Forecast of Air Quality Based on Ozone by Decision Trees and Neural Networks

Conference Paper · October 2012

DOI: 10.1007/978-3-642-37807-2_9

Ivan Olmos Pineda
Benemérita Universidad Autónoma de Puebla
21 PUBLICATIONS **10** CITATIONS
SEE PROFILE

David Pinto
Benemérita Universidad Autónoma de Puebla
176 PUBLICATIONS **1,047** CITATIONS
SEE PROFILE

Helena Gomez Adorno
Universidad Nacional Autónoma de México
47 PUBLICATIONS **531** CITATIONS
SEE PROFILE

Yuridiana Alemán
Benemérita Universidad Autónoma de Puebla
17 PUBLICATIONS **52** CITATIONS
SEE PROFILE

Forecast of Air Quality Based on Ozone by Decision Trees and Neural Networks

Nahun Loya, Iván Olmos Pineda, David Pinto,
Helena Gómez-Adorno, and Yuridiana Alemán

Benemérita Universidad Autónoma de Puebla
Puebla, México
{nahun.loya,helena.adorno,yuridiana.aleman}@gmail.com,
{iolmos,dpinto}@cs.buap.mx

Abstract. In this paper we explore models based on decision trees and neural networks models for predicting levels of ozone. We worked with a data set of the Atmospheric Monitoring System of Mexico City (SIMAT), which includes measurements hour by hour, between 2010 to 2011. The data come from of three meteorological stations: Pedregal, Tlalnepantla and Xalostoc in Mexico city. The data set includes 8 parameters: four chemical variables and four meteorological variables. Based on our results, it's possible to predict ozone levels with these parameters, with an accuracy of 94.4%.

Keywords: Decision Trees, C4.5, Neural networks, Ozone.

1 Introduction

Big cities such as Los Angeles, Tokyo, Moscow, and Mexico City have serious problems of air pollution. These cities monitoring the air quality in the troposphere, with the aim to detect ozone concentration emitted by inhabitant [1], and record its progress. Globally this pollutant is one of the most important, which is a triatomic molecule of oxygen [2].

The study of gas concentration is crucial, because if some positive variations of levels of ozone are detected, should suggest that controls on emissions are having a positive effect in the environment, and the absence of such trends suggest the need to change that controls.

This work aims to build models for predict the air quality based on the ozone as primary pollutant. We consider a set of chemical and atmospheric attributes that authors and experts of air pollution have shown to be predictors for air quality. The prediction is done using models of multilayer neural networks and decision trees, using the Weka toolkit as data mining software [3].

The main objective of this work is to find a good model based on neural networks and decision trees, with the aim to predict air quality. We worked with the following attributes in our experiments: Ozone (O_3), Carbon monoxide (CO), Nitrogen Dioxide (NO_2), Sulfur dioxide (SO_2), Temperature (TMP), Relative humidity (RH), Speed wind (WSP), and Wind direction (WDR). The

I. Batyrshin and M. González Mendoza (Eds.): MICAI 2012, Part I, LNAI 7629, pp. 97–106, 2013.

data considered in this work comes from of the Atmospheric Monitoring System of Mexico City (SIMAT, for its acronym in Spanish).

We proposed a methodology where a descriptive statistics for determining the general behavior of the data is initially performed. Also, we consider a pre-processing phase, including: a data cleaning, a data integration, and a data reduction. Then, a training phase is performed, taking as input the attributes selected in the previous phase. As result of this process, some classification models for predict the air quality were obtained. These models were tested with a cross validation.

This paper is organized as follows: section 2 shows several studies that have been conducted to predict levels of air quality using statistical tools, and other schemes such as neural networks. Section 3 describe the case of study, as well as the locations of weather stations considered in this study. Section 4 describes how the integration of data is done and the data clean is performed. Finally, section 5 presents the results obtained in our experiments.

2 Related Work

Several works have been conducted to predict the air quality with respect to the ozone. For example, Seinfeld *et al.*[4] shows how it is possible to measure tendencies of ozone levels based on estimators, such as: maximum daily, and average of the maximum daily in a period of 3 days.

In Mexico, there are attempts of many different scientific fields and institutions, which are trying to evaluate the ozone pollution [5], [6], [7]. In general, these works try to predict areas with high risk for inhabitants of the metropolitan area in Mexico City (ZMVM), using extreme values of pollutants.

Another works use data from only one monitoring site in Mexico City, where high levels of ozone are recorded and analyzed Garfias*et al.*[8]. In such investigations, authors proposed three different models for predict concentration of ozone based on 19 semi-annual observations.

On the other hand, there are different statistical techniques for predicting levels of ozone. For example, Aguirre *et al.*[9] shows the importance of neural networks in this task, where a multilayer perceptron model was used to predict maximum levels of ozone in the Pais Vasco, Spain.

Barai *et al.*[10] proposed neural networks to predict air quality, which works with a limited number of data, and this model is capable to work with noise. They proposed different models of neural networks for predicting air quality, with a very acceptable accuracy, using only two data sets: US Environmental Protection Agency (US. EPA) and Tata Energy Research Institute (TERI). These works are good examples of what it is possible to do with neural networks for predicting air pollutants.

In this paper, we explore important approaches with the aim to learn patterns that predict levels of ozone, such as multilayer perceptron neural networks, and classification trees models, including C4.5, and Random forest.

In the next section are explained details of the dataset used in this work.

3 Case of Study

In Mexico city several meteorological stations were created, which reports levels of pollutants hour by hour. In this study we focus in the dataset that come from of the SIMAT network, including data between the period of January 2010 to December 2011. This study includes data from three meteorological stations: Pedregal, Tlalnepantla and Xalostoc, which are strategically located in the metropolitan area of Mexico city. The size of the dataset is 420,480 records, including 8 attributes, recollected in 2 years of study. In this work we consider a set of attributes described by Seinfeld [4], which are predictors of ozone levels. These attributes are shown in Table 1. As we can see, the first column represents pollutants (Ozone O_3, Carbon monoxide CO, Nitrogen Dioxide NO_2, Sulfur dioxide SO_2) and atmospheric variables (Temperature TMP, Relative humidity RH, Speed wind WSP and Wind direction WDR). The highest values admitted by each pollutant according to the Official Mexican Norm (NOM-1993)[11] are present in the second column. The third column shows the measurement units for each variable [2]. Finally, last column represents the number of missing values by attribute.

Table 1. Set of abbreviations used in this paper

Pollutant/ Atmospheric variable	Value of the Standard on accordance with (NOM-1993)	Unit of Measure	Missing Values
Ozone (O_3)	0.11 ppm	Parts per million (ppm)	0
Carbon monoxide (CO)	11 ppm	Parts per million (ppm)	1088
Nitrogen Dioxide (NO2)	0.21 ppm	Parts per million (ppm)	1011
Sulfur Dioxide (SO2)	0.13 ppm	Parts per million (ppm)	792
Temperature (TMP)		Celsius Grades (C)	758
Relative Humidity (RH)		Percent (%)	369
Wind speed (WSP)		Meters over second (m/s)	925
Wind Direction (WDR)		North Grados	7925

Based on the data described above, we propose a methodology for buil–ding models with ANNs and decision trees. The next section describe the data preprocessing phase implemented in this work.

4 Data Preprocessing

In this section we present a data overview based on descriptive statistics, in order to give a general data perspective. This analysis let us to propose a data integration phase, and how missing values can be estimated.

4.1 Descriptive Statistics

Descriptive statistic is an important mathematical tool because allow us to understand the data behavior. With the aim to known trends of air pollutants, we compute statistical tendencies of each pollutant, which are shown in Table 2. In this table, "Min", "Max" and "Mean" are the minimum, maximum and mean of each pollutant per each meteorological station.

Table 2. Data summary per each meteorological station

	Pedregal (DSN=189)			Tlalnepantla (DSN=121)			Xalostoc (DSN=86)		
Attribute.	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.
O_3	0.000	0.033	0.182	0.000	0.026	0.183	0.000	0.024	0.150
CO	0.000	0.515	2.900	0.000	0.920	5.100	0.000	1.102	11.30
NO_2	0.001	0.025	0.115	0.004	0.033	0.161	0.000	0.034	0.138
SO_2	0.000	0.005	0.097	0.000	0.009	0.283	0.000	0.007	0.143
TMP	1.00	16.08	31.50	1.60	17.92	36.00	1.200	17.50	33.80
RH	0.000	45.83	95.00	1.00	45.46	100.00	1.00	46.80	100.00
WDR	0.000	-	360	0	-	360	0	-	360
WSP	0.000	1.80	7.60	0.00	2.00	8.20	0.00	2.0	10.90

In this table, we can see that Pedregal is the most contaminated meteorological station in this study. We can see that 189 times the NOM-1993 with respect of ozone levels was overcoming (DSN). In Tlalnepantla, the meteorological station recorded 121 times where the NOM-1993 was exceeding. In Xalostoc, located at the north of the Mexico City, the meteorological station recorded the lowest rate of contamination in the observation period, with only 86 events. Moreover, in this analysis is evident that most observations show that levels of air pollutants are acceptable. Note that these observations not include atmospheric variables such as the wind direction, because measures of central tendency such as mean and median are not good estimators.

One of the main objectives in this work is to predict with a good confidence level the air quality for the zone area surrounding each meteorological station considered in this work. In the following sections we present how it is possible to apply technics of data mining with the goal to discover such tendencies, including phases such as data cleaning and data integration.

4.2 Preprocessing Phase: Data Cleaning, Missing Values and Data Integration

The input database have missing values, non-relevant information (chemical variables), data from other stations not considered in this study, data in different unit measures, and noise. Therefore, all these problems were overcoming with different tools implemented in the R software [12]. Moreover, each database instance was categorized as good, regular, bad, very bad, and highly bad, according with the NOM-1993, and according with the ozone levels. In Table 3 are shown the ranges used in this work.

Table 3. Ozone level ranges used to categorize each instance

O_3	*Class*	Qualifier air quality
0.000-0.055	Green	Good
0.056-0.110	Yellow	Regular
0.111-0.165	Orange	Bad
0.166-0.220	Red	Very Bad
> 0.220	Purple	Highly bad

Although it is known that the selected classifiers can work with noise and missing values, we implement a phase where these problems are solved. Since data is continuous with respect of time, we implement a simple technique based on a linear interpolation, with the aim to detect outlier values and predict missing values. In both cases, if a value v_n is atypical or unknown, then it is computed based on the neighbors v_{n-1} and v_{n+1}.

4.3 Attribute Selection

Before to proceed with the training process, it is necessary to know the attributes that are more relevant for that process. In this phase, we propose to use an attribute evaluation technique based on the chi-square metric, where is computed the importance of each attribute considering the chi-square value with respect to the class attribute. The attribute with a value closer to 1 is the attribute that provides more information for the prediction class. In Table 4 are shown the values obtained in this process. With these results, we have a good idea of the importance of each attribute.

Table 4. Results using a χ^2 test for each meteorological station

Attribute	PEDREGAL	TLALNEPANTLA	XALOSTOC
HOUR	3.8+-0.4	3.2+-0.7	3.3+-0.4
CO	3.1+-0.5	2.9+-0.9	2.0+-0.0
NO_2	5.9+-0.3	3.7+-1.3	6.1+-0.8
SO_2	2.1+-0.3	4.8+-1.6	6.5+-2.2
TMP	1.0+-0.0	1.0+-0.0	1.0+-0.0
RH	5.1+-0.3	5.6+-0.4	5.0.+-0.6
WDR	7.6+-0.4	7.9+-0.3	7.3+-0.4
WSP	7.4+-0.4	6.9+-0.3	4.8+-0.7

On the other hand, in our analysis we can see that the dataset has a imbalance of classes. The imbalance produces poor results for predicting tasks. Then, we implement a phase where this problem was overcoming.

4.4 Removing the Class Imbalance

At this stage in our process, the original dataset has been preprocessed, removing noise, inconsistencies, etc. The next step consist of to perform a stratified sampling, where the strata are the seasons: spring, summer, fall and winter. This process try to preserve the distribution of the data with respect to the time. This process was implemented using a script in "awk", where the program select an instance in the dataset with respect to a probability p (this value is defined by the user). Also, the balancing of the classes is preserved with respect to the total sample size: if an instance to be selected belongs to a class where have been achieved the maximum number of instances, then the instance is not selected. With this simple rule, we preserve the balancing of the classes.

For our experiments, three different samples per each meteorological station were generated, with the aim to perform cross validation. In the next section we show the results obtained in our experiments.

5 Experimental Results

In our experiments, we generate three balanced samples of the original data set from each meteorological station, each of one with 42000 records. In our experiments, we worked with two types of machine learning algorithms, such as neural networks (multilayer perceptron), and decision trees (C4.5 and the Random forest). We used the Weka tool kit in our experiments.

The training of the neural network was performed with different configurations, all of them with 10 fold cross validation. Some parameters considered in these experiments include a learning rate = 0.3, momentum = 0.2. Different topologies were tested, with 1, 2, and 3 hidden layers (HL), with 3, 4, 5, 8, and 9 neurons ($\#Neu$) per layer. Also, we consider a topology with a hidden layer, and a $= \frac{No.Att+No.Class}{2}$ neurons, where $No.Att$ is the number of attributes, and $No.Class$ represents the number of classes. Further, we experimented with different epochs: 500, 1000, and 2000. Our results are shown in Table 5, where each entry is the percentage accuracy.

On the other hand, we explored with two algorithms based on decision trees, such as: C4.5 and Random Forest. We also used the implementation available in Weka. Just as in neural networks, we perform a search of parameters for each algorithm. In the case of C4.5, we used the following parameters: $factor =$ 0.05, $MinNumObj = 2$ and $Unpruned = FALSE$. Furthermore, in the case of the Random Forest algorithm, the parameters used in our experiments were: $MaxDepth = 10$, $Debug = False$, $NumTrees = 50$, $Seed = 1$. Results with these algorithms are shown in Table 6.

As we can see in Tables 5 and 6, the best results are obtained with the classification trees. According with the Table 5, the best result for the Pedregal station is with a multilayer perceptron neural network, with one hidden layer and 8 neurons, where the accuracy is 87.8%. For the case of Tlalnepantla station, the best result was obtained with a similar configuration: one hidden layer and 9 neurons, with an accuracy of 86.7%. Finally, with a identical topology, we obtain

Table 5. Results obtained with a Multilayer Perceptron Neural Network

		1 HL.			**2 HL.**			**3 HL.**		
#Neu.	**Epochs.**	Ped.	Tla.	Xal.	Ped.	Tla.	Xal.	Ped.	Tla.	Xal.
3	500	84.9	81.7	89.6	83.3	82.0	87.6	83.5	81.3	88.2
	1000	85.1	81.9	89.4	84.5	81.6	87.7	83.6	80.4	88.3
	2000	85.3	81.1	89.3	83.4	81.2	87.7	84.5	80.0	88.1
4	500	87.3	85.0	91.9	83.9	84.3	90.1	84.8	83.7	90.5
	1000	86.7	85.2	91.8	83.6	84.1	90.3	84.1	83.3	90.9
	2000	87.1	85.1	92.0	84.3	84.4	91.8	84.7	83.5	90.7
5	500	87.3	85.2	92.1	84.9	85.6	90.7	83.8	84.6	90.9
	1000	86.6	85.1	92.0	84.9	84.6	91.2	84.5	84.4	90.9
	2000	86.4	85.0	91.9	84.8	84.8	91.0	84.7	85.1	90.7
8	500	88.7	85.6	93.2	86.5	85.1	92.2	85.3	84.2	91.4
	1000	87.7	85.1	93.2	87.3	85.7	91.8	85.0	85.0	90.7
	2000	85.0	85.5	93.2	86.0	85.4	91.8	85.4	83.9	91.4
9	500	87.6	86.7	93.6	86.5	85.3	91.9	84.2	85.2	91.7
	1000	88.6	86.4	93.5	86.7	84.3	92.1	84.5	84.7	92.1
	2000	88.0	86.1	93.3	85.9	85.4	92.2	84.9	86.0	91.9
Avg.	500	86.4	85.0	92.6	86.2	85.1	91.5	82.9	84.5	91.1

Table 6. Results with the C4.5 and the Random Forest algorithms

Algorithm	Ped.	Tla.	Xal
C4.5	91.6	88.2	93.3
Random Forest	92.3	89.6	94.4

an accuracy of 85.3% for the case of Xalostoc station. It is evident that those results are consistent for all cases.

On the other hand, the best results obtained with decision trees were with the Random Forest algorithm, as is shown in Table 6. The accuracy achieved in these experiments is 92.3% for Pedregal station, 89.6% for Tlalnepantla, and 94.4% for Xalostoc. However, based on our experiments, the global accuracy of C4.5 was very similar to the Random Forest algorithm.

Remember that those results were obtained from data separated by seasons. These models can be used as a reference for predicting the behavior of air pollution with respect to seasons, but if we can predict in other timescales, then are necessary more experiments.

In this way, we proposed organize data per hour in order to have more precision at the moment to predict the pollution per day. Hence, we generate new balanced data samples from the preprocessed dataset. In those new experiments, we used again the decision trees algorithms, and the multilayer perceptron neural network, but with a topology with one hidden layer only. This decision is based on the previous results, because with more hidden layers did not improve the final accuracy.

Table 7. Results computed with the C4.5, the Random Forest and the Multilayer Perceptron algorithms, the data are obtained hour by hour

	C4.5.			Random Forest			MLP. 8 Neu.			MLP. 9 Neu.		
HOUR	PED	TLA	XAL	PED	TLA	XAL	PED	TLA	XAL	PED	TLA	XAL
1	99.73	99.86	99.73	99.59	99.86	99.59	99.73	99.86	99.73	99.73	99.86	99.73
2	99.86	100	99.86	99.86	100	99.86	99.86	100	99.86	99.86	100	99.86
3	99.86	100	99.86	99.86	100	99.86	99.86	100	99.86	99.86	100	99.86
4	100	100	100	100	100	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100	100	100	100	100	100
6	100	100	100	100	100	100	100	100	100	100	100	100
7	100	100	100	100	100	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100	100	100	100	100	100
9	100	100	100	100	100	100	100	100	100	100	100	100
10	98.35	99.45	98.08	97.81	99.45	97.81	97.81	99.45	97.67	97.94	99.45	97.53
11	89.44	90.12	87.52	88.75	90.26	88.75	90.12	91.08	89.85	89.44	90.95	89.30
12	75.31	80.52	75.58	76.95	80.11	76.95	76.95	82.58	75.31	76.13	82.17	75.99
13	66.80	70.51	65.98	70.92	70.10	70.92	69.68	72.02	72.57	70.10	71.47	72.98
14	71.06	69.96	68.86	69.96	69.14	69.96	71.88	71.47	68.18	74.49	70.10	70.10
15	70.78	73.39	67.76	73.39	71.60	73.39	69.68	73.25	69.14	70.23	72.70	69.41
16	67.22	73.25	70.64	72.57	71.74	72.57	70.42	73.48	70.10	71.74	72.66	71.10
17	73.66	78.19	74.35	76.54	73.25	76.54	73.39	75.03	75.99	72.98	73.94	76.82
18	74.35	81.62	85.60	86.69	80.38	86.69	81.62	81.48	84.64	80.25	82.44	85.46
19	89.44	94.79	94.79	95.47	94.51	95.47	88.20	93.96	94.24	88.48	94.10	93.83
20	93.83	99.04	97.39	97.39	99.04	97.39	94.10	99.04	96.43	94.10	99.04	96.57
21	98.22	99.45	98.77	98.63	99.35	98.63	98.08	99.45	98.77	98.08	99.45	98.77
22	98.90	100	99.59	99.59	100	99.59	98.90	100	99.59	98.90	100	99.59
23	99.73	100	99.73	99.73	100	99.73	99.73	100	99.73	99.73	100	99.73
24	99.73	100	99.73	99.73	100	99.73	99.73	100	99.73	99.73	100	99.73

In Table 7 are shown the results. We observed in these new experiments that in the early hours and the last hours of each day, the accuracy achieved by the algorithms are very high (close to 100%). However, for noon the accuracy sensible down, around 66%. These results are comprehensible: on one side, in hours between 12 am and 7 am, the human activity is low, then many factories, contamination produced by vehicles, and other factors are low. Because of this, in this lapse of time the air pollution is low, and then the accuracy obtained by the classifiers is high. After that, many human activities increase considerably, and then many contaminants are released into the atmosphere. Generally, it is required a period of 2 or 3 hours for pollutants build up in the atmosphere. Result of the above, between 12 pm and 5 pm, the air pollution is highly variable, depending of many factors, including natural factors (atmospheric conditions), and related with human activities. Because of this, the accuracy down sensible.

Finally, the last results suggest that the behavior of air pollution is moving in three stages. Because of this, we proposed new experiments where data is clustered in blocks of 8 hours: 12 a.m. to 8:59 a.m., 9 a.m. to 5:59 p.m., and 6 p.m. to 11:59 p.m. The results obtained in our experiments are shown in Table 8.

Table 8. Results obtained with data clustered in periods of 8 hours

	C4.5.			Random Forest.			MLP. 9 Neu.		
Period	PED	TLA	XAL	PED	TLA	XAL	PED	TLA	XAL
1-8 HRS	99.93	99.98	99.93	99.93	99.98	99.93	99.93	99.98	99.93
9-17HRS	80.09	81.35	80.68	81.84	82.51	82.90	81.17	82.20	81.38
18-24HRS	93.62	96.58	96.65	94.46	96.46	96.77	94.31	96.50	96.61

If we compare the results obtained in Table 7 and Table 8, we can see that the accuracy obtained for noon is increased. This mean that the new distribution of data is favorable for the classifiers, and then they can predict with more precision the air pollution. In these experiments we can see that the accuracy obtained between 9 a.m. and 17:59 p.m. is above to 80%, and the accuracy for the others periods is very high. The best results are obtained with the Random Forest algorithm, but are very close between them.

Another interesting result that can be seen from the results is the fact that the accuracy reported by the two different models (neural networks and decision trees) is very similar. As conclusion of this, it is possible to use any of the two models.

6 Conclusions and Future Work

In this research we explored models based on neural networks and decision trees capable to predict levels of ozone as an air pollutant.

We explored three different ways of building classification models for each meteorological station: the first one based on seasons observations, the second one with a observation hour by hour, and the third one with an observation in periods of 8 hours per day. In general and based on our results, we observed that the multilayer perceptron neural network and algorithms such as C4.5 and Random Forest are capable to predict the ozone with a similar accuracy.

However, it is possible to conclude that the third model, where data was organized in clusters of 8 hours, let us to build predicts with a very good accuracy. In this case, both neural networks or decision trees can be used to predict the ozone levels.

We know that there are many variables that were not considered in this study, and that may influence for the ozone levels in Mexico city, such as volcanic ash emitted by the Popocatepetl (an active volcano located 60 km southeast of Mexico city), solar radiation, particles floating into the air with less than 10 micrograms, pluvial precipitations, between others. As future work, we can include this variables, with the aim to increase the global accuracy. Moreover, it is possible to explore others approaches for predicting the ozone levels, such as support vector machines and bayesian networks.

References

1. Reyes, H., Vaquera, H., Villaseñor, J.: Estimate of tendencies in high levels of urban ozone using the quantiles of the distribution of generalized extreme values (gev) (2007) (en revisin)
2. Simat, D.: Biblioteca Virtual, México (2012), http://www.sma.df.gob.mx/simat
3. Mark Hall, Eibe Frank, G.H.B.P.P.R.I.H.W.: The weka data mining software: An update; sigkdd explorations. Master's thesis, Tesis de Maestría en Ingeniería Ambiental-UNAM, vol.11(1) (2012)
4. Seinfeld, J.: Comittee on tropospheric ozone formation and measurement; Board on Environmente Studies and Toxicology; Board on Atmospheric Sciences and Climate; Commission on Geosciencies, Environment and Resources; National Research Council, Rethinking the on ozone problem in urban and regional air Pollution. National Academic Press, Washington (1991)
5. Bravo, H.: La contaminación atmosférica por ozono en la zona metropolitana de la ciudad de méxico: evolución histórica y perspectivas. IX Comisión Nacional de los Derechos Humanos (1992)
6. I.N.E.G.I.: Estadísticas del Medio Ambiente. Semarnap, México (1999)
7. Molina, L.T.: The impacts of magacities on air pollution, environmetal aspects of urbanization, Goteborg Sweden (2004)
8. Garfias, M., Audry, J., Garfias, F.: Ozone trend analysis at pedregal station in the metropolitan area of mexico city. J. Mex. Chem. Soc. 49(4), 322–323 (2005)
9. Barron, E.A.A.A.L.J.R.: A sistem for forecast of the maximun ozone levels. Atmospheric Environment 38, 4689–4699 (2004)
10. Barai, S., Dikshit, A., Sharma, S.: Neural Network Models for Air Quality Prediction: A Comparative Study. In: Saad, A., Dahal, K., Sarfraz, M., Roy, R. (eds.) Soft Computing in Industrial Applications. ASC, vol. 39, pp. 290–305. Springer, Heidelberg (2007)
11. D.F., G.: Gaceta Oficial No. 129, pp. 1–14 (2006)
12. Proyect, R.: Biblioteca Virtual, USA (2012), http://www.r-project.org/