Unsupervised system to classify SO_2 pollutant concentrations in Salamanca, Mexico

J.M. Barrón-Adame , M.G. Cortina-Januchs , A. Vega-Corona , D. Andina

ABSTRACT

Salamanca is cataloged as one of the most polluted cities in Mexico. In order to observe the behavior and clarify the influence of wind parameters on the Sulphur Dioxide (SO_2) concentrations a Self-Organizing Maps (SOM) Neural Network have been implemented at three monitoring locations for the period from January 1 to December 31, 2006. The maximum and minimum daily values of SO_2 concentrations measured during the year of 2006 were correlated with the wind parameters of the same period. The main advantages of the SOM Neural Network is that it allows to integrate data from different sensors and provide readily interpretation results. Especially, it is powerful mapping and classification tool, which others information in an easier way and facilitates the task of establishing an order of priority between the distinguished groups of concentrations depending on their need for further research or remediation actions in subsequent management steps. For each monitoring location, SOM classifications were evaluated with respect to pollution levels established by Health Authorities. The classification system can help to establish a better air quality monitoring methodology that is essential for assessing the effectiveness of imposed pollution controls, strategies, and facilitate the pollutants reduction.

1. Introduction

Air pollution is a very complex phenomenon that poses significant threats to human health and the environment throughout the developed and developing countries (Chak & Xiaohong, 2008). Air pollution is caused by both natural and man-made sources. Major man-made sources of ambient air pollution include industries (Pal, Kim, Hong, & Jeon, 2008), transportation (Bignal, Ashmore, & Headley, 2008; Joumard, 2009), power generation (Younger, Morrow-Almeida, Vindigni, & Dannenberg, 2008), unplanned urban areas (Joumard, Lamure, Lambert, & Tripiana, 1996), etc. Therefore, the issue of air quality is receiving more attention as an increasing fraction of the countries population are now living in urban areas and are in demand of a cleaner environment (EPA, 2000; PNUMA, 2007; WHO, 2006).

Meteorology is well known to be an important factor contributing to air quality (Arain et al., 2007; Elminir, 2005; Mandurino et al., 2009; Seaman, 2000; Sousa et al., 2008). It is extremely important to consider the effect of meteorological conditions on atmospheric pollution, since they clearly influence dispersion capability in the atmosphere. It is well known that severe pollution episodes in the urban environment are not usually attributed to sudden increases in the emission of pollutants, but to certain meteorological conditions which diminish the ability of the atmosphere to disperse pollutants (Nadir & Selici, 2008; Turias, Gonzlez, Martfn, & Galindo, 2006). The frequency distribution of air pollutant concentration is useful in understanding the characteristics of air quality. It can be used to estimate how frequently a critical concentration level is exceeded (Ozden, Dogeroglu, & Kara, 2008). However, the concentrations of air pollutants usually vary randomly and are correlated with several factors such as types of fuels consumed, geographical and topographical peculiarities, town planning and meteorological factors, etc. (Demirci & Cuhadaroglu, 2000). Air quality management and information systems are required to control air pollutants and provide proper actions, controlling strategies and a better and safe environment for future generation (Bhanarkar, Goyal, Sivacoumar, & Chalapati Rao, 2005; Kurt, Gulbagci, Karaca, & Alagha, 2008; Lumbreras, ValdTs, Borge, & Rodrfguez, 2008). Thus, a thorough understanding of the meteorological field is fundamental to predicting and understanding air pollution in urban areas (Lee, Kim, Kim, & Lee, 2007).

Many clustering techniques can be used to determine the nature groups of similar objects (Du, 2010; Warren Liao, 2005). On atmospheric science, Hanna et al. (2001) showed ground ozone (O_3) concentrations in association with the mixing depth and wind field patterns and documented that meteorological fields may increase uncertainty in air quality (Hanna & Davis, 2002; Hanna, 2000). Yu and Chang (2001) analyzed the PM_{10} time series in

---

* Corresponding author.
  E-mail address: badamem@salamanca.ugto.mx (J.M. Barrón-Adame).

Taiwan from July 1993 to June 1998, and delineated the PM_{10} concentrations into five air quality basins by hierarchical clustering (Tai-Yi & Len-Fu, 2001). In another study presented by Turalloglu, Nuhoglu, and Bayraktar (2005), the relationship between daily average Total Suspended Particulate (TSP) and Sulphur Dioxide (SO_2) concentrations with meteorological factors for 1995–2002 winter seasons was statistically analyzed using the stepwise multiple linear regression analysis for Erzurum City. They have shown that, higher TSP and SO_2 concentrations are strongly related to colder temperatures, lower wind speed, higher atmospheric pressure and weakly correlated with rain and higher relative humidity (Turalloglu et al., 2005). More recently, Riccio, Giunta, and Chianese (2007) apply a trajectory classification of PM_{10} aiming to identify the role exerted by meteorology in the Naples urban area (Southern Italy). They identify and evaluate the effects of eight clusters on air quality (Riccio et al., 2007). Rimetz-Planchon, Perdrix, Sobanska, and BrTmard (2008) investigated also for PM_{10} polluted episodes with meteorological situations in an urban and industrialized coastal site of the southern part of the North Sea, representative of a typical harbor for trade. They explain the spatio-temporal variability of PM_{10} at the urban scale and identify Air Quality (AQ) regimes related to PM_{10} levels and local weather conditions applied to the air quality database of Dunkerque in 2002 (Rimetz-Planchon et al., 2008). Kim Oanh, Chutimon, Ekbordin, and Supat (2005) have also developed an automated scheme to classify the synoptic meteorological conditions governing over Northern Thailand. Because a quantitative approach utilizes a variety of meteorological variables for the classification of synoptic patterns, it involves intensive statistical data treatment, normally accomplished in the literature by a combination of the Principal Component Analysis (PCA) and clustering techniques (Kim Oanh et al., 2005). In the pre-mentioned reviews, traditional statistical clustering techniques were used for classification of environmental data. In recent years, the considerable progress has been in the developing of Artificial Neural Network (ANN) models for air quality (Cortina-Januchs, Barrón-Adame, Vega-Corona, & Andina, 2009; Gardner & Dorling, 1998). The Self-Organizing Maps (SOM) (Kohonen, 1990), an ANN with unsupervised learning is the other commonly used clustering algorithm in environmental data (Andina, Jevtić, Marcano, & Barrón-Adame, 2007). SOM is suitable for data classification because of its visualization property (Alvarez-Guerra, Gonzlez-Piuela, AndrTs, Galn, & Viguri, 2008; Seo & Obermayer, 2004; Vesanto & Alhoniemi, 2000). For example, the SOM has been used to identify patterns in satellite imagery in oceanography (Richardson, Risien, & Shillington, 2003); to visualize and cluster volcanic ash (Ersoy, Aydar, Gourgaud, Artuner, & Bayhan, 2007); or to estimate the risk of insect species invasion associated with geographic regions (Watts & Worner, 2009).

In this study, the suitability of SOM for classifying and interpreting the air quality and level of SO_2 concentrations in Salamanca city was investigated to implement an air pollution system. The results were compared to pollution levels for SO_2 established by Health Authorities. We have selected SOM as the best method for the following reasons:

- Unsupervised nature: Since the trajectory clusters are not known, unsupervised learning is required for trajectory clustering, which can be achieved by SOM.
- Classification/clustering power of neural networks: SOM is a neural network, which is a well-known powerful classification/clustering tool.
- Topological learning structure: In the training process, not only the winning neuron, but also neighboring neurons learn from the training data depending on their distance to the winning neuron, which is known as one of the most important aspects of SOM.

2. Material and method

2.1. Features of study area

Salamanca is a city in the Mexican state of Guanajuato with a population of approximately 234,000 inhabitants and located some 350 km to the northwest of Mexico city (INEGI, 2005).

Salamanca is cataloged as one of the most polluted cities in Mexico (Vega López, 2006). Although environmental management in Mexico began in 1971 with the Law to Prevent and Control Environmental Pollution, in the last decade Mexico began with true efforts to generate and compile environmental information. The National Institute of Ecology (INE), a decentralized organization of the Ministry of the Environment and Natural Resources (SEMARNAT, 2008), oversees policy-making decisions for air quality, solid and hazardous waste management, environmental impact assessment, global climate change, ozone depletion, wildlife management and natural reserves (INE, 2008).

In Salamanca, the Program to Improve the Air quality (ProAire) is composed of measures that affect transportation, industry, service sector, natural resources, health, and education. The ProAire program contemplates the urgent and immediate reduction of pollutant emissions when measurements of these pollutants register levels above those established by Health Authorities. When first ProAire concluded in 2000, environmental authorities undertook a longer, ambitious air quality improvement program ProAire 2002–2010. However, accurate measures were needed to determine how improving air quality would improve health and reduce health expenditures so that the new pollution control strategies could be evaluated (Gurjar, Butler, Lawrence, & Lelieveld, 2008; Jehng-Jung & Ming-Ru, 2006; Monteiro, Miranda, Borrego, & Vautard, 2007).

In our study the established ProAire limits by Health Authorities are taken as references to select the best SOM structure to classify the SO_2 concentrations correlated with wind fields. Table 1 shows the Established ProAire limits for SO_2.

### 2.2. Air pollutants and meteorological data

The main causes of pollution in Salamanca are due to fixed emission sources such as Chemical Industry and Power Generation, SO_2 being one the most important pollutant in air (IEEG, 2008; INE, 2008). Currently, in Salamanca an Automatic Environmental Monitoring Network (AEMN) is installed in which time series of criteria pollutant and meteorological parameters are obtained. Fig. 1 shows the AEMN distribution in Salamanca.

In our study, we have considered the maximum and minimum daily SO_2 concentration during the period of 2006 to train a SOM Neural Network in each monitoring station. Pollutant concentration have associated their correlated wind parameter measured simultaneously per minute. In this case, we considerer that Hourly and Daily mean do not represent an appropriate distribution of pollutants to train a Neural Network because this can be affected

Table 1
Established environmental contingency levels for SO_2 by Environmental Autorities in Salamanca, (a) *Pre-contingency* (b) *Phase I* and (c) *Phase II*.

Pollutant	Level activation	Level annulment
(a) Environmental Pre-contingency		
SO_2	$\geqslant$ 145 ppb and <225 ppb	<145 ppb
(b) Environmental Contingency Phase I		
SO_2	$\geqslant$ 225 ppb and <305 ppb	<145 ppb
(c) Environmental Contingency Phase II		
SO_2	$\geqslant$ 305 ppb	<145 ppb
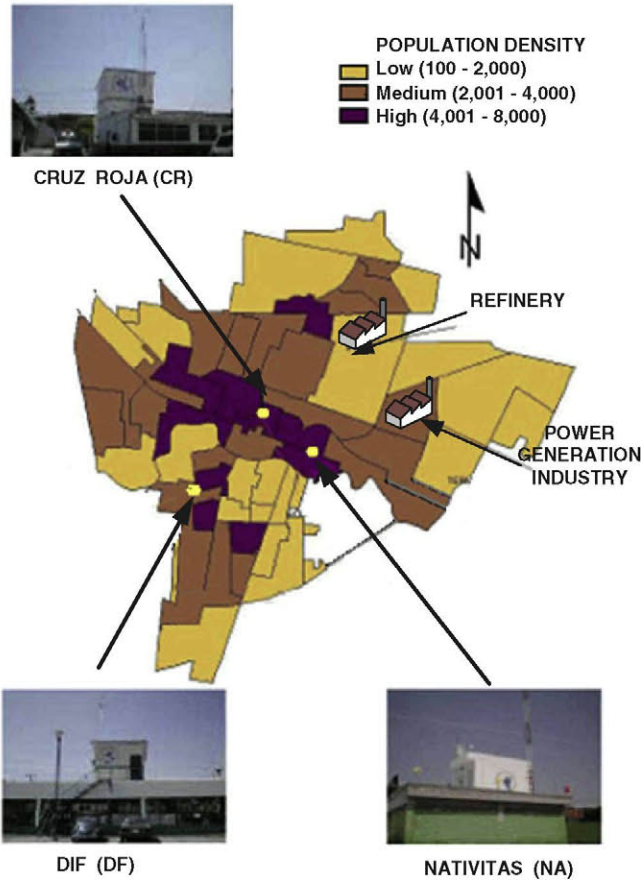
Fig. 1. Distribution of the Automatic Environmental Monitoring Network (AEMN) in Salamanca, Cruz Roja; DIF and Nativitas.

by extreme values (low or high) in the set. Table 2 summarizes the length of training vectors.

2.3. Self-Organizing Maps (SOM)

The basic SOM Neural Network consists of the input layer, and the output (Kohonen) layer which is fully connected with the input layer by the adjusted weights (prototype vectors). The number of units in the input layer corresponds to the dimension of the data. The number of units in the output layer is the number of reference vectors in the data space. In SOM, the high-dimensional input vectors are projected in a nonlinear way to a low-dimensional map (usually a two-dimensional space), and SOM can perform this transformation adaptively in a topologically ordered fashion. Therefore, the neurons are placed at the nodes of a two-dimensional lattice. Every neuron of the map is represented by an n dimensional weight vector (prototype vector), $\theta = [\theta_1, \ldots, \theta_n]$, where n denotes the dimension of the input vectors. The prototype vectors together form a codebook. The units (neurons) of the map are connected to adjacent ones by a neighborhood relation, which indicates the topology of the map. The rectangular topology was

Table 2
Number of patterns in the three-dimensional training vectors.

Station	Maximum minimum		Daily mean		Hourly mean	
	SO_2	PM_{10}	SO_2	PM_{10}	SO_2	PM_{10}
Cruz Roja	1506x3	1506x3	260x3	251x3	6552x3	6103x3
DIF	1560x3	1560x3	269x3	260x3	6338x3	6338x3
Nativitas	1284x3	1350x3	214x3	225x3	5572x3	5688x3
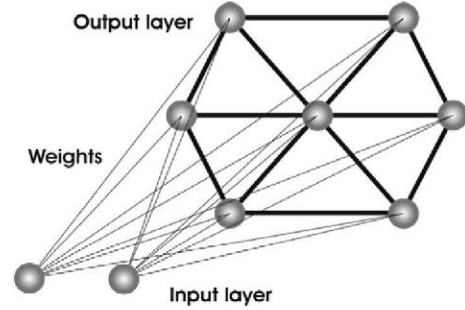
Fig. 2. A basic hexagonal SOM topological neighborhood.

used in this study. SOM can adjust the weight vectors of adjacent units in the competitive layer by competitive learning. Fig. 2 shows a hexagonal SOM topology.

In the training (learning) phase, the SOM forms an elastic net that folds onto the "cloud" formed by the input data. Similar input vectors should be mapped close together on the nearby neurons, and group them into clusters. SOM is an unsupervised classification which is used to cluster a data set based on statistics only, and can be trained by an unsupervised learning algorithm in which the network learns to form its own classifications of training data without external help. The SOM is trained iteratively. The learning steps are as follows (Hsin-Chung & Guor-Cheng, 2002):
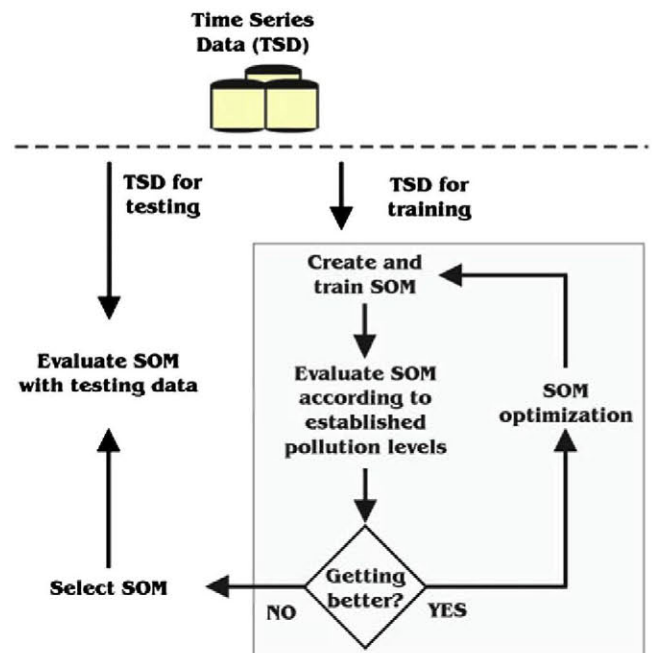
Step 1. Initialize randomly the weight vectors, $\theta_j(0)$, drawn from the input dataset and set $t = 0$.

Step 2. Present an input vector x to the network and compute the Euclidean distance, d_j, between a sample of input vectors and all the prototype vectors at iteration t

$$d_j = \|X_j - \theta_j(t)\| \tag{1}$$

Step 3. Find the winner unit c (best matching unit, BMU) which has the minimum Euclidean distance:

$$U_c = min\{d_j\} \tag{2}$$

Fig. 3. Clustering process.

Step 4. Update the connecting weight vectors of all neurons:

$$\theta_{i(t+1)} = \theta_j(t) + \eta(t)h_{cj}(t)[x(t) - \theta_j(t)] \tag{3}$$

Step 5. Increase time t to $t + 1$. If $t < T$ then go to step 2, otherwise stop the training.

Here, t is the time of iteration and T is a predefined number of iterations, respectively; $x(t)$ is an input vector randomly chosen at time t; $\eta(t)$ is the learning rate and is a decreasing function of time; $h_{cj}(t)$ is called the neighborhood function.

The neighborhood function will decrease in time. The topological distance $r = \|r_j - r_c\|$ is calculated between unit j and winner unit c. The most commonly used neighborhood function is the Gaussian:

$$h_{cj}(t) = exp\left(-\frac{\|r_j - r_c\|^2}{2\sigma^2(t)}\right) \tag{4}$$

where $\sigma(t)$ is called the neighborhood radius.

Both the learning rate and neighborhood radius decrease monotonically during training, and the $\eta(t)$ will converge towards 0. The learning is broken down into two phases: the ordering phase and tuning phase. In the ordering phase, the neighborhood radius decreases linearly from 5 to 1, and the value of 1 was maintained over the tuning phase.

Fig. 3 illustrates the clustering process. In the first step, the data are separated into two groups: training and testing. A SOM with four neurons is created and trained using a training dataset. Clustering results are compared with pollution levels established by Health Authorities. Another SOM is created with an additional neuron and trained. The evaluation criterion is compared. The number of neurons in SOM is increased until the evaluation criterion is achieved. The SOM with the best evaluation results is selected and the testing dataset is clustered using the best SOM. We stop the training process when all neurons in SOM structure have a difference of 1% of each variable in the feature space and will be
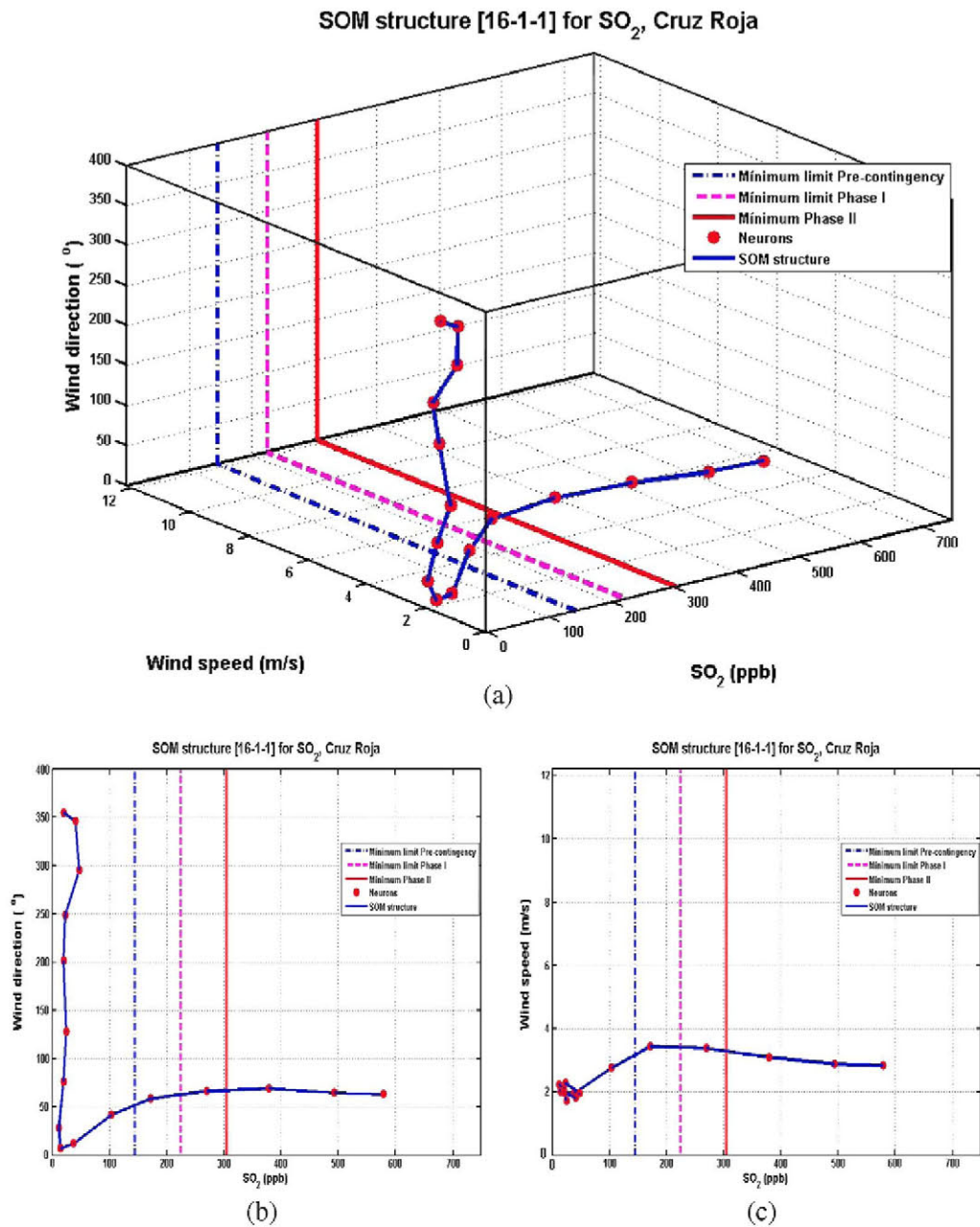

Fig. 4. Optimum SOM Neural Network with [16 × 1 × 1] structure to classify SO₂ pollutant concentrations correlated with wind parameters in Cruz Roja station.

considered in the error classification. Finally, the evaluation criterion values are reported.

Each pattern (pollutant concentration and meteorological variables) can be represented as a point in a 3-dimension space and its projection on the 1D lattice using an SOM has been used to detect similar or different behavior among patterns during the analysis period. Patterns with a similar behavior can be expected to be projected onto the same neuron, while patterns with different behavior will tend to be assigned to different neurons in the SOMs. An optimal mapping would be the one that preserves on the 1D lattice, in the most faithful fashion, the exist ing distances in the 3-dimensions space.

3. Experimental results

In the experiments, The SOM structures start with four neurons, and the number of neurons is increased one by one. A 1-dimensional SOM structure is used and number of neurons is increased only in one direction, such as $4 \times 1 \times 1, 5 \times 1 \times 1, 6 \times 1 \times 1, \ldots, 20 \times 1 \times 1$ (a total of 17 structures). Error classification level is computed by Mean Absolute Error (MAE) as:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |X_i - Y_i| \qquad (5)$$

where X_i and Y_i are the observed and estimated value at i time, and N is the total number of observations. The best SOM will be the one with the smallest error classification level.

Figs. 4–6 display the neuron positions in the feature space created with maximum and minimum daily SO_2 concentrations.

For Cruz Roja station (see Fig. 4), a $[16 \times 1 \times 1]$ SOM structure performs better to classify SO_2 pollutant concentrations correlated with wind parameters. Fig. 4(a) displays the 3D space for SO_2 and wind parameters. Fig. 4(b) displays the view of SO_2 and wind direction, and finally Fig. 4(c) displays the view of SO_2 and wind speed.

(a)

(b)

(c)
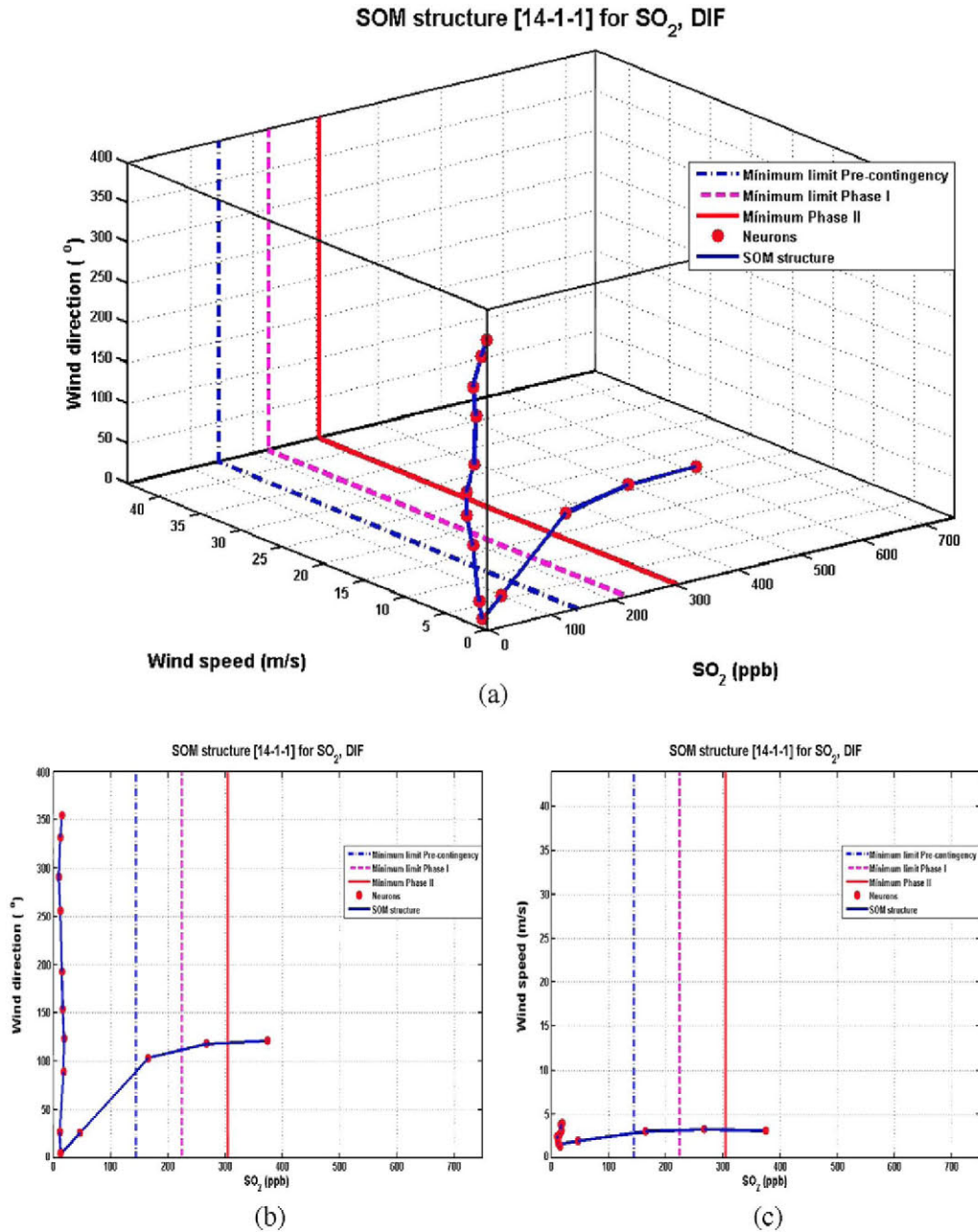
Fig. 5. Optimum SOM Neural Network with $[14 \times 1 \times 1]$ structure to classify SO_2 pollutant concentrations correlated with wind parameters in DIF station.
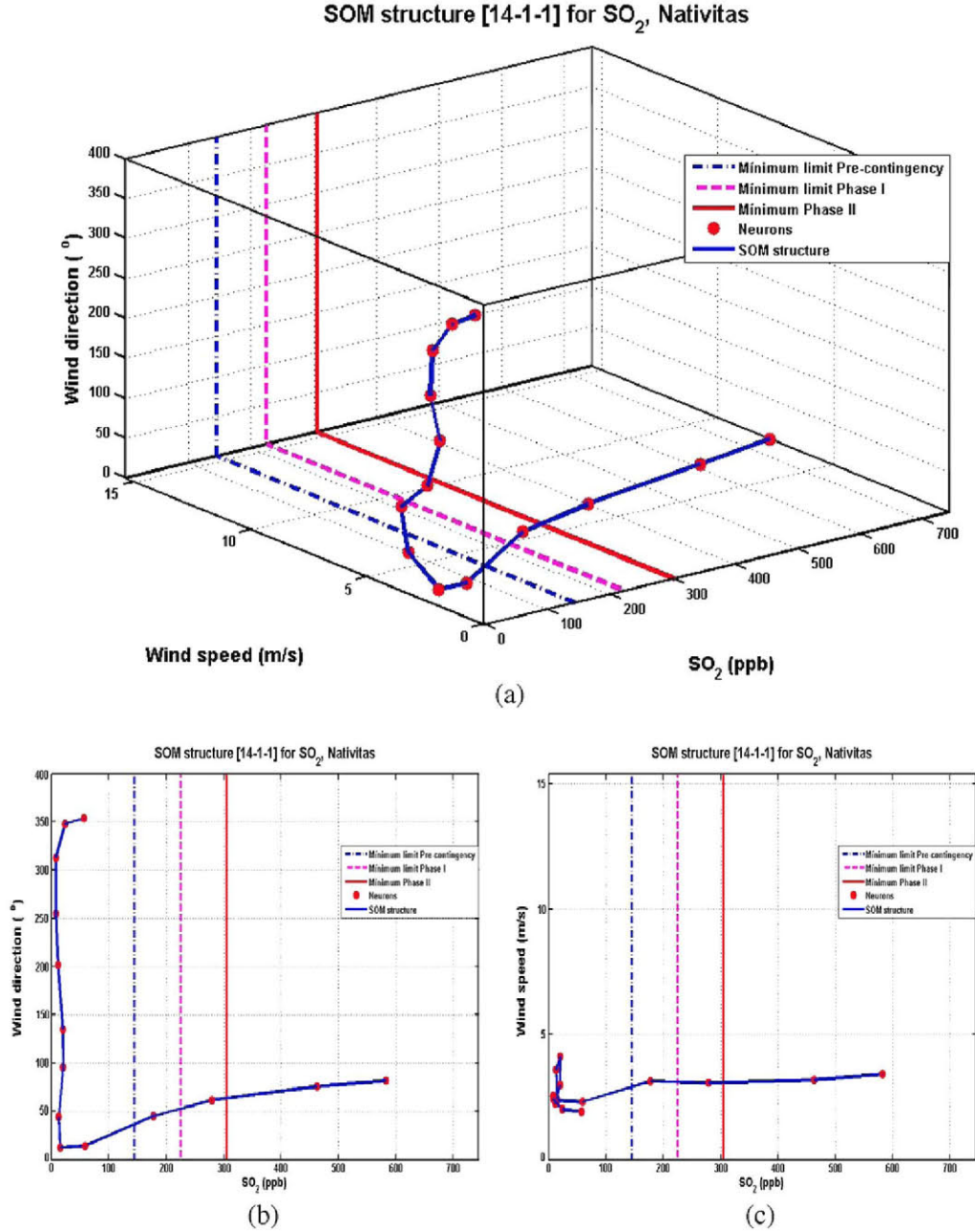
SOM structure [14-1-1] for SO₂, Nativitas

Fig. 6. Optimum SOM Neural Network with [14 × 1 × 1] structure to classify SO₂ pollutant concentrations correlated with wind parameters in Nativitas station.

Neurons 1–11 classify low SO₂ pollutant concentrations and are considered in Non-contingency level. Neuron 12 is considered in Pre-contingency level. Neuron 13 is considered in *Phase I* level. Finally, neurons 14, 15 and 16 classify the highest SO₂ pollutant con centrations and are considered in *Phase II* level. According to SOM structure, neurons 2 and 11 in Non-contingency level have high probability to past to Pre-contingency level. High SO₂ pollutant con centrations are presented with wind directions between 50° and 100° and wind speed between 2 and 4 m/s.

In the case of DIF station (see Fig. 5), a [14 × 1 × 1] SOM structure performs better to classify SO₂ pollutant concentrations correlated with wind parameters. Fig. 5(a) displays the 3D space for SO₂ and wind parameters. Fig. 5(b) displays the view of SO₂ and wind direction. Finally, Fig. 5(c) displays the view of SO₂ and wind speed. Neurons 1–11 classify low pollutant concentrations and are considered in Non-contingency level. Neuron 12 is considered in

Pre-contingency level. Neuron 13 is considered in *Phase I* level. Finally, Neuron 14 classify the highest SO₂ pollutant concentrations and is considered in *Phase II* level. High pollutants concentrations of SO₂ are presented in wind directions between 100° and 150° and wind speed between 2 and 4 m/s.

For Nativitas station (see Fig. 6), a [14 × 1 × 1] SOM structure performs better to classify SO₂ pollutants concentrations correlated with wind parameters. Fig. 6(a) displays the 3D space for SO₂ and wind parameters. Fig. 6(b) displays the view of SO₂ and wind direction and finally, Fig. 6(c) displays the view of SO₂ and wind speed. Neurons 5–14 classify the low pollutant concentrations and are considered as Non-contingency level. Neuron 4 is considered as Pre-contingency level. Neuron 3 is considered as *Phase I*. Finally, Neurons 1 and 2 are considered as Phase II level and classify the highest SO₂ pollutant concentrations. According to SOM structure, neuron 5 in Non-contingency level has the high

Table 3

SOM neuron position in the feature space for the three monitoring stations. The inverter order in Nativitas station is due to the aleatory SOM neuron order initialization.

Neuron positions									
Neuron	Cruz Roja			DIF			Nativitas		
	SO_2 ppb	Dir. (°)	Vel. (m/s)	SO_2 ppb	Dir. (°)	Vel. (m/s)	SO_2 ppb	Dir. (°)	Vel. (m/s)
1	20.5	1.9	354.8	15.1	1.2	354.8	582.7	3.3	81.6
2	40.8	1.8	346.3	12.5	1.7	332.0	462.5	3.1	75.6
3	47.2	1.9	295.3	9.5	2.4	290.8	279.1	3.0	61.5
4	22.7	2.2	248.7	12.2	2.3	256.1	177.7	3.1	44.9
5	20.1	2.0	201.8	15.1	2.7	192.6	58.3	2.2	13.8
6	24.6	1.7	127.8	16.5	3.8	153.9	15.4	2.3	12.8
7	19.7	2.0	76.7	18.9	3.9	123.7	12.8	3.5	44.7
8	12.1	2.2	28.2	18.2	3.1	89.5	20.1	4.1	95.3
9	15.0	2.0	7.1	11.5	1.8	26.7	20.0	2.9	135.2
10	37.6	1.9	12.3	13.1	1.6	4.6	11.5	2.2	201.6
11	103.2	2.7	42.1	46.8	1.9	26.6	7.6	2.5	255.3
12	171.8	3.4	58.8	165.2	3.0	103.1	7.6	2.4	313.0
13	270.2	3.3	66.7	267.4	3.2	117.9	23.7	2.0	347.9
14	379.2	3.1	69.5	374.0	3.1	121.0	57.0	1.9	354.1
15	493.0	2.8	64.7	–	–	–	–	–	–
16	578.9	2.8	63.1	–	–	–	–	–	–
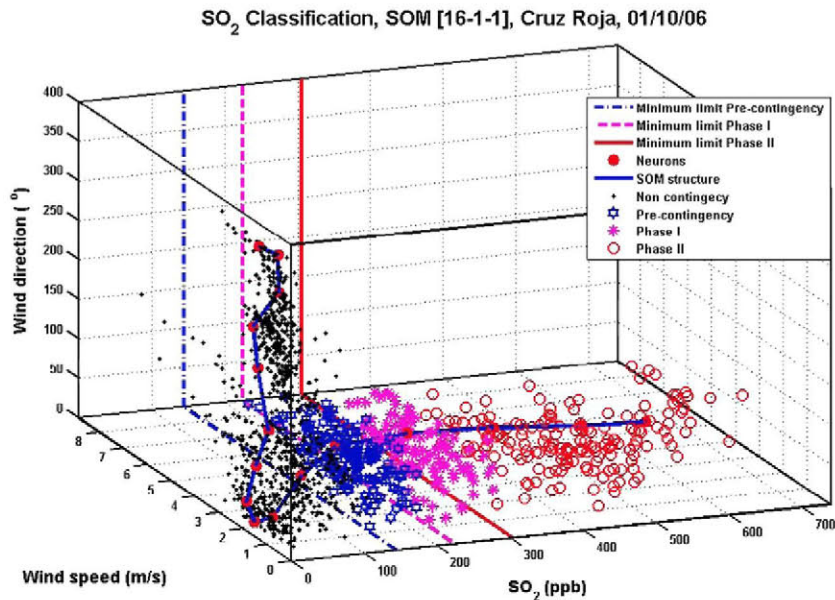
Fig. 7. SO_2 classification for October 1, 2006 in Cruz Roja station with a [16 × 1 × 1] SOM Neural Network structure where the • represents the SO_2 concentrations in Non-contingency, ★ represent the Pre-contingency, o represent the *Phase I* and ∗ represent the *Phase II* pollution levels, respectively.

probability to past to Pre-contingency level. High concentrations of SO_2 are presented with wind directions between 50° and 100° and wind speed between 3 and 4 m/s.

In the three locations, the SOM parameters in training/learning process for SO_2 were: rectangular SOM grid, distance from a home neuron to any other neuron was the Euclidean distance, *Ordering phase learning rate* = 0.9 and the number of training epochs were stopped when all neurons have a difference of 1% with previous SOM structures.

Table 3 summarizes the SOM neuron position in the feature space. Continuos lines separate the established contingency levels by Health Authorities. The inverter order in Nativitas station is due to the aleatory SOM neuron order initialization. Section 4 shows three classification examples of SO_2 pollutant concentrations, one for monitoring station.

4. Test results

After to train and tested many SOM Neural Network structures with different neurons in the output layer, Figs. 7–9 show the

visualization and classification results for one tested day in each monitoring station with the best trained SOM. The symbol • represents the SO_2 concentrations in *Non-contingency*, ★ represents the *Pre-contingency*, o represents the *Phase I* and ∗ represents the *Phase II* pollution levels, respectively. Tested days for SO_2 concentrations were: October 1 for Cruz Roja, November 18 for DIF and December 13 for Nativitas stations, respectively in 2006.

Table 4 shows our obtained results, such as; the SOM topology, the number of neurons for each contingency level and the MAE classification error (in percent) for the particular tested day of SO_2 in Cruz Roja, DIF and Nativitas stations, respectively. The ±1% in MAE error is the considered difference for to stop the SOM training process.

5. Conclusion

The aim of this paper was to classify the SO_2 pollutant concentrations correlated with wind parameters in order to implement a system to identify possible risk health in Salamanca.
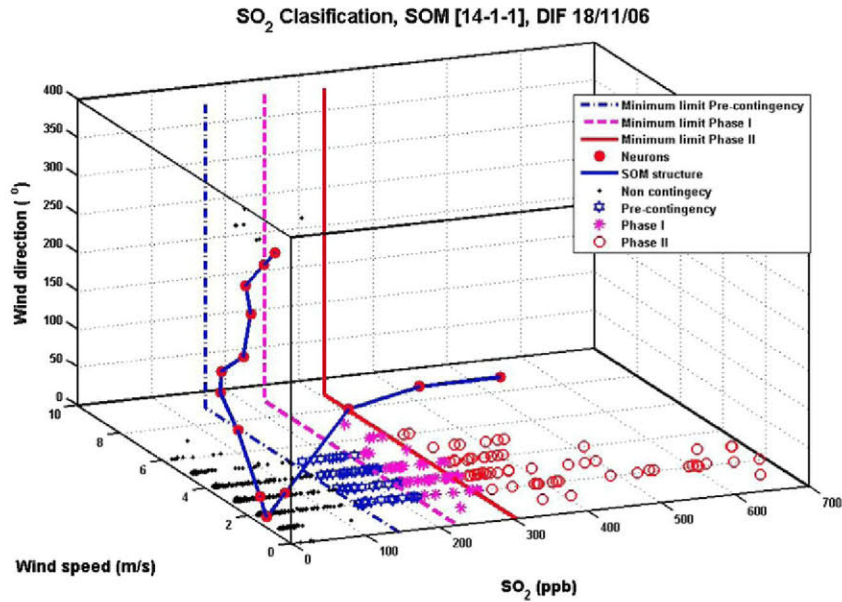
Fig. 8. SO₂ classification for October 1, 2006 in Cruz Roja station with a [16 × 1 × 1] SOM Neural Network structure where the • represents the SO₂ concentrations in Non-contingency, ★ represents the Pre-contingency, ○ represents the *Phase I* and ✳ represents the *Phase II* pollution levels, respectively.

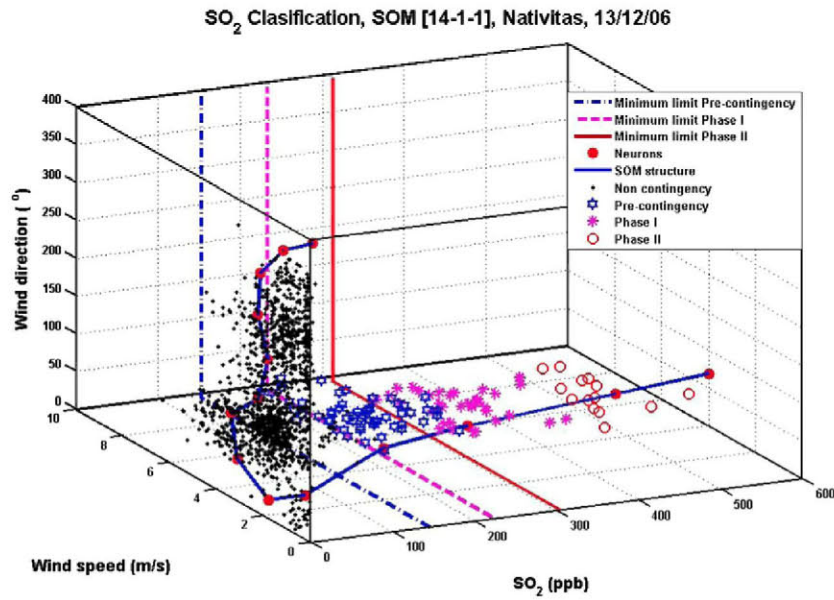**Fig. 9.** SO₂ classification for October 1, 2006 in Cruz Roja station with a [16 × 1 × 1] SOM Neural Network structure where the • represents the SO₂ concentrations in Non-contingency, ★ represents the Pre-contingency, ○ represents the *Phase I* and ✳ represents the *Phase II* pollution levels, respectively.

Table 4
SOM classification error, topologies and cluster in contingency levels obtained for each SO₂ Environmental Monitoring Network in Salamanca; A means *Non-contingency*, B *Pre-contingency*, C *Phase I* and D *Phase II*.

SOM classification error						
Monitoring station	SOM [x-y-z] topology	Neurons in level				MAE error (%) 1
		A (•)	B (★)	C (○)	D (✳)	
Cruz Roja	[16–1–1]	1–11	12	13	14 to 16	4.0
DIF	[14–1–1]	1–11	12	13	14	1.5
Nativitas	[14–1–1]	5–14	4	3	1 and 2	2.2

Meteorological parameters (wind direction and wind speed) that determine the source and emission rate of the pollutants were taken into account in the proposed system as decision factor in classification process. The advantage of system is the continuos pollution analysis per minute in the city. Experimental results have shown a correlation between pollutants and meteorological variables. Maximum SO_2 concentrations were identified in the wind direction of the fixed industry, but according to related studies, it is important to consider the effect of meteorological parameters on atmospheric pollution to understand the influence dispersion capability and the frequency distribution of air pollutant concentrations.

In SOM classification process, several SOM Neural Network structures (topologies) have been tested and trained in order to obtain a minimum SO_2 classification error. In each Environmental Monitoring location a training vector (from January to September, 2006) was build with the maximum and minimum daily SO_2 concentrations.

Obtained results show a process to classify the SO_2 pollutant concentrations. Implemented process consider meteorological variables as decision factors which allows to assess the air pollution. Studies are required to incorporate science based measures in air quality. On the basis of the present study, it can be concluded by means of SOM clustering analysis that meteorological variables (wind speed and wind direction) are important parameters influencing the air pollution behavior in Salamanca.

Proposed model will help researchers and policy-makers to select better air pollution control projects and practical insights into how to effectively and efficiently implement environment policies not only to human health but also other areas. The resulting data are expected to be useful not only for the future air pollution control applications in other studies, but also for the improvement of monitoring and evaluation systems building air quality management strategies.

Additionally, SOM Neural Network not only shows a classification process but also identify the emission source direction and wind speed threshold to measure high SO_2 concentrations. Presented results confirm that SOM is a very useful tool to cluster pollutant concentrations. The clusters were validated with SO_2 concentration and evaluated with established Environmental Laws. Very promising results are obtained.

Acknowledgment

References

Alvarez-Guerra, M., Gonzlez-Piuela, C., AndrTs, A., Galn, B., & Viguri, J. (2008). Assessment of self-organizing map artificial neural networks for the classification of sediment quality. *Environment International, 34*(6), 782–790.

Andina, D., Jevtić, A., Marcano, A., & Barrón-Adame, J. M. (2007). Error weighting in artificial neural networks learning interpreted as a metaplasticity model. In *IWINAC '07: proceedings of the 2nd international work-conference on the interplay between natural and artificial computation, Part I* (pp. 244–252).

Arain, M., Blair, R., Finkelstein, N., Brook, J., Sahsuvaroglu, T., Beckerman, B., et al. (2007). The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies. *Atmospheric Environment, 41*(16), 3453–3464.

Bhanarkar, A., Goyal, S., Sivacoumar, R., & Chalapati Rao, C. (2005). Assessment of contribution of so2 and no2 from different sources in jamshedpur region, india. *Atmospheric Environment, 39*(40), 7745–7760.

Bignal, K. L., Ashmore, M. R., & Headley, A. D. (2008). Effects of air pollution from road transport on growth and physiology of six transplanted bryophyte species. *Environmental Pollution, 156*(2), 332–340.

Chak, K. C., & Xiaohong, Y. (2008). Air pollution in mega cities in China. *Atmospheric Environment, 42*(1), 1–42.

Cortina-Januchs, M., Barrón-Adame, J. M., Vega-Corona, A., & Andina, D. (2009). Prevision of industrial so2 pollutant concentration applying anns. In *7th IEEE international conference on industrial informatics* (pp. 510–515).

Demirci, E., & Cuhadaroglu, B. (2000). Statistical analysis of wind circulation and air pollution in urban trabzon. *Energy and Buildings, 31*(1), 49–53.

Du, K. (2010). Clustering: A neural network approach. *Neural Networks, 23*(1), 89–107.

Elminir, H. K. (2005). Dependence of urban air pollutants on meteorology. *Science of the Total Environment, 350*(1–3), 225–237.

EPA (2000). US Environmental Protection Agency. Air Pollution Control Orientation Course. Tech. rep. <www.epa.gov/>.

Ersoy, O., Aydar, E., Gourgaud, A., Artuner, H., & Bayhan, H. (2007). Clustering of volcanic ash arising from different fragmentation mechanisms using Kohonen self-organizing maps. *Computers Geosciences, 33*(6), 821–828.

Gardner, M., & Dorling, S. (1998). Artificial neural networks (the multilayer perceptron) – A review of applications in the atmospheric sciences. *Atmospheric Environment, 32*(14–15), 2627–2636.

Gurjar, B., Butler, T., Lawrence, M., & Lelieveld, J. (2008). Evaluation of emissions and air quality in megacities. *Atmospheric Environment, 42*(7), 1593–1606.

Hanna, S. R., Lu, Z., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., et al. (2001). Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain. *Atmospheric Environment, 35*(5), 891–903.

Hanna, J. C. (2000). Making science relevant to environmental policy. *Environmental Science & Policy, 3*(1), 21–30.

Hanna, S. R., & Davis, J. M. (2002). Evaluation of a photochemical grid model using estimates of concentration probability density functions. *Atmospheric Environment, 36*(11), 1793–1798.

Hsin-Chung, L., & Guor-Cheng, F. (2002). Estimating the frequency distributions of pm10 and pm2.5 by the statistics of wind speed at Sha-Lu, Taiwan. *Science of the Total Environment, 298*(1–3), 119–130.

IEEG (July 2008). Institute of Ecology of Guanajuato State. Environmental Contingency programs. <www.ecologia.guanajuato.gob.mx/aire_limpio.php>.

INE (2008). National Institute Ecology. General Direction for Research on the Urban and Regional Pollution. Air Quality Research: Criteria Pollutants. <www.ine.org.mx>.

INEGI (2005). National Institute of Geography and Statistics. Population and Housing Census 2, 2005. <www.inegi.org.mx>.

Jehng-Jung, K., & Ming-Ru, H. (2006). Utilizing multiobjective analysis to determine an air quality monitoring network in an industrial district. *Atmospheric Environment, 40*(6), 1092–1103.

Joumard, R. (2009). Air pollution related to transport. *Atmospheric Environment, 43*(5), 985.

Joumard, R., Lamure, C., Lambert, J., & Tripiana, F. (1996). Air quality and urban space management. *Science of the Total Environment,* 57–67. highway and Urban Pollution.

Kim Oanh, N., Chutimon, P., Ekbordin, W., & Supat, W. (2005). Meteorological pattern classification and application for forecasting air pollution episode potential in a mountain–valley area. *Atmospheric Environment, 39*(7), 1211–1225.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE, 78*(9), 1464–1480.

Kurt, A., Gulbagci, B., Karaca, F., & Alagha, O. (2008). An online air pollution forecasting system using neural networks. *Environment International, 34*(5), 592–598. assessment of Urban and Regional Air Quality and its Impacts – Selected papers submitted to the third international symposium on air quality management at urban, regional and global scales (AQM2005) and 14th IUAPPA regional conference held in Istanbul on September 26–30, 2005.

Lee, S.-H., Kim, Y.-K., Kim, H.-S., & Lee, H.-W. (2007). Influence of dense surface meteorological data assimilation on the prediction accuracy of ozone pollution in the southeastern coastal area of the Korean Peninsula. *Atmospheric Environment, 41*(21), 4451–4465.

Lumbreras, J., ValdTs, M., Borge, R., & Rodrfguez, M. (2008). Assessment of vehicle emissions projections in Madrid (Spain) from 2004 to 2012 considering several control strategies. *Transportation Research Part A: Policy and Practice, 42*(4), 646–658.

Mandurino, C., & Vestrucci, P. (2009). Using meteorological data to model pollutant dispersion in the atmosphere. *Environmental Modelling & Software, 24,* 270–278.

Monteiro, A., Miranda, A., Borrego, C., & Vautard, R. (2007). Air quality assessment for portugal. *Science of the Total Environment, 373*(1), 22–31.

Nadir, I., & Selici, A. (2008). Investigating the impacts of some meteorological parameters on air pollution in balikesir, turkey. *Environmental Monitoring and Assessment, 140,* 267–277. doi:10.1007/s10661-007-9865-1.

Ozden Dogeroglu, T., & Kara, S. (2008). Assessment of ambient air quality in Eskisehir, Turkey. *Environment International, 34*(5), 678–687.

Pal, R., Kim, K.-H., Hong, Y.-J., & Jeon, E.-C. (2008). The pollution status of atmospheric carbonyls in a highly industrialized area. *Journal of Hazardous Materials, 153*(3), 1122–1135.

PNUMA (2007). PNUMA. PersPectivas del medio ambiente mundial 4.

Riccio, A., Giunta, G., & Chianese, E. (2007). The application of a trajectory classification procedure to interpret air pollution measurements in the urban

area of Naples (southern Italy). *Science of the Total Environment, 376*(1–3), 198–214.

Richardson, A., Risien, C., & Shillington, F. A. (2003). Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography, 59*(2–3), 223–239.

Rimetz-Planchon, J., Perdrix, E., Sobanska, S., & BrTmard, C. (2008). Pm10 air quality variations in an urbanized and industrialized harbor. *Atmospheric Environment, 42*(31), 7274–7283.

Seaman, N. L. (2000). Meteorological modeling for air-quality assessments. *Atmospheric Environment, 34*(12-14), 2231–2259.

SEMARNAT (2008). Ministry of the environment, natural resources and fisheries. <www.semarnat.gob.mx/Pages/inicio.aspx>.

Seo, S., & Obermayer, K. (2004). Self-organizing maps and clustering methods for matrix data. *Neural Networks, 17*(8–9), 1211–1229.

Sousa, S., Martins, F., Pereira, M., Alvim-Ferraz, M., Ribeiro, H., Oliveira, M., et al. (2008). Influence of atmospheric ozone, pm10 and meteorological factors on the concentration of airborne pollen and fungal spores. *Atmospheric Environment, 42*(32), 7452–7464.

Tai-Yi, Y., & Len-Fu, W. C. (2001). Delineation of air-quality basins utilizing multivariate statistical methods in Taiwan. *Atmospheric Environment, 35*(18), 3155–3166.

Turalloglu, F. S., Nuhoglu, A., & Bayraktar, H. (2005). Impacts of some meteorological parameters on so2 and tsp concentrations in Erzurum, Turkey. *Chemosphere, 59*(11), 1633–1642.

Turias, I., Gonzlez, F. J., Martfn, M., & Galindo, P. (2006). A competitive neural network approach for meteorological situation clustering. *Atmospheric Environment, 40*(3), 532–541.

Vega López, E. (Noviembre 2006). Gaceta Oficial del Distrito Federal. +rgano de Gobierno del Distrito Federal.

Vesanto, J., & Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks, 11*(3), 586–600.

Warren Liao, T. (2005). Clustering of time series data – A survey. *Pattern Recognition, 38*(11), 1857–1874.

Watts, M. J., & Worner, S. (2009). Estimating the risk of insect species invasion: Kohonen self-organising maps versus k-means clustering. *Ecological Modelling, 220*(6), 821–829.

WHO (2006). World Health Organization. Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide, Global update 2005: Summary of risk assessment. Tech. rep., WHO.

Younger, M., Morrow-Almeida, H. R., Vindigni, S. M., & Dannenberg, A. L. (2008). The built environment, climate change, and health: Opportunities for co-benefits. *American Journal of Preventive Medicine, 35*(5), 517–526.

Yu, T.-Y., & Chang, L.-F. W. (2001). Delineation of air-quality basins utilizing multivariate statistical methods in Taiwan. *Atmospheric Environment, 35*(18), 3155–3166.