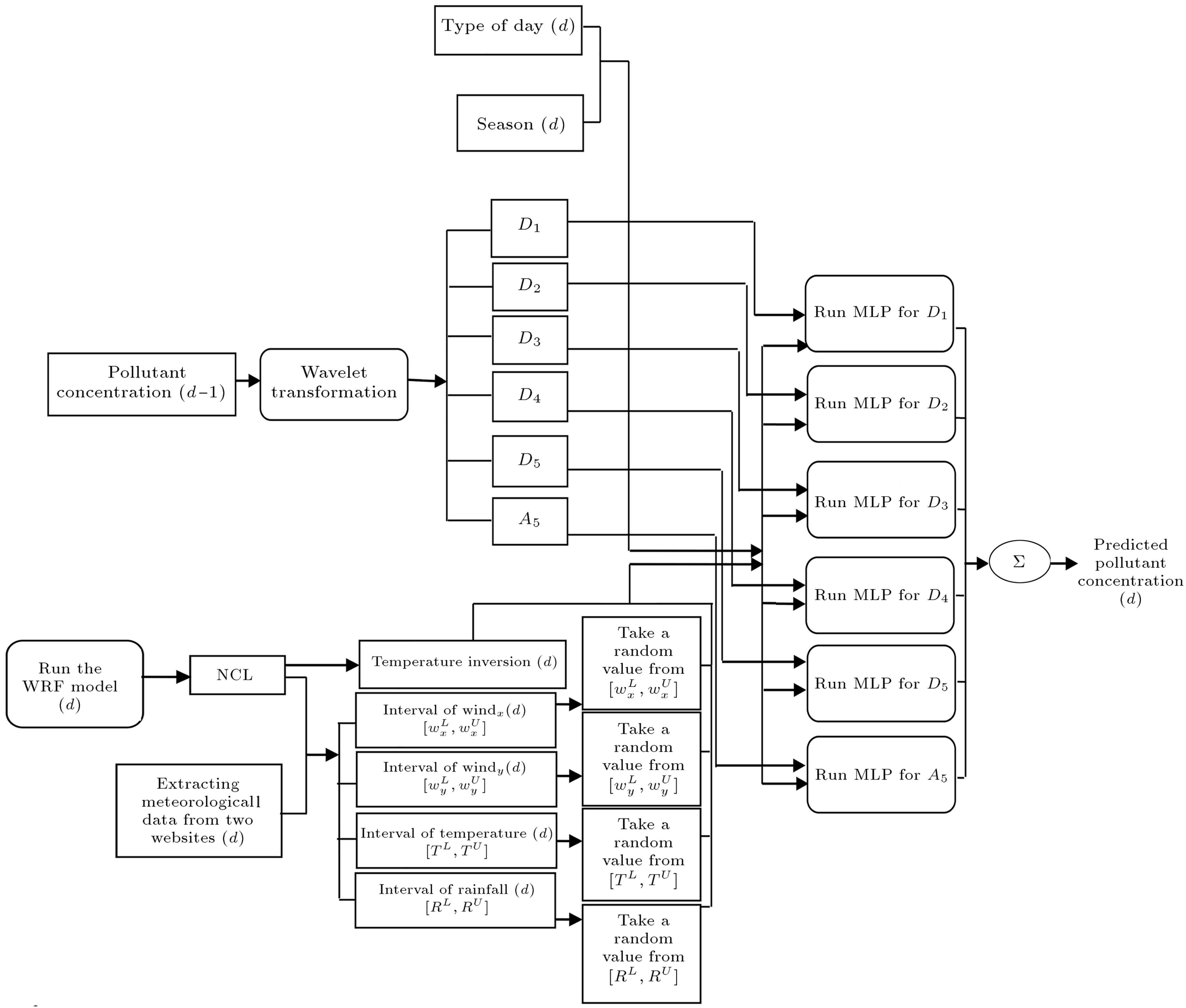

Figure 3. The general overview of the proposed model.

results of all predictions at all levels are finally summed up and the $PM_{2.5}$ concentration is predicted one the day ahead.

The general overview of the proposed model in online mode is depicted in Figure 3. In the following, more explanation about each of the involved tools corresponding to the developed algorithm is given.

2.2.1. Neural network

The MLP neural network was utilized in the proposed model. The activation functions in hidden and output layers are logistic and linear functions, respectively. In addition, Levenberge-Marquardt (LM) learning algorithm for training the network was employed. Moreover, the early stopping method was used to avoid overfitting the model with the training data. In the early stopping technique, the error in the validation data set in the training process was taken into account to monitor the training procedure. The training procedure stops when the validation error for a certain number of iterations increases. The weights and biases are adjusted once the error is minimized. To determine the optimal topology of the neural network, i.e., the number of hidden layers and neurons in each layer, all network topologies were tested using the validation data and the best scenario was finally extracted. In particular, this model comprised a maximum of two hidden layers, each containing a maximum of 20 neurons.

2.2.2. Discrete wavelet transformation

Due to the high variability of the time series of $PM_{2.5}$ concentration, it is quite difficult to ensure an accurate

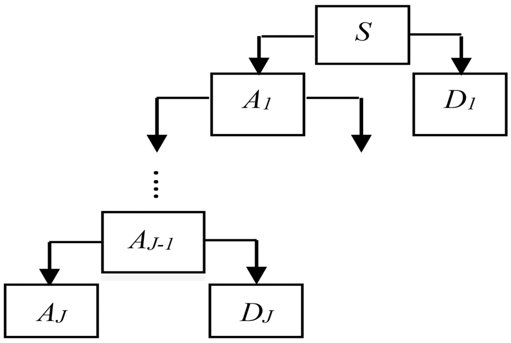

Figure 4. Details and approximations achieved from discrete wavelet transform at J levels.

prediction. The solution to this problem involves decomposing the time series with high variability into sub-series with lower variability. Then, the prediction strategy can be applied to each level with lower variability. Finally, the results of all levels are summed up. To this end, discrete wavelet transformation is used in this study.

In the composition process, the original time series is decomposed into detailed coefficients D_j at different levels ($j = 1, 2, \cdots, J$), and an approximate residual coefficient A_J through high pass and low pass filters [32,33]. In this approach, the number of points was halved at each level so that the number of points at level j is half of that at level $(j-1)$. Therefore, to reconstruct the primary time series, a reconstruction algorithm was employed [32,33]. Here, the number of points at each level should be equal to the original numbers. Figure 4 shows the approximations and details at J levels. The original signal was reconstructed using the approximation and the details presented in Eq. (2).

Further, Symlets sym6 wavelet was utilized since it could provide the lowest variability at each decomposition level. To determine the appropriate number of levels, all possible scenarios were taken into account. Since the number of points at each level was halved in the decomposition approach, the maximum number of levels was $\log_2 N$ in which N was the amount of data. According to 469 available data sets, the maximum number of levels was 8. The MLP neural network was employed to test all possible scenarios through a set of test data for each possible level (i.e., $J = 1, \cdots, 8$). The results of the MAE index showed that the five-level scenario was the best choice. Figure 5 presents the results of the five-level wavelet decomposition of the data for $PM_{2.5}$ concentrations where S gives the original time series corresponding to the available date.

$$S(n) = D_1(n) + D_2(n) + \cdots + D_J(n) + A_J(n). \quad (2)$$

To convert the original time series to wavelet representation, a strategy involving the application of the whole data with moving windows using 364 previous items and the next-day values was adopted. To apply the wavelet transformation to the decomposition of $PM_{2.5}$ concentration on the day previous to the prediction in the online mode, we considered 364 previous days and added the actual day to the end of the series.

2.2.3. WRF meteorological model

As already described in the previous sections, this study aims to develop a model for one-day ahead

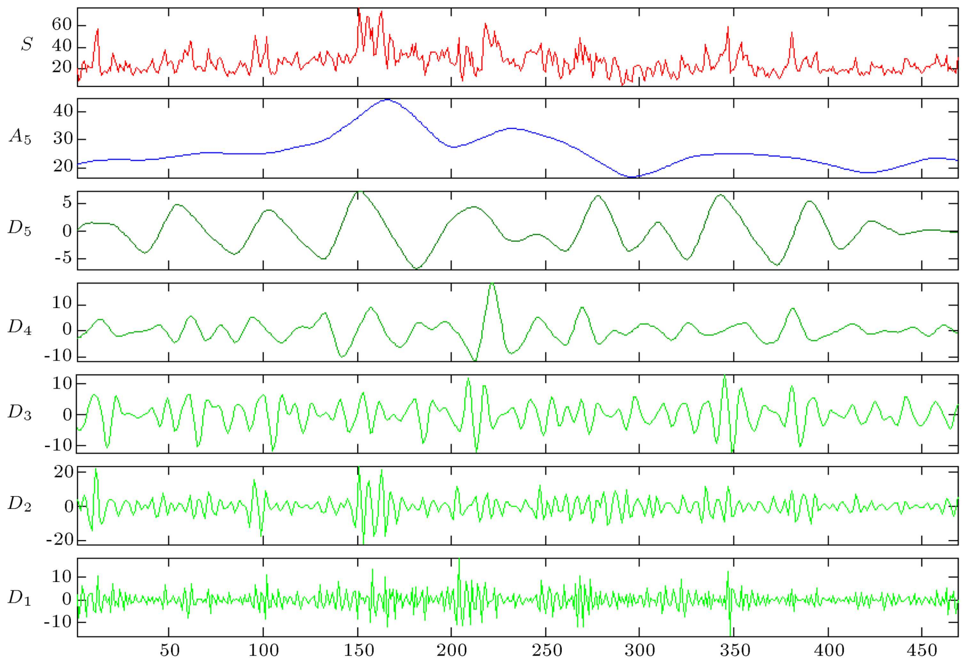

Figure 5. The wavelet decomposition of the original time series Sof $PM_{2.5}$ concentration at 5 levels.

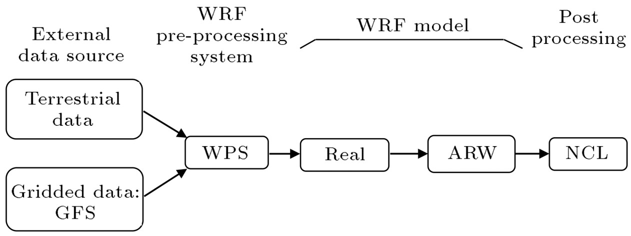

Figure 6. The main structure of the Weather Research and Forecasting (WRF) model.

prediction of $PM_{2.5}$ concentration. To this end, a set of input parameters, including wind in two directions, temperature, rainfall, temperature inversion, type of day, and season of the year, is required for daily prediction and $PM_{2.5}$ concentration for the previous day. It is clear that the actual values for some parameters, including wind speed in two directions, temperature, rainfall, and temperature inversion, are not available for daily prediction; therefore, the predicted values for these parameters should be taken into consideration. To this end, the WRF meteorological model was utilized to predict these parameters for one day ahead. WRF is a state-of-the-art atmospheric modeling system designed for both meteorological research and numerical weather prediction. It offers a host of options for atmospheric processes and can run on a variety of computing platforms. The WRF model configuration is shown in Figure 6. Initially, terrestrial data (such as terrain, land use, and soil types) and gridded data (such as Global Forecast System (GFS) data) are required to run the WRF model. The WRF Preprocessing System (WPS) is a set of three programs whose collective role is to prepare the necessary input for the real program to conduct a real-data simulation. Advanced Research WRF (ARW) solver is the key component of the modeling system, which is composed of several initialization programs for idealized and real-data simulations as well as the numerical integration program. Several programs are supported for post-processing, i.e., NCAR Graphics Command Language (NCL). Graphic Tools facilitate visualization of the model output; hence, NCL was employed in the present work. The details of WRF model are described in the user guide [34]. In this study, to extract the required meteorological data over the Mashhad domain, WRF model is configured with two nested domains with horizontal resolutions of 5 and 15 km and grid points of 100×99 and 85×67, respectively. The domains were efficiently run together using two-way grid nesting in WRF. The vertical structure of the atmosphere was resolved with 30 vertical levels extending up to 5000 Pa. The GFS data of $0.25° \times 0.25°$ and a vertical resolution of 27 pressure levels were used to define the initial and boundary conditions. After running WRF, the NCL post-processing tool was employed to translate the meteorological data from WRF output to the required format. Finally, activated physical schemes in the WRF model are as follows: microphysics, long-wave radiation, shortwave radiation, surface layer, land surface, planetary boundary layer, cumulus parameterization, fluxes, cloud effect, and number of soil layers in the land surface model.

2.2.4. Uncertainty analysis and Monte Carlo simulations

The uncertainty of the model results from the ambiguity of the input data. Among the eight input parameters of the model, the day and season of the year are not subject to uncertainty and variability. However, the other six input parameters, i.e., wind in two directions, temperature, rainfall, $PM_{2.5}$ concentration, and temperature inversion, cannot be predicted with absolute certainty. Therefore, uncertainty in the input data cannot be eliminated from the model. To achieve a more accurate prediction, Monte Carlo simulation was used. Besides, uncertainty associated with measuring the $PM_{2.5}$ concentration was not considered in this study because it could not improve the results of the model. The mechanism of the Monte Carlo method is based on the idea of taking a sample random population and estimating the desired outputs from this sample [35]. The expected value for a function g of the uncertain variables of x with a probability density function of $f_x(x)$ is given by Eq. (3):

$$E(g(x)) = \int_{x \in X} g(x) f_x(x) dx. \tag{3}$$

In particular, x represents the input parameters, including the temperature, wind in two directions or the volume of the rainfall, and $g(x)$ represents the predicted value of pollutant concentration ($PM_{2.5}$) resulting from the proposed model. By taking n samples of x, i.e., $(x_1, x_2, \cdots, x_n)$ into consideration, the Monte Carlo estimation can be achieved using Eq. (4):

$$\tilde{g}_n(x) = \frac{1}{n} \sum_{i=1}^{n} g(x_i). \tag{4}$$

It was assumed that each uncertain parameter had a uniform distribution over a given interval. In particular, for each uncertain parameter, the lower and upper bounds of this interval were set using minimum and maximum values, respectively, obtained from the WRF model and two other websites, namely www.wunderground.com and www.timeanddate.com. The algorithm uses an iterated algorithm to set the appropriate number of samples (n). At the first step, the model is run 10 times and the average values for the output are considered as the predicted $PM_{2.5}$

concentration. Essentially, in a run of the model for each uncertain parameter, the algorithm starts by taking a random sample from the corresponding interval to be used as the input parameter. Then, the number of samples is doubled (i.e., 2*10), and the algorithm is run again. This process continues as long as the mean difference in the $PM_{2.5}$ concentration for two consecutive iterations is greater than 0.01.

3. Results and discussion

In this section, the effectiveness of the proposed model is evaluated. As mentioned earlier, the MLP neural network should be applied at each decomposition level. To this end, 20% of the data (see Section 2.1) were randomly selected to be used as the testing set. Furthermore, the testing data were uniformly taken from different seasons of the year. The remaining data were considered as the training and validation sets. The MLP model was run 100 times and in each run, all the training and validation data sets were randomly selected. Since the initial weights of the MLP were randomly set, the network was trained 10 times per run to select the best network. Finally, the network with the least mean square error over the validation data was selected as the working one. The application of the early stopping criterion [36] in the neural network ensured obtaining similar results for both training and testing data sets.

To evaluate the results, two following approaches are utilized:

- Different measures including the MAE, RMSE, correlation coefficient (R), and Index of Agreement (IA), all reported in Eqs. (5)–(8), respectively, were employed:

$$MAE = \frac{1}{n}\left(\sum_{i=1}^{n} |a_i - p_i|\right), \tag{5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} |a_i - p_i|^2}, \tag{6}$$

$$R = \frac{R_{pa}}{std(p)std(a)}, \tag{7}$$

$$IA = 1 - \frac{\sum_{i=1}^{n} (a_i - p_i)^2}{\sum_{i=1}^{n} \left(|a_i - \overline{a}| + |p_i - \overline{a}|\right)^2}. \tag{8}$$

In these relations, n is the number of instances; p_i and a_i represent the predicted and observed values, respectively. In addition, $\overline{a}$ is the average of the observed data, std the standard deviation, and R_{pa} the covariance value between the predicted and observed data sets.

- The error is usually defined in terms of the *band error*. This criterion gives the difference between the observed and predicted intervals where the observed and predicted values are reduced. The range of pollution values is normally divided into five equal intervals. In this paper, the bands for $PM_{2.5}$ concentration are [0–22], [23–44], [45–66], [67–88], and [89–110]. According to the data, the number of $PM_{2.5}$ concentration in each band is 238, 234, 28, 6, and 1, respectively. For example, the band error in the measured and predicted pair (21, 24) is reported as +1, since 21 and 24 fall into the first and second intervals, respectively. The same measure is used in [5,37,38].

The proposed model was evaluated using the MAE, RMSE, R, IA, and *band error* criteria, the results of which are reported in Tables 3 and 4, respectively. The results shown in Table 3 revealed that a combination of wavelet transformation and MLP (MLP+W) could significantly improve the performance of that achieved by the MLP.

Table 4 reports the output of the *band error* criterion. In this table, the numbers in parentheses give the error percentage, while the numbers before the parentheses indicate the number of samples. The results also pointed to the superiority of the MLP+W model to the MLP model due to its lower error. In particular, the errors for MLP and MLP+W models were 30.85% and 27.66%, respectively, and the mean error was only ±1 band.

The scatter diagram of the results obtained by the MLP+W model compared to the measured data is plotted in Figure 7. In addition, Figure 8 graphically illustrates the results of forecasting the test data obtained by the MLP+W model. A comparison between

Table 3. The results (MAE, RMSE, R, IA) of models on the testing data.

	MLP	**MLP+W**
MAE	6.090	5.397
RMSE	8.216	7.060
R	0.661	0.749
IA	0.806	0.850

Table 4. The results (band error) of models on the testing data.

	MLP	**MLP+W**
±1 band	29 (30.85)	26 (27.66)
±2 band	0 (0.0)	0 (0.0)
±3 band	0 (0.0)	0 (0.0)
±4 band	0 (0.0)	0 (0.0)
Total	**29 (30.85)**	**26 (27.66)**

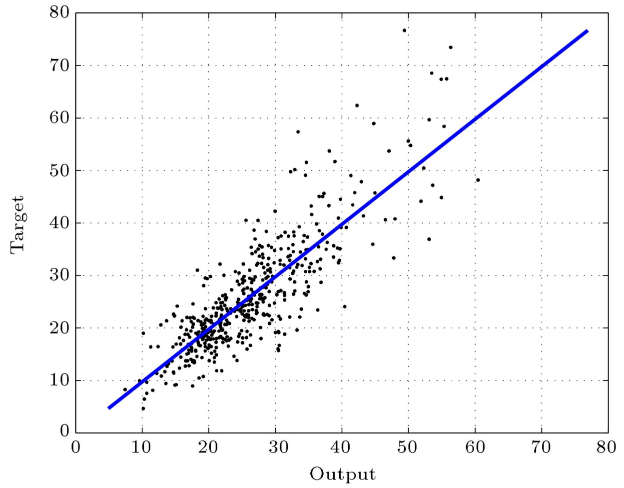

Figure 7. The scatter diagram of $PM_{2.5}$ concentration obtained by MLP+W and measured data.

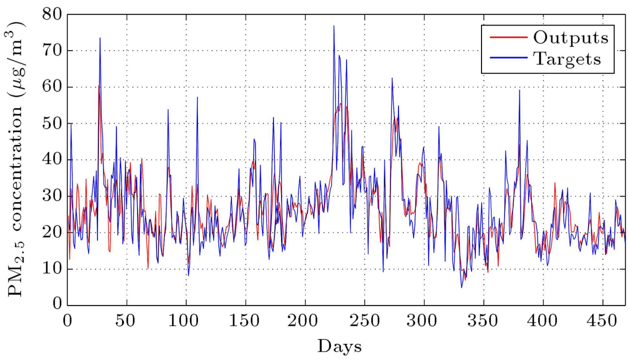

Figure 8. The results of prediction of the testing data obtained by MLP+W: the predicted (red line) and measured (blue line) values of $PM_{2.5}$ concentration.

MLP+W and MLP models demonstrated that some of the high peaks missed by the MLP model were almost anticipated by the hybrid model.

To determine the distribution of the prediction errors, the histogram of the errors for the results in the MLP+W model is shown in Figure 9. In this diagram, the horizontal and vertical axes represent the error value and number of samples, respectively.

The results of the proposed model in the online mode for day-ahead prediction of 19 days in May and 16 days in November 2016 are shown in Tables 5 and 6, respectively. According to Table 5, the proposed model in the online mode achieved the highest accuracy. According to Table 6, only 13-day prediction (from 35 days) is not classified in the desired category. However, in the 22 remaining days, the predicted values are in agreement with the reality. The scatter diagram of the prediction data resulting from the proposed model in the online mode and the measured data is shown in Figure 10.

Table 5. The results (MAE, RMSE, R, IA) of the proposed model in online mode.

	Online model
MAE	5.323
RMSE	6.783
$\boldsymbol{R}$	0.831
IA	0.894

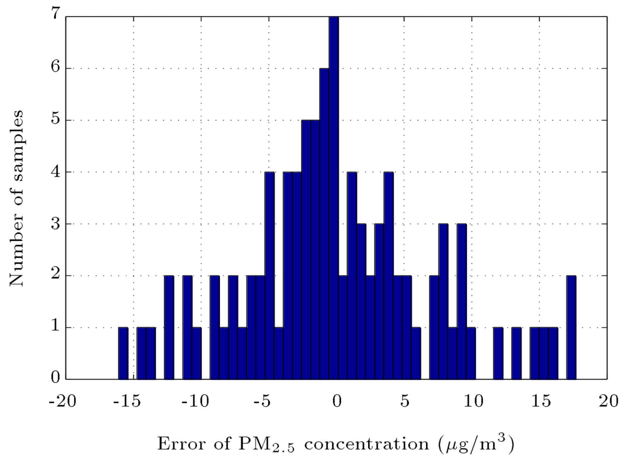

Figure 9. The histogram of errors for $PM_{2.5}$ concentration obtained by the MLP+W model.

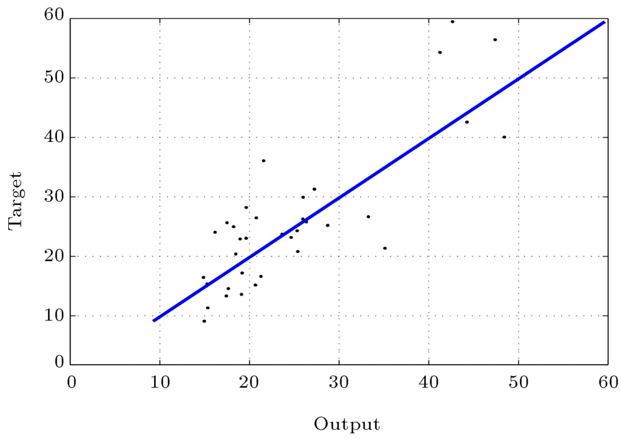

Figure 10. The scatter diagram of $PM_{2.5}$ concentration obtained by the proposed model and the measured data in the online mode.

Table 6. The results (band error) of the proposed model in online mode.

	Online model
±1 band	13 (37.14)
±2 band	0 (0.0)
±3 band	0 (0.0)
±4 band	0 (0.0)
Total	**13 (37.14)**

4. Conclusions

The present study proposed a hybrid model for one-day ahead prediction of the daily average concentration of particulate matter smaller than 2.5 micrometers in online mode. This model was established based on the Weather Research and Forecasting (WRF) meteorological model, Monte Carlo simulation, wavelet transform, and Multi-Layer Perceptron (MLP) neural network. In this model, the decomposition of the time series of pollutant concentration into several levels with lower variability could increase the accuracy of the final forecast. Moreover, the proposed model did not need comprehensive information about air pollutants and pollution sources. In addition, the proposed model was capable of considering the nonlinear relation among various parameters. One of the main advantages of the proposed model was its proper implementation in the real world since it utilized predictable input parameters.

To evaluate the performance of the model in real world, the proposed model was run in online mode for 35 days. In this regard, the values for such parameters as temperature, rainfall, wind in two directions, and temperature inversion were obtained using the WRF meteorological model for one day ahead. Since these parameters could not be predicted with specific certainty, Monte Carlo simulation was employed to improve the accuracy of the prediction. The results of the prediction of the period under study in the online mode pointed to the good performance of the proposed model.

References

1. Molina, M.J. and Molina, L.T. "Megacities and atmospheric pollution", *Air. Waste. Manag. Assoc.*, **54**, pp. 644–680 (2004).
2. Kolehmainen, M., Martikainen, H., and Ruuskanen, J. "Neural networks and periodic components used in air quality forecasting", *Atmos. Environ.*, **35**, pp. 815–825 (2001).
3. Zhang, Y., Bocquet, M., Mallet, V., et al. "Real-time air quality forecasting, part I: History, techniques, and current status", *Atmos. Environ.*, **60**, pp. 632–655 (2012).
4. U.S. EPA "Guidelines for developing an air quality (Ozone and $PM_{2.5}$) forecasting program", U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, North Carolina. EPA-456/R-03-002 (2003).
5. Feng, X., Li, Q., Zhu, Y., et al. "Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation", *Atmos. Environ.*, **107**, pp. 118–128 (2015).
6. Hrust, L., Klaić, Z.B., Križan, J., et al. "Neural network forecasting of air pollutants hourly concentrations using optimised temporal averages of meteorological variables and pollutant concentrations", *Atmos. Environ.*, **43**, pp. 5588–5596 (2009).
7. Perez, P. and Gramsch, E. "Forecasting hourly $PM_{2.5}$ in Santiago de Chile with emphasis on night episodes", *Atmos. Environ.*, **124**, pp. 22–27 (2016).
8. Donnelly, A., Naughton, O., Broderick, B., et al. "Short-term forecasting of nitrogen dioxide (NO_2) levels using a hybrid statistical and air mass history modelling approach", *Environ. Model. Assess.*, **22**, pp. 231–241 (2016).
9. Fernando, H.J.S., Mammarella, M.C., Grandoni, G., et al. "Forecasting PM_{10} in metropolitan areas: efficacy of neural networks", *Environ. Pollut.*, **163**, pp. 62–67 (2012).
10. Genc, D.D., Yesilyurt, C., and Tuncel, G. "Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere", *Environ. Monit Assess.*, **166**, pp. 11–27 (2010).
11. Vlachogianni, A., Kassomenos, P., Karppinen, A., et al. "Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM_{10} in Athens and Helsinki", *Sci. Total. Environ.*, **409**, pp. 1559–1571 (2011).
12. Perez, P. and Reyes, J. "An integrated neural network model for PM_{10} forecasting", *Atmos. Environ.*, **40**, pp. 2845–2851 (2006).
13. Arhami, M., Kamali, N., and Rajabi, M.M. "Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations", *Environ. Sci. Pollut. Res.*, **20**, pp. 4777–4789 (2013).
14. Osowski, S. and Garanty, K. "Forecasting of the daily meteorological pollution using wavelets and support vector machine", *Eng. Appl. Artif. Intell.*, **20**, pp. 745–755 (2007).
15. Luna, A.S., Paredes, M.L.L., de Oliveira, G.C.G., et al. "Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil", *Atmos. Environ.*, **98**, pp. 98–104 (2014).
16. Brunelli, U., Piazza, V., Pignato, L., et al. "Two-days ahead prediction of daily maximum concentrations of SO_2, O_3, PM_{10}, NO_2, CO in the urban area of Palermo, Italy", *Atmos. Environ.*, **41**, pp. 2967–2995 (2007).
17. Boznar, M., Lesjak, M., and Mlakar, P. "A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain", *Atmos. Environ.*, **27B**(2), pp. 221–230 (1993).
18. Perez, P. and Reyes, J. "Prediction of maximum of 24-h average of PM_{10} concentrations 30h in advance in Santiago, Chile", *Atmos. Environ.*, **36**, pp. 4555–4561 (2002).

19. Jiang, D., Zhang, Y., Hu, X., et al. "Progress in developing an ANN model for air pollution index forecast", *Atmos. Environ.*, **38**, pp. 7055–7064 (2004).

20. Lu, H.C., Hsieh, J.C., and Chang, T.S. "Prediction of daily maximum ozone concentrations from meteorological conditions using a two-stage neural network", *Atmos. Res.*, **81**, pp. 124–139 (2006).

21. Paschalidou, A.K., Karakitsios, S., Kleanthous, S., et al. "Forecasting hourly PM_{10} concentration in Cyprus through artificial neural networks and multiple regression models: implications to local environmental management", *Environ. Sci. Pollut. Res.*, **18**, pp. 316–327 (2011).

22. Wang, P., Liu, Y., Qin, Z., et al. "A novel hybrid forecasting model for PM10 and SO_2 daily concentrations", *Sci. Total. Environ.*, **505**, pp. 1202–1212 (2015).

23. Gao, M., Yin, L., and Ning, J. "Artificial neural network model for ozone concentration estimation and Monte Carlo analysis", *Atmos. Environ.*, **184**, pp. 129–139 (2018).

24. Siwek, K. and Osowski, S. "Improving the accuracy of prediction of PM_{10} pollution by the wavelet transformation and an ensemble of neural predictors", *Eng. Appl. Artif. Intell.*, **25**, pp. 1246–1258 (2012).

25. Bai, Y., Li, Y., Wang, X., et al. "Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions", *Atmos. Pollu. Res.*, **7**, pp. 557–566 (2016).

26. Bidokhti, A.A., Shariepour, Z., and Sehatkashani, S. "Some resilient aspects of urban areas to air pollution and climate change, case study: Tehran, Iran", *Scientia Iranica, A*, **23**(5), pp. 1994–2004 (2016).

27. http://www.razavimet.ir

28. http://www.scats.com.au

29. http://www.weather.uwyo.edu

30. http://www.raob.com

31. http://www.epmc.mashhad.ir

32. Misiti, M., Misiti, Y., Oppenheim, G., et al., *Wavelet Toolbox*, MathWorks, Natick (1996).

33. Mallat, S. "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Transactions PAMI*, **11**, pp. 674–693 (1989).

34. User's Guide for the Advanced Research WRF (ARW) Modeling System Version 3.9. Available online: http://www.mmm.ucar.edu/wrf/users/docs/user_guide_V3/contents.html (accessed 2017).

35. Bouloukza, I., Mourad, M., Medoued, A., et al. "Multi-objective optimization design and performance evaluation of slotted Halbach PMSM using Monte Carlo method", *Scientia Iranica, D*, **25**(3), pp. 1533–1544 (2018).

36. Sarle, W.S. "Stopped training and other remedies for overfitting", *Proceedings of the 27th Symposium on the Interface of Computer Science and Statistics*, pp. 352–360 (1995).

37. Kurt, A. and Oktay, A.B. "Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks", *Expert. Syst. with Appl.*, **37**, pp. 7986–7992 (2010).

38. Domańska, D. and Wojtylak, M. "Explorative forecasting of air pollution", *Atmos. Environ.*, **92**, pp. 19–30 (2014).

Biographies

Yasaman S. Sadabadi received her BSc and MSc degrees in Industrial Engineering from Sadjad University of Technology and Ferdowsi University, Mashhad, Iran in 2013 and 2016, respectively. Her research interests include prediction of pollutant concentration.

Majid Salari received his BSc and MSc degrees in Applied Mathematics from Ferdowsi University of Mashhad, Iran in 2004 and 2006, respectively. In addition, he received his PhD in Automations and Operations Research from University of Bologna, Italy in 2010. In 2010, he joined the Department of Industrial Engineering as an Assistant Professor at Ferdowsi University of Mashhad and became an Associate Professor in 2016. His current research interests include combinatorial optimization, applied operation research, evolutionary computation, heuristic search, logistics, distribution, and vehicle routing.

Reza Esmaili received his PhD degree in Climatology. He has been the Administrative Manager of Environmental Pollution Monitoring Center of Mashhad (second metropolis of Iran) since 2014. He has conducted several scientific researches on urban bio-climatology and tourism climate indices in Iran during his PhD program. At the moment, most of his recent investigations are focused on air pollution due to his executive work experiences. His current research interests include urban environment pollution, air pollution, modelling of air pollution, climate risks, and urban climatology.