Ozone Concentration Prediction using Artificial Neural Networks

CAMELIA GAVRILA*
Technical University of Civil Engineering of Bucharest, 66, Pache Protopopescu Blvd, 021414, Bucharest, Romania

The aim of this paper is to determine a mathematical model which establishes the relationship between ozone levels together with other meteorological data and air quality. The model is valid for any season and for any area and is based on real-time data measured in Bucharest and its surroundings. This study is based on research using artificial neural networks to model nonlinear relationships between the concentration of immission of ozone and the meteorological factors: relative humidity (RH), global solar radiation (SR), air temperature (TEMP). The ozone concentration depends on following primary pollutants: nitrogen oxides (NO, NO_2), carbon monoxide (CO). To achieve this, the Levenberg-Marquardt algorithm was implemented in Scilab, a numerical computation software. Performed sensitivity tests proved the robustness of the model and its applicability in predicting the ozone on short-term.

Keywords: air pollution, troposphere ozone, artificial neuronal networks, hyperbolic tangent model

Existing models use different statistical or mathematical methods [1-8] and consider various parameters such as NO_2, NOx [3-4]. The newest models are based on neural network models [6] and Bayesian network models [9-10].

The applications of neural networks might vary a lot, but the prediction and process control is one of the most elaborate. This, together with genetic algorithms, can predict future states of a process, such as the estimation of ozone immission (O_3) depending on the primary pollutants: nitrogen oxides (NO, NO_2), carbon monoxide (CO) and meteorological factors, such as relative humidity (RH), the global solar radiation (SR), air temperature (TEMP).

Neural networks have several advantages, such as

-They can learn from examples.

-They tolerate defects in the sense that they can treat incomplete datasets

-They can solve nonlinearity problems and

-Once trained, they can make predictions and generalizations at a high computing speed.

An important application for the neural networks is applied in modeling and system identification [11].

They have been successfully used in several control applications of systems, robotics, pattern recognition, medicine, weather forecasting, energy systems, and optimization problems, signal processing, social sciences etc.

Artificial neural network models are connectionist models, which contain parallel processing units, called neurons. These are simplified models of the biological nervous system. Neural calculations are performed on a network composed of interconnected neurons, having two fundamental features: its architecture and behavior over time (dynamic behavior). Another important difference between neural models is represented by the type of the learning algorithm that determines when and in which manner the weights of synaptic connections are changing.

The Levenberg-Marquardt method, presented in [12], is one of the currently mostly used methods and is considered one of the fastest and most effective way of training neural networks, which can also be easily implemented in the SPSS environment.

The paper [8] presents an exponential model that determines the relationship between ozone levels and other independent variables (meteorological and air quality data) using the algorithm Levenberg-Marquardt, applied for a certain period of the year and for certain areas of Bucharest.

The purpose of this article is to determine a more general mathematical model which can be applied for every season and for every area of Bucharest. To achieve this goal we used real time measurements from various areas of the city and its surroundings, provided by the National Network of Air Quality Monitoring.

This paper is based on research using artificial neural networks to model nonlinear relationships between the concentration of the ozone immission and other pollutants and meteorological parameters. Our model might be used to predict O_3 concentrations in locations where there is no data recorded by the weather stations.

Experimental part

Materials and methods

The model consists in a multilayer neural network with error propagation. It can be used for any multi-layer network with differentiable activation functions, as it is a supervised training method based on the gradient descending method, which adjusts the weights to reduce the error [11].

The training algorithm with error propagation at the end can be implemented in various ways depending on the method of calculation error. At this moment the most effective is the Levenberg-Marquardt method. This method avoids the calculation of Hessiene matrix, approximating it as follows:

$$H = J^T . J \quad (1)$$

where J is the Jacobian matrix, containing the first order derivatives of the function representing the neural network error.

The description of the Levenberg-Marquardt training method is presented in [12]. It is currently the mostly used method and is considered one of the fastest and most efficient methods of training neural networks. Moreover, this method can be easily implemented in SPSS [13] or Scilab [14] environments.

The data used in this article are real time measurements provided by the National Network of Air Quality Monitoring for different areas of the city of Bucharest and its surroundings [15].

The input parameters are: nitrogen oxides (NO, NO_2), carbon monoxide (CO), relative humidity (RH), global solar radiation (SR) and air temperature (TEMP) and the output parameter is the concentration of the ozone immission (O_3).

*email: cgavrila2014@gmail.com; Phone: (+40)724684846

Time	NO	NO2	CO	RH	SR	TEMP	O3 mas
0	0.3329	0.6518	0.3725	0.6162	0.0000	0.4117	0.3242
1	0.3360	0.5170	0.3127	0.6515	0.0001	0.3366	0.3252
2	0.1862	0.2937	0.1141	0.7476	0.0003	0.2572	0.3283
3	0.0924	0.1791	0.0599	0.8132	0.0003	0.1881	0.3237
4	0.0236	0.0606	0.0210	0.8655	0.0005	0.1274	0.2987
5	0.0000	0.0000	0.0000	0.9312	0.0003	0.0677	0.2691
6	0.1083	0.1295	0.0450	0.9824	0.0010	0.0236	0.1407
7	0.2824	0.2840	0.1809	1.0000	0.0163	0.0000	0.0422
8	0.7022	0.4344	0.5931	0.9349	0.0482	0.0550	0.0257
9	1.0000	0.5329	1.0000	0.7981	0.0831	0.1738	0.1036
10	0.9047	0.6443	0.9291	0.6666	0.1188	0.3110	0.2782
11	0.7733	0.5826	0.8509	0.5023	0.1582	0.4615	0.4577
12	0.7070	0.5306	0.8468	0.3715	0.2011	0.6016	0.6171
13	0.6760	0.4823	0.8558	0.1978	0.8141	0.7739	0.7626
14	0.6682	0.4486	0.8024	0.0987	1.0000	0.8813	0.8692
15	0.6426	0.4799	0.7879	0.0347	0.9352	0.9513	0.9356
16	0.5813	0.4510	0.7768	0.0000	0.8389	0.9967	0.9993
17	0.6006	0.5442	0.8633	0.0145	0.6674	0.9979	0.9591
18	0.6302	0.5674	0.8783	0.0211	0.3659	0.9843	0.9479
19	0.6262	0.6944	0.8676	0.0801	0.2402	0.9207	0.8311
20	0.6577	0.8245	0.8681	0.2104	0.0901	0.8028	0.6705
21	0.6491	0.9968	0.7737	0.3189	0.0469	0.6887	0.4859
22	0.6827	0.9752	0.6655	0.4325	0.0002	0.5902	0.3620
23	0.5497	0.8028	0.5793	0.5439	0.0003	0.4993	0.3586
24	0.3329	0.6518	0.3725	0.6162	0.0000	0.4117	0.3242
min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0257
max	1.0000	0.9968	1.0000	1.0000	1.0000	0.9979	0.9993

Table 1
THE ARITHMETICAL MEAN OF REAL-TIME DATA BETWEEN 2011 AND 2014 IN DIFFERENT AREAS IN BUCHAREST

The average data for the hours of the day is presented in table 1. The data is normalized.

It can be observed in figure 1 that the concentrations for nitrogen dioxide have a 2-peak profile (one during the evening and one in the morning), which is a normal behavior for a city with the size of Bucharest. The diurnal variations for nitrogen and carbon monoxide are not so important.

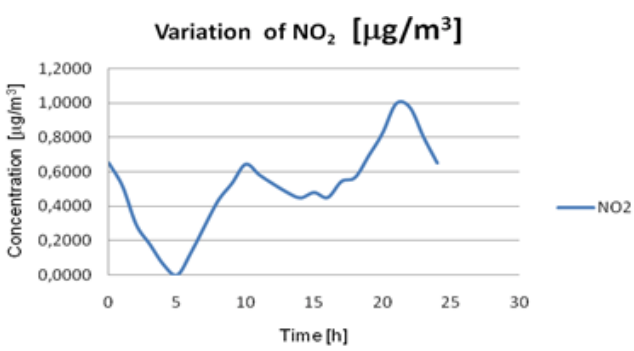

Fig. 1. NO_2 Variation

Results and discussions

The data used were transformed to be normalized and after were introduced in SPSS to build an artificial neural network of the Radial Basis Function type (fig. 2).

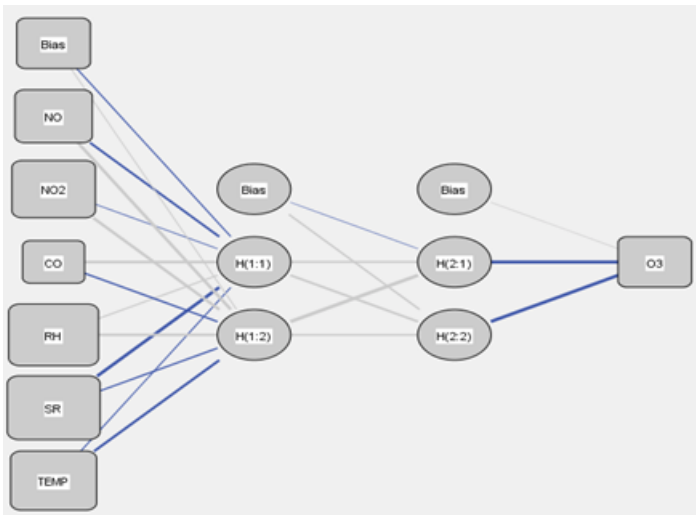

Fig. 2. Neural Network Model for Predicting Concentrations of O_3 Immission

Analyzing the obtained neural network in SPSS can be observed that it suggests an optimal activation function represented by a hyperbolic tangent, not only for the hidden levels as well as for the output.

To determine this function, we developed a program in the free software Scilab, which uses the Levenberg-Marquardt algorithm, described in [8].

Given the results obtained by analyzing the non-dimensional measured data presented in table 1, using artificial neural networks, we obtained a valid mathematical model for every season and every area, (fig. 3):

$$\begin{aligned} O_3 &= -2.59 \cdot th(NO) + 1.02 \cdot th(NO2) + 0.83 \cdot th(CO2) + \\ &+ 0.20 \cdot th(RH) - 1.22 \cdot th(SR) - 1.33 \cdot th(TEMP) - \\ &- 3.60 \cdot e^{-NO} + 2.21 \cdot e^{-NO2} + 1.13 \cdot e^{-CO} + 0.98 \cdot e^{-RH} - \\ &- 1.68 \cdot e^{-SR} - 2.65 \cdot e^{-TEMP} + 4.23 \end{aligned} \quad (2)$$

The model validation was done using other sets of input-output data, (table 2). The results confirm the possibility of using artificial neural networks in order to predict the concentration of ozone in locations where there is no meteorological data. It also proves the accuracy of such a model, (fig. 4).

The main goal of the model is to maximize its ability to generalize and this feature is connected to the sensitivity of the model. This rule applies for most applications in general, and for air quality in particular.

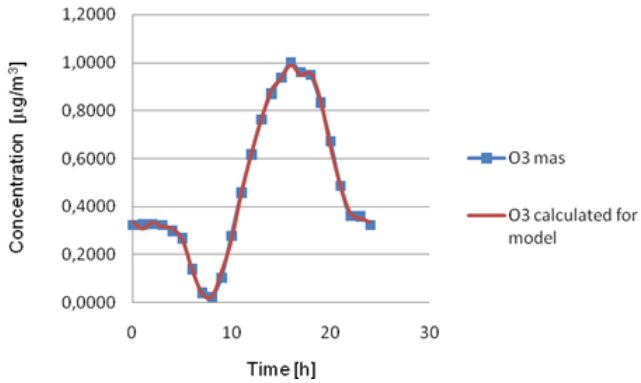

Fig. 3. Measured and Estimated Ozone, using the Hyperbolic Tangent Function

How the model sensitivity is influenced by input data changes is important and our numerically evaluation method is presented below.

We perturbed the input values (table 1) with a percentage varying up to 50% and we validated the model obtained with the values measured at the table 2.

To analyze the model sensitivity two statistical indices of the Environmental Protection Agency (EPA - USA), [16], have been used:

- the MNBE (Mean Normalized Bias Error) – with a confidence threshold of 5 - 15% calculated by the formula:

Table 2
THE ARITHMETICAL MEAN OF REAL-TIME DATA FROM 2014 AND 2015 IN OTHER AREAS IN BUCHAREST

Time	NO	NO2	CO	RH	SR	TEMP	O3 mas	O3 calc
0	0.3562	0.6771	0.3994	0.6222	0.0000	0.4060	0.4122	0.3562
1	0.3380	0.5104	0.3420	0.6300	0.0001	0.3318	0.3364	0.3380
2	0.1829	0.2817	0.1244	0.7390	0.0002	0.2528	0.2565	0.1829
3	0.0955	0.1649	0.0663	0.8100	0.0002	0.1843	0.1855	0.0955
4	0.0297	0.0591	0.0248	0.8634	0.0003	0.1259	0.1273	0.0297
5	0.0000	0.0000	0.0000	0.9362	0.0002	0.0652	0.0658	0.0000
6	0.1087	0.1172	0.0376	0.9832	0.0008	0.0234	0.0232	0.1087
7	0.2959	0.2824	0.1874	1.0000	0.0142	0.0000	0.0000	0.2959
8	0.7297	0.5033	0.6108	0.9402	0.0445	0.0511	0.0513	0.7297
9	1.0000	0.6518	1.0000	0.8043	0.0763	0.1656	0.1677	1.0000
10	0.9130	0.7663	0.9416	0.6637	0.1097	0.3047	0.3111	0.9130
11	0.7615	0.6799	0.8590	0.5064	0.1454	0.4521	0.4629	0.7615
12	0.6896	0.5836	0.8389	0.3722	0.1854	0.5915	0.6042	0.6896
13	0.6752	0.5416	0.8505	0.1951	0.7811	0.7676	0.7670	0.6752
14	0.6528	0.4772	0.7822	0.0971	1.0000	0.8788	0.8698	0.6528
15	0.6314	0.5068	0.7779	0.0311	0.9367	0.9534	0.9466	0.6314
16	0.5618	0.4537	0.7461	0.0000	0.8380	1.0000	0.9986	0.5618
17	0.5910	0.5415	0.8431	0.0099	0.6732	0.9958	1.0000	0.5910
18	0.6212	0.5835	0.8763	0.0238	0.3706	0.9781	0.9874	0.6212
19	0.6281	0.6959	0.8796	0.0742	0.2480	0.9193	0.9315	0.6281
20	0.6439	0.8070	0.8517	0.1975	0.0929	0.8013	0.8153	0.6439
21	0.6552	0.9936	0.7998	0.3129	0.0116	0.6823	0.6935	0.6552
22	0.6940	1.0000	0.6969	0.4305	0.0001	0.5803	0.5895	0.6940
23	0.5960	0.8742	0.6217	0.5349	0.0002	0.4923	0.5003	0.5960
24	0.3562	0.6771	0.3994	0.6222	0.0000	0.4060	0.4122	0.3562

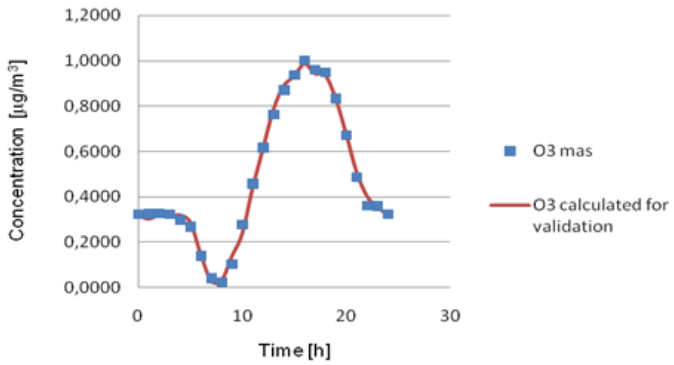

Fig. 4. Measured and Estimated Ozone for Real Data in Table 2

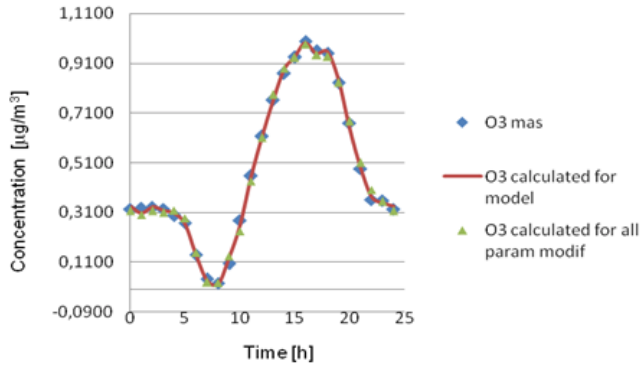

Fig. 5. The Robustness of the Model to Predict Ozone Concentration

$$MNBE = \frac{1}{N}\sum_{t=1}^{N}\frac{c_{mas.}(x,t)-c_{ev.}(x,t)}{c_{mas.}(x,t)}$$

- and the MNGE (Mean Normalized Gross Error) – with a confidence threshold of 30 - 35% and calculated with the formula:

$$MNGE = \frac{1}{N}\sum_{t=1}^{N}\frac{|c_{mas.}(x,t)-c_{ev.}(x,t)|}{c_{mas.}(x,t)}$$

The results obtained for the calculated indices are for the MNBE a 10% confidence threshold and for MNGE a 31% confidence threshold.

Our tests show that small differences of input data provide rather small changes in the output values. This is a feature for rather good ability to generalize. The robustness of our values and the good applicability of the model to predict ozone are presented in figure 5.

Conclusions

The goal of this paper is to determine a mathematical model, which shows the relationship between ozone levels together with other meteorological data and air quality. The applicability is valid for any season and for any area. The model is based on real time data measured in Bucharest and its surroundings.

The European legislation imposes thresholds for pollutant concentrations. Unfortunately, these are systematically exceeded. Some of the main causes are the traffic and the industrial influence.

Using artificial neural networks, we mathematical modeled the air quality for predicting ozone concentrations in 24-hours. For this we used exogenous variables such as ozone precursors: NO, NO_2, CO and other meteorological data, e.g. global solar radiation, relative humidity and temperature. Our model performed well for the tested period and it had a fair accuracy. We validate the model using independent data from different locations in Bucharest and its surroundings and for other periods of time. To verify the generalization ability of our mathematical model, we numerically evaluated its sensitivity to variations in input data and the tests obtained pretty good results.

References

1. GAVRILA, C., VARTIRES, A., DAMIAN, A., Rev. Chim. (Bucharest), **67**, no. 6, 2016, p. 1080.
2. MIRCEA, I., FALUP, O., IVAN, R., IONEL, I., Rev. Chim.. (Bucharest), **66**, no. 2, 2015, p. 247.
3. NENCIU, F., VAIREANU, D.-I., Rev. Chim.(Bucharest), **65**, no. 5, 2014, p. 565.
4. MARCU, N., CRISTEA, M., MEGHISAN, G.M., DASCALU, D., NASTA, L.N., Rev. Chim. (Bucharest), **67**, no. 6, 2016, p. 1195.
5.BRIGGS, D.J., HOOGH, C., GULLIVER, J., WILLS, J., ELLIOTT, P., KINGHAM, S., SMALLBONE, K., Science of The Total Environment, 253, nr. 1–3, 2000, p. 151.
6. YI, J., PRYBUTOK, V.R., Environmental Pollution, 92, nr. 3, 1996, p. 349.
7. GAVRILA, C., ARDELEAN, F., COMAN, A., BURCHIU, E., Mathematical Modeling in Civil Engineering, nr 2, 2013, p. 25.
8. GAVRILA, C., COMAN, A., GRUIA, I., ARDELEAN, F., VARTIRES, A., Romanian Journal of Physics, 61 nr. 5-6, 2016, p. 1067.
9. GAVRILA, C., TEODORESCU, N., GRUIA, I., Journal of Water Science and Technology: Water Supply, 13, nr. 4, 2013, p. 883.
10. GAVRILA, C., FRUNZULICA, R., PATRASCU, R., U.P.B. Sci. Bull., Series A, 76, nr. 1, 2014, p. 3.
11. DUMITRACHE, I., CONSTANTIN, N., DRAGOICEA, M., Reele neurale. Identificarea si conducerea proceselor, Ed. MatrixRom, Bucuresti, 1999.
12. HAGAN, M.T., MENHAJ, M., IEEE Transactions on Neural Networks and Learning Systems, nr. 5, 1994, p. 989.
13. *** IBM-SPSS (Statistical Package for the Social Sciences)-Trial version 20.0.0 (www.spss.ie/software/index.html).
14. *** Scilab free software, (http://www.scilab.org/).
15.*** The National Network of Air Quality Monitoring (http://www.calitateaer.ro/valori.php).
16. Environmental Protection Agency (EPA - USA) (https://www.epa.gov/lead/lead-laws-and-regulations)

Manuscript received: 22.09.2016