Self-Organizing Maps to Validate Anti-Pollution Policies

Ángel Arroyo[1], Álvaro Herrero[1], Verónica Tricio[2], Emilio Corchado[3]

[1]Grupo de Inteligencia Computacional Aplicada (GICAP), Departamento de Ingeniería Civil, Escuela Politécnica Superior, Universidad de Burgos, Burgos, Spain.
{aarroyop, ahcosio}@ubu.es
[2]Departmento de Física, Facultad de Ciencias, Universidad de Burgos, Burgos, Spain.
vtricio@ubu.es
[3]Departamento de Informática y Automática, University of Salamanca, Salamanca, Spain.
escorchado@usal.es

Abstract. This study presents the application of Self Organizing Maps to air-quality data in order to analyze episodes of high pollution in Madrid (Spain capital city). The goal of this work is to explore the dataset and then compare scenarios with similar atmospheric conditions (periods of high NO_2 concentration): some of them when no actions were taken and some when traffic restrictions were imposed. The main pollutants recorded at these stations along eleven days during four time intervals in years from 2015 to 2018, are analyzed in order to determine the effectiveness of the anti-pollution measures. The visualization of trajectories on the Self-Organizing Map let us clearly see the evolution of pollution and consequently evaluate the effectiveness of the measures applied before, after and during the protocol activation time.

Keywords. Air quality, time evolution, Self-Organizing Maps, trajectories, data visualization.

1. Introduction

In recent years, our knowledge of atmospheric pollution and our understanding of its effects have advanced greatly. Systematic measurements in any country are fundamental due to the health risks caused by high levels of atmospheric pollution. Measurement stations acquire data continuously and, in the case of Spain, these data are available for further study and analysis thanks to the open data policy of public institutions [1]. In the City of Madrid, an Integral Air Quality System (IAQS) [2] was developed in order to monitor the levels of emissions of the main pollutants. To fulfill its objective, the IAQS is constituted by three subsystems: surveillance system, prediction and Information System. The IAQS comprises policies with associated actions to be taken during episodes of high pollution by Nitrogen Dioxide (NO_2) [3]. According to European regulation, the maximum value of concentration for this pollutant is 200 $\mu g/m^3$ averaging in a period of an hour and 40 $\mu g/m^3$ averaging in a period of a year [4]. In the city centers of many European capital cities (such as Paris, London, etc.) these limits are exceeded when there is no rain and wind, and there are high emissions from road traffic. European capitals are developing protocols and defining actions to restrict traffic in large cities in periods of high air pollution, in order to protect the health of citizens. The city of Madrid is trying to control the high levels of air pollution specially produced by the traffic emissions, developing the IAQS, aimed at knowing the levels of atmospheric pollution in the city in real time. A part of this integral plan is the set of measures to be adopted during episodes of high levels of NO_2 [3]. According to the severity of the situation, four scenarios are defined; the Scenario I consist on informing the population and the agents involved, the speed limit in the M-30 (one of Madrid ring-roads) and the accesses to the city (both directions) from the M-40 (another Madrid ring-road) are reduced to 70 km/h, and the use of public transport is promoted. Scenario II comprises the activation of the Environmental Health Alert System and the prohibition of vehicles owned by non-residents to park in the areas of the Regulated Parking Service (RPS) all over the city. When the warning level is exceeded during two consecutive days, the Scenario III is activated; in addition to the measures adopted during Scenario II, it is added the restriction of circulation in the interior area of the M-30 road for 50% of all vehicles. Furthermore, the non-circulation of empty taxis (except Ecotaxis and Eurotaxis) in the interior area of the M-30 road is recommended. The Scenario IV is considered when the warning level is exceeded during three consecutive days or when the alert level is reached. The measures associated to this scenario are the most restrictive ones, comprising the mandatory restriction on the circulation of taxis (except Ecotaxis and Eurotaxis) in the interior area of the M-30 road and a set of measures to promote public transport.

Scant attention has been devoted to the problem of forecasting and analyzing short periods of high air pollution by NO_2 in big cities in previous work; [5] proposes a network air quality diagnosis of Madrid city center by taking into account both transport exhaust emissions and population exposure levels. The paper aims at identifying air pollution network hotspots in Madrid city center, but does not assess the effectiveness of the anti-pollution measures implemented, as present work does. In [6], the application of dimensionality reduction [7] and clustering techniques [8] to episodes of high pollution in Madrid City (Spain) is presented in order to visually check the effectiveness of the protocols for traffic restrictions during episodes of high NO_2 levels. In [9], the convenience of quantile regression to predict extreme concentrations of NO_2 is investigated. Using data from the city of Madrid, including NO_2 concentrations as well as meteorological measures, models of quantile regression to predict extreme NO_2 concentrations are built. In [10], a prototype has been developed in order to support short-term prediction of NO_2 maximum concentration levels in Athens, Greece. The prototype is based on a case-based reasoning approach combining heuristic and statistical techniques. The prototype performance is compared with that of a Back Propagating Neural Network (BPNN) [11] and a decision tree.

The Self-Organizing Maps (SOMs) [12] have been used for the analysis of air pollution in many studies. In [13], an air quality modelling which can forecast urban air quality for the next day using airborne pollutant is developed, considering meteorological and timing variables. Hourly airborne pollutant and meteorological averages collected during the years 1995–1997 were analysed in order to identify air quality episodes having typical and the most probable combinations of air pollutant and meteorological variables. This modelling was done using the SOM, the Sammon's mapping [14] and fuzzy distance metrics. In [15], the effects of long-range transport patterns of air masses to the regional PM profile in a megacity, Istanbul (Turkey), are studied. Five-day hourly backward trajectories were obtained by the HYSPLIT [16] model for selected episodic events in 2008. Self-Organizing Maps (SOM) was used to cluster these trajectories.

Present study focuses on the comparative analysis of the environmental pollution in the center of Madrid City, during four-periods in years from 2015 to 2018 with similar meteorological conditions. These meteorological conditions are characterized by a high stability, due to the practical absence of wind and rain, together with a very dry climate. These weather conditions favour high concentrations of pollutants such as NO_2, PM10, CO and SO_2 among other usual pollutants in high traffic situations. Protocols for the control of pollution during episodes of high NO_2 emissions had not yet been approved during the first time period in year 2015, while they were in force in the other three periods these protocols were applied (years 2017 and 2018). The underlying idea is to assess the effect of such protocols by comparing the pollutant levels in similar conditions.

Unlike the time window in a previous work [6] by the authors, a wider time window is used for data analysis in this study (from 2015 to 2018), and hence four episodes of high NO_2 concentrations are studied and with a larger number of samples in each of the them. In order to do that, the SOM with the extension of trajectory data, are applied to the air quality data acquired in two locations described in Section 3. With the application of the aforementioned SOM trajectory extension (technique not applied in any of the previous works presented in this section), the study aims to analyze the evolution of the air pollution over the elapsed time: during the days prior to the activation of the protocols described above, during the days of its activation and on subsequent days. It can be observed how the data samples are grouped according to different levels of pollution throughout the eleven days analysed for each one of the four episodes. In previous study, through the application of dimensionality reduction [7] and clustering techniques [8], it was observed the data grouping with similar levels of air quality, but without considering the time evolution.

The rest of this paper is organized as follows. Section 2 presents the techniques and methods that are applied. Section 3 details the real-life case study that is addressed in present work, while Section 4 describes the experiments and results. Finally, Section 5 sets out the main conclusions and future work.

2. Self-Organizing Maps

The Self-Organizing Maps (SOM) [12] [17] is a biologically plausible method for visualizing high dimensional data onto a low dimensional display.

A SOM consists of components called nodes or neurons. Associated with each neuron, there is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of neurons is a regular spacing in a hexagonal or rectangular grid. It is composed of a discrete array of L nodes arranged on an N-dimensional lattice and it maps these nodes into a D-dimensional data space while maintaining their ordering. The dimensionality, N, of the lattice is normally less than that of the data.

Typically, the array of nodes is one or two-dimensional, with all nodes connected to the N inputs by an N-dimensional weight vector. The self-organization process is commonly implemented as an iterative on-line algorithm. An input vector $\mathbf{x}$ is presented to the network and a winning node c is chosen whose weight vector W_c has the smallest Euclidean distance [18] from the input:

$$c = \arg\min_i \left(\left\| \mathbf{x} - W_i \right\| \right) \tag{1}$$

The SOM is a vector quantizer, and data vectors are quantized to the reference vector in the map that is closest to the input vector. The weights of the winning node and the nodes close to it are then updated to move closer to the input vector. The neighbourhood of node i is the set of nodes denoted by $N(i)$ that are close enough to be influenced by the node i whenever it is the winner. Therefore, if the winner is c, then the weights of the nodes $i \in N^{(c)}$ will be updated during training. The amount by which the neighbours are updated is determined by the neighbourhood function, h_{ci}, which is a function of the Euclidean distance between the winner c and the other nodes in its neighbourhood i. This function is normally a Gaussian function [19]. There is also a learning rate parameter, η, which usually decreases as the training process progresses. The weight update rule is as follows:

$$\Delta W_i = \eta h_{ci} \left[\mathbf{x} - W_i \right], \forall i \in N^{(c)} \tag{2}$$

When this algorithm is sufficiently iterated, the map self-organizes to produce a topology-preserving mapping of the lattice of weight vectors to the input space based on the statistics of the training data. Each weight vector lies approximately at the centre of its Voronoi region [20], which holds the subset of points in the data space that are closer to this vector than any other in the map.

2.1. Neighborhood functions

The Neighborhood functions used in Section 4 are as follows:

- Gaussian function

$$h_{ci}(t) = e^{-d_{ci}^2/2\sigma_t^2} \tag{3}$$

Where σ_t is the neighborhood radius at time t, and $d_{ci} = |r_c - r_i|$ is the distance between map units c and i.

- Bubble function

$$h_{ci}(t) = 1(\sigma_t - d_{ci}) \tag{4}$$

Where σ_t is the neighborhood radius at time t, and $d_{ci} = |r_c - r_i|$ is the distance between map units c and i, and $1(x)$ is the step function.

2.2. U-Matrix

The U-matrix (unified distance matrix) visualizes distances between neighboring map units, and thus shows the cluster structure of the map: high values of the U-matrix indicate a cluster border, uniform areas of low values indicate clusters themselves. Each component plane shows the values of one variable in each map unit. On top of these visualizations, additional information can be shown: labels, data histograms and trajectories [21].

The SOM with trajectories function is available in the SOM MATLAB toolbox [22]. This function launches a "comet" trajectory visualization. It also allows to visualize the distribution of samples in neurons over time, this distribution is reflected graphically in the U-matrix as can be seen in section 4 of present paper.

3. Real-life Case Study

In present study, pollutant data recorded in two different places in the city of Madrid (Spain) are analyzed. Hourly data from two different time intervals with similar conditions of high air pollution have been selected; in the first one (comprising 11 days) no actions against pollution were taken in Madrid City while in the other three (comprising 11 days each of them), the previously explained Scenarios were activated.

The two stations selected for this study are:

1. Madrid 1. “Plaza del Carmen” Station. Geographical coordinates: 3º42'17" W; 40º25'16" N; 657 meters above sea level (masl). Data acquisition station characterized as background urban station.
2. Madrid 2. “Escuelas Aguirre” Station. Geographical coordinates: 3º40'52" W; 40º25’32" N; 672 masl. Data acquisition station characterized as urban traffic station.

Figure 1 shows the location of the two selected stations that have been studied in present paper.

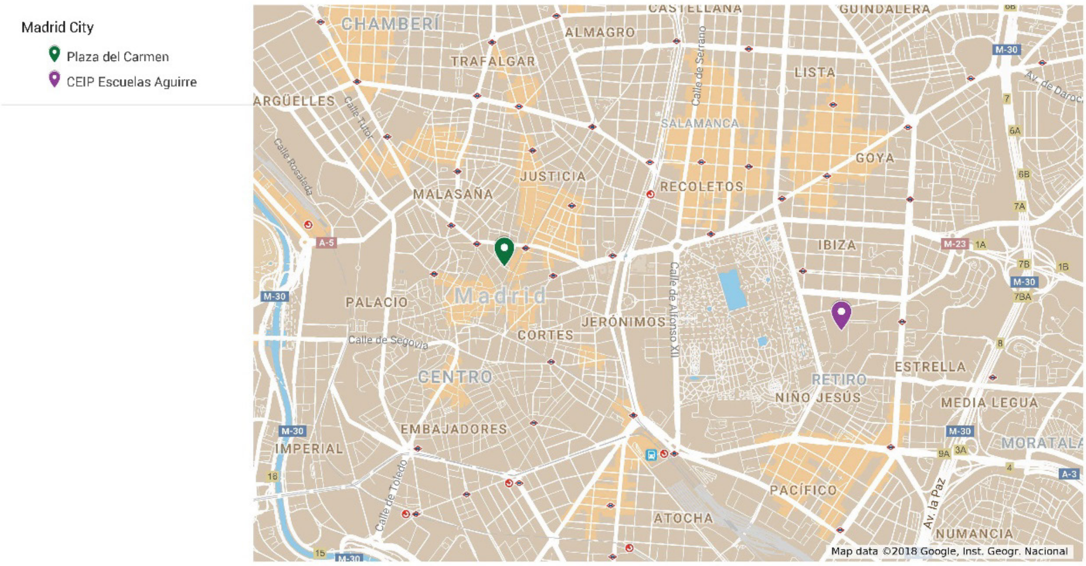


Figure 1 Location of the two selected stations in Madrid, by Google Maps.

These stations have been selected from the Madrid network of measurement stations due to two main reasons: both of them are located in the M-30 road where protocols for the air pollution control during episodes of high NO_2 are activated, and the two of them record information about the same pollutants, that are:

- Nitrogen Dioxide (NO_2) - µg/m³, primary pollutant. From the standpoint of health protection, nitrogen dioxide has set exposure limits for long and short duration [23].
- Sulphur Dioxide (SO_2) - µg/m³, primary pollutant. It is a gas. It smells like burnt matches. It also smells suffocating. Sulfur dioxide is produced by volcanoes and in various industrial processes. In the food industry, it is also used to protect wine from oxygen and bacteria [23].
- Carbon Monoxide (CO) - µg/m³, primary pollutant. Is an odorless, colorless gas formed by the incomplete combustion of fuels. When people are exposed to CO gas, the CO molecules will displace the oxygen in their bodies and lead to poisoning [23].
- Ozone (O_3) - µg/m³, secondary pollutant. Ozone is an odorless, colorless gas composed of three oxygen atoms. It occurs both in the Earth's upper atmosphere and at ground level. It can be "good" or "bad" for people's health and for the environment, depending on its location in the atmosphere [23].

From the timeline point of view, data are selected from four different time intervals; in [24] the list of high NO_2 episodes in Madrid city is available. The data from the four days prior to the entry into force of the protocols and the data of some days after the end of the protocols have been selected, until completing eleven days for each episode. The four episodes analysed in present study are:

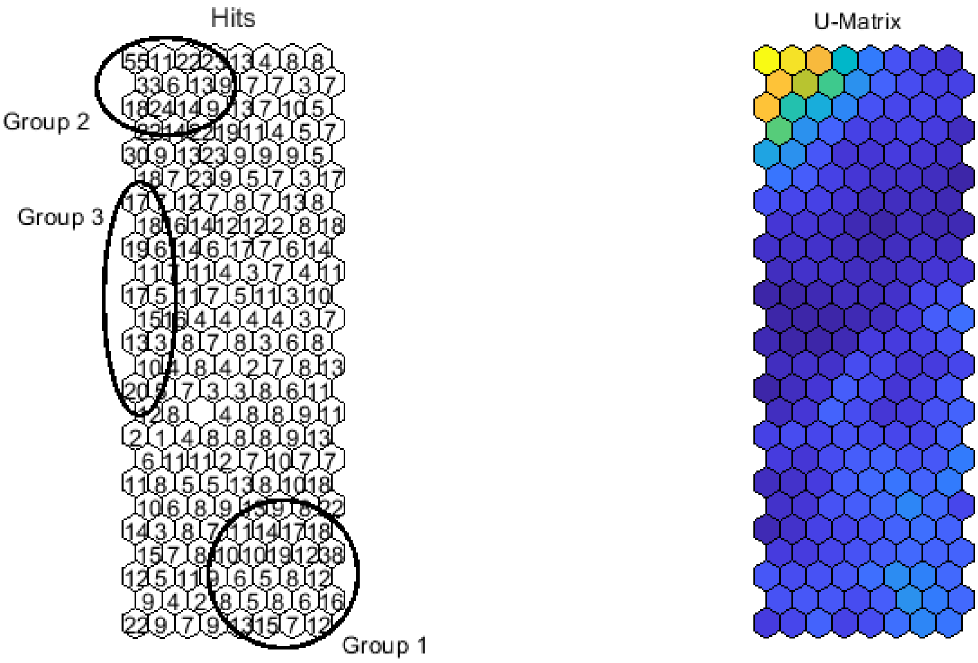
1. January, 5th to 15th, 2015. During these days, some characteristics of high environmental pollution determined by a very dry meteorology and lack of wind were present [25]. The protocols in the IAQS were not activated as they were approved in March 2015.
2. March, 6th to 16th, 2017. During those days, the environmental conditions were very similar to those in the 2015 period. Protocol actions with both Scenarios (I and II), above described, were activated on Friday (March, 10th) and Saturday (March, 11th).
3. October, 20th to 30th, 2017. The atmospheric conditions that have characterized those days have been of high stability [25]. Protocols actions were activated from Tuesday (October, 24th) to Saturday (October, 30th). Both Scenarios (I and II), were applied in this episode of high NO_2 concentration.
4. January, 20th to 30th, 2018. Again, a high stability in the climate (low wind and very low humidity), has favoured conditions of high atmospheric pollution. Protocols actions were activated from Tuesday (January, 23th) to Wednesday (October, 24th). Only the Scenario I was activated on this episode.

The episode numbered as 1. has been selected as the last episode of high levels of NO_2 prior to the entry into force of the current regulations. The other episodes (2., 3. and 4.) have been selected because high levels of NO_2 in the two selected stations were registered and Scenarios I and II were activated. Up to now there have been no episodes where Scenarios III and IV have been activated. Data about the four pollutants were recorded with an hourly frequency (from 1:00 to 24:00), so there is a total of 2,089 samples (24 samples, per each one of the eleven days in each one of the four episodes). Data missing or corrupted are omitted. All data from these six variables were normalized for the study. This data set has been divided into two subsets, one for each measuring station.

4. Results and Discussion

The SOM, described in Section 2, was applied to the case study presented in Section 3 and the results are discussed below. To obtain the results presented in this section, multiple tests have been performed with different parameter values for the SOM; initialization: random and linear, training algorithm: batch and sequential, number of neurons: 50, 80, 100, 120, 150 and 200, neighborhood function: Gaussian, cut Gaussian, bubble, and Epanechikov function.

As a first step, the Hits and the U-matrix visualizations of the obtained SOM for the whole dataset (samples corresponding to both stations described in Section 3) are presented in Figure 2. The SOM was selected as the best one, obtained with the following parameter values: random initialization, batch training algorithm, 200 neurons, and Gaussian neighborhood function. In following subsections, data are split according to the station, for a fine-grained analysis.



(a) (b)

Figure 2 (a) Hits and (b) U-matrix visualizations for the complete dataset.

From Fig. 2, three different groups of neurons can be identified; Group 2 contains most of the samples with higher levels of air pollution, corresponding mainly to the days before and the first days of the protocol activation. Many of the neurons with lower values of pollution are found in Group 3 and correspond to the first days of the analyzed data set (where the maximum levels of NO_2 in the air had not been reached) and to the final days of episodes, because of the positive effect of the traffic control measures. Group 1 contains a high concentration of data, which have low-average levels of air pollution, these levels can be recorded in nightly shots on in any of the analyzed days.

Figure 3 shows the results of applying SOM trajectory to the complete dataset corresponding to both stations.

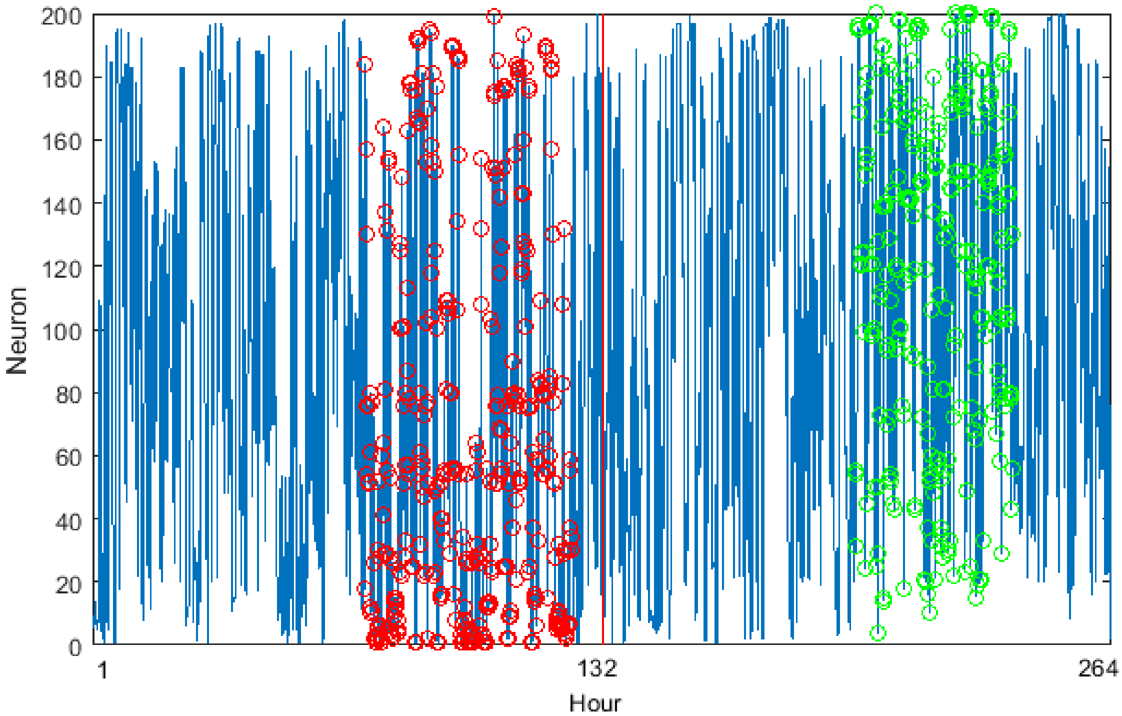


Figure 3 SOM with trajectories, both station

In Figure 3, the X-axis represents the time in hours for the eleven days analyzed (a total of 264 hours). The first sample at the beginning of the X-axis corresponds to the 1:00 on the first day and the last one corresponds to the 24:00 on the 11th day. The Y-axis represents each one of the 200 neurons in the SOM, numbered from bottom to top. The actions taken according to Scenarios I and II begin on the fourth day in the morning, except for the first episode in which these protocols were not yet approved, and are activated between 2 and 5 days, depending on the evolution of weather conditions and the effectiveness of traffic restriction protocols. With this information it can be deduced that the samples (colored in red) located to the left of the horizontal red line are previous to the beginning of the activation of the protocols and the samples at the right side (colored in green) in Figure 2 corresponds to the activation of the protocols, as well as a few days later.

A high number of samples on the right, are located around neurons at the upper side of Figure 2, while those on the left, are much more distributed around neurons in the lower part of the figure. This fact would graphically illustrate the positive influence of the activation of these measures of traffic restrictions.

Figure 4 shows the U-matrix with the trajectories for both stations. It corresponds to the samples selected and colored in Figure 3. By means of the U-matrix, it is visually identified that the Group 2 in Figure 2.a, corresponds in Figure 4 mainly whit data colored in red and samples assigned to Groups 1 and 3 in Figure 2.a are most of them colored in green.

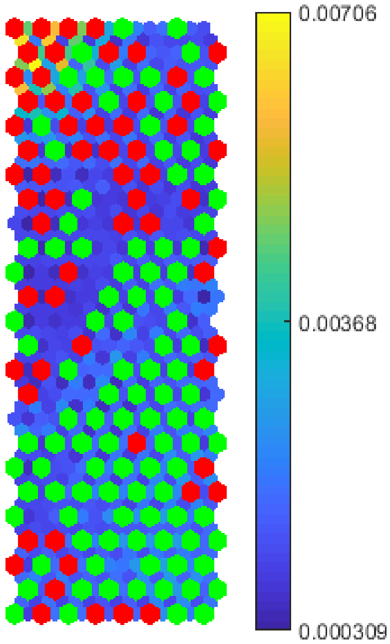
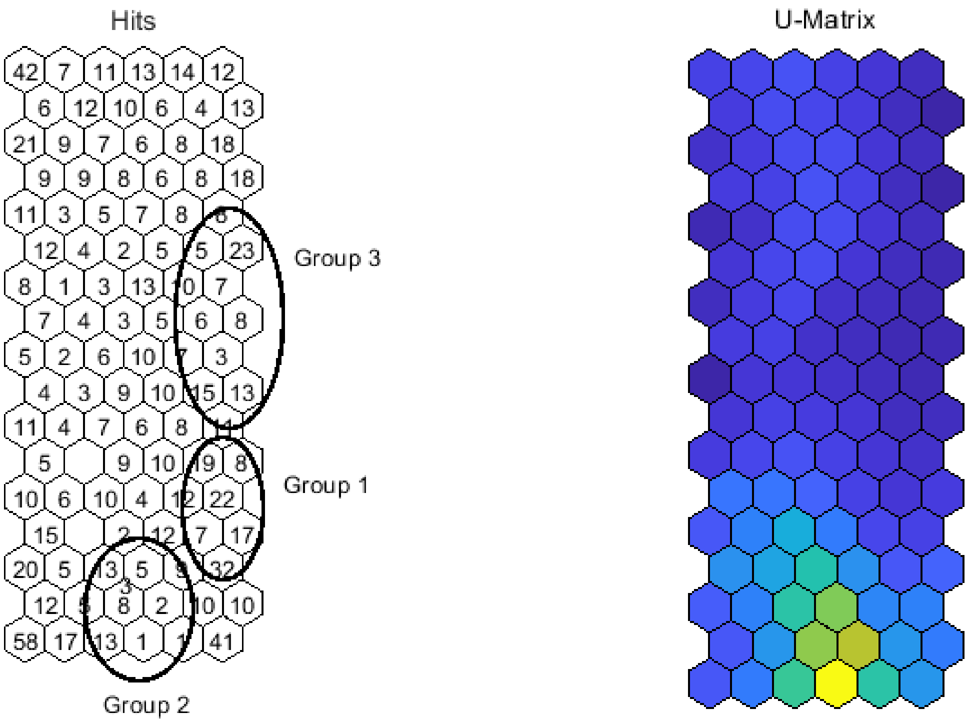


Figure 4 SOM with trajectories on the U-matrix, both stations

4.1. “Plaza del Carmen” Station

New experiments were run for the dataset only containing data from the “Plaza del Carmen” station. The Hits and the U-matrix visualizations corresponding to the best SOM mapping for this dataset are presented in Figure 5. Results were obtained with the following parameter values: Random initialization, batch training algorithm, 100 neurons and bubble neighborhood function.



(a) (b)

Figure 5 (a) Hits and (b) U-matrix for “Plaza del Carmen” station.

In Figure 5.a concentration of samples with the highest levels of air pollution are gathered in Group 2. Group 3 contains most of the samples with lower levels of air pollution and in Group 1.a high number of samples with moderate levels of air pollution are grouped, which can correspond to periods with low levels of traffic (at night) or to the days when the protocols for the traffic control are beginning to achieve their goals. These results are similar to those shown in Figure 2.a.

Figure 6 shows the results of applying SOM trajectory to the subset of data corresponding to the "Plaza del Carmen" station.

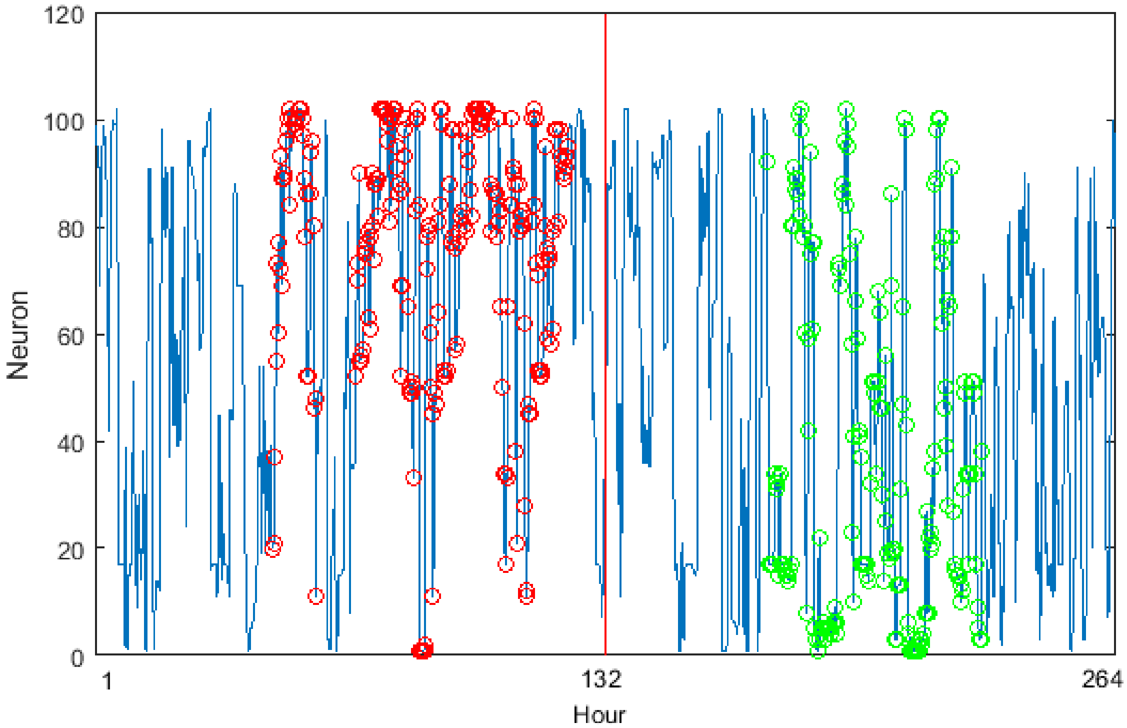


Figure 6 SOM with trajectories, "Plaza del Carmen" station

A high number of samples on the right side (in green), are located around neurons at the upper side of Figure 6, while those on the left side, are much more distributed around neurons in the lower part of the figure. This fact would graphically illustrate the positive influence of the activation of the measures of traffic restrictions. The pattern of time evolution visualized in Figure 3 (both stations) is repeated in Figure 6.

Figure 7 shows the U-matrix with the trajectories for the "Plaza del Carmen" station. It corresponds to the samples selected and colored in Figure 4. Thanks to this figure, it can be seen that the Group 2 in Figure 5.a, corresponds in Figure 7 mainly whit data colored in red and samples assigned to Groups 1 and 3 in Figure 5.a now are most of them colored in green.

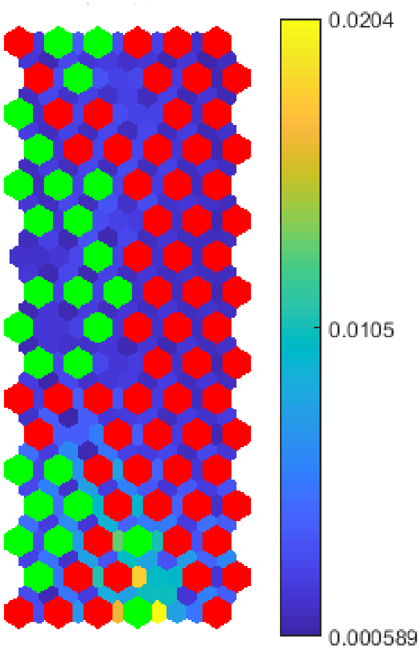


Figure 7 SOM with trajectories on the U-matrix, "Plaza del Carmen" station

Figure 8 shows the results of applying SOM trajectory on each of the original features with normalized pollutant information (NO_2, SO_2, CO, O_3), to the subset of data corresponding to the "Plaza del Carmen" station.

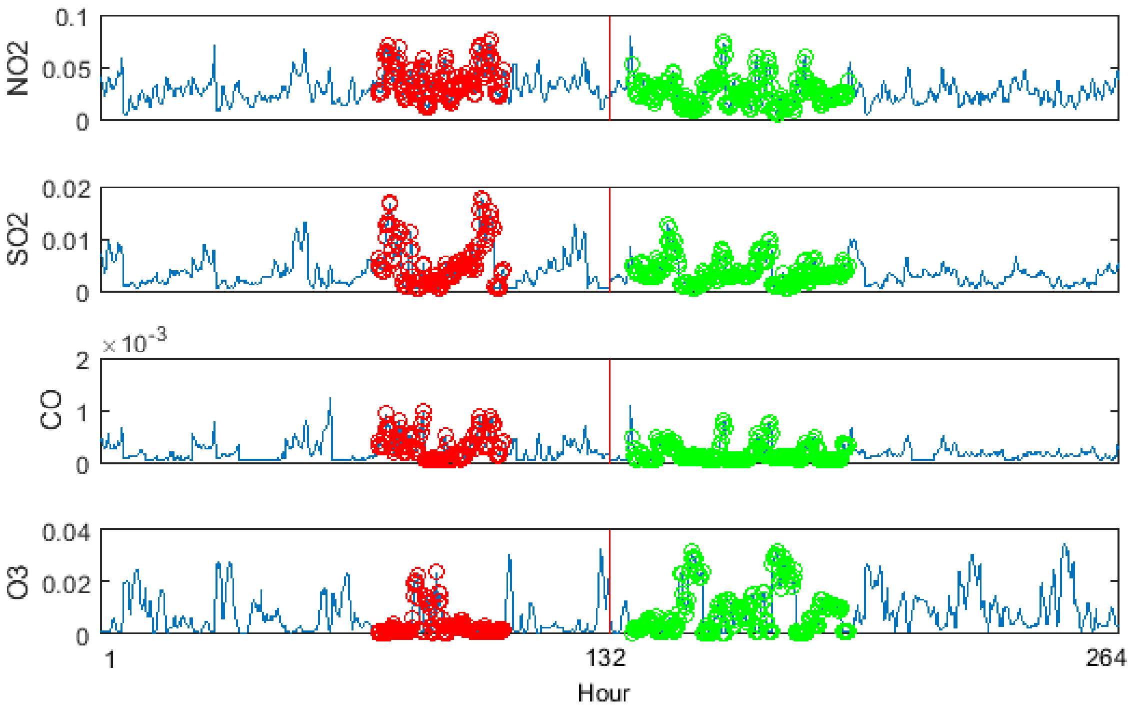


Figure 8 SOM with trajectories for each of the four pollutant features, “Plaza del Carmen” station

In Figure 8, the behavior of the 4 selected pollutants can be analyzed separately. It can be seen that the pollutants whose values decreased (in general terms) in the right part of Figure 8 are NO_2, SO_2 and CO. This is very positive because, in the center of the cities, these pollutants depend mostly of the traffic emissions. This fact, once again, demonstrates the effectiveness of traffic control protocols set up in Madrid city in episodes with high levels of NO_2. In the case of the O_3 pollutant, it can be said that its behavior may depend more on the variation of weather conditions than of the air quality, only the NO_2 influences the levels of O_3 and in a deferred way over time, which leads to explain that the O3 levels increase in the right part of the image. The meteorological conditions usually change in the last days analyzed (right part of Figure 8), these changes in the meteorology usually leads to the end of the episode and its traffic control restrictions.

Figure 9 shows SOM trajectories on the U-matrix corresponding to the “Plaza del Carmen” subset of data, for each of the four components separately.

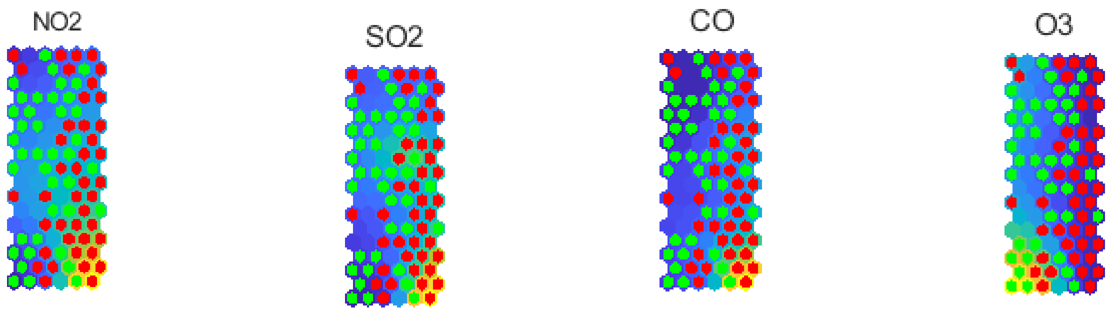


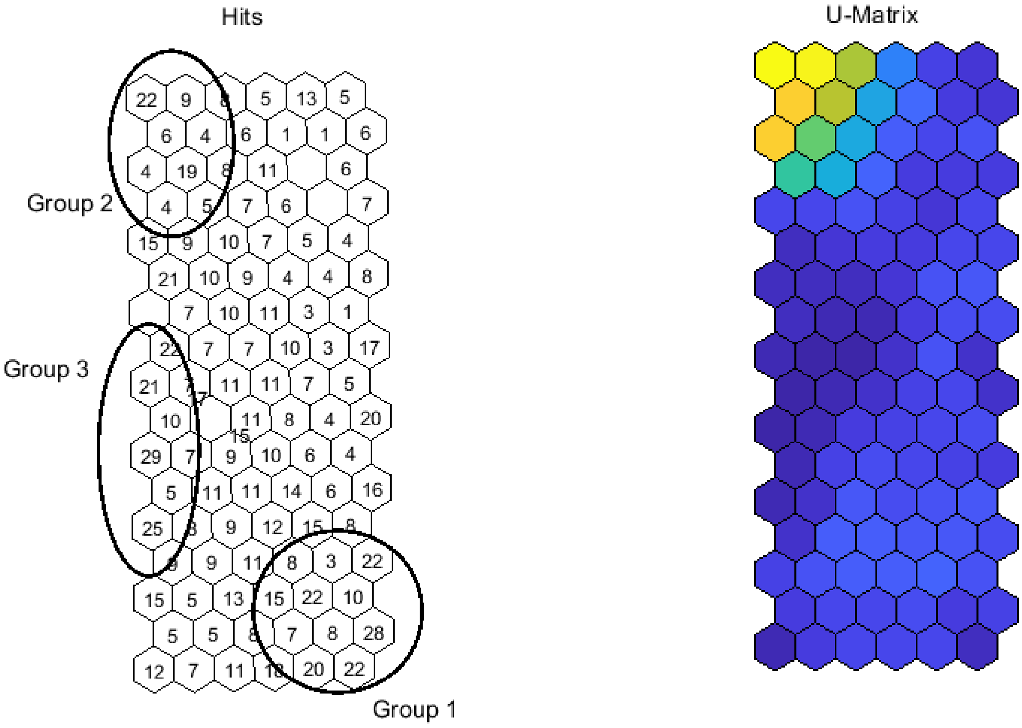
Figure 9 SOM with trajectories on the U-matrix for each of the four features, “Plaza del Carmen” station

The most remarkable characteristic of Figure 9 is the fact that for NO_2, SO_2 and CO pollutants are colored in red those areas corresponding to Group 2 in Figure 5 (a) which corresponds to the set of samples with higher levels of air pollution. It’s different in the case of O_3 pollutant, as it has been mentioned previously this pollutant does not depend as much on air pollution as the other three.

4.2. “Escuelas Aguirre” Station

The Hits and U-matrix visualizations corresponding to the “Escuelas Aguirre” station are shown in Figure 10. Among all the results, it was selected and obtained with the following parameter values: random initialization, batch training algorithm, 100 neurons and Gaussian neighborhood function.

The U-matrix corresponding to the "Escuelas Aguirre" dataset shows a mapping similar to that shown in Figures 3.a and 5.a. Samples with higher levels of air pollution are located in Group 2 and correspond to the days before the entry into force of the protocols of contamination; many of the samples with lower levels of air pollution in most of the components and in Group 1 are concentrated an important amount of samples low and medium levels of air pollution specially in NO_2, SO_2 and CO are located in Group 3.



(a) (b)

Figure 10 (a) Hits and (b) U-matrix for "Escuelas Aguirre" station.

Figure 11 shows the results of SOM trajectories to the "Escuelas Aguirre" dataset. In Fig. 11, it can be highlighted that in the right part of the figure, on the days corresponding to the entry into force of the protocols for the control of road traffic, the data samples tend to be distributed in neurons different from those of the first days (left part of the image). This fact is consistent with the results from the previous station (Figure 6).

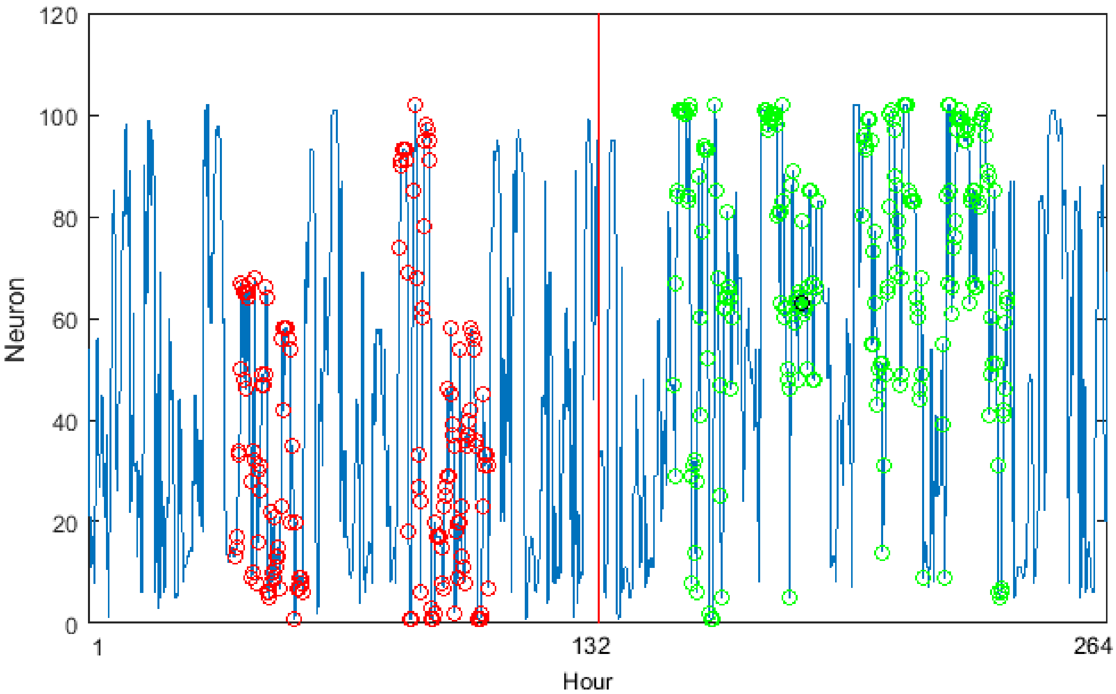


Figure 11 SOM with trajectories, “Escuelas Aguirre” station

Figure 12 shows the U-matrix corresponding to the SOM trajectory of “Escuelas Aguirre” subset of data.

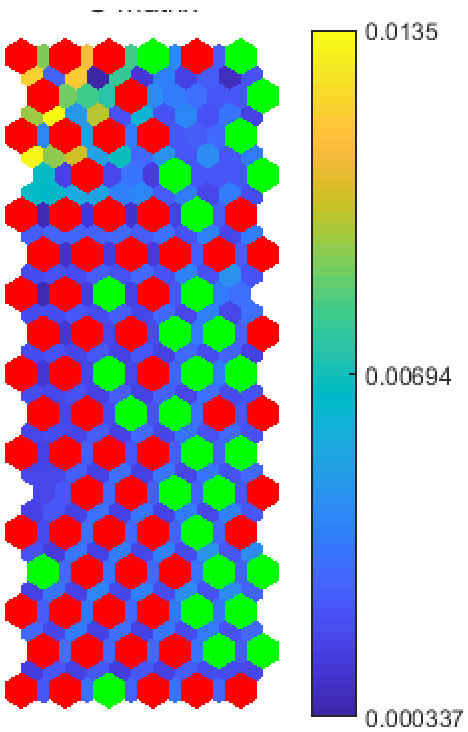


Figure 12 SOM with trajectories on the U-matrix, “Escuelas Aguirre” station

The trajectories generated in the U-matrix for the location of "Escuelas Aguirre" show positive results, since in the upper-left part of the Figure 12 the samples with higher levels of air pollution are concentrated and the corresponding neurons are colored in red.

Figure 13 shows the results of applying SOM trajectory for each of the four components with pollutant normalized information (NO_2, SO_2, CO, O_3), to the dataset corresponding to the “Escuelas Aguirre” station.

Analyzing the evolution over time of each of the 4 components, except for some pollution peaks which happened in moments of high road traffic in the center of the city, a clear improvement is observed in the right part of the image for levels of NO_2, SO_2 and CO pollutants. It can be seen a significant decrease in NO_2, which is the main target of these measures involving traffic control, especially considering that it is a station categorized as "urban traffic" station. As in Figure 8, the

O_3 levels barely changes or even increases its levels, this is because it is more susceptible to changes in meteorology than air pollution levels (only influenced by the NO_2 pollutant), this is consistent with the results analyzed in previous case study, Figure 8.

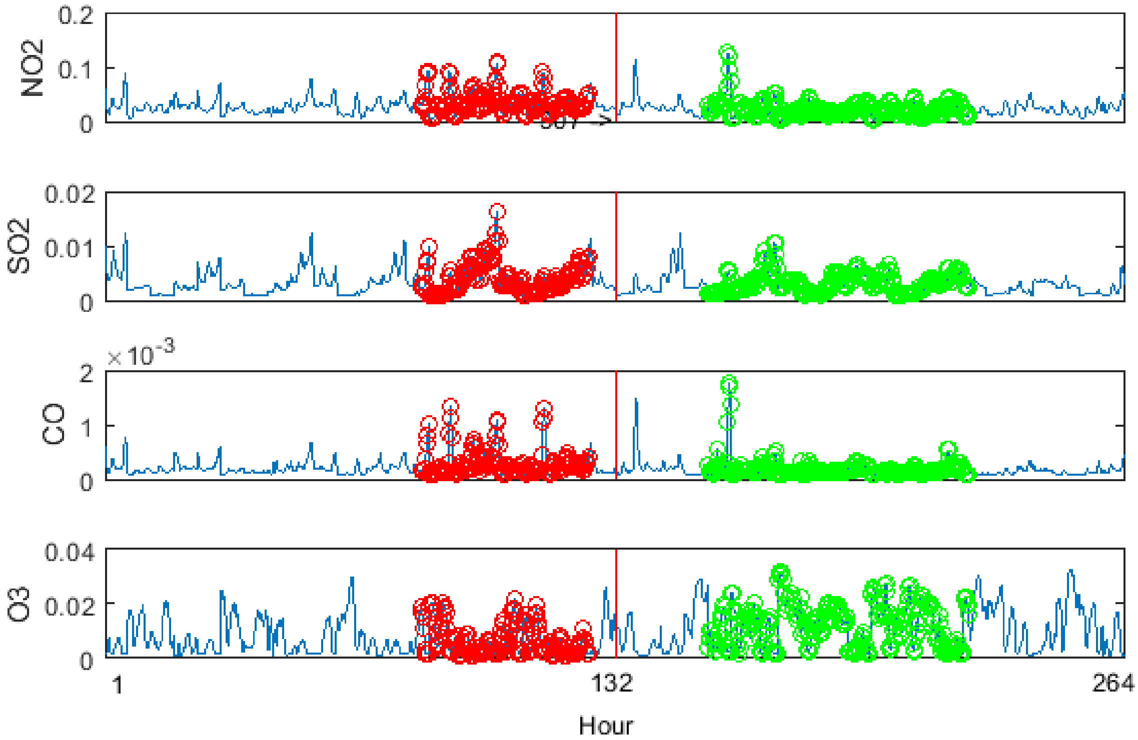


Figure 13 SOM with trajectories for each of the four pollutant features, "Escuelas Aguirre" station

Figure 14 shows the U-matrix corresponding to the SOM trajectory of "Escuelas Aguirre" dataset for each of the four components separately.

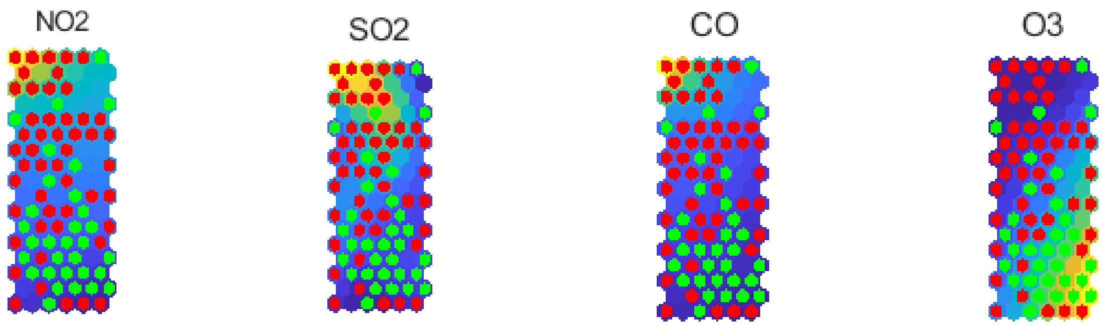


Figure 14 SOM with trajectories on the U-matrix for each of the four components, "Escuelas Aguirre" station

In the case of NO_2, SO_2 and CO pollutants, the areas with the highest pollution values are located in the upper left part of the Figure 14, identified as Group 2 in Figure 10 and these areas of high air pollution gathers data colored in red.

5. Conclusions and Future Work

Main conclusions derived from obtained results can be divided into two groups; firstly, those regarding to the analysis of air quality conditions in the considered case study. Secondly, those related to the performance of the SOM applied in the case study.

Since the entry into force of the protocols for the control of road traffic in the center of Madrid in episodes of high levels of NO_2, in several occasions during the years 2017 and 2018 it has been necessary to activate these protocols. Sometimes only the Scenario I was activated while in other episodes the Scenarios I and II were activated. The elapsed time of these measures is variable, from a couples of days to almost a week. Considering the results presented in Section 4, it can be concluded that throughout the activation of these protocols the general levels of air pollution are reduced (Figures 6 and 11) and especially the levels of NO_2, SO_2 and CO (Figures 8 and 13). It's important to highlight that in the results presented in Section 4 there are not an excessively clear aggregation, this is due to the fact that the data of an episode previous to the approval of the protocols with three episodes in which the protocols were in force are combined. Comparing the results of "Plaza del Carmen" with "Escuelas Aguirre", in the first case a greater decrease is obtained when applying the traffic control protocols in Figure 8 respect to Figure 13. If we observe the evolution of the components independently, the results are equally positive for both locations (Figures 8 and 11), it is important to highlight the significant reduction in NO_2 levels in the "Escuelas Aguirre" station, which is a very important hit as it is categorized as "urban traffic" station. It is worth highlighting the different behavior of O_3 compared to the other three other pollutants, much less influenced by the air quality variability, except by the deferred influence over time of the NO_2 pollutant. This fact makes very interesting the analysis by independent components.

Regarding to the application of the techniques presented in Section 2, the results have been extremely satisfactory. The SOM and its trajectories extension has proved to be a very useful analysis tool to visualize the evolution over time of a data set. Previous studies have applied techniques such as dimensionality reduction and clustering to analyze datasets about air quality, but the time variable was omitted. Applying SOM with trajectories let us analyze pollutions taking into account the important time dimension.

Future work will focus on extending the proposed analysis to other European big cities such as Barcelona, Paris or London where similar episodes of high pollution are happening. On the other hand, the trajectories SOM tool will be compared to some other visualization techniques that can deal with the time component.

References

[1] Government of Spain - Aporta Project, http://administracionelectronica.gob.es

[2] Council of Madrid City - Air Quality Integral System, http://www.mambiente.munimadrid.es/opencms/opencms/calaire/SistemaIntegral/concepto.html

[3] Council of Madrid City - Scenarios for the control of emissions during high periods of NO2 concentration in the air, http://www.mambiente.munimadrid.es/opencms/opencms/calaire/ServCiudadanos/ProtocoloNO2.html

[4] European Union - European Commission Environment, http://ec.europa.eu/environment/air/quality/standards.htm

[5] Prada, F. P., Monzon, A.:Identifying Traffic Emissions Hotspots for Urban Air Quality Interventions: The Case of Madrid City (No. 17-05015) (2017).

[6] Arroyo, A., Tricio, V., Herrero, A., Corchado.: Analysing the Effect of Recent Anti-pollution Policies in Madrid City Through Soft-Computing. Advances in Intelligent Systems and Computing Springer 649, 286-295 (2017).

[7] Van der Maaten, L. J. P., Postma, E. O., Van den Herik, H.J.:Dimensionality reduction: A comparative review. Journal of Machine Learning Research 10, 1-41 (2009).

[8] Jain, A. K., Murty, M. N., Flynn, P. J.: Data Clustering: A Review. ACM computing surveys (CSUR) 31(3), 264-323 (1999).

[9] Aznarte, J.L.: Probabilistic forecasting for extreme NO2 pollution episodes. Environ Pollut. 229, 321-328 (2017).

[10] Kalapanidas, E., Avouris, N.: Short-term air quality prediction using a case-based classifier. Environmental Modelling & Software 16(3), 263-272 (2001).

[11] Buscema, M.: Back Propagation Neural Networks. Substance Use & Misuse 33(2), 233-70 (2009).

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

[13] Kolehmainen, M., Martikainen, H., Hiltunen, T., Ruuskanen, J.: Forecasting Air Quality Parameters Using Hybrid Neural Network Modelling. Environmental Monitoring and Assessment 65(1–2), 277–286 (2000).

[14] Sammon, J.W.: A Nonlinear Mapping for Data Structure Analysis. IEEE Transactions on Computers 18(5), 401-409 (1969).

[15] Karaca, F., Camci, F. :Distant source contributions to PM10 profile evaluated by SOM based cluster analysis of air mass trajectory sets. Atmospheric Environment 44(7), 892-899 (2010).

[16] Air Resource Laboratoy - HYSPLIT project, https://www.arl.noaa.gov/hysplit/hysplit/

[17] Kohonen, T.: Self-Organisation and Associative Memory. Springer Series in Information Sciences. 8 (1998).

[18] Danielsson, P.E.: Euclidean distance mapping. Computer Graphics and Image Processing 14(3), 227-248 (1980).

[19] Wolfram MathWorld - Gaussian Function, http://mathworld.wolfram.com/GaussianFunction.html

[20] Voronoi, G.: Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites. Journal für die reine und angewandte Mathematik 133, 97-178 (1908).

[21] Horenko, I.: On clustering of non-stationary meteorological time series. Dynamics of Atmospheres and Oceans 49(2–3) 164-187 (2010).

[22] Laboratory of Computer and Information Science - SOM Toolbox, http://www.cis.hut.fi/projects/somtoolbox/

[23] PubChem – PubChem Compounds, https://pubchem.ncbi.nlm.nih.gov/

[24] Council of Madrid City - List of Episodes of High NO2, http://www.mambiente.munimadrid.es/opencms/opencms/calaire/Episodios/Informes_episodios.html~~.~~

[25] Council of Madrid City - Air quality annual reports, http://www.mambiente.munimadrid.es/opencms/opencms/calaire/Publicaciones/Memorias.html