

# Adaptive system of forecasting the air quality

W. Kamiński & E. Tomczak

*Technical University of Lodz,*

*Faculty of Process and Environmental Engineering, Poland*

## Abstract

In the study the adaptive system of forecasting pollutant levels has been proposed. The data for calculations were obtained from an automatic monitoring station located in a Polish city. The system was presented on the example of the suspended particulate matter concentration  $PM_{10}$ . The proposed method is based on the application of the neural networks namely RBF type which evolves from the data set compiled for the period of a few months. The range of concentrations is divided into five classes of air quality. On the basis of this model one may forecast the level of pollution a few days ahead.

*Keywords: radial basis function, air quality, prediction.*

## 1 Introduction

The life and industrial activity of man contribute to the occurrence of too high pollution levels in large agglomerations. This phenomenon frequently leads to the occurrence of pollutant concentrations in excess of standard limits. Apart from the meteorological prognosis there is more and more frequently provided the prognosis concerning the air pollution level due to the fact that higher concentrations of certain substances may lead to health problems of humans and animals. The indices characterizing the pollution level are more often used for this purpose. Dividing appropriately the range of the possible pollutant concentrations, for instance,  $O_3$ ,  $CO$ ,  $NO_x$ ,  $PM_{10}$ , etc. into sub-ranges and naming them, for example, a low, acceptable, high, emergency level, one puts the information into the form which is more understandable for a reader.

One of the frequently applied methods in air pollution problems was the radial basis function network RBF [1–7]. RBF network approximates an unknown mapping function  $F: R^N \rightarrow R$  according to the superposition [8]:



$$y = F(x) = \sum_{k=1}^M w_k \Phi \left( \frac{\|x - c_k\|}{s_k} \right) \quad (1)$$

in which radially symmetric basis functions  $\Phi_k$ ,  $k=1, \dots, M$  are linearly combined in proportion determined by adjustable weight coefficients  $w_k$ . The constants  $c_k \in R^N$ ,  $s_k \in R$  are centers of the basis functions and scaling factors of the Euclidean norm  $\|x - c_k\|$ , respectively. There are five parameters to be determined in the RBF network so that it can be used as a viable tool for performing the required data processing task:

1. the number of basis functions  $M$ ,
  2. the type of the basis functions  $\Phi$ ,
  3. scaling factors  $s_k$ ,
  4. localisation of centers  $c_k$ ,
  5. set of weights  $w_k$ .
- $k=1, \dots, M$ .

The first four parameters define the sampling pattern of the input data space  $x \in R^N$  by means of symmetric kernel functions that play the role of local receptors for the input data field. Whereas, the set of weights  $w_k$  is responsible for providing an optimum contribution of each of the basis functions, to the network output, so that it results in best approximation of the desired target mapping function. Accordingly, the design task of an RBF network can be broken up into two phases: determination of the network receptive field (parameters 1–4), that is followed by tuning (e.g., by means of training) the output layer weights.

## 2 The aim, territory and scope of research

The testing ground for research to test the neural prognostic models was the city of Lodz (population: 800000) located in the centre of Poland. Lodz is located on the gentle, western slope of Lodz Heights, a northern part of Lodz Upland. Denivelations in city boundary are quite large. The highest altitude exceeds 275 m (NE outskirts of the city), and the lowest is 170 m (SW outskirts of the city). In the centre of Lodz surface features are not so varied.

The active surface in Lodz is diversified mainly due to a varied density, character and height of buildings. Downtown building complex is very compact and weakly diversified in height. Mainly flat roofs made upper active surface monotonous. Monotony of active surface in the intensive building complex area is even higher due to lack of larger hydrographical objects having very small values of ground roughness coefficient and specific radiation and heat balances. The especially strong wind field modification in Lodz is due to the urban building complex. High ground roughness influences wind speed decrease. Town planning structures cause very high diversification of direction and wind speed.



It has a significant influence on forming the air pollution emission fields. Small and dispersed areas of high and low green in the centre of the city do not contribute to the significant diversification of the active surface. Annual cycles of anthropogenous emission of heat and air pollution are very distinct and they are an effect of seasonal heat energy demand. Lodz is located a relatively large distance from other large cities and urban-industrial agglomerations. Considering this distance and structure of winds over Poland, the effects of a large cities influence on the deposition field of pollution in Lodz is relatively small.

The data for calculations were taken from a cool season only, i.e. November through March, when particulate matter concentration in the air was relatively high. Total input data included meteorological parameters and concentrations of  $PM_{10}$  in the air comprised 544 days.

Five classes of particulate matter concentrations  $PM_{10}$  (for maximum daily values) were formed. The range of classes was related to the concentrations identified as valid air quality standards, taking into account multiplicity of standard concentrations, e.g. 50%, 100%, 150%, over 200% of the standard value.

The input data of the model are: the maximal and mean daily concentration of  $PM_{10}$  in the air on the day  $n$  and the following meteorological parameters on the day  $n+1$ :

1. maximal and minimal (extreme) daily air temperature,
2. daily sum of rainfall,
3. daily sum of Sun radiation,
4. wind velocity,
5. velocity of vertical air flow considering the direction of motion,
6. mean daily atmospheric pressure.

The classes' width was declared in accordance with the effective standards of the air quality (and multiplication factor of these standards, Table 1). Allocation to the class was on the basis of maximal  $PM_{10}$  concentration on the day  $n+1$ .

### 3 Method of air quality forecasting

The pollution level forecast was obtained using RBF networks, i.e. kernel orthonormalization procedure [8]. The method proposed includes the consecutive addition of the function  $\Phi_k$  and searching the localization of center  $c_k$  and scaling factor  $s_k$  using evolutionary algorithm. The weight  $w_k$  is calculated using the least square method. By consecutive addition the basis function  $\Phi_k$ , differing in terms of the center localization and scaling factor the model is adjusted to the data with the accuracy assumed.

In adaptive forecasting the new data obtained from the measurements are introduced and, simultaneously, the oldest data in the same size are removed from the training set.

In further procedure the following hypothesis was assumed that the vectors constructed on the canters of basis functions and the central point of the data set,

prior to the addition of new data, and, after do not change. Simultaneously, the scaling factors are conserved.

In the calculations the Gauss function was assumed as basis functions of RBF network. In this way, the adjustment of the model to the new data requires correction of centers  $c_k$  and calculation of new weights  $w_k$   $k=1, \dots, M$ . In the method proposed it is not necessary to search again for new centers and scaling factors of function  $\Phi_k$  using evolutionary algorithm.

Table 1: Characteristics of input data used in air quality forecasting.

Parameter	Classes of PM <sub>10</sub> concentration [ $\mu\text{g}/\text{m}^3$ ]				
	I	II	III	IV	V
Range of classes [ $\mu\text{g}/\text{m}^3$ ]	1-49	50-99	100-149	150-199	>200
Number of cases	92	211	117	77	43
Mean max daily PM10 concentration [ $\mu\text{g}/\text{m}^3$ ]	37.8	53.6	78.8	97.9	160.9
Mean daily PM10 concentration [ $\mu\text{g}/\text{m}^3$ ]	27.9	42.2	61.5	75.3	121.7
Mean daily sum of rainfall [mm]	2.02	0.85	0.6	0.37	0.11
Mean daily sum of radiation [ $\text{W}/\text{m}^2$ ]	616	1062	1231	1271	1490
Mean daily wind velocity [m/s]	4.51	3.43	2.88	2.65	2.19
Mean daily vertical velocity [m/s]	0.0089	-0.085	-0.098	-0.128	-0.14
Mean daily maximal temperature [ $^{\circ}\text{C}$ ]	4.58	2.93	2.42	0.61	-0.93
Mean daily minima temperature [ $^{\circ}\text{C}$ ]	0.4	-1.24	-2.31	-4.46	-7.44
Mean daily atmospheric pressure [hPa]	986	989	989	991	991

Let us denote by  $x(1:n, 1:p)$  the initial input set vectors ( $n$  – the input vector size;  $p$  –size of data) and by  $x_n(1:n, 1:L)$  the new data vectors ( $L$  – size of new data), and by  $c^*(1:n, 1:M)$  the adjusted centers vectors. The vectors, prior to the introduction of the new data, produced on the basis of centers of basis functions and the central point of the data set will have the following coordinates:

$$x_c(1:n, k) = \frac{\sum_{i=1}^p x(1:n, i)}{p} - c^*(1:n, k) \quad (2)$$

$$k = 1 \div M$$



The new localizations of the basis function centers may be derived from the following relation:

$$c(1:n, k) = \frac{\sum_{i=1}^L x_n(1:n, i) - \sum_{i=1}^L x(1:n, p-i+1)}{p} + c^*(1:n, k) \quad (3)$$

where:

- $c$  – the new coordinates of the centers of basis functions,
- $c^*$  – the former coordinates of basis functions,
- $L$  –size of new data.

The introduction of new data to the input set requires the calculation of new coordinates of basis functions in accordance with relationship (3) and the calculation of new weights using the least square method.

#### 4 Quality of forecasting

At the first step, the data from the consecutive 100 days were taken and approximated using the RBF neural network, i.e. kernel orthonormalization. Next, the consecutive data were added (size  $L$ ) simultaneously removing the oldest data of the same size. The centers of basis functions were adjusted and, subsequently, new weights were calculated. In this case dependent variable is the air quality class. The quality of prediction depends upon size of the new data which are introduced. The comparison of the results is presented in Table 1.

The results of air quality class prediction were arranged in the three- category system:

- category I – good results: difference between the calculated and real class is equal to 0 (full conformity of the forecast with reality),
- category II – acceptable results (correct): difference between the calculated and real class is  $\pm 1$  (the forecast differs from reality by one concentration class),
- category III – unacceptable (erroneous) results: difference between the calculated and real class is greater than +1 and smaller than  $-1$  (the forecast differs from reality by more than one concentration class).

As it may be inferred from Table 1, category III was not present in all the cases under scrutiny.

Table 2: Accuracy of prediction vs. number of new data added.

Size of new data	category I	category II
20	10	10
10	7	3
5	3	2
3	2	1
2	2	0
1	1	0



## 5 Concluding remarks

The proposed method, i.e. the adaptive prediction of the air quality classes, is based on the networks of RBF type.

The initial data enables to define the localization of basis functions centers, scaling factors and weights. The actualization of the data set allows to adjust adaptively the network to the new data. The adaptation of networks to the new data was based on the hypothesis of invariance of the vectors constructed on the centers of basis functions and the central point of the data set. The method is simple and does not require many calculations.

## Acknowledgement

This work was financed from funds allocated to science in the years 2008–2010 as a part of the research project.

## References

- [1] Bator, R., Sieniutycz, S., Application of artificial neural network for emission prediction of dust pollutants, *International Journal of Energy Research* 30 (13), pp. 1023-1036, 2006
- [2] Cao, A.-Z., Tian, L., Cai, C.-F., Zhang, S.-G., Research and prediction of atmospheric pollution based on wavelet and RBF neural network, *Journal of System Simulation* 18 (5), pp. 1411-1413, 2006
- [3] Ibarra-Berastegi, G., Short-term prediction of air pollution levels using neural networks, *WIT Transactions on Ecology and the Environment* 86, pp. 23-31, 2006
- [4] Kaminski, W., Tomczak, E., Prediction of carbon monoxide concentration near roads by means of artificial neural networks, *Advances in Air Pollution* 11, pp. 763-772, 2002
- [5] Lu, W. Z.; Wang, W. J.; Wang, X. K.; Xu, Z. B.; Leung, A. Y. T., Using Improved Neural Network Model to Analyze RSP, NO<sub>x</sub> and NO<sub>2</sub> Levels in Urban Air in Mong Kok, Hong Kong, *Environmental Monitoring and Assessment* Volume: 87 (3), pp. 235 – 25, 2003
- [6] Sun, G., Hoff, S.J., Zelle, B.C., Nelson, M.A., Forecasting daily source air quality using multivariate statistical analysis and radial basis function networks, *Journal of the Air and Waste Management Association* 58 (12), pp. 1571-1578, 2008
- [7] Wang, S., Yu, D.L., Adaptive RBF network for parameter estimation and stable air-fuel ratio control, *Neural Networks* 21 (1), pp. 102-112, 2008
- [8] Kaminski W., Strumillo P., Kernel Orthonormalization in Radial Basis Function Neural Networks, *IEEE Trans. Neural Networks*, vol.8, no.5, 1177-1183, 1997.

