2D Mapping by Kohonen Networks of the Air Quality Data From a Large City†

Neva Grošelj,*,‡ Jure Zupan,‡ Silvia Reich,§ Laura Dawidowski,|| Darío Gomez,|| and Jorge Magallanes||

National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia, Escuela de Ciencia y Tecnología, Universidad de San Martín, San Martín, Argentina, and Unidad de Actividad Química, Comisión Nacional de Energía Atómica, San Martín, Argentina

Received October 29, 2003

The 15-variable environmental data (7 concentrations: CO, SO_2, O_3, NO_x, NO, NO_2, particulate matter smaller than 10 μ (PM_{10}), and 8 weather data: cloudiness, rainfall, insolation factor (Isf_i), temperature, pressure at two locations, and wind intensity with direction) in a period of 45 days with 1-h intervals were extracted from a larger database of concentrations recorded in minute intervals for the same time period. The monitoring site was located in the City of Buenos Aires in a relatively heavy traffic crossroad of two avenues. The data required special pretreatment where the hourly content of rain, wind intensity, wind velocity, and cloudiness were concerned. The new variable named insolation factor (relative UV radiation) calculated on the basis of the general meteorological data, the geographic position of the monitoring site, cloudiness, date, and the time of the recording was composed. The relative intensity of UV radiation was modeled by a Gaussian function, multiplied by a cloudiness factor. Based on the 14-variable input and the 1-variable output (ozone) data, first, the clustering of all 980 data records was made. The top map clustering showing the ozone concentration was related to the maps of all 14 variables. The link between O_3 clusters, NO_2, and Isf_i weight levels is shown and discussed. As a preliminary result of this study some of the most interesting correlations between the maps and remaining variables are given.

1. INTRODUCTION

One of the main goals of pollution monitoring and acquiring of various environmental data is to generate new facts and knowledge about the correlation between the concentrations of different pollutants[1−5] (primary and secondary) at a given time and the meteorological data for several hours in advance. These can already be forecast quite precisely. By establishing and knowing the strong correlations or even mathematical models between the data on pollution concentrations and the weather data it is possible to predict various critical situations regarding the excess of threshold limits of several secondary pollutants which enables the authorities to act accordingly. Many authors have already studied the behavior of ozone in relation to other pollutants and weather data. Most of the references are focused on the use of different modeling methods such as Multiple Regression Analysis (MRA),[6] Partial Least Squares regression (PLS),[7] and Error Back-propagation neural networks.[8] These methods are applicable to the modeling and analysis of any data where an effect (for example, pollution of the city or damage to plants) is caused by a number of variables that have a nonlinear influence. To produce better models, some authors have tried to improve the information content of input variables by mapping the original ones into the space spanned by eigenvectors, found from Principal Component Analysis (PCA)[6,9] and by nonlinear mapping using Kohonen Self-organized Maps (SOM).[10−12] In atmospheric science Artificial Neural Networks (ANN) are mainly employed as alternatives to standard and/or complementary modeling techniques to the standard ones.[13] Among various ANN techniques and designs Multi-Layer Perceptron (MLP) with the Error Back-propagation learning procedure are used most often.[5,14−17] A good insight into the beginning of MLP applications in atmospheric science can be found in the publication by Gardner et al.[18] Due to their highly flexible designs the ANNs applications not only are an alternative to standard modeling techniques but also are used for the visualization and preprocessing of data as well. Kolehmainen and co-workers[10,11] have used SOMs and MLP as alternative modeling techniques for the prediction of residual NO_2 concentrations. Magallanes et al.[12] used SOMs for nonlinear scaling of data.

About 1000 data records, each presented as a 15-dimensional vector (concentrations of 7 pollutants: CO, SO_2, O_3, NO_x, NO, NO_2, and PM_{10}, and 8 weather related variables: cloudiness, rainfall, insolation factor, temperature, pressure at two locations, and wind intensity with direction),[1,19−21] were used in our study. We tried to find the correlations between the pollutants and weather conditions with the use of unsupervised ANN—the Kohonen and Counterpropagation neural networks.[22−24]

2. EXPERIMENTAL SECTION

2.1. Acquisition of Data. The monitoring site was located in the city of Buenos Aires, the capital of Argentina (population of about 4 million), which expands on a flat terrain on the right shore of the La Plata river (34°38′ S,

* Corresponding author phone: +386 1 4760279; fax: +386 1 4760 300; e-mail: neva.groselj@ki.si.

† The paper is dedicated to Professor G. W. A. (Bill) Milne, our friend of many years, former Editor-in-Chief of the *Journal of Chemical Information and Computer Sciences*. We have shared many common research and social interests, and we hope to do so in the future.

‡ National Institute of Chemistry.
§ Universidad de San Martín.
|| Unidad de Actividad Química.

10.1021/ci030418r CCC: $27.50
Published on Web 02/17/2004

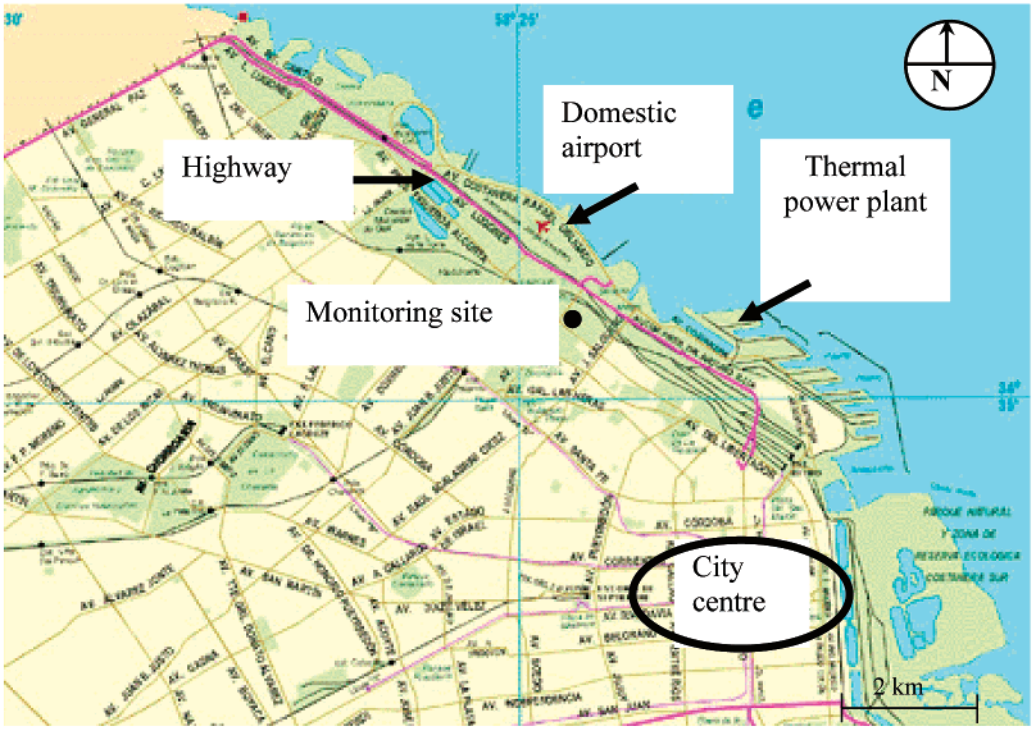

Figure 1. The location of the monitoring site.

58°28′ W) covering a surface of about 200 km^2. The layout of the northern part of Buenos Aires City showing the location of the monitoring site and relevant air pollutant sources is shown in Figure 1.

The measurement cabin was installed on an open space located at about 4 km from the city center. The place is located at the junction of two main avenues, Sarmiento and Figueroa Alcorta. Both avenues hold an important volume of relatively fast traffic for all days of the week. Figueroa Alcorta Avenue runs from SE to NW and holds the traffic running from the city center to the periphery. This traffic increases toward the evening. On the other hand, Sarmiento Avenue runs from NE to SW and holds the incoming traffic to the city center. This traffic is much heavier during morning hours. The highway surrounding the city reaches by the center and runs almost parallel to Figueroa Alcorta at about one kilometer from the monitoring site. Two important air pollution sources are located nearby, the domestic airport (about 1 km in the NNW direction) and the Puerto power plants (about 2 km toward the east).

At the monitoring site, one-minute concentrations of the following air pollutants were measured: carbon monoxide (CO); nitrogen oxides, discriminated as nitric oxide (NO) and nitrogen dioxide (NO_2); ozone (O_3); and sulfur dioxide (SO_2). Hourly average concentrations of suspended particulate matter less than 10 μm (PM_{10}) were also registered. The zero calibration of the equipment was made with high quality synthetic air. Air pollutants were measured with state of the art equipment and methods as summarized in Table 1.

The selection of monitoring location, the operation and maintenance of systems (including calibration), and data validation were carried out according to an internal proceeding based on the QA/QC guidelines of the World Health Organization.[25]

The National Weather Service of Argentina provided hourly surface meteorological data from the domestic airport station. This information includes the following: wind direction and intensity, barometric pressure at two locations (sea level and measurement site), cloud coverage, and ambient temperature. These data were processed by the international QA/QC criteria for hourly meteorological data. Daily rainfall is also included in the data set.

Table 1: Monitoring Methods and Employed Devices

pollutant	method	equipment
CO	nondispersive infrared absorptiometry	HORIBA, APMA-360
NO and NO_2	chemilluminiscence	HORIBA, APNA-360
O_3	ultraviolet absorption	HORIBA, APOA-360
SO_2	ultraviolet fluorescence	HORIBA, APSA 360A
PM_{10}	β-attenuation	MET ONE, BAM-1020

2.2. Preprocessing of Data. *Normalization of Data.* In the time span from August 11, 2001 0.0 a.m. to September 24, 2001 2.0 p.m., there should be 1071 records measured exactly per 1 h. In the mentioned period 91 times either all weather data were missing or the concentrations were not available due to the technical malfunctioning of the instruments. Additionally, due to the failure in the sampling system 63 hourly data (about 6%) of the PM_{10} variable were not available. To retain the PM_{10} data, which is a very important factor influencing the health conditions in large cities,[26,27] we have kept this variable in the process despite the missing values. The missing values were marked with a special key. This key (−9999.) enables our adapted Kohonen and Counter-propagation Neural Network Software[28] to handle the missing data.[29] Hence, the study is made with the 980 time records (= 1071−91) for which all values for 13 variables are available while for the PM10 variable only 917 values are known and 63 are missing.

Due to the fact that multivariate handling of data depends on the normalization of variables we have for all our work applied two different normalizations to all 15 variables x_i,

$i = 1...15$. One in a continuous space

$$x_{ij}^{\text{new}} = \frac{x_{ij}^{\text{old}} - \bar{x}_i}{\sigma_i} \quad (1)$$

and one limited to the interval from 0 to 1

$$x_{ij}^{\text{new}} = \frac{x_{ij}^{\text{old}} - x_{i,\min}^{\text{old}}}{x_{i,\max}^{\text{old}} - x_{i,\min}^{\text{old}}} \quad (2)$$

In the continuation we are showing only the results obtained by the first normalization because they turned out to be superior to the results obtained with data normalized by eq 2.

The PM_{10} values were normalized only for the available data, while the −9999. key for missing values was kept on place.

Amount of Rain. Due to the fact that only the daily amount of rain is given, while for all other variables hourly recordings are available, we prepared four data sets with different interpolations of the rain during the particular 24 h. In the first data set the amount of rain was not considered at all; therefore, 13 instead of 14 input variables described this set. In the second data set, we distributed the daily amount of rain r_{daily} between the 24 h of the corresponding day (from 9 a.m. to 8 a.m. the next day). Each hourly amount of rain r_i being proportional to the hourly cloudiness$_i$ is given in the following way:

$$r_i = \frac{\text{cloudiness}_i \times r_{\text{daily}}}{\sum_{j=9\text{ a.m.}}^{8\text{ a.m. next day}} \text{cloudiness}_j} \quad i=1...24 \quad (3)$$

In the third and fourth data set the hourly amount of rain r_i was assigned equally to all hours in which cloudiness was between 5 and 8 or only to the ones equal to 8, respectively.

All four data files consist of 980 records. The objects in the first one were described with 13-dimensional, while the remaining three were with 14-dimensional input vectors. To study the influence of the rain on the pollutants more accurately, we further divided the second file (with rain distributed according to formula 3) into two parts. The first subfile contains all data in which the calculated rain is equal to zero, and the second subfile contains the remaining data. This means that 13-dimensional vectors and the second subfile containing all rainy day data with 14-dimensional vectors described the first subfile.

It is of course possible that on the rainy day during one or several hours the cloudiness was zero, and consequently (eq 3) the hourly amount of rain was zero.

Insolation Factor (Isf). Knowing that UV radiation is a very important factor in the chemical reaction in which O_3 is produced,[1,30] we tried to incorporate an approximation for the intensity of UV radiation. The basic shape of the Gaussian function consists of 12.00 h as the central point and a 3.00 h standard deviation. The intensity of the Gaussian shape was adapted to the hourly cloudiness$_i$ with the factor (1.0− 0.1 × cloudiness$_i$) in the following relation:

$$\text{Isf}_i = (1.0 - 0.1 \times \text{cloudiness}_i)e^{-(i-12/3)^2} \quad i = 1...24 \quad (4)$$

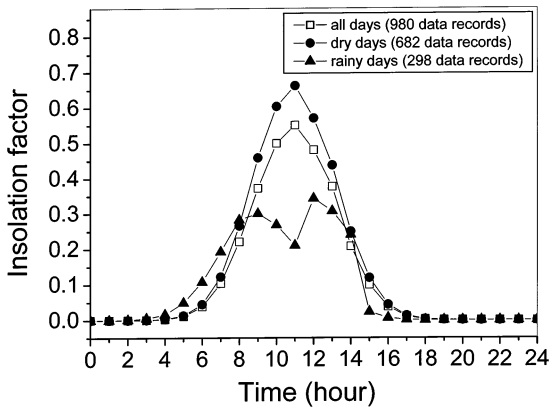

Figure 2. Hourly averages of the insolation factor (Isf$_i$) as a function of time for the following: all days (980 data records) (□), dry days (682 data records) (●), and rainy days (298 data records) (▲).

If cloudiness$_i$ is zero, then the insolation factor Isf$_i$ follows the Gaussian shape exactly. If, on the other hand, the cloudiness$_i$ is different from zero, then the Isf$_i$ is shielded accordingly. Isf$_i$ could be adapted to the geographical altitude of Buenos Aires and to the appropriate months of recording (early spring August and September).[19] However, for a preliminary investigation, the inclusion of this correction seems to be overabundant. The reason for introducing Isf$_i$ is explained later on. Figure 2 shows the difference between Isf$_i$ on rainy (file with 298 data records) and dry days (file with 682 data records). Each point in Figure 2 represents the average of the insolation factor from all data records of the particular file at the same hour. The drop of the Isf$_i$ during the rainy days can be clearly seen. The reason is that the Isf$_i$ is a function of the cloudiness (see eq 4), which as it turns out was on the average the largest around 11 a.m., which consequently makes the transmittance of the UV radiation through the clouds smaller.

Wind Velocity and Direction. The standard coding of wind direction in intervals of 10 radial degrees starting from the north (N) toward the east (E) is not a suitable input variable for mathematical models. The wind directions close to N (between NNW and NNE) are coded with values between 34 and 2, which does not comply with the criterion *"small change in the variable causes small change in the response"*. To satisfy it, the wind direction was decomposed into *x* and *y* components, *x* pointing to the E and *y* pointing to the N, respectively. By combining these two directions with the available wind intensity it is possible to evaluate two new variables w_x and w_y suitable for the proposed use.

2.3. Kohonen Self-Organized Maps of Various Variables. The generation of self-organizing maps of weights on different levels of the Kohonen network is well-known and described in the literature[22,23] and tutorials.[29] The self-organized maps are made within one layer of neurons each having *m* weights (in our case $m = 7 + 8 = 15$ variables). The Kohonen layers in this study were composed of 900 neurons placed in a (30 × 30) layout (Figure 3).

The most excited neuron $\mathbf{W}_e$ for each object $\mathbf{X}_s$ is uniquely chosen by the minimum distance criterion between all neurons $\mathbf{W}_j$ ($j = 1,..., N \times N$) and the object $\mathbf{X}_s$

$$\mathbf{W}_e \leftarrow \{\min d(\mathbf{X}_s, \mathbf{W}_j)\} \quad (5)$$

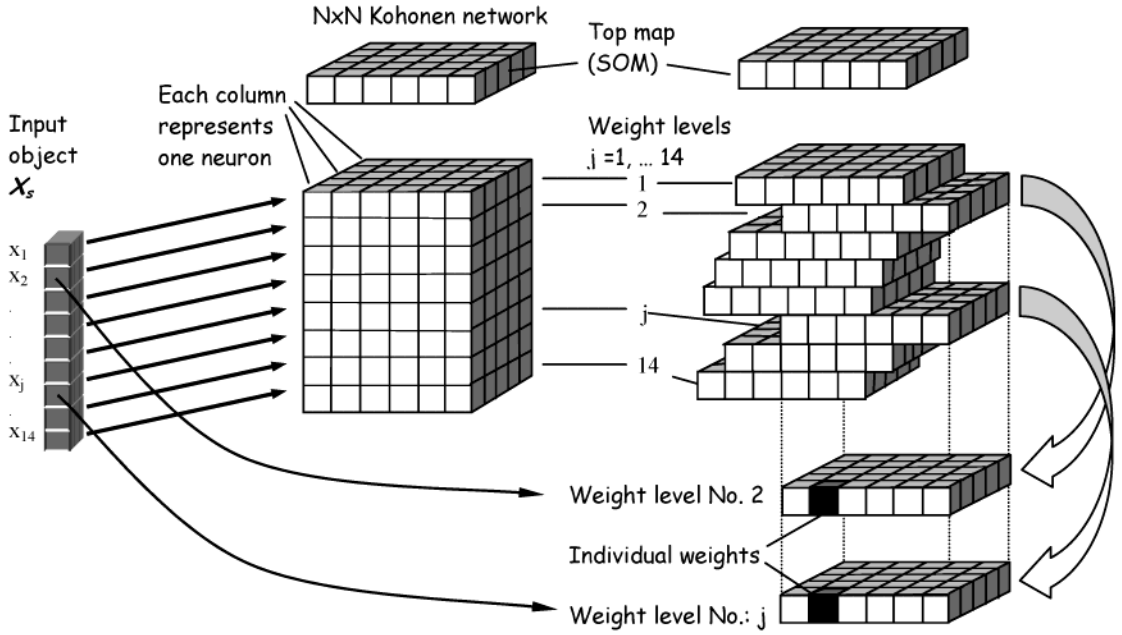

Figure 3. The neurons in the Kohonen layer are presented as columns with the same number of weights as there are variables in $\boldsymbol{X}_s$. The weights w_{ji} (i = 1,...N × N) receiving the same variable j form the weight layer j. All weight maps and the SOM are superimposed over each other.

where

$$d^2(\mathbf{X}_s,\mathbf{W}_s) = \frac{1}{m}\sum_{i=1}^{m}(x_{is} - w_{ij})^2 \qquad (6)$$

The distance in eq 6 is given per input variable (m = 13). The substitution of the absolute distance with the relative one (distance per variable: the inclusion of the factor 1/m in eq 6) is necessary because a number of objects have missing values. The use of the relative distance per variable makes the Kohonen clustering possible although some values of the variable PM10 are missing. This correction of distance calculation is described more in detail in ref 29.

The correction on the ith weight of the jth neuron $\mathbf{W}_j$ after the excited neuron $\mathbf{W}_e$ has been chosen by the object $\mathbf{X}_s$ = (x_{s1},x_{s2},..., x_{sm}) is carried out using the equation

$$\Delta w_{ji} = \eta\left(1 - \frac{d_j}{d_{\max} + 1}\right)(x_{si} - w_{ji}^{\text{old}}) \qquad (7)$$

where η is the learning rate, and d_j is the topological distance (d_j = 0,1, ..., $d_{\max}$) between the neuron $\mathbf{W}_j$ and the excited neuron $\mathbf{W}_e$. During the learning procedure the learning rate η was decreasing linearly from 0.5 to 0.01.

Although the goal of this study was the preliminary exploration of data (finding relations between the input variables) and not the generation of a model, we have treated the ozone data as the target data and input them on the output side virtually achieving the counterpropagation neural network instead of the Kohonen one. The correction of ozone weights in this output layer was evaluated according to the similar equation to eq 7:

$$\Delta w_{ji}^{\text{ozone}} = \eta\left(1 - \frac{d_j}{d_{\max} + 1}\right)(x_{si}^{\text{ozone}} - w_{ji}^{\text{ozone,old}}) \qquad (8)$$

The weights w_{ji}^{ozone} were not considered in eq 6 for the determination of the most excited neuron position.

The range of neurons around the excited neuron $\mathbf{W}_e$, where the corrections of weights $\mathbf{W}_j$ were still carried out, was limited by the topological distance $d_{\max}$, which was diminishing from N to 0 during the training as well. In our case the objects are described by 14 active variables composed of concentration and meteorological data. The concentration of O_3 is regarded as the nonactive variable that does not enter the distance calculations and forms the top map (SOM). The top map (SOM) and all weight maps were obtained after 600 epochs of training. One epoch of the training process was completed after all 980 data records (concentrations and weather data) were sent through the network once, and all necessary weight corrections after the input of each data record were made in the network. The interested reader can find a more detailed description of the Kohonen map formation in the textbooks.[24]

3. RESULTS AND DISCUSSION

The decrease of the amount of ozone with the increasing NO_2 concentration in the atmosphere is in accordance with the chemistry of these pollutants in the air. It is explained by the following reactions:[31]

$$\mathrm{NO_2(g)} \xrightarrow{h\nu} \mathrm{NO(g) + O(g)} \qquad (9)$$

Namely, the nitrogen dioxide, NO_2, dissociates to NO and oxygen when exposed to a bright light (UV, λ = 328−286 nm) (photoinitiated reaction). The reaction, which follows the photoinitiated reaction of NO_2, is the formation of O_3. This reaction is written as[32]

$$\mathrm{O(g) + O_2(g) \rightarrow O_3(g)} \qquad (10)$$

The oxygen atom is extremely reactive and readily attaches to a molecule of O_2 forming ozone O_3. Of course, the

Table 2: Correlation Coefficient between Weather Related Variables and Ozone, NO_2 Concentrations

weather related variables	O_3	NO_2
lsf_i	0.22	−0.16
pressure (town)	0.19	−0.06
pressure (see)	−0.17	−0.01
wind direction (V_x)	0.12	−0.08
air temperature	−0.09	0.08
cloudiness	−0.04	0.04
rain	−0.04	−0.01
wind direction (V_y)	0.02	0.14

chemistry of the pollutants in the atmosphere is much more complex and cannot be described only by the above-mentioned chemical reactions. For our preliminary studies, with the usage of these basic reactions, we just tried to represent the mutual behavior of ozone, NO_2, and Isf_i. It was of great interest to us whether the newly composed variable Isf_i (eq 4) could be used in the study, i.e., how much it is correlated to the concentration of ozone and to some other variables. Indeed the absolute values of the calculated correlation coefficients between the normalized concentrations of ozone and all other 14 variables ($r_{ozone,var}$) are low. The highest being 0.46 with NO_2, followed by four almost equal correlations of 0.28, 0.26, 0.26, and 0.25 with NO_x, CO, NO, and SO_2, respectively. The correlation coefficients between the ozone and weather data are even lower (from 0.02 to 0.19). This indicates that either there is no correlation whatsoever or there is a variation of correlations between ozone and different groups of weather conditions (cloudy/clear days, rainy/dry, windy/calm, UV radiation, etc.).

To test these two alternatives, we have introduced the insolation factor (Isf_i), which enables us to (a) to introduce the amount of UV radiation, which depends on the cloudiness, (b) to divide the amount of the rainfall through the rainy days more accurately, and (c) to take the cloudiness into account separately. Since the correlation coefficient between Isf_i and the ozone concentration (0.22) is greater than any of the weather variables with ozone, we consider its introduction justified (Table 2):

Regarding the facts that the highest correlations between the pollutants and ozone and between the weather variables and ozone are associated with NO_2 and Isf_i, respectively, the maps of these three variables are shown in Figure 4.

To obtain even better correlations we have applied the Kohonen mapping of all variables, separately on the data of rainy and dry days. In the following we will discuss only the relations between the top map (ozone) and two weight maps (NO_2, Isf_i) (Figure 4B,C). In each case the maps were obtained by a 30 × 30 Kohonen network, using the corresponding data sets, described in the section Preprocessing of Data, and the Kohonen training procedure was described in the section Kohonen Self-Organized Maps of Various Variables. All maps shown in Figure 4 were produced using the same parameters (600 epochs of training, triangular interpolating scheme,[24] with a linearly decreasing learning rate η from 0.5 to 0.01, and the same random initialization).

The Kohonen maps obtained by the complete data set, which includes all 980 14-dimensional data records (rainy and dry days), are presented in Figure 4A. The maps obtained by the data in which the calculated rain (eq 3) is equal to zero are shown in Figure 4B. Each of the 682 data records in this file are formally represented by the same 14 variables. Because the hourly rain fall variable is zero in all cases, effectively, the maps were produced with 13 variables only. In the third case (Figure 4C) the maps are obtained from the 298 records from the rainy days. To be more precise, in this set only those records for which the hourly rain was different from zero (eq 3) was taken into account. Namely, there were several cases when on a rainy day the cloudiness was zero and corresponding data records were considered in case B.

On the left side of Figure 4 the top maps of the ozone and on the right side the weight maps of NO_2 and Isf are shown. Each map is composed of 900 weights colored according to their relative value shown in the scale at the bottom of Figure 4.

For all three variables (concentrations of O_3, NO_2, and Isf) the correlation coefficients r_{map} between the weights in the corresponding weight levels (maps) are calculated and shown on the edges of the triangles between the maps.

In general the correlations r_{map} are low, but the visual comparison between the ozone top map and all other weight maps reveals that the maps shown in Figure 4 reveal the best correlation (O_3/ NO_2, O_3/Isf, NO_2/Isf). This is a good example of how the visualization with the help of the Kohonen self-organizing maps improves the analysis of complex data.

In Figure 4, the inverse proportionality between O_3 and NO_2 in Kohonen maps can be seen in all cases. The calculated correlation coefficient is negative ($r_{map} = -0.43$), while the correlation between O_3 and Isf_i is positive, with the value $r_{map} = 0.20$ (Figure 4A). These observations are in accordance with the correlations between the experimental data reported above. However, if the data are separated into the dry and rainy days, parts B and C of Figure 4, respectively, the responses in Kohonen maps and consequently the correlations calculated between the weights in each of the maps (r_{map}) are different from those given in Figure 4A. It can be clearly seen that all three correlation coefficients for the rainy days in Figure 4C are significantly higher than those for the dry days (Figure 4B). The difference can be caused by the different behavior of two factors (amount of PM_{10} particles and direction of the wind) within dry and rainy days, respectively. The inspection of the original data showed that the concentrations of PM_{10} particles for rainy days (maximum concentration is 0.09, average concentration is 0.021) is lower than the concentrations for dry days (maximum concentration is 0.12, average concentration is 0.026). We can assume that PM_{10} particles contribute to a different behavior of the pollutants for both cases. The distributions of the wind direction for the rainy and dry days were calculated. The examination of the calculated data showed that the percentage of south and west winds is approximately the same (15 ± 2)% for rainy or dry days. On the other hand, on the rainy days there were 39% more of the northern winds versus 31% of the east winds, while for the dry days the situation was almost exactly the opposite. Therefore, the influence of the winds cannot be excluded.

In accordance with the chemical reactions, it is normal to observe a negative correlation between NO_2 and O_3. To analyze this in detail, the average hourly concentrations of NO_2, O_3, and the insolation factor for the period of 45 days

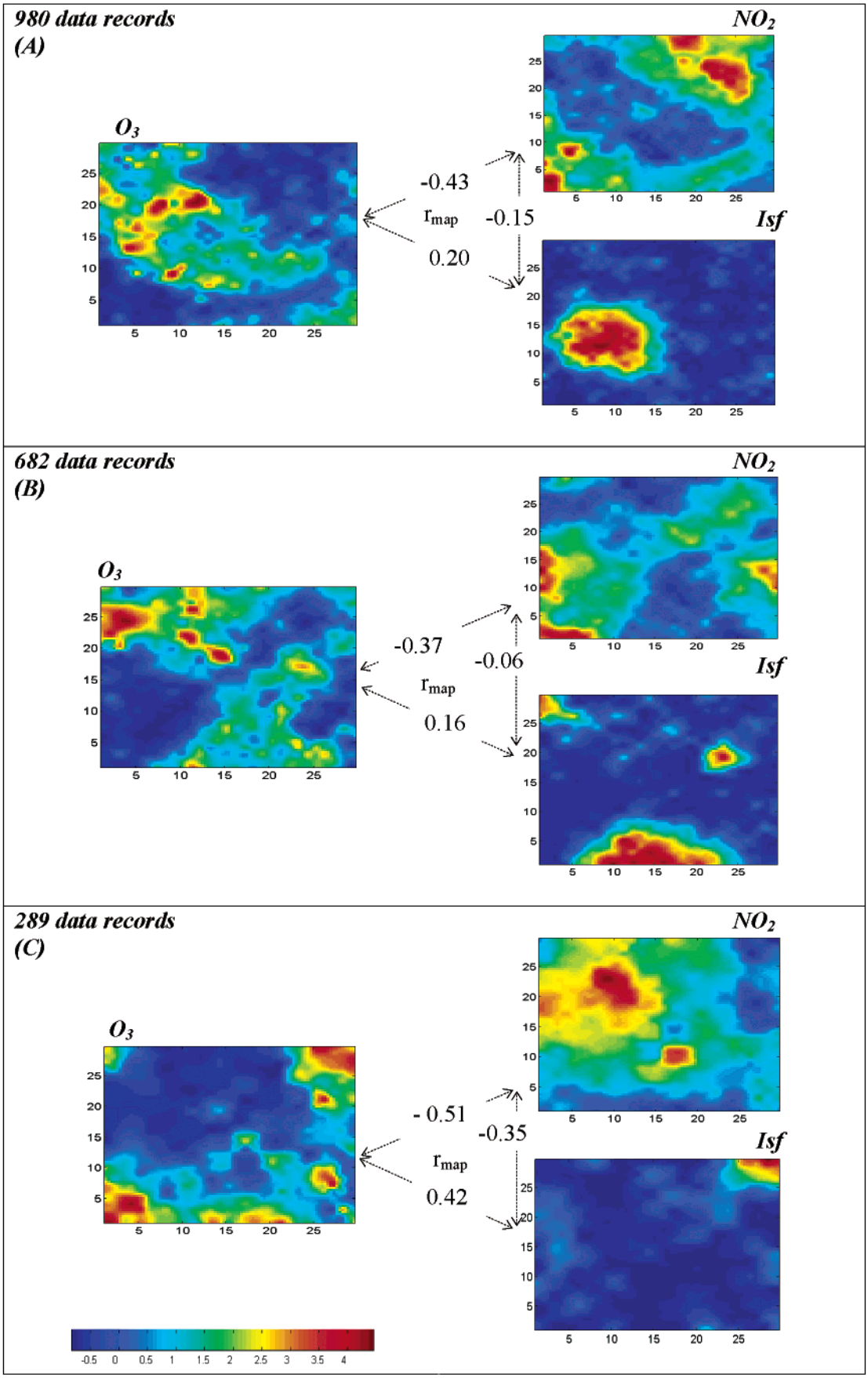

Figure 4. Nine Kohonen maps of three variables are shown: (A) Kohonen maps trained by all 980 data records, (B) Kohonen maps trained by 682 data records (only dry days), and (C) Kohonen maps trained by 289 data records (only rainy days). On the left side the top maps of the ozone and on the right side the weight maps of NO_2 and Isf_i are presented. The color bar represent particular concentrations (relative values for Isf_i) with upper and lower limits given in parentheses: O_3 (4.25×10^{-4}—3.14×10^{-2} ppm), NO_2 (3.30×10^{-5}—5.40×10^{-2} ppm), and Isf_i (3.66×10^{-3}—1).

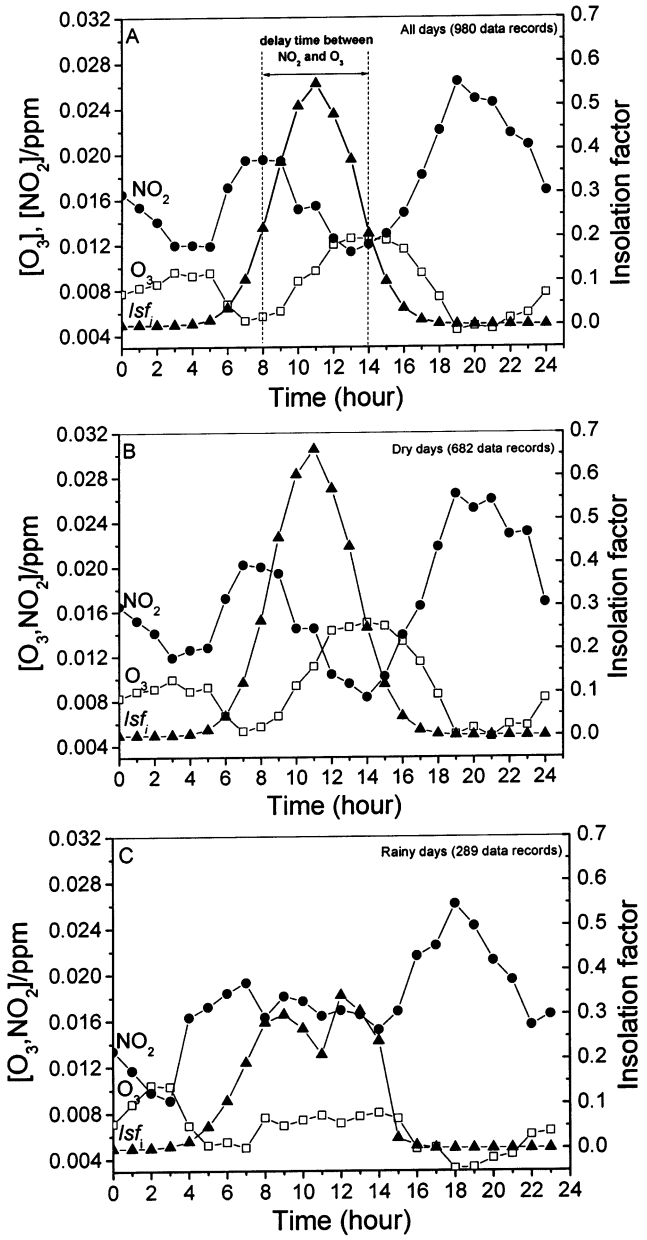

Figure 5. Hourly averages of (O_3), (NO_2), and Isf_i as a function of time for all days (980 data records) (A), for dry days (682 data records) (B), and for only rainy days (289 data records) (C). Average hourly concentrations are marked with the following: (□) average hourly concentrations of O_3, (●) average hourly concentrations of NO_2, and (▲) average hourly relative values of Isf_i.

were calculated from the raw database (Figure 5A). Figure 5 shows that the concentration of NO_2 starts increasing with the increasing traffic at about 5 a.m. It is interesting that on rainy days it starts earlier. With the sunrise (at about 7 a.m.), UV radiation increases and the decomposition of NO_2 begins.

The present study showed that there is a considerable difference in the correlation factors between the ozone, NO_2, and Isf_i when measured on rainy or sunny days. Due to the high nonlinearity of the processes one may argue that not only the amount of rain but also the frequency events will be important as well. To study the frequency of rain on the concentration of the pollutants a longer period of measurements should be performed.

Unfortunately, in the period of 45 days (winter 2001, August 11th to September 24th) we have 13 rainy days distributed in only six rainy and seven sunny intervals, which is not enough to study the influence of the rain frequency.

4. CONCLUSIONS

From the curves in Figures 5A−C (hourly averages over the entire period) high correlations between the concentrations of ozone and NO_2 are obtained, which indicates together with PM_{10} and the wind data that it is possible to generate a good model for the prediction of the ozone concentrations from the acquired data. On the other hand, the comparison of various weight maps (see Figures 4A−C) show that models for predicting the concentration of ozone will be far from simple or even linear. The high and low values of the ozone weight maps (left maps in Figures 4A−C) are highly irregular, so nonlinear modeling will be necessary in order to capture the main trends of such nonlinear maps. Despite this, a relatively good visual correlation of the two-dimensional weight maps of all data records on the rainy day data can be observed which indicates that good models with different artificial neural network architectures (most probably using the error back-propagation) should be tried. Additionally, the time delay or time-shift between various meteorological data and concentrations of pollutants should be taken into account when constructing good and reliable models. We have demonstrated that the Kohonen weight maps can help us in determining the proper time delays for the further studies toward the final model. The search for the proper time delays between different concentrations of air pollutants (NO, NO_2, O_3, etc.) will be the next step in our study.

ACKNOWLEDGMENT

This work was supported within the frame of the Scientific and Technological Cooperation Program coordinated by SETCIP of Argentina and the MSZS of Slovenia (Project ES/PA/00-EXIII/005). The authors kindly acknowledge the National Meteorological Service for providing meteorological data and Club de Amigos for permitting the installation of the monitoring station on its premises. H. Bajano, A. Ponso, and C. Rickert of the Atomic Energy Commission of Argentina undertook the monitoring campaign. Slovenian authors (N.G. and J.Z.) acknowledge the financial support by the Ministry for Education, Science and Sport of Slovenia through the program P-104-508. At the same time they would like to thank Mrs. Tanja Cegnar, MSc., from the Hydrometeorological Institute of Slovenia, for valuable discussions.

REFERENCES AND NOTES

(1) Graedel, T. E.; Crutzen, P. J. *Scientific American Library*; New York, 1995; pp 10−57.
(2) Clapp, L. J.; Jenkin, M. E. Analysis of the relationship between ambient levels of O_3, NO_2 and NO as a function of NO_x in the UK. *Atmos. Environ.* **2001**, *35*, 6391−6405.
(3) Made, D.; Ozolins, G.; Peterson, P.; Webster, A.; Orthofer, R.; Vandeweerd V.; Gwynne, M. Urban Air Pollution in Megacities of the World. *Atmos. Environ.* **1996**, *30*, 681−686.
(4) Liu, S. C.; Trainer, M.; Fehsenfeld, F. C.; Parrish, D. D.; Williams, E. J.; Fahey, D. W.; Hubler, G.; Murphy, P. C. Ozone production in the rural troposphere and the implication for regional and global ozone distributions. *J. Geophys. Res.* **1987**, *92*, 4191−4207.
(5) Comrie, A. C. Comparing neural networks and regression models for ozone forecasting. *J. Air Waste Manage.* **1997**, *47*(6), 653−663.
(6) Rohli, R. V.; Hsu, S. A.; Blanchard, B. W.; Fontenot, R. L. Short-Range Prediction of Tropospheric Ozone Concentrations and Exceedances for Baton Rouge, Louisiana. *Weather Forecast* **2003**, *18*(2), 371−383.

(7) Hadjiinski, L.; Geladi, P.; Hopke, P. A comparison of modelling nonlinear systems with artificial neural networks and partial least squares. *Chemometr. Intell. Lab*. **1999**, *49*, 91−103.
(8) Abdul-Wahab, S. A.; Al-Alawi, S. M. Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks, *Environ. Modell. Softw*. **2002**, *17*, 219−228.
(9) Roadknight, C. M.; Balls, G. R.; Mills, G. E. Dominic Palmer-Brown, Modelling complex environmental data, IEEE Transactions on Neural Network. *IEEE T Neural Network* **1997**, *8*(4), 852−862.
(10) Kolehmainen, M.; Martikainen, H.; Ruuskanen, J. Neural networks and Periodic Components used in air quality forecasting. *Atmos. Environ.* **2001**, *35*, 815−825.
(11) Kolehmainen, M.; Martikainen, H.; Hiltunen, H.; Ruuskanen, J. Forecasting Air Quality Parameters Using Hybrid Neural Network Modelling. *Environ. Monit. Assess.* **2000**, *65*(1−2), 277−286.
(12) Magallanes, J. F.; Zupan, J.; Gomez, D.; Reich, S.; Dawidowski, L.; Grošelj, N. The Mean Angular Distance Among Objects and Its Relationships with Kohonen Artificial Neural Networks. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1403−1411.
(13) Karppinen, A.; Kukkonen, J.; Eloläde, T.; Konttinen, M.; Koskentalo, T.; Rantakrans, E. A Modelling System for Prediction Urban Air Pollution: Model Description and Applications in the Helsinki Metropolitan Area. *Atmos. Environ.* **2000**, *34*, 3723−3733.
(14) Boznar, M.; Lesjak, M.; Mlakar, P. A neural network-based method for the short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. *Atmos. Environ.* **1993**, *B27*(2), 221−230.
(15) Elizondo, D.; Hoogenboom, G.; McClendon, R. W. Development of neural network model to predict daily solar radiation. *Agric. For. Meteorol.* **1994**, *7*, 115−132.
(16) Rege, M. A.; Tock, R. W. A simple neural network for estimating emission rates of hydrogen sulfide and ammonia from single point sources. *J. Air Waste Manage.* **1996**, *46*, 953−962.
(17) Yi, J.; Prybutok, R. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environ. Pollut.* **1996**, *92*, 349−357.
(18) Gardner, M. W.; Dorling, S. R. Artificial Neural Networks (The multilayer Perception) − A Review of Application in the Atmospheric Sciences. *Atmos. Environ*. **1998**, *32*, 2627−2636.
(19) Vanicek, K.; Frei, T.; Litynska, Z.; Schmalwieser, A. *UV Index for the Public*; COST-713 Action, UVB Forecasting, Brussels 2000.
(20) Mcleod, W. L.; Butz, R. G. Riskpro-Environmental-Polution Modelling. *J. Chem. Inf. Comput. Sci.* **1991**, *31*(3), 427−429.
(21) Hushon, J. M.; Powell, J.; Town, W. G. Summary of the History and Status of the System-Development for the Environmental Chemicals Data and Information Network (ECDIN). *J. Chem. Inf. Comput. Sci.* **1983**, *23*(1), 38−43.
(22) Kohonen, T. *Self-Organization and Associative Memory, 3rd ed.;* Springer-Verlag: Berlin, FRG, 1989.
(23) Hecht-Nielsen, R. Counterpropagation Networks. *Appl. Optics* **1987**, *26*, 4979−4984.
(24) Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design*; Wiley-VCH: Weinheim, 1999.
(25) United Nations Environment Programme-World Health Organization. *GEMS/AIR Methodology Review Handbook Series, Volume 1. Quality Assurance in Urban Air Quality Monitoring;* UNEP-WHO: Nairobi, 1994.
(26) Le Tertre, A.; Medina, S.; Samoli, E.; Forsberg, B.; Michelozzi, P.; Boumghar, A.; Vonk, J. M.; Bellini, A.; Atkinson, R.; Ayres, J. G.; Sunyer, J.; Schwartz, J.; Katsouyanni, K. Short-term effects of particulate air pollution on cardiovascular diseases in eight European cities. *J. Epidemiol. Commun. H.* **2002**, *56*(10), 773−779.
(27) Dominici, F.; McDermott, A.; Zeger, S. L.; Samet, J. M. National maps of the effects of particulate matter on mortality: Exploring geographical variation. *Environ. Health Persp.* **2003**, *111*(1), 39−43.
(28) Zupan, J. KCTRF-Program for Kohonen and Counterpropagation ANN. KIDP 1452, 1994.
(29) Zupan, J.; Nović, M.; Ruisanchez, I. Kohonen and counter-propagation ANNs in analytical chemistry. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 1−23.
(30) Björn, L. O.; Callaghan, T. V.; Gehrke, C.; Johanson, U.; Sonesson, M. Ozone depletion, Ultraviolet radiation and plant life. *Chemosph: Global Chang. Sci.* **1999**, *1*, 449−454.
(31) Finlanson-Pitts, B. J.; Pitts, J. N., Jr. *Atmospheric Chemistry;* 1986; Chapter 1, p 30.
(32) Finlanson-Pitts, B. J.; Pitts, J. N., Jr. *Atmospheric Chemistry;* 1986; Chapter 7-D-2a, p 419.

CI030418R