Unsupervised system to classify $\text{SO}_2$ pollutant concentrations in Salamanca, Mexico

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

ARTICLE INFO

Keywords:
Air pollution
Meteorological variables
Artificial neural networks
Self-Organizing Maps (SOM)
Clustering

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 ($\text{SO}_2$) 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 $\text{SO}_2$ 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 (Arañ et al., 2007; Elminir, 2005; Mandurino et al., 2009; Seamans, 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, González, Martín, & Galindo, 2006). The frequency distribution of air pollutant concentrations is useful in understanding the characteristics of air quality. It can be used to estimate how frequently a critical concentration level is exceeded (Orden, 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 environment for future generation (Bhanarkar, Goyal, Sivacoumar, & Chalapati Rao, 2005; Kurt, Gulbagci, Karaca, & Alagha, 2008; Lumbreras, Valdés, Borge, & Rodríguez, 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 ($\text{O}_3$) 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 $\text{PM}_{10}$ time series in
Taiwan from July 1993 to June 1998, and delineated the PM10 concentrations into five air quality basins by hierarchical clustering (Tai-Yi & Len-Fu, 2001). In another study presented by Turalaloglu, Nuhoglu, and Bayraktar (2005), the relationship between daily average Total Suspended Particulate (TSP) and Sulphur Dioxide (SO2) 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 SO2 concentrations are strongly related to colder temperatures, lower wind speed, higher atmospheric pressure and weakly correlated with rain and higher relative humidity (Turalaloglu et al., 2005). More recently, Riccio, Giunta, and Chianese (2007) apply a trajectory classification of PM10 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 (Ricció et al., 2007). Rimetz-Planchon, Perdrix, Sobanska, and Brémard (2008) investigated also for PM10 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 space-temporal variability of PM10 at the urban scale and identify Air Quality (AQ) regimes related to PM10 levels and local weather conditions applied to the air quality database of Dunkerque in 2002 (Rimetz-Planchon et al., 2008). Kim Oanh, Chutimon, Ekborin, 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-Jamuchis, 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, Gonzalez-Pluera, AndriTs, Galin, & 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 (Ersosy, Aydar, Gourgaud, Artur, & Bayhan, 2007); or to estimate the risk of insect species invasion associated with geographic regions (Watts & Warner, 2009).

In this study, the suitability of SOM for classifying and interpreting the air quality and level of SO2 concentrations in Salamanca city was investigated to implement an air pollution system. The results were compared to pollution levels for SO2 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 Lópe, 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 (SEMAR-NAT, 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; Jele-Mung & 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 SO2 concentrations correlated with wind fields. Table 1 shows the Established ProAire limits for SO2.

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. SO2 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 SO2 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 consider that Hourly and Daily mean do not represent an appropriate distribution of pollutants to train a Neural Network because this can be affected

<table>
<thead>
<tr>
<th>Pollutant</th>
<th>Level activation</th>
<th>Level annulment</th>
</tr>
</thead>
<tbody>
<tr>
<td>SO2</td>
<td>≥145 ppb and &lt;225 ppb</td>
<td>&lt;145 ppb</td>
</tr>
<tr>
<td>SO2</td>
<td>≥225 ppb and &lt;305 ppb</td>
<td>&lt;145 ppb</td>
</tr>
<tr>
<td>SO2</td>
<td>≥305 ppb</td>
<td>&lt;145 ppb</td>
</tr>
</tbody>
</table>
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), \( \mathbf{v} = [v_1, \ldots, v_n] \), 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 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, \( \mathbf{v}_i(0) \), drawn from the input dataset and set \( t = 0 \).

**Step 2.** Present an input vector \( \mathbf{x} \) to the network and compute the Euclidean distance, \( d_j \), between a sample of input vectors and all the prototype vectors at iteration \( t \)

\[
d_j = ||\mathbf{x} - \mathbf{v}_j(t)||
\]

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

\[
U_c = \min\{d_j\}
\]

---

**Table 2**

<table>
<thead>
<tr>
<th>Station</th>
<th>Maximum</th>
<th>Minimum</th>
<th>Daily mean</th>
<th>Hourly mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \text{SO}_2 )</td>
<td>( \text{PM}_{10} )</td>
<td>( \text{SO}_2 )</td>
<td>( \text{PM}_{10} )</td>
</tr>
<tr>
<td>Cruz Roja</td>
<td>1506x3</td>
<td>1506x3</td>
<td>260x3</td>
<td>251x3</td>
</tr>
<tr>
<td>DIF</td>
<td>1560x3</td>
<td>1560x3</td>
<td>260x3</td>
<td>260x3</td>
</tr>
<tr>
<td>Nativitas</td>
<td>1284x3</td>
<td>1350x3</td>
<td>214x3</td>
<td>225x3</td>
</tr>
</tbody>
</table>
Step 4. Update the connecting weight vectors of all neurons:
\[ \theta_{i(t+1)} = \theta_{i(t)} + \eta(t) h_{i(j)}[x(t) - \theta_{i(t)}] \] (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; \( h_{i(j)}(t) \) is called the neighborhood function.

The neighborhood function will decrease in time. The topological distance \( r = \| r_i - r_c \| \) is calculated between unit \( j \) and winner unit \( c \). The most commonly used neighborhood function is the Gaussian:
\[ h_{i(j)}(t) = \exp\left(-\frac{\| r_i - r_c \|^2}{2 \sigma^2(t)}\right) \] (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 x 1 x 1] structure to classify SO\(_2\) 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 existing 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:

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

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.

![SOM structure [14-1-1] for SO2, DIF](image1)

(a)

![SOM structure [14-1-1] for SO2, DIF](image2)

(b)

![SOM structure [14-1-1] for SO2, DIF](image3)

(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.
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 concentrations and are considered in Phase II level. According to SOM structure, neurons 2 and 11 in Non-contingency level have high probability to pass to Pre-contingency level. High SO₂ pollutant concentrations 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 x 1 x 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 x 1 x 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 1–11 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.

<table>
<thead>
<tr>
<th>Neuron</th>
<th>Cruz Roja</th>
<th>DIF</th>
<th>Nativitas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SO$_2$ ppb</td>
<td>Dir. (°)</td>
<td>Vel. (m/s)</td>
</tr>
<tr>
<td>1</td>
<td>20.5</td>
<td>1.9</td>
<td>354.8</td>
</tr>
<tr>
<td>2</td>
<td>40.8</td>
<td>1.8</td>
<td>346.3</td>
</tr>
<tr>
<td>3</td>
<td>47.2</td>
<td>1.9</td>
<td>295.3</td>
</tr>
<tr>
<td>4</td>
<td>22.7</td>
<td>2.2</td>
<td>248.7</td>
</tr>
<tr>
<td>5</td>
<td>20.1</td>
<td>2.0</td>
<td>201.8</td>
</tr>
<tr>
<td>6</td>
<td>24.6</td>
<td>1.7</td>
<td>127.8</td>
</tr>
<tr>
<td>7</td>
<td>18.7</td>
<td>2.0</td>
<td>76.7</td>
</tr>
<tr>
<td>8</td>
<td>12.1</td>
<td>2.2</td>
<td>28.2</td>
</tr>
<tr>
<td>9</td>
<td>15.0</td>
<td>2.0</td>
<td>7.1</td>
</tr>
<tr>
<td>10</td>
<td>37.6</td>
<td>1.9</td>
<td>12.3</td>
</tr>
<tr>
<td>11</td>
<td>103.2</td>
<td>2.7</td>
<td>42.1</td>
</tr>
<tr>
<td>12</td>
<td>171.8</td>
<td>3.1</td>
<td>58.8</td>
</tr>
<tr>
<td>13</td>
<td>270.2</td>
<td>3.3</td>
<td>69.7</td>
</tr>
<tr>
<td>14</td>
<td>379.2</td>
<td>3.1</td>
<td>69.5</td>
</tr>
<tr>
<td>15</td>
<td>493.0</td>
<td>2.8</td>
<td>64.7</td>
</tr>
<tr>
<td>16</td>
<td>578.9</td>
<td>2.8</td>
<td>63.1</td>
</tr>
</tbody>
</table>

Fig. 7. SO$_2$ classification for October 1, 2006 in Cruz Roja station with a [16 x 1 x 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.

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 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. Continuous 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

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.
SO₂ Classification, SOM [14-1-1], DIF 18/11/06

Fig. 8. SO₂ classification for October 1, 2006 in Cruz Roja station with a [16 x 1 x 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.

SO₂ Classification, SOM [14-1-1], Natl vitas, 13/12/06

Fig. 9. SO₂ classification for October 1, 2006 in Cruz Roja station with a [16 x 1 x 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.

<table>
<thead>
<tr>
<th>Monitoring station</th>
<th>SOM [x-y-z] topology</th>
<th>Neurons in level</th>
<th>MAE error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A (•)</td>
<td>B (◆)</td>
</tr>
<tr>
<td>Cruz Roja</td>
<td>[16–1–1]</td>
<td>1-11</td>
<td>12</td>
</tr>
<tr>
<td>DIF</td>
<td>[14–1–1]</td>
<td>1-11</td>
<td>12</td>
</tr>
<tr>
<td>Natl vitas</td>
<td>[14–1–1]</td>
<td>5-14</td>
<td>4</td>
</tr>
</tbody>
</table>
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 factors which allows to assess the air 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 SO₂ classification error. In each Environmental Monitoring location a training vector (from January to September, 2006) was used with the maximum and minimum daily SO₂ concentrations.

Obtained results show a process to classify the SO₂ 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 environmental 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 SO₂ concentrations. Presented results confirm that SOM is a very useful tool to cluster pollutant concentrations. The clusters were validated with SO₂ concentration and evaluated with established Environmental Laws. Very promising results are obtained.

Acknowledgment

The authors thank the Environmental Authorities: Patronage for the air Quality Monitoring (Salamanca) and the Institute of Ecology of Guanajuato (INEE) for supplying the measured data. And: The National Council for Science and Technology (CONACYT) in Mexico, The Computational Intelligence Laboratory (LabINCo). (University of Guanajuato, Mexico), The Group for Automation in Signals and Communications (GASC). (Technical University of Madrid, Spain) for the help provided to complete this study.

References


