Application of radial basis functions compared to neural networks to predict air pollution

C. Capilla
Polytechnic University of Valencia, Spain

Abstract

This paper studies the application of radial basis functions to predict nitrogen oxides 24 hours in advance. The forecast interval was chosen for practical regulatory reasons. The two study areas are in Valencia (Spain), where these pollutants have reached critical levels, and there has been a significant connection between them and several health problems. The models use nitrogen oxide concentrations, traffic, meteorological data, and periodic components (sine and cosine terms for the daily and weekly cycles) as hourly inputs. In one monitoring station the most accurate nitric oxide predictions were obtained when the radial basis function model included all these variables as inputs. In this site the forecast evaluation criteria gave better results for nitrogen dioxide prediction than for nitric oxide. In the other monitoring station, better predictions were obtained for nitric oxides than for nitrogen dioxide. There were differences in the forecasts accuracy between sites. The results are compared with the forecasts obtained with multilayer perceptron neural networks. Nitrogen dioxide predictions were more accurate with the multilayer perceptron approach at one of the sites.
Keywords: urban air quality, nitrogen oxides, neural networks, radial basis functions, multilayer perceptron.

1 Introduction

Air quality is a major concern in urban areas. Local administrations manage air pollutants monitoring networks and analyze their information in order to prevent health problems. Indicators of urban air quality are ozone, nitrogen oxides, sulphur dioxide or atmospheric particulates [1, 2]. Pollutants levels predictions are useful

WIT Transactions on Ecology and The Environment, Vol 198, © 2015 WIT Press
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/AIR150041

when evaluating the effectiveness of the plans introduced by the administration to reduce critical events. Tools for forecasting have to model temporal variations and to include complex non-linear relationships between meteorology, traffic and pollutants. The link between climatology and pollutions plays an important role in air quality variability.

Deterministic approaches to predict pollution levels, are not useful in coastal areas [3]. They are more appropriate over extensive areas such as whole regions and large cities. They require precise data from the emission and transportation of pollutants, and meteorological conditions. Statistical approaches are required when the complexity of a problem increases, and the theoretical understanding decreases due to ill-defined interactions between systems.

Statistical models establish relationships between input variables (predictors) and output variables (pollutants levels), without detailing the causes and effects in the formation of pollutants. Seasonality, trends and autocorrelations of pollutants are often analyzed using ARIMA time series models [4, 5]. These methods are limited by their weakness when modeling non-linear temporal variations. The possible presence of chaotic dynamics in pollutant concentrations allows the application of non-linear time series [6]. Classification and regression trees have also been used to study pollutants variability [7, 8].

The application of neural networks has been shown to be an effective alternative to more traditional statistical techniques in the air quality research area [9, 10]. The neural network models can be trained to approximate virtually any smooth, measurable function [11], and they make no prior assumptions concerning the data distribution. They can be trained to accurately generalize when presented with new, unseen data and can model highly non-linear functions [12]. The neural network approach which has been most applied to analyze atmospheric pollutants variability, is the multilayer perceptron (MP) [10]. This method was used by Ibarra-Berastegi *et al.* [13] and the best performance of these models was obtained for the prediction of nitrogen dioxide (NO_2) 1 hour ahead. Caselli *et al.* [14] showed that the multivariate regression models gave less accurate results than MP methods, and failed when fitting spiked high values of pollutant concentrations. They also applied a radial basis function (RBF) network, which could predict the pollutant trend but the mean relative error was higher than with a MP network. Compared with traditional neural networks, the RBF method can not only produce more accurate results, but also achieve simpler network architecture and faster training speed [15]. More recently, Wang *et al.* [16] established a RBF model to estimate the impact of meteorology indicators on sulfur dioxide. The proposed model gave satisfactory results in forecasting the data.

The objective of this study is to investigate for the first time the capability of the RBF approach to predict nitrogen dioxide (NO_2) and nitric oxide (NO) hourly concentrations in Valencia (Spain). The main goal is to estimate levels 24 hours ahead at two different locations. Shorter time forecasts are of minimal value for the air quality management purposes of the local administration. NO_2 and NO are critical air pollutants in Valencia [17]. They are a consequence of motor vehicle emissions [18]. Tenias *et al.* [19] showed a significant connection between a 10 $\mu g/m^3$ increase in NO_2 level and the relative risk of asthma emergency visits in

this city. Daily levels of NO_2 in Valencia are also associated with cardiovascular admissions [20, 21]. NO_2 is precursor of secondary pollutants that are related to photochemical smog and acid rain. Ambient air NO_2 is in large part originated by the oxidation of NO. Therefore the link between climate and these pollutants plays an important role on their variability, and has to be taken into account when selecting optimal pollutant reduction strategies. Recent research [22] showed that MP networks performed better than multiple regression models to forecast NO_2 values. In this paper RBF models are designed and compared with MP networks, to establish the most efficient forecasting tool. Meteorological and traffic variables, pollutants concentrations, and seasonal components are used as predictors.

2 Material and methods

2.1 Study area and dataset

The study is located in the urban area of Valencia (Spain). The climatology and structure are Mediterranean. There are around one million inhabitants. The air pollution monitoring network is managed by the local government since 1995. It measures pollution variables in the whole urban area. Mass concentrations of nitrogen oxides are determined using the chemiluminescence method. Pollutants concentrations are expressed in $\mu g/m^3$. The volumes are standardized at a temperature of 293°K and a pressure of 101.3 kPa. The traffic network of the local municipality obtains the number of vehicles (NV) circulating every hour at locations close to the pollution monitoring sites.

The data set are hourly observations from the air pollution and traffic networks. The study considers two monitoring stations (Pista Silla and Viveros), where high pollution episodes were registered during the period 2002–2005. The limit value of NO_2 for the protection of human health in a calendar year was exceeded at Pista Silla, in 2003, 2004 and 2005. The highest annual NO_2 mean was observed in 2003. This station also measures wind speed (WS, m/s), wind direction (WD, degrees), temperature (T, °C), solar radiation (SR, W/m^2), relative humidity (RH, %) and pressure (P, mbar). At Viveros, WS, WD, T and SR observations were provided by the National Institute of Meteorology, which manages a meteorological station close to the air pollution station. Pista Silla station is in a roadside site located a few meters from a motorway, and Viveros is in an avenue close to the city centre. The distance between them is 2.6 km. Traffic density is high at both sites. The matrix of data (hourly measurements) had 18339 entries for Pista Silla (years 2003–2005) and 16221 for Viveros (years 2002–2004). Table 1 contains averages, coefficients of variation and maximum values of pollutants, meteorological and traffic variables.

The source activity (e.g. traffic) and periodic variations in nature (e.g. photochemical reactions in the atmosphere), mainly contribute to NO_2 and NO concentrations. Periodic components are observed in the time series at the week and daily levels [22] at the two stations.

Table 1: Descriptive analysis of the variables.

Station	Variable	Mean	CV	Maximum
Pista Silla	NO_2	58.8	0.51	249
	NO	52.0	1.14	624
	WS	1.1	0.82	8.6
	WD	187.8	0.57	360
	T	18.7	0.36	38.2
	RH	60.8	0.25	92
	P	1022.2	0.01	1044.7
	SR	153.3	1.60	947
	NV	2945.7	0.57	38712
Viveros	NO_2	36.72	0.67	238
	NO	19.6	1.93	596
	WS	1.8	0.72	11.9
	WD	168.3	0.69	360
	T	18.9	0.34	38.2
	SR	170	1.53	1033.3
	NV	1088.7	0.66	13456

2.2 Neural networks

The neural networks are composed of three layers of neurons: the input, the hidden and the output layers. They are applied to eight models which have different number I of predictors X_i or neurons in the input layer. The predictors are pollutants concentrations, meteorological parameters, traffic variable or seasonal components (sine and cosine terms for the daily and weekly cycles). The models output Y is the prediction of NO_2 or NO concentrations 24 hours in advance; therefore the number of neurons in the output layer is equal to 1. Table 2 shows the models that are analyzed.

Table 2: Models analyzed.

Model	Output variable	Input variables
1	$(NO_2)_{t+24}$	$Meteorology_t$, $traffic_t$, $(NO_2)_t$
2	$(NO_2)_{t+24}$	$Meteorology_t$, $traffic_t$, $Seasonality_{t+24}$,$(NO_2)_t$
3	$(NO_2)_{t+24}$	$Meteorology_t$, $traffic_t$, $(NO_2)_t$, NO_t
4	$(NO_2)_{t+24}$	$Meteorology_t$, $traffic_t$, $Seasonality_{t+24}$,$(NO_2)_t$, NO_t
5	NO_{t+24}	$Meteorology_t$, $traffic_t$, NO_t
6	NO_{t+24}	$Meteorology_t$, $traffic_t$, $Seasonality_{t+24}$,NO_t
7	NO_{t+24}	$Meteorology_t$, $traffic_t$, NO_t, $(NO_2)_t$
8	NO_{t+24}	$Meteorology_t$, $traffic_t$, $Seasonality_{t+24}$, NO_t, $(NO_2)_t$

The number of neurons H in the hidden layer of the MP method, is determined by experimentation, training the neural networks with values of H from 5 to 30. Greater values of H do not give a better performance. The MP networks are trained with two backpropagation algorithms: the scaled conjugate gradient algorithm (SCG) and the Levenberg–Marquard algorithm (LM). The output Y can be expressed as follows:

$$Y^o = f^o(b^o + \sum_{j=1}^{H} w_j^o f_j^h\,(b_j^h + \sum_{i=1}^{I} w_{ij}^h X_i)) \tag{1}$$

where o denotes the elements of the output layer and h indicates the elements of the hidden layer. w_j^o is the weight that connects the neuron j of the hidden layer with the neuron of the output layer, and w_{ij}^h is the weight that connects the neuron i of the input layer with the neuron j of the hidden layer. b^o is the bias of the neuron of the output layer, and b_j^h is the bias of neuron j of the hidden layer. f^o is the transfer function of the neuron of the output layer. In this work the linear transfer function has been applied for f^o. f_j^h is the transfer function of neuron j of the hidden layer. The most widely used f_j^h are the hyperbolic transfer function (tansig) and the logarithmic sigmoid function (logsig):

$$tansig(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2}$$

$$logsig(x) = \frac{1}{1+e^{-x}} \tag{3}$$

RBF networks may require more neurons than standard feed-forward backpropagation networks, 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 vectors are available. Each neuron in the hidden layer of the RBF method calculates the Euclidean distance of the input vector X to its own centre c_j. The calculated distance is transformed via a transfer function $\emptyset_j$. The result of this operation in each neuron of the hidden layer is multiplied by a weighting factor w_j^o and summed with the other neurons weighted results, to give the predicted output value Y. The final output Y can be expressed:

$$Y^o = \sum_{j=1}^{H} w_j^o\, \emptyset_j(\|X - c_j\|, \sigma_j) \tag{4}$$

$\|X - c_j\|$ denotes the Euclidean distance. The Gaussian function is widely used as a non-linear transfer function. σ_j is the bandwidth of this function. In the training process, the steepest gradient descent learning process is used to adjust the appropriate settings of the parameters (e.g. weights, centers, and bandwidths).

Overtraining occurs when the neural networks memorize the patterns introduced to it and it is not capable of identifying new situations. The early stopping technique can be used to avoid this problem [23]. In this method the data set is separated into three subsets: the training set, the validation set and the test set. The training set is used to update the network weights and biases. During the training, the validation set is used to guarantee the generalization capability of the model, and training should stop before the error on the validation set begins to rise. The test set is a new set used to check the generalization of the MP. In this

work, the models are trained on data from the first year. Data from the second year are used as the validation set, and observations from the third year are the test data set. The computations are performed with the *Neural Network Toolbox* of MATLAB.

2.3 Evaluation criteria

Four evaluation criteria are obtained to compare the performance of the MP and RBF methods with the eight models, to forecast the test data set. The correlation coefficient r between the forecasted values Y_f and the observations Y quantifies the global description of the model. The root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - Y_{fi})^2}{n}} \tag{4}$$

where n is the number of observations in the test data set. The mean absolute error (MAE)

$$MAE = \frac{\sum_{i=1}^{n}|Y_i - Y_{fi}|}{n} \tag{5}$$

An expression of accuracy of predictions as a percentage can be computed with the mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - Y_{fi}}{Y_i}\right| \tag{6}$$

3 Results and discussion

Table 3 contains the predictions results at Pista Silla with the MP networks that had the best performance. In all cases the most accurate forecasts were obtained with the Levenberg–Marquard backpropagation algorithm.

Table 3: MP predictions at Pista Silla.

Model	Output	Transfer function	n_h	r	RMSE	MAE	MAPE
1	NO_2	tansig	14	0.59	20.30	16.49	0.51
2	NO_2	tansig	14	0.63	19.35	15.38	0.45
3	NO_2	logsig	30	0.56	20.48	16.32	0.48
4	NO_2	logsig	10	0.65	20.49	16.55	0.50
5	NO	logsig	10	0.61	45.12	27.80	1.31
6	NO	tansig	12	0.65	43.27	26.98	1.19
7	NO	logsig	10	0.59	45.26	28.39	1.42
8	NO	tansig	12	0.65	43.05	26.56	1.19

NO_2 predictions at time t+24 are better with model 2 in terms of the RMSE, MAE and MAPE values. This model includes as predictors the meteorological indicators, the traffic volume, the seasonal components and the NO_2 concentration at time t. The best value of the correlation coefficient between observations and predictions, is obtained with the model that also includes NO concentration at time t. The prediction of NO at time t+24 is more accurate with the model 8, which includes all the predictors. The forecast errors are greater for NO than for NO_2. Table 4 shows the best RBF network predictions at Pista Silla for the eight models.

Table 4: RBF predictions at Pista Silla.

Model	Output	bandwidth	r	RMSE	MAE	MAPE
1	NO_2	1	0.55	21.4	16.60	0.39
2	NO_2	2.25	0.66	19.8	15.34	0.34
3	NO_2	1	0.55	21.5	16.61	0.38
4	NO_2	2.25	0.66	19.7	15.31	0.35
5	NO	2.25	0.56	47.1	27.60	1.19
6	NO	2.25	0.65	43.2	26.30	1.15
7	NO	2.25	0.54	47.6	29.40	1.45
8	NO	2	0.65	43.1	26.80	1.20

With this neural networks models, the best NO_2 predictions for time t+24, were obtained with model 4 in terms of r, RMSE and MAE. This model includes all the predictors as inputs. With model 2 the MAPE is slightly smaller; in this case the model does not use NO concentration at time t as predictors. The comparison of values in Tables 3 and 4, indicates that the RBF forecasts are better for NO_2 than the MP ones, when considering the r, MAE and MAPE values, The RMSE is smaller with MP network and model 2.

The best NO prediction at time t+24, is attained with the RBF network when applied to model 6, in terms of r, MAE and MAPE. With model 8 the RMSE is smaller. In both cases seasonal components are included. This pollutant forecast is worse than NO_2 forecast when considering the RMSE, MAE and MAPE values. The comparison with the Table 3, shows that the NO prediction is better with the MP applied to model 8, than any of the models with the RBF network.

The best NO_2 and NO prediction results at Viveros are given in Tables 5 and 6. Table 5 corresponds with the MP networks computations.

With models 1, 5, 6, 7 and 8, the MP performance was better with the Levenberg–Marquard backpropagation algorithm. Models 2, 3 and 4 worked more accurately with the scaled conjugate gradient learning method. At this site, the r values of NO_2 predictions are very small. Model 1 (predictors are meteorology, traffic and NO_2 observations) has the best r, MAE and MAPE. The inclusion of seasonality (model 2) only improves the RMSE indicator. However, the NO predictions perform better when the models incorporate seasonal cycles. When the inputs are these cycles, meteorology, traffic and NO levels, the results have smaller MAE and MAPE. If the model also considers NO_2 levels as inputs, r and RMSE

Table 5: MP predictions at Viveros.

Model	Output	Transfer function	n_h	r	RMSE	MAE	MAPE
1	NO_2	tansig	7	0.12	23.39	18.30	0.93
2	NO_2	tansig	10	0.07	23.05	18.31	0.98
3	NO_2	tansig	7	0.05	23.38	18.57	0.97
4	NO_2	logsig	7	0.01	23.40	18.77	0.97
5	NO	logsig	5	0.49	33.78	19.86	3.52
6	NO	logsig	16	0.52	33.19	19.03	3.16
7	NO	tansig	10	0.49	33.94	19.26	3.20
8	NO	tansig	10	0.65	33.10	19.10	3.32

indicators are better. NO predictions have higher r values than NO_2 predictions, but the latter have smaller values of the other three indicators. In Pista Silla site, MP networks (Table 3) have more accurate predictions for NO_2 and NO than at Viveros site.

Table 6 shows the forecasts results with the RBF networks at Viveros, for the two pollutants and the eight models.

Table 6: RBF predictions at Viveros.

Model	Output	bandwidth	r	RMSE	MAE	MAPE
1	NO_2	1	0.55	18.6	14.6	0.67
2	NO_2	1.75	0.62	17.6	13.7	0.62
3	NO_2	2.25	0.55	18.6	14.6	0.67
4	NO_2	1.75	0.61	17.5	13.7	0.62
5	NO	1.75	0.49	33.9	18.5	2.8
6	NO	2	0.53	33.2	18.3	2.9
7	NO	1	0.50	33.6	18.1	2.6
8	NO	2	0.53	33.2	18.1	2.7

NO_2 predictions are much better with this method than with MP networks, as the four criteria indicate. Models 2 and 4, which include daily and weekly seasonality, have higher r coefficients, and smaller RMSE, MAE and MAPE values. NO forecast are more accurate with RBF method than with MP networks, when considering MAE and MAPE results, but r coefficients and RMSE parameters are quite similar with the two procedures. Including seasonality (models 6 and 8) improves r and RMSE statistics, being slightly better model 8 with all the predictors. The comparison with Pista Silla site shows that RBF networks had lower MAPE indicators at Viveros for NO_2 predictions. Moreover, RMSE and MAE of NO forecast at Viveros were better than at Pista Silla with RBF networks, but r and MAPE values were worse at the second station.

4 Conclusions

In this work, a comparison of multilayer perceptron and radial basis functions networks is made. The aim is to predict hourly levels of two pollutants, 24 hours in advance at two locations of an urban area. Different models are studied, depending on the number of inputs they consider. The inclusion of daily and weekly seasonality improves the quality of predictions. In the first site, the RBF method performs better when meteorology, traffic, seasonal cycles, and NO_2 and NO levels are included. At the second site, however, the best NO_2 prediction is attained when NO levels are not included as inputs. In this site RMSE and MAE of NO_2 forecasts are smaller than at the first site, where r and MAPE values are better.

The MP networks only performed better tan RBF networks when predicting NO_2 at the second station, and NO concentration was not considered as predictor. In all the other cases RBF methods resulted more accurate than MP networks.

References

[1] Senger, S.U., Estimation of the reduction of the concentration of air pollutants until the year 2010 in relation to the UN ECEE Göteborg protocol dated 1. December 1999. *Umweltwissenschaft und Schadstoff Forschung*, **12(3)**, pp. 152–156, 2000.

[2] Chauhan, A.J. & Johnston, S.L., Air pollution and infection in respiratory illness. *British Medical Bulletin*, **68(1)**, pp. 95–112, 2003.

[3] Rye, P.J., Modelling photochemical smog in the Perth region. *Mathematical and Computer Modelling*, **21(9)**, pp. 111–117, 1995.

[4] Box, G.E.P., Jenkins, G.M. & Reinsel, G.G., *Time Series Analysis. Forecasting and Control*, 4th edition, Wiley: Hoboken, New Jersey, USA, 2008.

[5] Prada-Sanchez, J.M., Febrero-Bande, M., Cotos-Yanez, T., Gonzalez-Manteiga, W., Bermudez-Cela, L. & Lucas-Dominguez, T., Prediction of SO_2 pollution incident near a power station using partially linear models and an historical matrix of predictor-response vectors. *Environmetrics*, **11(2)**, pp. 209–225, 2000.

[6] Kocak, K., Saylan, L. & Sen, L., Nonlinear time series prediction of ozone concentration in Istanbul, *Atmospheric Environment*, **34(8)**, pp. 1267–1271, 2000.

[7] Ryan, W.F., Forecasting ozone episodes in the Baltimore metropolitan area, *Atmospheric Environment*, **29(17)**, pp. 2387–2398, 1995.

[8] Gardner, M.W. & Dorling, S.R., Statistical surface ozone models: an improved methodology to account for non-linear behaviour, *Atmospheric Environment*, **34(1)**, pp. 21–34, 2000.

[9] Shalkoff, R., *Pattern Recognition: Statistical Structural and Neural Approaches*, Wiley: New York, USA, 1992.

[10] Gardner, M.W. & Dorling, S.R., Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences, *Atmospheric Environment*, **32(14-15)**, pp. 2627–2636, 1998.
[11] Hornik, K., Stinchcombe, M. & White, H., Multilayer feedforward networks are universal approximators, *Neural Networks*, **2(5)**, pp. 359–366, 1989.
[12] Bishop, C.M., *Neural Networks for Pattern Recognition*, Claredon Press: Oxford, UK, 1995.
[13] Ibarra-Berastegi, G., Elias, A., Barona, A., Saenz, J., Ezcurra, A. & Diaz de Argandoña, J., From diagnosis to prognosis for forecasting air pollution using neural networks: Air pollution monitoring in Bilbao, *Environmental Modelling & Software*, 23(5), pp. 622–637, 2008.
[14] Caselli, M., Trizio, L., De Gennaro, G. & Ielpo, P., A simple feedforward neural network for the PM_{10} forecasting: Comparison with a radial basis function networks and a multivariate linear regression model. *Water, Air and Soil Pollution*, **201(1)**, pp. 365–377, 2009.
[15] Lu, W.Z., Wang, W.J., Wang, X.K., Yan, S.H. & Lam, J.C., Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong, *Environmental Research*, **96(1)**, pp. 79–87, 2004.
[16] Wang, C., Zhang, W., Wang J. & Zhao, W., The prediction of SO_2 pollutant concentration using a RBF neural network, *Applied Mechanics and Materials*, **55–57**, pp. 1392–1396, 2011.
[17] European Communities, *Europe´s Environment: the Fourth Assessment*. European Environment Agency, Office for Official Publications of the European Community, Luxembourg, 2007.
[18] Ballester, F., Iñíguez, C. & García, F. *ENHIS-1 project: WP5 health impact assessment. Local city report Valencia*. Available at http://www.apheis.org/CityReports2005/Valencia.pdf, 2005.
[19] Tenias, J.M., Ballester F. & Rivera, M.L., Association between hospital medical emergency visits for asthma and air pollution in Valencia, Spain. *Journal of Occupational and Environmental Medicine*, **55(8)**, pp. 541–547, 1998.
[20] Ballester, F., Tenias, J.M. & Pérez-Hoyos, S., Air pollution and emergency hospital admissions for cardiovascular diseases in Valencia, Spain, *Journal of Epidemiology and Community Health,* **55(1)**, pp. 57–65, 2001.
[21] Ballester, F., Rodríguez, P., Iñíguez, C., Sáez, M., Daponte, A. *et al.*, Air pollution and cardiovascular admissions association in Spain: result within the EMECAS project, *Journal of Epidemiology and Community Health,* **60(4)**, pp. 328–336, 2006.
[22] Capilla, C., Multilayer perceptron and regression modelling to forecast hourly nitrogen dioxide concentrations. *WIT Transactions on Ecology and the Environment*, **183**, pp. 39–48, 2014.
[23] Sarle, W.S., Stopped training and other remedies for overfitting. *Proceedings of the 27th Symposium on the Interface*, 1995.