Environment Protection Engineering

Vol. 44 2018 No. 3

DOI: 10.5277/epe180311

NADJET DJEBBRI[1], MOUNIRA ROUAINIA[1]

PREDICTION OF INDUSTRIAL POLLUTION BY RADIAL BASIS FUNCTION NETWORKS

Atmospheric pollution has been receiving a significant interest for several decades since industries cause more and more pollution. Thanks to the development of many prediction techniques, scientists and industries are focusing more on pollution prediction. The aim of this work is to predict the two pollutant concentrations (NO_x and CO) in industrial sites by a modified radial basis function (RBF) based neural network. The modification considered the spread parameter *h* of the activation function in the RBF network. In order to get the best network, the variations of this parameter for three cases were considered. In the first case, only pollutants concentrations variables were used, while in the second one, only the meteorological variables were utilized. In the third case, pollutants' concentrations were connected with meteorological variables. Based on calculation errors, the best model that ensures the best monitoring of pollutants concentration could be identified.

1. INTRODUCTION

A continuous increase in industrial pollution and environmental degradation has become a major concern for the international community, leading to greater attention on threats. Many countries have already introduced laws to limit and report emission from a large spectrum of commercial and industrial facilities. The prediction of air pollutants has become an important task for the control and emergency management in the case of pollution incidents. The prediction of industrial pollution is a phenomenon that received a special interest for a very long time. This accelerated the development of the prediction methods and provided scientific data based on these techniques. The artificial neural networks (ANN) based methods are widely used in air quality monitoring which heavily relies on the local meteorological conditions and the concentrations of pollutant.

Radial basis networks can require more neurons than standard feed forward back propagation networks (Fig. 1), but often they can be designed in a fraction of the time it takes to train standard feed forward networks. They work best when many training

[1]University 20 août, 1955 Skikda, Algeria, corresponding author N. Djebbri, e-mail address: n_djebbri@yahoo.fr

vectors are available [1]. Binet et al. [2] presented a radial basis function (RBF) neural network method for estimating $PM_{2.5}$ concentrations based on sparse observed inputs. Liu et al. [3] built an emission prediction model for compressed natural gas (CNG)/diesel dual fuel engine (DFE) based on RBF neural network for analyzing the effect of the main performance parameters on the CO, NO_x emissions of DFE. Kyriaki et al. [4] used a radial basis function neural network system, which was classifying countries based on their emissions of carbon, sulfur and nitrogen oxides, and on their gross national income.

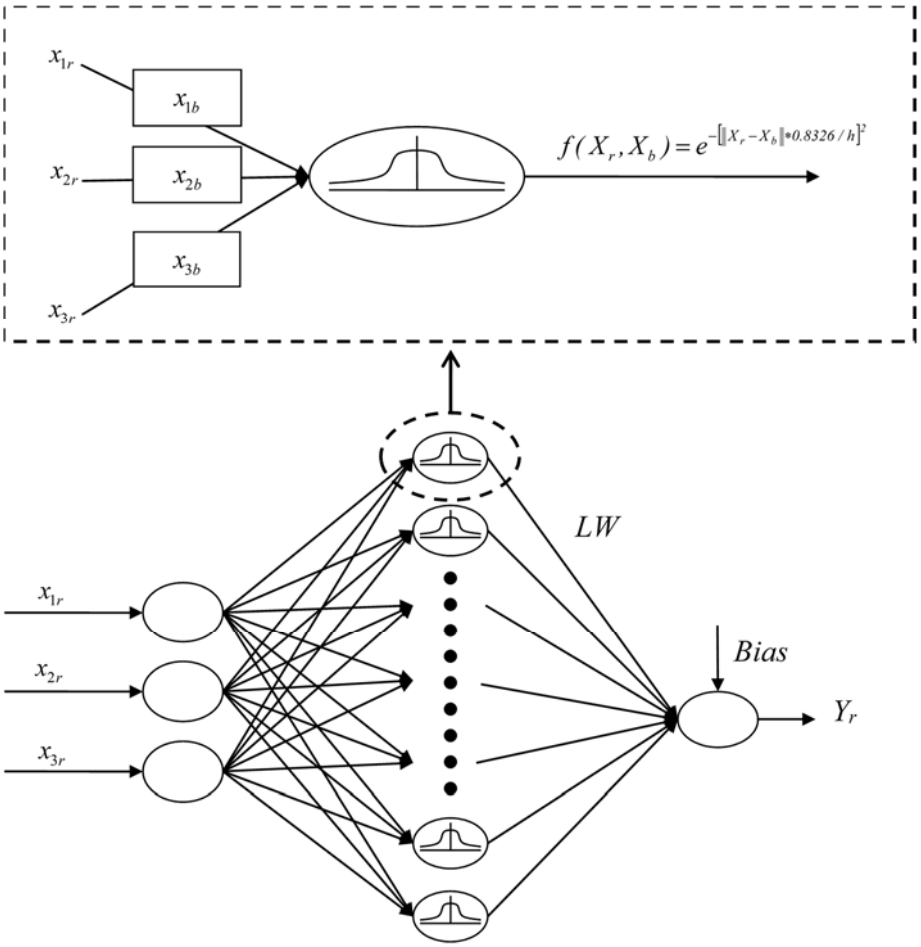

Fig. 1. Schematic of an RBF network

Bo et al. [5] proposed a method for predicting gas content based on the RBF neural network optimized by a genetic algorithm. Shourong et al. [6] combined the RBF neural network with time series on CO_2 emissions to make a forecast of its emissions in China. Chuanbao and Fuwu [7] described an approach for replacing the engine out NO_x sensor

with a radial basis function neural network (RBFNN) based NO_x perception. Zheng and Shang [8] selected the parameters PM_{10}, SO_2, NO_2, temperature, pressure, humidity, wind direction and wind speed as the influence factors, while the prediction models based on RBF neural network were constructed.

The aim of this paper is to predict the NO_x and CO pollutants concentration in industrial sites by a modified radial basis function network (RBF). The modification we considered concerns the spread parameter h of the activation function in the RBF network. The variations of this parameter in three cases were considered in order to get the best network in each case, and then these three cases were tested until getting the best case. These three cases are characterized as follows: in the first case (RBF1) only pollutants concentrations variables were used, while in the second one (RBF2), only the meteorological variables were utilized. In the third case (RBF3), pollutants concentrations were combined with the meteorological variables.

2. RBF NETWORK METHOD

An RBF is a three-layer network, with only one hidden layer (Fig. 1). The number of neurons in the hidden layer is equal to the number of historical observations of predictors (successors). In fact, each neuron in the hidden layer represents a pair of historical observations of predictors/dependents. The output of each neuron is actually the contribution of the historical observation in estimating the real-time event [9].

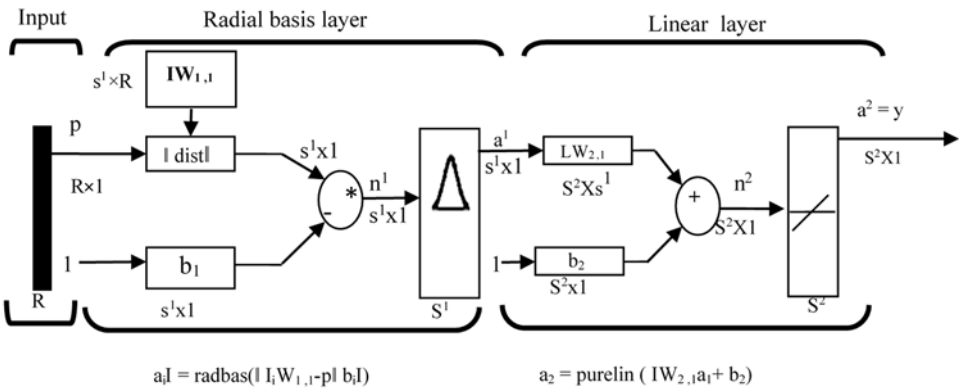

Fig. 2. RBF network architecture

Radial basis network consists of two layers (Fig. 2): a hidden radial basis layer of S^1 neurons and an output linear layer of S^2 neurons, where R is the number of elements in the input vector, a_{il} is ith element of a_1 where $I_iW_{1,1}$ is a vector consisting of the ith row of $IW_{1,1}$.

The ||dist|| box in Fig. 2 accepts the input vector p and the input weight matrix $IW_{1,1}$, and produces a vector having S^1 elements. The elements are the distances between the input vector and vectors $I_iW_{1,1}$ formed from the rows of the input weight matrix [1].

According to Fig. 1, the RBF uses a Gaussian performance function. The input to this function is the Euclidian distance between each input to the neuron and the specified vector of the same size of the input [10, 11]. The Gaussian function uses the following relation:

$$f(X_r,\ b) = \mathrm{e}^{-I^2} \tag{1}$$

$$I = \frac{0.8326\left\|X_r - X_b\right\|}{h} \tag{2}$$

where X_r – the network input with unknown output, X_b – observed inputs in time or location b, and h – spread. The output of the function approaches 0 to 1, when $\left\|X_r - X_b\right\|$ approaches a large value to 0, respectively. The value of the output between those limits depends on h.

The general form of calculating a dependent variable (Y_r) by predictor X_r is then

$$Y_r = LWf(X_r,\ X_b) + \mathrm{Bias} \tag{3}$$

where LW and Bias – weight matrix of connections from the hidden layer to the output layer and bias matrix of the output layer, respectively. When an RBF network is developed, LW and bias matrices are calculated by solving the system of equation of

$$T_b = LWf(X_r,\ b) + \mathrm{Bias} \tag{4}$$

where T_b is the target associated with the bth observation [12].

To determine the best spread parameter for interpolation of concentrations using the RBF network to predic NO_x and CO concentrations a try-and-error approach was used. Various values of spread starting from 0.1 ending to 6 with 0.1 incremental steps were used. It should be noted that since the radial basis network acts as the exact estimator function, the application of different spread values within the calibration set results in a similar averaged error of approximately zero. Therefore, to examine the performance of the network in a practical manner, the approach of cross-validation is used. In this approach, each pair of input/output is omitted from the n observation of data set once and the other $n - 1$ pairs of data are used to estimate the omitted one. This iteration is repeated n times, and the averaged simulation error for all n pairs of data is considered as the indicator of the real performance of the network. The algorithmic steps used to

describe the approach applied to find out which spread (h) minimizes the average error of concentration estimation are shown in Fig. 3.

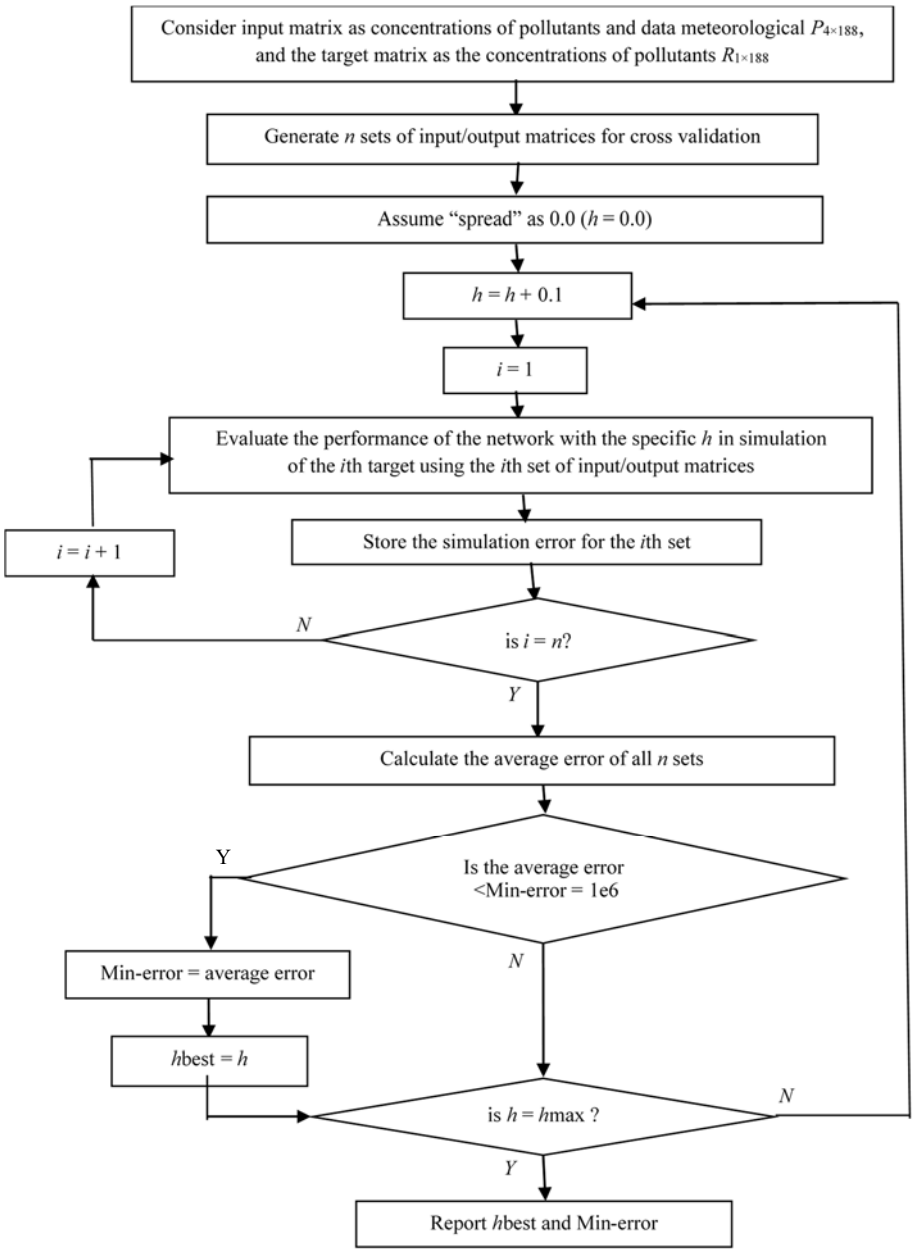

Fig. 3. Algorithm of the suitable spread (h)

The coefficient of determination (R^2) explains how much of the variability in the input data can be explained by the fact that they are related to the observed values or how close the points are to the line. R^2 takes on values between 0 and 1, with values closer to 1 implying a better fit [13–15]. It is given by

$$R = \sqrt{\frac{\sum_{i=1}^{n}\left(O_i - \bar{O}\right)^2 - \sum_{i=1}^{n}\left(O_i - P_i\right)^2}{\sum_{i=1}^{n}\left(O_i - \bar{O}\right)^2}} \tag{5}$$

The mean absolute error (MAE) is the average difference between predicted and actual data values. The MAE (Eq. 6) ranges from 0 to infinity and a perfect fit is obtained when MAE = 0.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|P_i - O_i\right| \tag{6}$$

The mean-squared error (MSE) is one of the most commonly used measures of success for numerical prediction. The smaller the MSE value, the better the performance of the model is [16–18]. Its value is computed by

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(P_i - O_i\right)^2 \tag{7}$$

where P_i and O_i are the predicted and observed concentrations and $\bar{O}$ represent the observation mean.

3. EXPERIMENTAL PROCEDURE

Site and data description. Gas Natural Liquefies (GNL) complex with a surface of 90 ha is located in Skikda industrial area, 6 km to the East from Skikda city center. It was built in 1972, evolved in 1980 and renovated in 2000. Our database taken from the industrial zone (GL1K) relies on the daily measurements from 188 groups of data containing:

- NO_x and CO pollutant concentrations in the SKIKDA area, during the period from October 2015 to April 2016,
- the meteorological variables, for the same area and period; the measured variables are: the speed and direction of the wind, relative humidity and temperature. The inputs and outputs are standardized in the interval of [0; 1] [19].

Method. To monitor the air quality and the prediction of industrial pollution, the RBF network based method considering two pollutants was applied. Our objective is to forecast the concentration of NO_x and CO pollutants. In this study, the best spread parameter h of the Gaussian activation function is selected by determining the min error of RBF model for the three cases RBF1–RBF3. The best spread (h) according to each case was considered to present the prediction of two pollutants (CO and NO_x) and their prediction errors. To identify the best model, the mean squared error (MSE) and the mean absolute error (MAE) were calculated for each of the three phases. The method was implemented in Matlab (7.7 version). The method of RBF neural network was applied according to Figs. 1–3 and the data of industrial site. The models' variables are given in Table 1.

Table 1

Variables of the model

Variable	Description
X_1	temperature
X_2	humidity
X_3	wind speed
X_4	direction of the wind
X_5	concentration of NO_x
X_6	concentrations of CO

4. RESULTS AND DISCUSSION

Figures 4–6 illustrate the changes of the minimal prediction error upon changing the spread h for the cases RBF1–RBF3. Table 2 shows the MAE values for these cases. The results show that the smallest MAE is obtained in the case of RBF3 (0.0534), while the best spread value h_{best} is equal to 5.8. Table 3 shows the best spread values at several intervals for the same case. Therefore, the best choice for spread value is 5.8.

Table 2

Values of h_{best} and MAE
for RBF1, RBF2 and RBF3

Models	h_{best}	MAE
RBF1	6.0000	0.0853
RBF2	1.2000	0.1207
RBF3	5.8000	0.0534

Figures 7–9 illustrate the pollutants prediction and the prediction errors at h_{best} for each case (RBF1, RBF2, and RBF3), while Table 4 shows the mean and the absolute prediction errors at h_{best}.

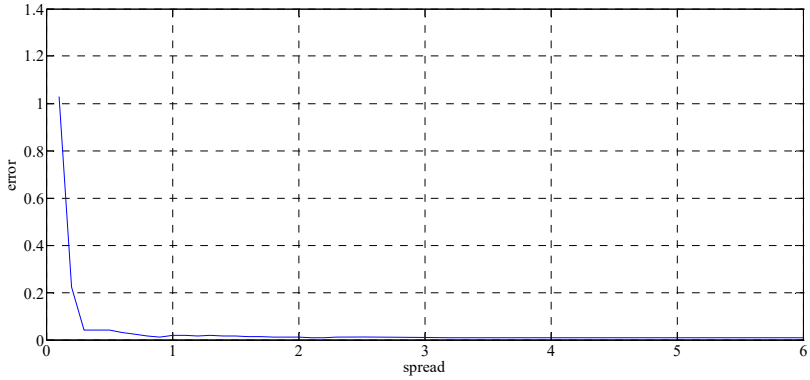

Fig. 4. Prediction error in terms of MAE for various values of the spread parameter h for the RBF1 (using only pollutant concentrations as inputs) for $h \in [0, 6]$

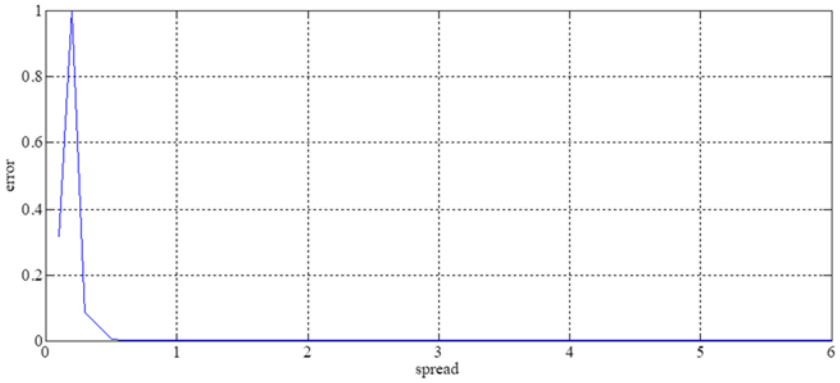

Fig. 5. Prediction error in terms of MAE for various values of the spread parameter h for the RBF2 (using only meteorological variables as inputs) for $h \in [0, 6]$

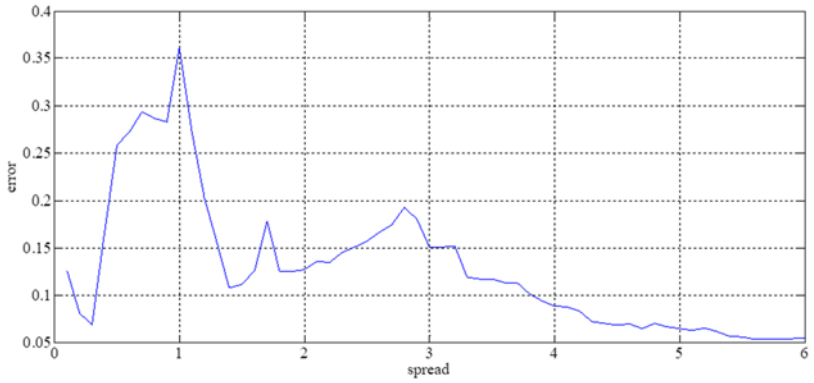

Fig. 6. Prediction error in terms of MAE for various values of the spread parameter h for the RBF3 (using pollutants concentrations combined with the meteorological variables as inputs) for $h \in [0, 6]$

a)

Table 3

Evolution of MAE according to spread at RBF3

Interval spread	Spread best	MAE
[0, 2]	0.8000	0.1128
[0, 4]	0.8000	0.1128
[0, 6]	5.8000	0.0534
[0, 8]	8.0000	0.0651

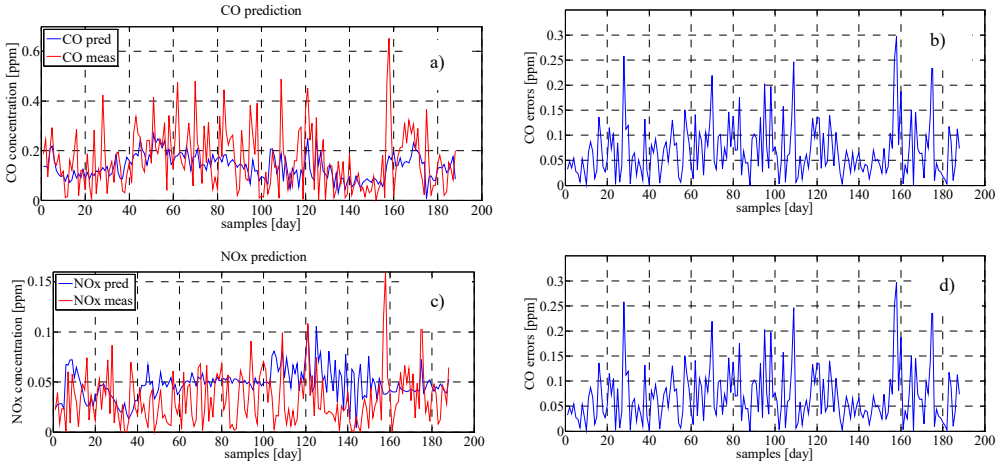

Fig. 7. Pollutants predictions and their errors by RBF1:
a) CO prediction at h_{best}, b) CO prediction errors at h_{best},
c) NO_x prediction at h_{best}, d) NO_x prediction errors at h_{best}

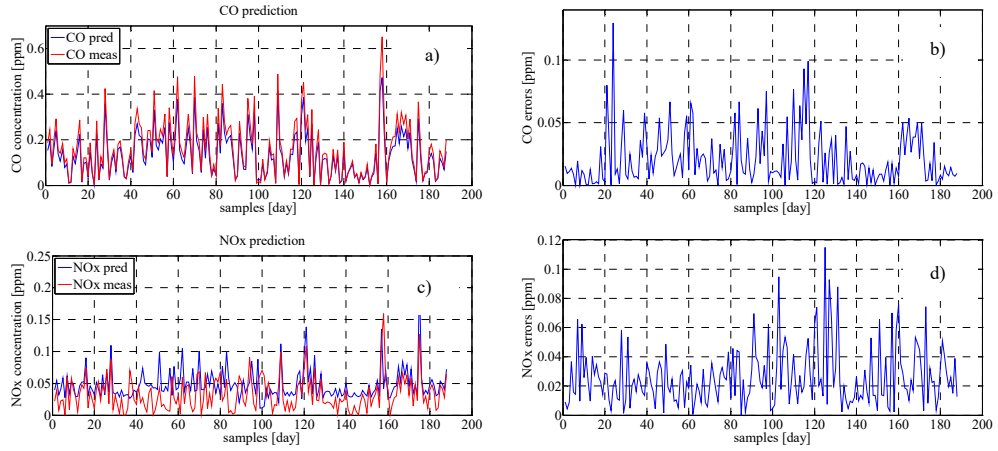

Fig. 8. Pollutants predictions and their errors by RBF2:
a) CO prediction at h_{best}, b) CO prediction errors at h_{best},
c) NO_x prediction at h_{best}, d) NO_x prediction errors at h_{best}

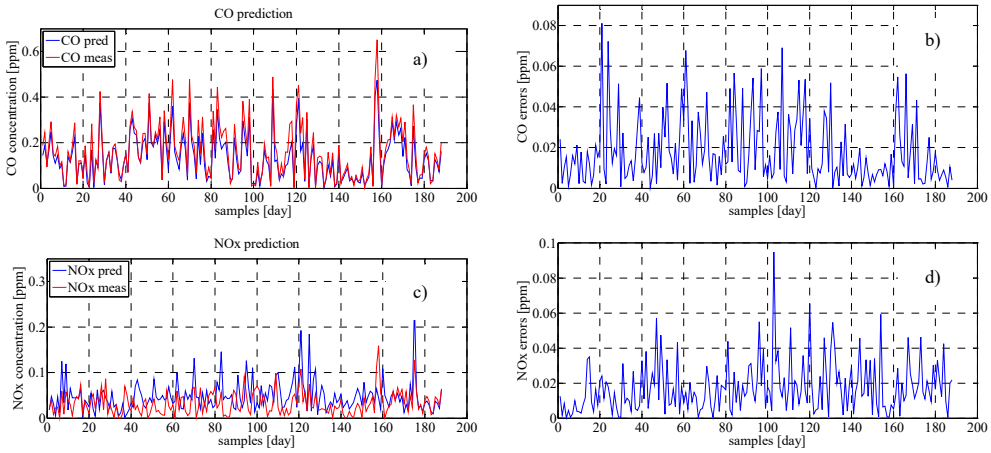

Fig. 9. Pollutants predictions and their errors by RBF3:
a) CO prediction at h_{best}, b) CO prediction errors at h_{best},
c) NO_x prediction at h_{best}, d) NO_x prediction errors at h_{best}

The curves of predicted and measured CO and NO_x concentrations are divergent in Figs. 7 and 8, while they are convergent in Fig. 9. The prediction errors for CO and NO_x are smaller in Fig. 9 than those in Figs. 7 and 8. We can easily conclude that the RBF3 model illustrated in Fig. 9 presents better performances than RBF1 and RBF2 models for prediction NO_x and CO.

Table 4

Mean squared errors and mean absolute errors for RBF1, RBF2 and RBF3

Model	NO_x concentration [ppm]			CO concentration [ppm]		
	MSE	MAE	R^2	MSE	MAE	R^2
RBF1	0.0014	0.0293	0.5398	0.0079	0.0695	0.2374
RBF2	0.0011	0.0261	0.3464	$9.0063.10^{-4}$	0.0209	0.9127
RBF3	$5.2243 \cdot 10^{-4}$	0.0174	0.5398	$6.0510 \cdot 10^{-4}$	0.0180	0.9349

All these results show the efficiency of the RBF3 based pollution prediction model and its accuracy comparing to the RBF1 and RBF2 based prediction. We can deduce from Table 4 that the values of the MSE and MAE for the two pollutants in the predictive model RBF3 are smaller than their values in predictive models RBF1 and RBF2 also closer to 0. The values of coefficient of determination (R^2) in (RBF3) are better than its values in the case of the RBF1 and RBF2 also close to 1. It can be concluded that the minimal MSE and MAE for the two pollutants are found in RBF3. Thus, using pollutant concentrations combined with meteorological variables leads to a better prediction.

5. CONCLUSION

The aim of this paper was to forecast the NO_x and CO pollutants concentrations by the use of the forward-forward retro propagation ANN model with a radial basis function (RBF). This particular network has a Gaussian activation function which changes with the variations of the spread parameter h. This change may alter the RBF network.

The test was performed for three various networks. The modification we considered concerns the spread parameter h of the activation function in the RBF network. The variation of this parameter in three cases was considered in order to get the best network in each case, and then these three cases were tested until getting the best case. The results show that the best spread and the best prediction are obtained in the third case (RBF3). Thus, it can be concluded that an efficient prediction of NO_x and CO concentrations and forecast will be performed by the use of the radial basis function neural network with a fixed value of the spread parameter h of 5.8, and a data set containing pollutants with previous concentrations combined with meteorological variables.

REFERENCES

[1] CHEN S., COWAN C.F.N., GRANT P.M., *Orthogonal least squares learning algorithm for radial basis function networks*, IEEE Trans. Neural Networks, 1991, 2 (2), 302.

[2] BIN Z., MIN W., NENG W., GAINES W.J., XIN F., YUQI T., *Spatial modeling of PM2.5 concentrations with a multifactorial radial basis function neural network*, Environ. Sci. Pollut., 2015, 3 (22), 10395.

[3] LIU Z., FEI S., *Study of CNG/diesel dual fuel engine's emissions by means of RBF neural network*, J. Zhejiang University Science, 2004, 5 (8), 960.

[4] KYRIAKI K., LAZAROS I., *Employing a radial basis function artificial neural network to classify western and transition european economies based on the emissions of air pollutants and on their income*, International Federation for Information Processing EANN/AIAI, 2011, 2 (364), 141.

[5] BO Z., JIANFENG S., *Gas content prediction based on GA-RBF neural network*, Chinese Control and Decision Conference, Xuzhou, China, 2010, 978 (1), 3104,

[6] SHOURONG L., RONGXI Z., XIN M., *The forecast of CO_2 emissions in China based on RBF neural networks*, 2nd International Conference on Industrial and Information Systems, Dalian, China, 2010.

[7] CHUANBAO L., FUWU Y., *Radical basis function neural network-based NO_x soft sensor technique*, International Conference on Electrical and Control Engineering, Yichang, China, 2011.

[8] Zheng H., Shang X., *Study on prediction of atmospheric PM2.5 based on RBF neural network*, Fourth International Conference on Digital Manufacturing and Automation, Qingdao, China, 2013.

[9] SHAHAB A., *Data-Driven Modeling. Using MATLAB® in Water Resources and Environmental Engineering*, Springer, Dordrecht 2014.

[10] YIRAN S., DING L.Y., YANTAO T., YAOWU S., *Air:fuel ratio prediction and NMPC for SIengines with modified Volterra model and RBF network*, Eng. Appl. Artif. Int., 2015, 45, 313.

[11] PIETRO Z., HAIBO C., MARGARET C.B., *Predicting real-time roadside CO and NO_2 concentrations using neural networks*, IEEE Trans. Int. Transp. Syst., 2008, 9 (3), 514.

[12] SURAJDEEN A.I., MOUSTAFA E., MOUHAMED A.H., AHMED A.A., *RBF neural network inferential sensor for process emission monitoring*, Control Eng. Pract., 2013, 21, 962.

[13] WILLMOTT C., *Some comments on the evaluation of the model performance*, Bull. Am. Meteor. Soc., 1982, 63 (11), 1309.

[14] WILLMOTT C., ACKLESON S., DAVIS R., FEDDEMA J., KLINK K., LEGATES D., O'DONNELL J., ROWE C., *Statistics for the evaluation and comparison of models*, J. Geophys. Res., 1985, 90 (5), 8995.
[15] ROESON S.M., STEYN D.G., *Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations*, Atm. Environ., 1990, 24 (2), 303.
[16] COMAN A., IONESCU A., CANDAUY., *Hourly ozone prediction for a 24-h horizon using neural networks*, Environ. Model. Soft., 2008, 23 (2008), 1407.
[17] JUNNINENA H., NISKAA H., TUPPURAINENC K., RUUSKANENA J., KOLEHMAINENA M., *Methods for imputation of missing values in air quality data sets*, Atm. Environ., 2004, 38, 2895.
[18] KOLEHMAINEN M., MARTIKAINEN H., RUUSKANEN J., *Neural networks and periodic components used in air quality forecasting*, Atm. Environ., 2001, 35, 815.
[19] NISSES M., *E-documentations techniques de complexe GL1K-Skikda*, Algeria, 2011.