

Application of artificial neural networks for classification and prediction of air quality classes

J. Skrzypski, K. Kamiński, E. Jach-Szakiel & W. Kamiński
*Technical University of Lodz,
Faculty of Process and Environmental Engineering, Poland*

Abstract

In this study, the results of investigations which enable an extension of the mathematical methods supporting air quality management in cities are presented. The actions were focused on the development of neural models of classification and prediction of the air quality classes (in respect of PM10 dust concentrations). The air quality class on a following day was predicted. The aim of modelling was to predict the air quality classes in the afternoon and in the evening when PM10 concentrations attained the daily maxima. The monitoring of PM10 concentration and the meteorological data for winter periods in 2004–2007 was used. The artificial neural network methods (ANN) with a simultaneous application of data compression methods were tested. The results of the air quality prediction are satisfactory. The accurate prognoses are predominant. The percentage of wrong prognoses is relatively small. The investigations confirm that neural prediction models allow good results to be obtained of the air quality class prediction. The results of the research prove that the tested models may be applied in the practice of air quality management in cities.

Keywords: classification, prediction, air quality, big cities, PM10, artificial neural network, modelling.

1 Introduction

The anthropogenic emission of air pollution contributes considerably to the ravage of the Earth. For decades particular attention has been paid to the protection of the stratospheric ozone layer and the protection of ecosystems



against the results of acidic substance deposition (acid rains). For 25 years air pollution has constituted an increasing hazard for the inhabitants of cities. The strong and permanent development of urbanized regions is a consequence of the global demographic explosion and the intensive industrialization of many cities. The secondary result is the concentration of municipal, industrial and transport emission sources [1]. Certain types of meteorological conditions and types of the atmospheric circulation as well as the specific features and peculiarities characteristic of the climate of all cities influence high concentrations of air pollution [2].

In cities the air quality standards usually are not met. Periodically, especially in city centres, high concentrations of air pollution generate a health risk for inhabitants of cities. An indispensable tool of health risk management in cities constitutes the air quality prognoses. The prognoses serve as a basis for building systems of warnings and alarms. Those prognoses are also a basis for announcing a decision which concerns the periodical reduction of activity of chosen companies in order to reduce the air pollution emission.

In the present study the results of the investigations aimed at increasing the effectiveness of ecological safety management in industrialized cities are demonstrated. This aim was accomplished by developing a new model of air quality prediction using the artificial intelligence methods and the input data compression. Artificial neural networks (ANNs) are a proved tool for description and prediction of the air pollution concentration fields in cities. ANN display an ability to generalize the dependence and minimize the influence of secondary factors. This allows us to use the incomplete data comprising measurement errors. The data compression methods were applied to generate the aggregated sets of meteorological data and the air pollution concentrations.

2 Territory and scope of research

The testing ground for research was the city of Lodz (population: 800,000 citizens) located in the centre of Poland. In the study the results of the examinations performed by the Regional Inspectorate of Environmental Protection in Lodz in 2004–2007 are used. The measurements are carried out continuously and automatically. The results are averaged in 1-hour ranges. In the study the monitoring data of air pollution concentration were taken in the period of cool seasons (November–March) when those concentrations and, thus, health risk are the highest. The neural models were tested in relation to PM10 because standard level for this pollution especially in downtown was frequently exceeded. The data from the monitoring station and meteorological station, both located in the city centre, were used.

The ANN with a simultaneous application of the data compression was subject to testing. ANNs were used for description and prediction of the air pollution concentration fields in cities [3–8].

The concentration of pollution in the air changes dynamically. Nevertheless, there occur noticeable seasons, weekly and daily cycles. In the case under consideration the aim of modelling was the air quality class prediction in the



afternoon and in the evening when PM10 concentrations attained daily maxima (Figure 1). Prediction of maximum concentrations is more difficult due to greater variety of concentrations on particular days. It is more difficult to obtain satisfactory effects of modelling. Despite this, the maximum concentration predictions are indispensable owing to an increased risk of conditions which are unbeneficial to the health of citizens. Development of prediction model enables to manage the health risk using warnings and alarms.

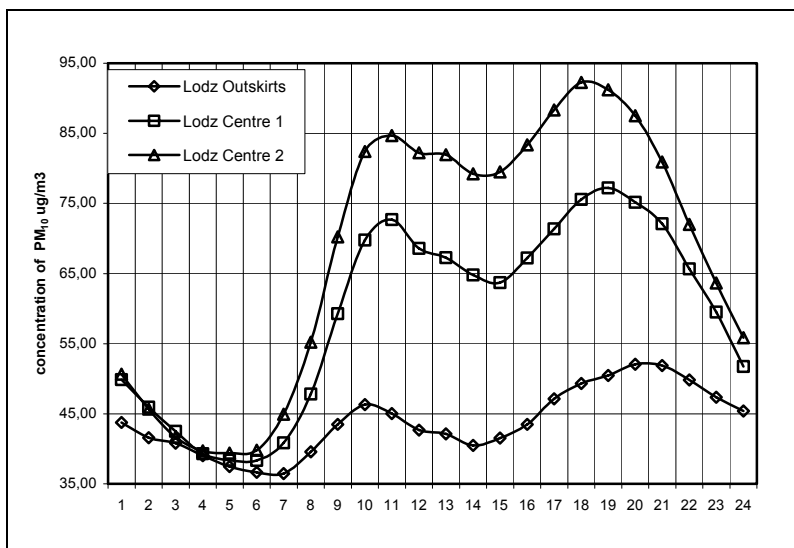


Figure 1: Daily changes of mean PM10 concentration in winter seasons 2004–2006 for monitoring stations localized in Lodz (the centre of the city) and on outskirts of the city.

3 Prediction models

In the study [9] the methodology allowing one to define the air quality classes in the city was presented. The application of self-organizing artificial neural networks (ANN) was used which allowed establishing the required classes only on the basis of information included in the input data set without any preliminary assumptions. The method was tested on the example of Lodz, the third Polish city with regard to the population. As a reference pollutant, determining the aerosanitary state of the air, suspended particulate mater PM10 was assumed. PM10 constitutes one of the most troublesome pollution in cool season when the general state of the air quality substantially influences the combustion of fuels for heating of buildings.

3.1 Input data for classification

As the input data for the air quality classification, the measurements of PM10 concentrations made in the period from November to March were used. In the cool season, the highest probability of concentrations which exceed the norm occurs. PM10 concentrations were added to values of the basic meteorological parameters predicted for the next day. The vectors of nine coordinates created in this way (Table 1) were subject to classification in order to create the air quality classes.

Table 1: Vector coordinates on the basis of which the prototypes were created.

No	Vector coordinate		max.	min.	unit
1.	PM10 mean concentration from the hs 16÷20	day n+1	443.4	5.2	µg/m ³
2.	PM10 mean daily concentration	day n	277.3	9.9	µg/m ³
3.	PM10 mean concentration from the hs 16÷20		443.4	5.2	µg/m ³
4.	daily radiation sum	day n+1	5112	27	W/m ²
5.	mean daily horizontal wind velocity		7.3	0.0	m/s
6.	mean horizontal wind velocity from the hs 16÷20		8.4	0.0	m/s
7.	mean daily maximum temperature		25.6	-18.9	°C
8.	mean daily minimum temperature		8.9	-24.0	°C
9.	mean temperature from the hs 16÷20		14.4	-22.8	°C

3.2 Prototypes generation

The input data can be allocated into subsets. Those subsets are represented by prototypes which constitute the structured subsets of meteorological data and data concerning the air pollution concentration.

The procedure of prototype generation uses the learning set. In the present study the learning set comprising 543 vectors was used. The structure of the applied SSN encompassed one neuron layer. Each of the neurons was represented by a set of synaptic weights corresponding to particular coordinates of learning vectors. Neuron weights display an ability to adapt in such a way that they can resemble the models presented. Modification of synaptic weights occurred in the so-called learning process which involves, from the input data set, the sequential and random presentation of the learning vectors to the network and the modification of neuron weights according to the appropriate scheme. In the work the operation analysis of two self-organizing network algorithms of the classical type: winner takes all (WTA) and of neuron gas (NG) was reported. Additionally, the influence of the number of processing elements in the network on the separability of the classes was obtained.

On completion of the learning process, the synaptic weight sets corresponding to the prototypes of the air quality classes were created.

3.3 Prototype generation effects

The qualitative analysis of the prototypes obtained was performed. For this purpose the assessment criterion, which allowed us to decide about the



separability of the air quality classes was applied. The air quality classes were treated as separable if the Euclidean distance between every two prototypes was greater than the sum of mean radii of subsets assign to those prototypes.

A detailed analysis of the results showed that the best results were obtained for the classical learning algorithm of the WTA and seven neurons in the network layer. The prototypes obtained as a result of the classification, corresponding to particular neurons in the trained neuron network, are presented in Table 2.

In Table 2 seven prototypes and the appropriate number of vectors from the input set assigned, are presented. In the presented case all of the subsets corresponding to the particular prototype were separable. This conclusion inspired the Authors to further analysis and creation of the air quality classes in reference to PM10 dust concentration during the daily maximum (the hrs 16-20). Based on the analysis of the first coordinate of the obtained prototypes, it was proposed to distinguish four air quality classes on the following day which are as follows:

- 1) Good air quality: PM10 concentrations smaller than $50 \mu\text{g}/\text{m}^3$ (classes I and II),
- 2) Deteriorated air quality: PM10 concentrations attain the value from 50 to $120 \mu\text{g}/\text{m}^3$ (class III),
- 3) Bad quality (hazardous state) PM10 concentrations greater than $120 \mu\text{g}/\text{m}^3$ (class IV and V),
- 4) Emergency state: PM10 concentrations greater than $230 \mu\text{g}/\text{m}^3$ (classes VI and VII).

The separated subsets were also obtained in case of the application of only three prototypes and WTA self-organizing algorithm. Nonetheless, for such the input data splitting the prototype of PM10 concentration on a day $n+1$, exceeding the emergency state defined by the Polish law, was not obtained. The results

Table 2: The prototypes, WTA algorithm, seven neurons in the network layer.

No	1	2	3	4	5	6	7	8	9		
coordinate:	C_{16_20}	C_{24hr}	C_{16_20}	R_{24hr}	V_{24hr}	V_{16_20}	T_{max_24hr}	T_{min_24hr}	T_{16_20}		
day:	$n+1$	n		$n+1$							
prototypes										population	
I	38.5	28.4	36.0	1042	3.9	3.8	4.0	-0.6	2.4	206	
II	50.7	75.6	101.8	893	3.4	3.5	2.8	-1.4	1.4	69	
III	74.0	48.8	61.4	1323	3.0	2.9	2.8	-1.7	1.6	157	
IV	114.3	175.6	224.7	1169	2.6	2.8	1.3	-5.0	0.1	14	
V	119.1	80.8	104.5	1063	2.7	2.6	-0.3	-5.2	-1.8	80	
VI	232.0	112.8	146.6	1366	2.0	2.0	-1.6	-8.1	-3.1	14	
VII	330.3	221.0	340.4	1433	1.4	2.0	-2.8	-11.1	-4.9	3	



Table 3: The prototypes, WTA algorithm, three neurons in the network layer.

No	1	2	3	4	5	6	7	8	9	
coordinate:	C _{16_20}	C _{24hr}	C _{16_20}	R _{24hr}	V _{24hr}	V _{16_20}	T _{max_24hr}	T _{min_24hr}	T _{16_20}	
day:	n+1	n	n	n+1	n+1	n+1	n+1	n+1	n+1	
prototypes										population
I	47.2	33.8	43.2	1112	3.6	3.6	3.8	-0.9	2.3	319
II	93.3	72.2	93.5	1048	2.9	2.9	1.4	-3.1	0.1	197
III	184.0	166.1	220.2	1232	2.2	2.5	-1.0	-7.6	-2.4	27

obtained by means of classification into three prototypes are summarized in Table 3.

For the aforementioned case, it was proposed to distinguish three air quality classes on a following day which are as follows:

- 1) Good air quality: PM10 concentrations smaller than 50 $\mu\text{g}/\text{m}^3$ (class I),
- 2) Deteriorated air quality: PM10 concentrations attain the value from 50 to 90 $\mu\text{g}/\text{m}^3$ (class II),
- 3) Bad air quality with a possibility of emergency state: concentration PM10 greater than 90 $\mu\text{g}/\text{m}^3$ (class III).

3.4 Methodology of the air quality prediction

The vectors (Table 1) after classification allow predicting the aero-sanitary state on a following day (n+1) on the basis of PM10 concentration on the present day (n) as well as the prognosis of the weather parameters for a day (n+1). The air quality prediction is based on a comparison of the vector encompassing two measurements of pollution concentration and six predicted meteorological parameters with the appropriate fragments of the prototypes (Figure 2). The comparison is aimed at finding the prototype for which the set of variables bears the strongest resemblance to the conditions contained in the vector for which the air quality class prognosis is produced.

The Euclidean distances between fragment of prototypes containing the coordinates from 2 to 9 and a tested vector with identical coordinates are calculated. The smallest distance points out the adequate, prototype. Next, based on the first coordinate of a selected prototype air quality classes (Table 2 and Table 3) is predicted on a following day.

4 The analysis of the air quality class prediction effects

The selection of the atmospheric air quality classes was evaluated for the data set on the basis of which the prototypes had been created. This allowed us to verify whether the predicted air quality class will correspond to the class to which the trained neural network will attribute the learning vector of nine coordinates.



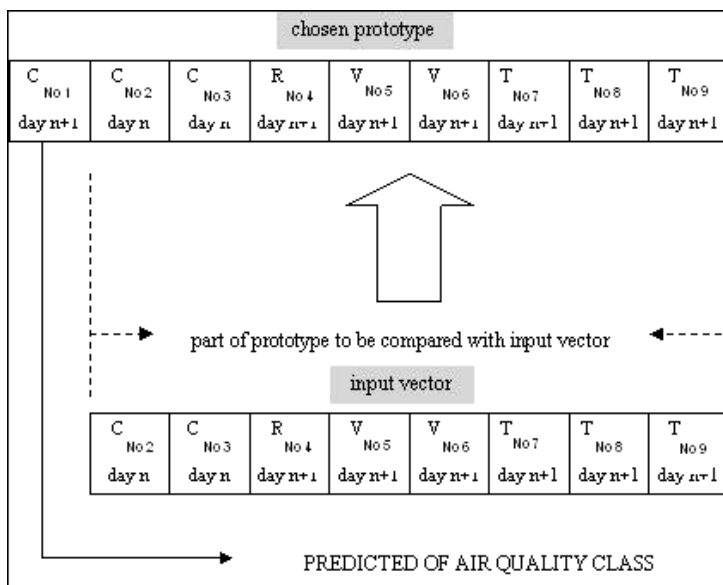


Figure 2: The air quality prediction on a day (n+1).

The following rules of the prognoses quality assessment were accepted:

- If the air quality class chosen for proposed procedure (minimum of the Euclidean distances between fragment of prototypes containing the coordinates from 2 to 9 and a tested vector) was in agreement with the class to which the trained neural network attributed a learning vector, then the prognosis was considered to be good.
- If the predicted air quality class differed by one class from the class attributed to a vector, the prognosis was regarded as acceptable.
- The other cases encompassing the prognosis discrepancy by more than one air quality class were treated as the wrong ones.

4.1 The analysis of prediction results for four air quality classes

The analysis of prognoses quality performed for this purpose proved that for the case of the seven prototypes and four air quality classes the following prognoses were obtained:

- For the prognoses of class I: 75.27% of good prognoses, 14.91% of the acceptable ones and 9.82% of the wrong ones;
- For the prognoses of class II: 82.17% of good prognoses, 17.83% of the acceptable ones and no erroneous prognoses were obtained.
- For the prognoses of class III: 48.94% of good prognoses, 28.72% of the acceptable ones and 22.34% of the wrong ones;

- For the prognoses of class IV: 41.18% of good prognoses, the same result of the acceptable prognoses was attained, 17.64% of the wrong prognoses.

Analyzing the results it may be stated that the efficiency of predicting of classes, which correspond to the bad quality of the air and to the possibility of the emergency state (class III and IV) does not exceed 50% on condition that only good prognoses are considered. Nevertheless, the percentage of wrong prognoses for those cases is relatively small and is equal to no more than 20%. Analyzing all the results obtained it may be concluded that the trained neural network makes it possible to predict the air quality classes on a following day in a satisfactory way, with the possibility of an error by one class in case of the increased PM10 concentrations.

4.2 The analysis of prediction results for three air quality classes

The quality assessment of prognoses for the three created prototypes (Table 3) and the three atmospheric air quality classes was performed. The analysis of the results proved that for this case 89.87% of good prognoses and 10.13% of acceptable prognoses were obtained. The wrong prognoses, in other words encompassing the discrepancy by more than one class, did not occur. Obviously, the reduction of the air quality class number to three contributes to the fact that in this case the acceptable prognoses should be treated as unsatisfactory ones. The analysis of the quality of prognoses proved that in case of three prototypes and three air quality classes the following prognoses were obtained:

- for class I: 92.79% of good prognoses and 7.21% of the acceptable ones;
- for class II: 86.29% of good prognoses and 13.71% of the acceptable ones;
- for class III: 81.48% of good prognoses and 18.52% of the acceptable ones.

The analysis of the results obtained for particular classes showed that in none of the aforementioned cases the wrong prognoses occurred. Very good possibilities of prognoses of PM10 concentrations exceeding the norm, equal to more than 80%, were found. Obviously, the application of the three air quality classes, contrary to the case in which four such cases were singled out, makes it impossible to distinguish the prognoses of the bad air quality in the city from the possibility of emergency states. It is only possible to predict the generally defined hazard state for the citizens.

5 Conclusions

In the present study the new methods of classification and prediction of the air quality classes in cities were tested. Two cases were analyzed. In case 1 seven prototypes and four air quality classes were applied. In case 2 three prototypes and three air quality classes were utilized. On the whole, the results of the air quality prediction of the maximum PM10 concentration are satisfactory.



Very good prediction results were obtained in case 2. None of the prognoses was wrong. Good prognoses were predominant (from over 80% of prognoses for the class of the highest concentrations to over 90% of prognoses for the class of the lowest concentrations). The remaining prediction results were regarded as acceptable. In case 1 the prediction results were slightly worse than their equivalents obtained in case 2 because the wrong prognoses occurred. The greatest number of the wrong prognoses (ca. 20%) was obtained in the classes of the highest PM10 concentrations. In the lower concentration classes the wrong prognoses did not occur or they occurred sporadically (below 10% of prognoses). The prognoses for the majority of the other days were appropriate and acceptable. The basic reason for the slightly deteriorated prediction results for case 1 is an increase of the number of prototypes (and, in consequence, the number of neurons).

With the expanded structure of the data used to create the separated prototypes the span of cases with high PM10 concentrations was insufficient to train the neural network. This problem will be examined in the future after the expansion of the input vector comprising the new data from the following years. The investigations confirm that the neural prediction models make it impossible to obtain good results of the air quality class prediction.

The results of the investigations prove that the tested models may be applied in practice of the air quality management in cities.

Acknowledgement

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

Appendix

Standard deviations of PM10 concentrations on a day $n+1$ for the case in which seven prototypes were created are presented in Table A1.

Table A1: Standard deviations of concentrations within 16-20 hours on a day $n+1$, seven neurons in the network layer.

Prototype	$C_{16,20}$	St. Dev.
	$[\mu\text{g}/\text{m}^3]$	$[\mu\text{g}/\text{m}^3]$
I	38.5	14,61
II	50.7	17,80
III	74.0	17,84
IV	114.3	55,61
V	119.1	21,03
VI	232.0	36,42
VII	330.3	109,83

Standard deviations of PM10 concentration for a day $n+1$ for the case in which three prototypes were created are presented in Table A2.



Table A2: Standard deviations of concentrations within 16-20 hours on a day n+1, three neurons in the network layer.

Prototype	C _{16,20}	St. Dev.
	[$\mu\text{g}/\text{m}^3$]	[$\mu\text{g}/\text{m}^3$]
I	47,2	19,66
II	93,3	35,85
III	184,0	95,65

References

- [1] Mayer H. Air pollution in cities, *Atmospheric Environment*, 33, pp. 4029-4037, 1999
- [2] Skrzypski J., Analysis and modelling of the system of air pollution concentration fields in big cities, in: *Environmental Engineering Studies*, Ed. By Paloski et al., Kluwer Academic/Plenum Publishers, New York, 2003
- [3] Lu W.Z., Wang W.J., Wang X.K., Yan S.H., Lam J.C., Potential assessment of neural network model with PCA/RBF approach for forecasting pollution trends in Mong Kok urban air, Hong Kong, *Environmental Research* 97 pp. 79-87, 2004
- [4] Niska H., Hiltunen T., Karppinen A., Ruuskanen J., Kolehmainen M., Evolving the neural network model for forecasting air pollution time series, *Engineering Applications of Artificial Intelligence*, 17, pp. 159-167, 2004
- [5] Kolehmainen M., Martikainen H., Ruuskanen J., Neural networks and periodic components used in air quality forecasting, *Atmospheric Environment*, 35, pp. 815-825, 2001
- [6] Viotti P., Liuti G., Di Genova P., Atmospheric urban pollution: application of an artificial neural network (ANN) to the city of Perugia, *Ecological Modelling*, 148, pp. 27-46, 2002
- [7] Kukkonen M., Niska H., Dorling S., Chatterton T., Foxall R., Cawley G., Extensive evaluation of neural network models for prediction of NO₂ and PM₁₀ concentrations, compared with deterministic modeling and measurements in central Helsinki, *Atmospheric Environment*, 37, pp. 4539-4550, 2003
- [8] Perez P., Reyes J., PM10 forecasting in Santiago, Chile, *WIT Press, Air Pollution XIII*, Ed. C.A. Brebbia, pp.33-38, 2005
- [9] Kamiński K., Skrzypski J., Kamiński W., Jach-Szakiel E., Air quality classification based on pm10 concentration and meteorological data by means of self-organizing neural networks, *Journal of the Air & Waste Management Association*, in press

