Atmospheric Environment 60 (2012) 37–50

Contents lists available at SciVerse ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Assessment and prediction of air quality using fuzzy logic and autoregressive models

José Juan Carbajal-Hernández [a,b,*], Luis P. Sánchez-Fernández [a], Jesús A. Carrasco-Ochoa [b], José Fco. Martínez-Trinidad [b]

[a] Center of Computer Research – National Polytechnic Institute, Av. Juan de Dios Bátiz S/N, Gustavo A. Madero, Col. Nueva. Industrial Vallejo, 07738 México D.F., Mexico[1]
[b] Computer Science Department, National Institute of Astrophysics, Optics and Electronics, Luis Enrique Erro #1, Santa Maria Tonantzintla, Puebla C.P. 72840, Mexico

HIGHLIGHTS

► We examine air quality parameter levels using a statistical index (Sigma).
► We model environmental interactions using a reasoning process.
► We predict air quality parameters using historical observations.
► We compare the proposed air quality index with the Mexican and the U.S. indices.
► Our results show a better performance compared against traditional methodologies.

ARTICLE INFO

Article history:
Received 14 June 2011
Received in revised form
28 May 2012
Accepted 1 June 2012

Keywords:
Artificial intelligence
Air quality assessment
Pattern processing
Prediction

ABSTRACT

In recent years, artificial intelligence methods have been used for the treatment of environmental problems. This work, presents two models for assessment and prediction of air quality. First, we develop a new computational model for air quality assessment in order to evaluate toxic compounds that can harm sensitive people in urban areas, affecting their normal activities. In this model we propose to use a Sigma operator to statistically asses air quality parameters using their historical data information and determining their negative impact in air quality based on toxicity limits, frequency average and deviations of toxicological tests. We also introduce a fuzzy inference system to perform parameter classification using a reasoning process and integrating them in an air quality index describing the pollution levels in five stages: *excellent*, *good*, *regular*, *bad* and *danger*, respectively. The second model proposed in this work predicts air quality concentrations using an autoregressive model, providing a predicted air quality index based on the fuzzy inference system previously developed. Using data from Mexico City Atmospheric Monitoring System, we perform a comparison among air quality indices developed for environmental agencies and similar models. Our results show that our models are an appropriate tool for assessing site pollution and for providing guidance to improve contingency actions in urban areas.

1. Introduction

The presence in the air of substances which involve risk, danger or serious problems to people's health is known as air pollution. The main sources of air pollution are industrial processes involving combustion (industry and automobile) (Bartra et al., 2007). Nowadays, efficient methods for assessment of air quality are needed in order to establish mechanisms for managing pollutant concentration and preventing illness for sensitive people (USEPA, 2009; SMA, 2009). The criterion for good air quality varies with the kind of ecosystem and is established in levels. Several methodologies for the assessment and monitoring of air pollutants have been implemented by organizations such as the *United States Environmental Protection Agency* (USEPA, 2009), the Pan American Health Organization (PAHO, 2009) and the Mexican Ministry of Environment (SMA, 2009) among others, all of them have developed indexes for air quality. USEPA, SMA, and other similar indexes exhibit several weak points, where some parameters in the index equations can dramatically influence the final score without valid

* Corresponding author. National Institute of Astrophysics, Optics and Electronics, Computer Science Department, Luis Enrique Erro #1, Santa Maria Tonantzintla, Puebla C.P. 72840, Mexico.
E-mail address: juancarvajal@sagitario.cic.ipn.mx (J.J. Carbajal-Hernández).
[1] Tel.: +222 2663100.

1352-2310/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.atmosenv.2012.06.004

Table 1
Air quality parameters and their importance to air pollution.

Air quality parameters	Importance on air pollution.
Ozone (O_3)	In earth's lower atmosphere, ground-level ozone is considered "bad" (Bell et al., 2004; Borja et al., 1996). Breathing ozone can trigger a variety of health problems including chest pain, coughing, throat irritation, and congestion. It can worsen bronchitis, emphysema, and asthma (Bell et al., 2004; NOM-020-SSA1-1993).
Sulfur dioxide (SO_2)	This is one of the causes for concern over the environmental impact in the use of fuels as power sources. Inhaling sulfur dioxide is associated with increased respiratory symptoms and disease, difficulty in breathing, and premature death (USEPA, 2009; WHO, 1987; NOM-022-SSA1-1993).
Nitrogen dioxide (NO_2)	This reddish-brown toxic gas has a characteristic sharp, biting odor, being one of the most prominent air pollutants. Inhalation of such particles may cause or worsen respiratory diseases such as emphysema, bronchitis it may also aggravate existing heart disease (USEPA, 2009; Dovilé, 2008; NOM-023-SSA1-1993).
Carbon monoxide (CO)	It is a colorless, odorless, non-irritating but very poisonous gas. The health threat from carbon monoxide at low levels is most serious for those who suffer from cardiovascular disease, such as angina pectoris. At much higher levels, carbon monoxide can be poisonous. (USEPA, 2000; NOM-021-SSA1-1993).
Particulate matter	The size of the particles is directly linked to their potential effects for causing health problems. Once inhaled, these particles can affect the heart and lungs and cause serious health effects (Arreola and González, 1999; NOM-025-SSA1-1993). USEPA and SMA group particle pollution into two categories: particles smaller than 10 and 2.5 μm (PM_{10} and $PM_{2.5}$).

justification. However, the most critical drawback of these indexes is that they cannot deal with uncertainty and subjectivity present in this complex environmental problem.

Alternative methodologies for assessing air quality using fuzzy logic, which introduce environmental levels in the respective index and have a more accurate air quality evaluation, have been proposed. (Upadhyaya and Dashore, 2010; Liu et al., 2009; Alhanafy et al., 2010; Sowlat et al., 2011). However those researches only compute one sample and do not evaluate statistically the affectations of sampling deviations in the final score of a measurement set.

Other methodologies have been applied to the analysis of environmental pollution increasing the accuracy of the results, such as artificial neural networks (Salazar, 2007), associative memories (Yañez et al., 2008), support vector machines (Wang et al., 2008), factor analysis (Bishoi et al., 2009) among others. All these methodologies have the same lack of a reasoning process.

Air quality requirements are based on the results of chemical toxicity tests. These tests measure people responses to defined quantities of specific compounds. Air quality parameters have certain toxicity limits, where low or high concentrations can be harmful for people (NADF-009-AIRE-2006; WHO, 2005). Following the negative situations generated by the combination of different air quality parameters, it is possible to implement a computational model that according to the limits and fluctuations of those parameters can be used for determining when a concentration is good or bad for people. This strategy would reduce potentially negative situations for populations. In this way, associated diseases would be reduced.

Our proposal is to develop novel computational models to effectively assess and predict air quality conditions, in order to reduce uncertainty and imprecision in decision-making tools. These models combine different signal processing techniques, as well as fuzzy logic and autoregressive models. Our work is presented in three steps: first, a new algorithm for assessment of toxicity levels in air quality parameters is proposed, this algorithm is called the Sigma operator. The second step consists in creating a new air quality index based on a fuzzy inference system. These two steps form a model to assess the air quality. Finally, air quality parameters are predicted through an autoregressive model in order to predict future air quality conditions using the developed fuzzy inference system.

In the present study, the proposed models are applied for analyzing the air quality of Mexico City and its Metropolitan Area, where air quality parameters are assessed for establishing an indicator for good or bad air quality.

2. Air quality analysis

An air pollutant is a substance in the air that can cause harm to humans and the environment (Bartra et al., 2007). Pollutants can be present as solid particles or gases, and are frequently monitored in order to avoid negative health effects in populations (USEPA, 2009; SMA, 2009). There are six common air pollutants that have been studied for defining air quality levels: ozone, sulfur dioxide, nitrogen dioxide, carbon monoxide and particulate matter smaller than 10 and 25 μm Table 1 describes the importance of the 6 most common air quality parameters used as indicators of air quality on Mexico City and its Metropolitan Area and their respective importance.

2.1. Toxicity levels

Air quality parameters present random perturbations that can be harmful for long time expositions. In order to classify the negative impact of those parameters, it is necessary to define the levels for optimal or harmful concentrations. Table 2 presents the classification levels of the air quality parameters defined according to air pollution standards proposed by the USEPA and the SMA (USEPA, 2009; NADF-009-AIRE-2006).

Table 2
Classification levels of air quality parameters.

IMECA levels						
Parameters	"Good"	"Regular"	"High"	"Very high"	"Extremely high"	Deviation
O_3 [ppm]	0.000–0.055	0.056–0.110	0.111–0.165	0.166–0.220	>0.220	0.0275
NO_2 [ppm]	0.000–0.105	0.106–0.210	0.211–0.315	0.316–0.420	>0.420	0.0525
SO_2 [ppm]	0.000–0.065	0.066–0.130	0.131–0.195	0.196–0.260	>0.260	0.0325
CO [ppm]	0.00–5.50	5.51–11.00	11.01–16.50	16.51–22.00	> 22.00	2.75
PM_{10} [$\mu g\ m^{-3}$]	0–60	61–120	121–220	221–320	>320	30
$PM_{2.5}$[$\mu g\ m^{-3}$]	0–15.4	15.5–40.4	40.5–65.4	65.5–150.4	>150.4	7.7
	"Excellent"	"Good"	"Regular"	"Bad"	"Dangerous"	
AQI	0–50	51–100	101–150	151–200	>200	25

In our study, the allowed deviations used to determine range bounds (where values can be considered closer or farther from a specified level) were fixed as the half of the respective levels defined for each air quality parameter.

2.2. Air quality levels

According to air quality standards (USEPA, 2009; NADF-009-AIRE-2006), the negative effects in health of air pollutants can be classified as follows:

- *Excellent*: suitable for conducing outdoor activities.
- *Good*: outdoor activities can be carried out; but possible discomfort in children, the elderly and people with illnesses can be present.
- *Regular*: outdoor activities should be avoided; greater health effects in the population, particularly in children and older adults with cardiovascular and/or respiratory problems such as asthma, can be present.
- *Bad*: greater adverse health effects in the general population, particularly children and older adults with cardiovascular and/or respiratory conditions such as asthma.
- *Dangerous*: health effects on the general population. Serious complications can be presented in children and older adults with cardiovascular and/or respiratory conditions such as asthma.

3. Air quality assessment model

In this section, we introduce an air quality assessment model; for a better understanding we divide the presentation in two parts. The first part (Subsection 3.1), presents the use of a Sigma operator to statistically assess air quality parameters using their historical data information and determining their negative impact in air quality based on toxicity limits, frequency average and deviations of toxicological tests. The second part (Subsection 3.2) presents a fuzzy inference system to perform parameter classification using a reasoning process and integrating them in an air quality index (AQI) describing the pollution levels in five stages: excellent, good, regular, bad and danger, respectively.

3.1. Sigma operator (σ)

In air quality assessment, toxicity ranges define the negative impact on health. However, measurements into a range can be close to another range and these deviations directly affect the air quality score. In order to estimate the effect in air condition of fluctuations and deviations of parameter concentrations, the Sigma operator provides a [0, 1] score for toxicity levels.

The Sigma operator (σ) calculates the sample average from a desired level and the deviations from the average of such levels (Table 2) as follows.

3.1.1. Index 1: α (Frequency)

The frequency (α) represents the percentage of individual parameters whose current concentration is out of their allowed limits (failed tests).

$$\alpha = \frac{m_f}{m_t} \tag{1}$$

where m_f is the number of failed tests and m_t is the number of total measurements.

3.1.2. Index 2: β (Amplitude)

The amplitude (β) represents the average deviation of a set of measurements that are close to other ranges or out of the assessed level. An asymptotic function establishes the result in a [0, 1] range for the frequency as follows:

$$\beta = \frac{d}{\sqrt{1+d}} \tag{2}$$

Where d is the average deviation and it can be computed with Eq. (3):

$$d = \frac{\sum_i^n e_i}{m_t} \tag{3}$$

where i: 1, 2, … n; n is the number of calculated deviations and e_i is the ith deviation of the set of measurements. Deviation e can be determined in two steps as follows:

a) When the value must not exceed the level:

$$e = \frac{m - l_a}{t_a - l_a} \tag{4}$$

where m is the value of the test; l_a is the upper limit of the range to evaluate; t_a is the upper tolerance.

b) When the value must not fall below the level:

$$e = \frac{l_b - m}{l_b - t_b} \tag{5}$$

where l_b is the lower limit of the evaluated range; t_b is the lower range tolerance.

For example, limits and tolerance for O_3 (first row in Table 2) concentrations in a *regular* condition can be defined according to Table 2 as $l_a = 0.0825$; $t_a = 0.1375$; $l_b = 0.0825$ and $t_b = 0.0275$.

3.1.3. Index 3: σ (Sigma operator)

The Sigma operator (σ) classifies a set of toxicological concentrations establishing a status level according to Table 2. The σ index is computed as follows:

$$\sigma = 1 - \alpha\beta \tag{6}$$

The value that σ takes can be interpreted as follows:

- If $0 < \sigma < 1$, the concentration set is classified inside the evaluated range.
- If $\sigma = 0$, the concentration set is classified totally outside the evaluated range

Fig. 1 shows an example of two signals (set of air quality parameter concentrations) that have been assessed using the Sigma operator. In this example, we evaluate particles smaller than 10 μm and ozone for *excellent* and *regular* levels.

3.2. Fuzzy inference system (FIS)

In this section, a fuzzy inference system is used in order to analyze the air quality condition in urban areas based on a reasoning process (rules). The main contribution of the fuzzy inference system

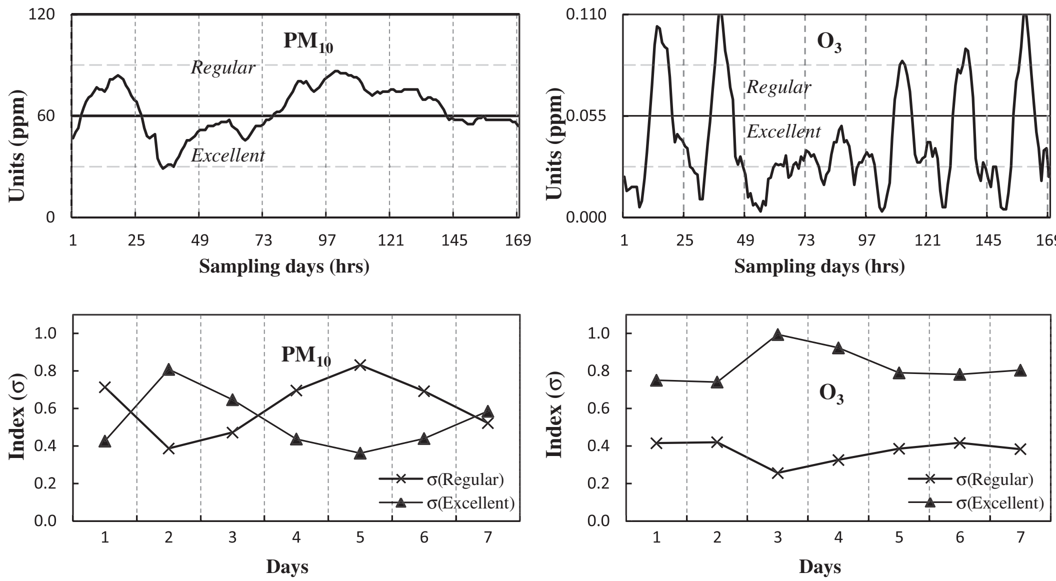

Fig. 1. A week of measurements was used to evaluate PM_{10} and O_3 concentrations using the Sigma operator for regular and excellent levels. It is important to remark that concentrations on the bounds deteriorate the level classification since limits used to define the Sigma operator are established in the half of the assessed level.

is to create an air quality index using the fuzzy theory, increasing the effectiveness of air quality assessment over traditional methodologies, and integrating particular concentration levels in a fuzzy index. Our hypothesis is that minimal deviations and changes in the air quality parameters directly affect the air quality status generating harmful tendencies in the assessment of air pollution.

Primary pollutants can be affected by external factors as the weather and these pollutants can produce secondary pollutants due to chemical processes. The fuzzy inference system takes into account the interactions between typical primary pollutants in different concentrations in order to detect and control effects of secondary pollutants.

It is important to consider that air quality parameters are defined by limits which must not be exceeded. When concentrations fall inside a range, the classification of these concentrations can be easily determined, but when concentrations fall on boundaries, the classification of air quality cannot be clearly established. Uncertainty allows a treatment of air quality parameters when their classification status is ambiguous, allowing to quantify concentrations that present offsets from a respective level. Subjectivity refers to specific interpretations of any aspect of experiences. The proposed index uses subjectivity in order to determine different problems, in the air pollutants, generated by different parameter conditions. This process is implemented using rules, which are included within a fuzzy inference system. Both, uncertainty and subjectivity are powerful techniques for an effective assessment of the air quality condition in our model.

Fuzzy inference is the process of formulating a mapping from a given input to an output using fuzzy logic (Fig. 2). This mapping provides a basis from which decisions can be made, or patterns could be discerned (Mo-Yuen, 1997). The process of fuzzy inference can be expressed in four phases: membership functions, inference rules (If-then rules), aggregation, and defuzzification (Zadeh, 1978; Ocampo et al., 2006; Soler, 2007).

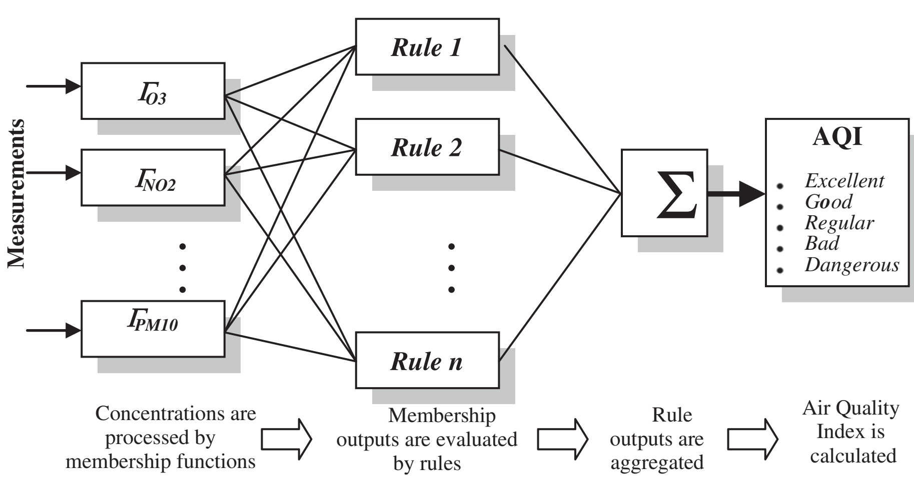

Fig. 2. Architecture of the Fuzzy Inference System applied to the air quality problem.

3.2.1. Membership functions

Membership functions (μ) transform real data into a value in [0, 1] value. There is not a specific method to build a membership function; the most common functions are triangular, rectangular, trapezoidal or Gaussian (Soler, 2007; Ocampo et al., 2006).

In our fuzzy inference system (FIS), we propose to use two types of membership functions: input membership functions for air quality parameters and output membership functions for air quality status. For input membership functions, the Sigma operator is used as membership function since it statistically evaluates air parameter concentrations in a [0, 1] range, using their respective toxicity levels as proposed in Table 2 (*good*, *regular*, *high*, etc.). It is important to remark, that each air quality parameter level corresponds to one membership function; in this case, five membership functions by parameter were implemented. Limits, parameters and ranges used to build the input membership functions based on the Sigma operator are shown in Table 3. On the other hand, membership functions for air quality conditions are built using linear functions, since they facilitate the defuzzification process and provide a good performance. Trapezoidal membership functions define the output transformation of the FIS (Ocampo et al., 2006; Zadeh, 1965), and they can be represented as in expression 7.

$$\mu(x, a, b, c, d) = \max\left\{\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right\} \quad (7)$$

where x is an air quality concentration parameter; a, b, c and d are membership parameters. In this work, we have developed five membership functions for the air quality according to the recommendations of the standards: *excellent*, *good*, *regular bad* and *dangerous* (USEPA, 2009; NADF-009-AIRE-2006). These functions were built using their respective air quality ranges and deviations defined in Table 2. The Fig. 3 shows the output membership functions used for computing the air quality and Table 3 shows the parameters used by the respective functions.

3.2.2. Inference rules (reasoning process)

In current air quality indices, when one air quality parameter takes a value (either low or high) out of the allowed limits, the index status changes irrespective of the excellent conditions of other parameters. On the other hand, when all air quality parameters are inside the same level, but some of them are on the bound, the index condition does not change. Additionally, some bad parameter concentrations can detonate harmful chemical reactions in other important parameters in air pollution and they are not considered in current air quality indices. These conditions for parameter concentrations will be taken into account in the rules of our FIS assessing the environmental condition of air pollution. In this sense, the fuzzy inference system will be able to detect potential crisis, bad changes on parameter concentrations or possible consequences in sensitive people only if the rules are built correctly.

Fig. 3. Air quality membership function.

In air quality assessment, there are expressions that are frequently used by experts: "if all parameter concentrations are in good levels, then the expected air quality is excellent" or "if the ozone level is good and nitrogen oxide is regular, and sulfur dioxide is good, and carbon monoxide is good, and particulate matter is good, then the expected air quality is good". In fuzzy language, those expressions could be enunciated as follows:

Rule 1: If $\mathbf{O_3}$ is *good* and $\mathbf{NO_2}$ is *good* and $\mathbf{SO_2}$ is *good* and **CO** is *good* and $\mathbf{PM_{10}}$ is *good* then **AQ** is *excellent.*
Rule 2: If $\mathbf{O_3}$ is *good* and $\mathbf{NO_2}$ is *regular* and $\mathbf{SO_2}$ is *good* and **CO** is *good* and $\mathbf{PM_{10}}$ is *good* then **AQ** is *good.*

In the same way, other rules can be enunciated. Robustness of the system depends on the number and quality of the rules. In this work, a set of 1025 rules were built and comprise the core of the FIS. The size of the set corresponds to the interactions and harmful combinations of the air quality parameters. In this example, we enunciate three more rules showing the main air conditions:

Rule 3: If $\mathbf{O_3}$ is *good* and $\mathbf{NO_2}$ is *good* and $\mathbf{SO_2}$ is *high* and **CO** is *good* and $\mathbf{PM_{10}}$ is *good* then **AQ** is *regular.*
Rule 4: If $\mathbf{O_3}$ is *good* and $\mathbf{NO_2}$ is *good* and $\mathbf{SO_2}$ is *good* and **CO** is *very high* and $\mathbf{PM_{10}}$ is *good* then **AQ** is *bad.*
Rule 5: If $\mathbf{O_3}$ is *good* and $\mathbf{NO_2}$ is *regular* and $\mathbf{SO_2}$ is *good* and **CO** is *good* and $\mathbf{PM_{10}}$ is *extremely high* then **AQ** is *dangerous.*

Inference rules are built considering air quality parameters combinations. According on how harmful a concentration is, the rule is built choosing as consequence an air quality condition. Also, if concentrations can disestablish other air quality parameters (primary or secondary parameters), the air quality condition is assigned. In this sense, the size of the rule set depends of how many rules are needed to describe the ecosystem dynamics. Standards define criteria clearly established for rule building; however, the success of the inference process depends on expert criteria, who define possible effects and crisis that an air quality parameter combination could generate.

Table 3
Parameters for membership functions used in the fuzzy inference system.

Indicator	"Good"			"Regular"				"High"				"Very high"				"Extremely high		
	$t_b = l_b$	l_a	t_a	t_b	l_b	l_a	t_a	t_b	l_b	l_a	t_a	t_b	l_b	l_a	t_a	t_b	l_b	$l_a = t_a$
O_3 [ppm]	0	0.027	0.082	0.027	0.082	0.082	0.137	0.082	0.137	0.137	0.192	0.137	0.192	0.192	0.247	0.19	0.247	∞
NO_2 [ppm]	0	0.052	0.157	0.052	0.157	0.157	2.62	0.157	2.62	2.62	0.367	2.62	0.367	0.367	0.472	0.368	0.472	∞
SO_2 [ppm]	0	0.032	0.097	0.032	0.097	0. 097	0.162	0.098	0.162	0.162	0.227	0.162	0.227	0.227	0.292	0.227	0.292	∞
CO [ppm]	0	2.25	7.75	2.25	7.75	7.75	13.75	7.75	13.75	13.75	18.75	13.75	18.75	18.75	24.25	18.75	24.25	∞
PM_{10} [μg m^{-3}]	0	30	90	30	90	90	150	90	150	150	250	150	250	250	350	250	350	∞
$PM_{2.5}$[μg m^{-3}]	0	7.7	23.1	7.7	23.1	23.1	48.1	23.1	48.1	48.1	73.1	48.1	73.1	73.1	158.1	73.1	158.1	∞
	$a = b$	c	d	a	b	c	d	a	b	c	d	a	b	c	d	a	b	$c = d$
AQI	0	25	75	25	75	75	125	75	125	125	175	125	175	175	225	175	125	250

Even though change of season or influences parameter concentration; the proposed fuzzy inference system does not consider temporality when it evaluates concentrations levels. In this sense, the fuzzy inference system does not depend directly of the season; it depends exclusively on the input information, which of course depends on the season.

Inference rules must be evaluated in order to determine the air quality condition level. In this work, rules based on *and* operators have been implemented; they can be evaluated using the following expression:

$$\mu_R = \min\left\{\mu^i_{O_3},\ \mu^j_{NO_2},\ \mu^k_{SO_2},\ \mu^l_{CO},\ \mu^m_{PM_{10}}\right\} \tag{8}$$

where i, j, k, l and m are the evaluated levels respectively; this expression represents the rule antecedent. Fig. 4 illustrates the operation of rules 1 and 2.

3.2.3. Aggregation

Once the rule set has been processed, an integration of the results must be done. Output membership functions for air quality are matched and truncated according the consequent of the rule as follows:

$$\mu_{R\_out} = \min\left\{\mu_R,\ \mu^l_{AQI}\right\} \tag{9}$$

where l is the selected membership function (*excellent, good, regular, bad or dangerous*); this expression represents the rule consequent. Therefore, all truncated membership functions (μ_{R_out}), for each rule, are aggregated by superposing the shapes creating one final membership function (μ_{out}). Fig. 4 shows the aggregation process.

3.2.4. Defuzzification

The final step of the fuzzy inference system is the defuzzification process, where the Air Quality Index is computed using the centroid method (Fig. 4). The centroid function is the most prevalent and physically appealing of all available methods for a defuzzification process (Ocampo et al., 2006; Mo-Yuen, 1997; Ross, 2004). The centroid method returns the center of area under the curve formed by the output fuzzy function according to expression 10:

$$AQI = \frac{\int x\mu_{out}(x)\mathrm{d}x}{\int \mu_{out}(x)\mathrm{d}x} \tag{10}$$

4. Air quality prediction model

In general, prediction models developed for air quality prediction are based on analyzing the current air quality status, and giving assumptions for future conditions (Wang et al., 2011; Alhanafy et al., 2010; Yong et al., 2008). In this case, a prediction assumption could be wrong, since their models are only based on interpretations without any background. A prediction must be done using a solid background of the air quality parameters.

In this work, predictions are based on past observations using historical measurements of air quality parameters, predicting particular concentrations in different time periods and assessing them using a fuzzy inference system. The main idea of this step is to effectively predict the air quality parameters. However, the prediction success is based on the assessment of the predicted concentrations using the fuzzy inference system proposed in Section 3. In this sense, the prediction process is linked to the fuzzy environment, providing a better performance over traditional models.

Main air quality indicators are used to implement contingency plans in order to prevent future crisis. However, decision making using those indicators is based on parameter concentrations measured instantly. In this sense, an opportune treatment of air quality can avoid harmful crisis if an accurate prediction of parameters and air assessment is implemented.

There are several methodologies for data prediction, artificial neural networks (ANN) is one of the most used models. However, in an ANN the most suitable topology for a specific problem cannot be determined, and usually it is determined by a trial and error. In this

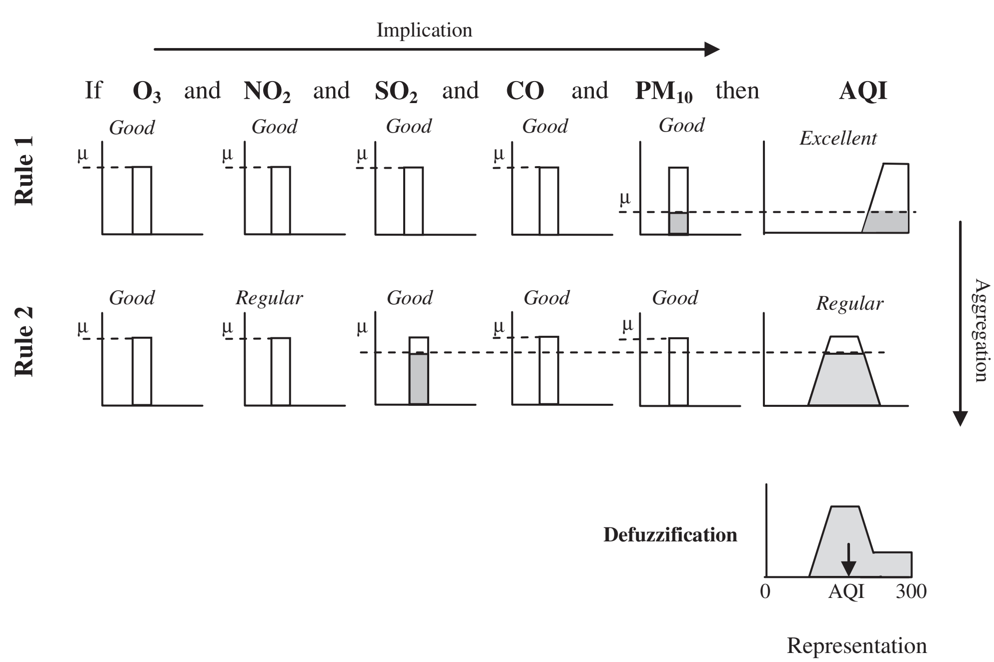

Fig. 4. Fuzzy inference diagram for the air quality scoring problem with five variables and two rules. Rules 1 and 2 were used to show the defuzzification process. The membership values (μ) of the 5 variables are used to truncate the AQI membership function assessed in the respective rule. All truncated functions (μ_{AQI}) are combined creating a final membership function (μ_{out}), which is used to determine the AQI by the centroid method.

section an autoregressive model (AR) is used due to its simplicity. An AR model is built by adding terms obtained from the original signal, and they can be implemented following a determined number of steps.

4.1. Preprocessing

Measurements always depend on the performance of electronic devices and human protocols. However, fluctuations can be due to changes in weather or changes in air quality. In order to deal with these fluctuations, before a prediction model is built (autoregressive model) some previous steps are needed to have a signal more suitable for modeling. In this work, smoothing and detrending of signals are used as preprocessing steps. The preprocessing is useful for the prediction, but not for the assessment of air quality. Even though, in preprocessing step some information is lost, it is irrelevant for air quality assessment. In general a signal is preprocessed in order to have a more suitable signal and to reduce the computational cost for further analysis. In this sense, preprocessing is not a result by itself; it is only a step that allows smoothing the input signal of the prediction process.

4.1.1. Smoothing

The signal of an air quality parameter contains fluctuations having several peak values and random behaviors. In order to reduce these effects, the smoothing process erases peaks using a digital filter. In this work, a moving average weighted filter is used for smoothing each air quality signal. This filter works using an average of signal points (measured concentrations) for producing new output points of the new filtered signal, as follows (Emmanuel, 1993):

$$y(n) = \sum_{i-0}^{K} b_i x(n-i) \tag{11}$$

where $x(n)$ is the air quality measured signal, $y(n)$ is the new smoothed signal, K is the filter order, b_i are the Spencer 15 terms coefficients (Kenney and Keeping, 1962), and they are defined as $1/320$ [−3, −6, −5, 3, 21, 46, 67, 74, 67, 46, 21, 3, −5, −6, −3]. Replacing the Spencer coefficients in Eq. (11) with $k = 14$, the smoothing using a moving average weighted filter is as in Eq. (12):

$$y(n) = -\frac{3}{320}x(n) - \frac{6}{320}x(n-1) - \frac{5}{320}x(n-2) + \ldots - \frac{3}{320}x(n-14) \tag{12}$$

Fig. 5 shows the smoothing process for the air quality signals.

4.1.2. Detrending

A signal usually contains some constant amplitude offset components or low frequency trends, those behaviors jointly with the fact that amplitudes of these trends sometimes are large could corrupt the results of time series modeling. (Chatfield, 2004; Shumway and Stoffer, 2000). Therefore, it is necessary to remove these trends before performing further analysis (Chapra and Canale, 1999). The trend is estimated as follows:

$$y(n) = a_0 + a_1 x(n) + e \tag{13}$$

where a_0 and a_1 are coefficients that represent the amplitude offset and the slope respectively, y is the output signal and e is the error between the modeled and the observed values. The coefficient a_1 and a_0 can be calculated using:

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \tag{14}$$

$$a_0 = \bar{y} - a_1\bar{x} \tag{15}$$

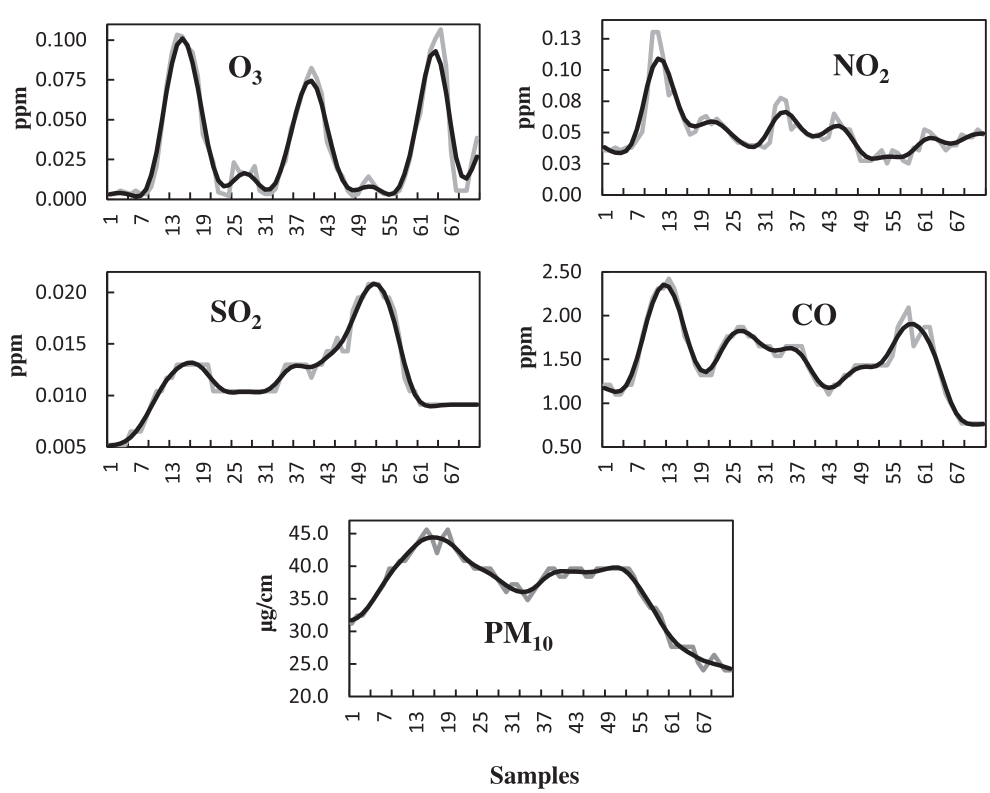

Fig. 5. Original and smoothed signal of the environmental variables using a moving average filter.

where n is the number of series points, and x_i is the ith measurement.

4.2. Prediction models

Air quality signals are built from series of measured values (observed values). Those measures are used to create a model which has a similar behavior. In this sense, air quality signals can be modeled using their own historical information. In this work, an autoregressive model (AR) is build using a set of measured parameter concentrations. In this sense, the AR model is able to predict a current value of the air quality parameter $x(t)$, based on a series of observed values $x(t-1)$, $x(t-2)$,..., $x(t-n)$ and a prediction error (De la Fuente and García, 1998; Brockwell and Davis, 1996), where n determines the number of past values used for predicting a new value (model order).

There is no straightforward way to determine the correct model order. There are several formal techniques for choosing the model order; in our study, the PHI criterion was the technique with best results for the AR models. The PHI criterion is a weighted estimation error based on the variation of a given signal with a penalty term when exceeding the optimal number of parameters to represent the signal (Chatfield, 2004; Emmanuel, 1993):

$$\mathrm{PHI} = e_n\left(1+\frac{(2n)\ln|\ln|L||}{L}\right) \tag{16}$$

where L is the number of points in the time series, n is the model order and e_n is the prediction error. In this case n varies to reach an optimum score. The AR models that we propose to describe the air quality signals using the model order (estimated with the *PHI* criterion) are:

$$O_3(t) = \sum_{i=1}^{20} a_i O_3(t-i) + e(t) \tag{17}$$

$$NO_2(t) = \sum_{i=1}^{20} d_i NO_2(t-i) + e(t) \tag{18}$$

$$SO_2(t) = \sum_{i=1}^{23} b_i SO_2(t-i) + e(t) \tag{19}$$

$$CO(t) = \sum_{i=1}^{22} g_i CO(t-i) + e(t) \tag{20}$$

$$PM_{10}(t) = \sum_{i=1}^{20} c_i PM_{10}(t-i) + e(t) \tag{21}$$

where $e(t)$ is the predicted error and a_i, b_i, c_i, d_i, g_i are the AR coefficients. Eqs. (17)–(21) represent the AR models for the air quality signal prediction. Each AR coefficient is calculated for a particular air quality parameter using its own information. The AR coefficients can be determined using the Yule–Walker equations (Dijkhof and Wensik, 2000) as follows:

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(n) \end{pmatrix} \tag{22}$$

where the γ operator can be calculated using the expected value as:

$$\gamma(h) = \langle x(t)x(t-h) \rangle \tag{23}$$

where x is an air quality signal.

Fig. 6 shows examples of the air quality signals reconstruction, where a total of 24 points (one day) were predicted using the proposed AR(n) models.

Once air quality parameter concentrations are modeled and predicted, the final step is to process those values using the fuzzy inference system; the final result is a predicted air quality assessment (P-AQI).

5. Experimental results

There are several models for air quality assessment. However, most of these models assess immediate measurements averaging frequently measured parameters without any previous treatment of the information. We can enunciate some of them for example, the metropolitan index for air quality (**I**ndice **M**etropolitano de la **C**alidad del **A**ire, IMECA in Spanish) (NADF-009-AIRE-2006), was developed for the Atmospheric Monitoring System (SIMAT, 2009), and shows the pollution level of the Mexico City and its Metropolitan Area in a [0, 250] range (see Table 2), this index uses some transforming equations and provides, as an overall evaluation, the air quality parameter with the highest result (SMA, 2009). The USEPA index is the base for most air quality indexes worldwide and it shows the air quality level in a [0, 500] range (see Appendix A). Additionally, Sowlat et al. (2011) proposed a fuzzy inference system based on linear inputs transformations where parameter concentrations are aggregated in a complete index. In this paper, we proposed another approach where inputs are no linear functions; moreover in our work we give a particular treatment to each parameter before process it by the inference system. In this section, we compare these four approaches (AQI, IMECA, USEPA and Sowlat et al.) in order to show the advantages of our proposal.

5.1. Study area

The research area encompassed the Federal District of Mexico and its Metropolitan Area (Fig. 7), which includes more than eight million inhabitants, (Fig. 7). The Federal District has an area of 1.485 km^2 and has a minimum altitude of 2200 m (7217 feet) above sea level, and is surrounded by mountains and volcanoes that reach elevations of over 5000 m high. This area is located in the Trans-Mexican Volcanic Belt located in the high plateaus of south-central Mexico (INEGI, 2009). Air pollution differences within Mexico City can be large due to meteorological factors, particularly wind direction.

5.2. Data information

The Mexico City Atmospheric Monitoring System (SIMAT, 2009) is committed to operate and maintain a trustworthy system for monitoring air quality in Mexico City and its Metropolitan Area (Fig. 7), made up by the Automatic Network for Atmospheric Monitoring (RAMA, 2009) and the Manual Network for Atmospheric Monitoring (REDMET, 2009), publishing their pollutant concentration information.

For our experiments, data information from 2008 year was obtained from RAMA and SIMAT databases. Although, concentration measurements have different time frequencies, these databases offer measurements each hour. IMECA results are given using averages of measurements in their respective period according to

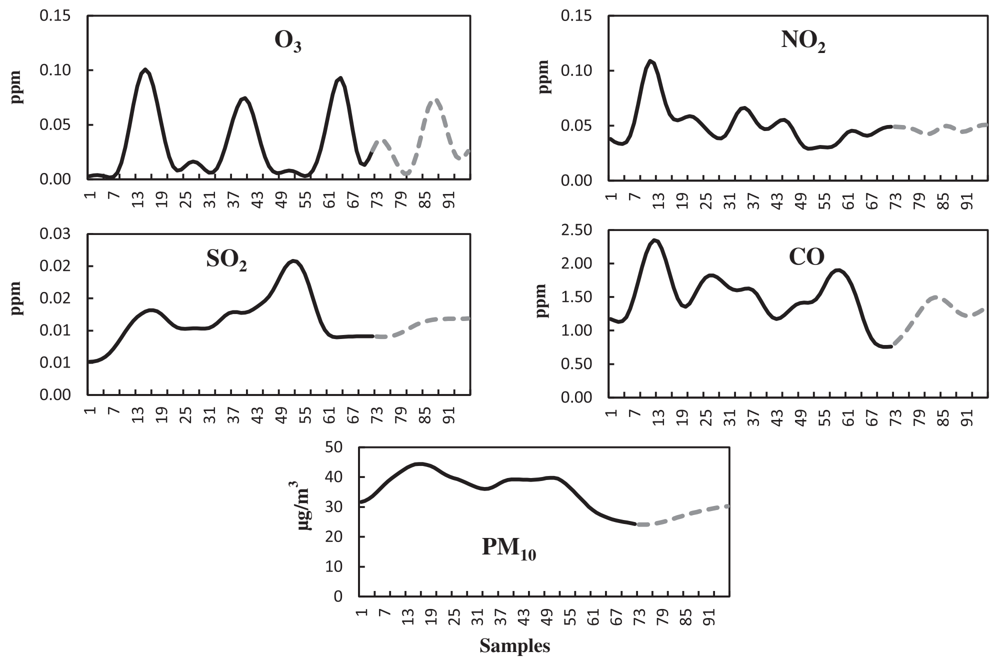

Fig. 6. Prediction of the environmental variables. The AR model predicts 24 measurements (1 day).

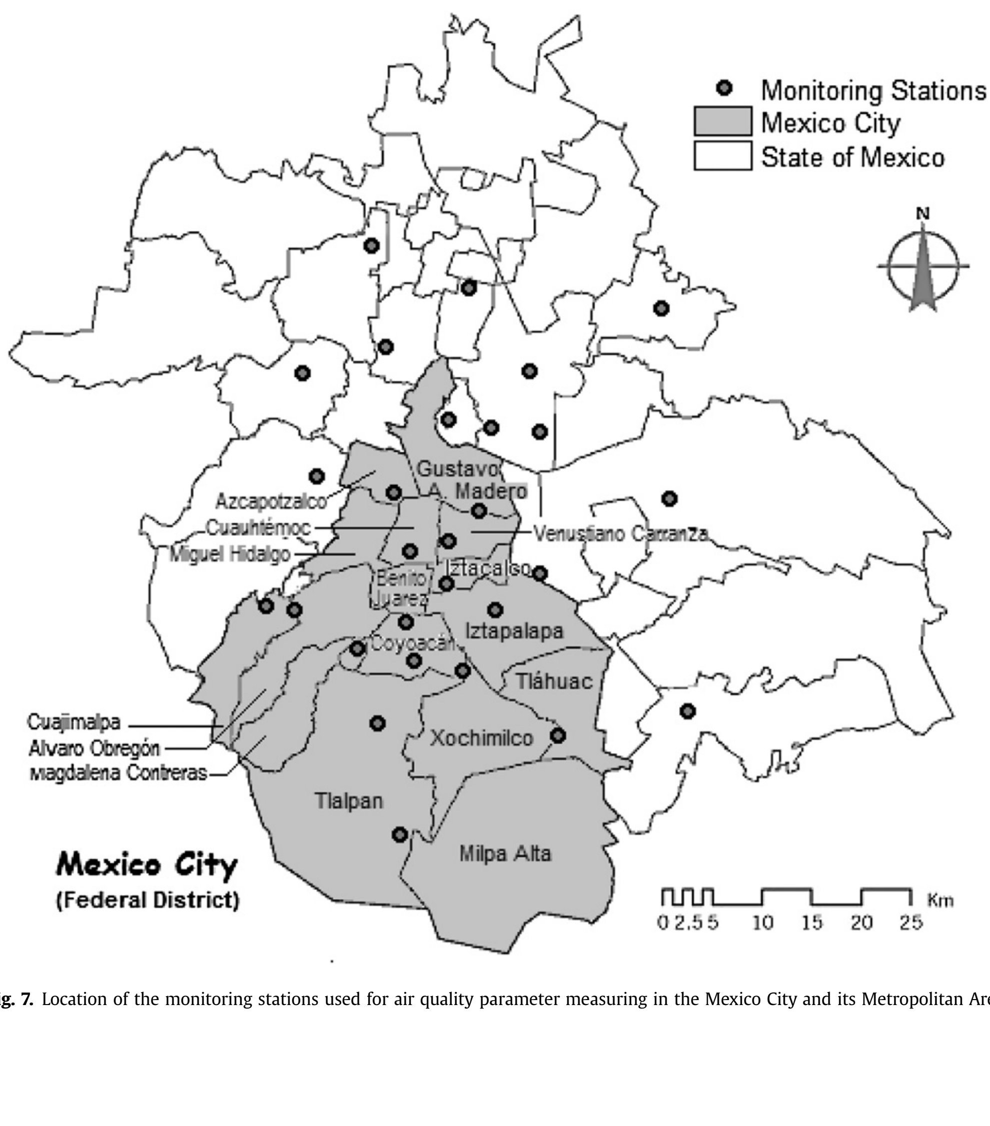

Fig. 7. Location of the monitoring stations used for air quality parameter measuring in the Mexico City and its Metropolitan Area.

the standards. Ozone and nitrogen dioxide averages are reported every hour; sulfur dioxide every 24 h; carbon oxide every 8 h and particulate matter every 24 h. Particles smaller than 2.5 μm are not measured by SIMAT.

5.3. Air quality assessment results

In our experiments we will assess the condition of the air quality in the Mexico City and its Metropolitan Area, through different air quality indexes, using data extracted from SIMAT public databases from January to March, 2008 (SIMAT, 2009). A better understanding of the environmental behaviors can be observed in Fig. 1, where it can be noticed that variable fluctuations appear every day.

Air quality has been assessed using a set of 24 measurements for each air quality parameter (one day of information). Following ranges and allowed deviations (see Table 3) it is possible determining a fuzzy input for the FIS using the σ index. A comparison of AQI against IMECA, USEPA and the index proposed by Sowlat et al. (2011) is shown in Fig. 8. According to the results showed in Fig. 8, air quality scores given by air quality indices in general are below 80, it means that those values are always considered as *excellent* and *good* air quality (51 and 100 levels respectively). However, the fuzzy environment and the treatment of data information using the Sigma operator in our proposal for air quality assessment (AQI), both affect directly the final score, reporting the worst air conditions, since AQI considers the amplitudes and deviations of the samples. The reasoning process provides a more accurate evaluation since it integrates all concentration levels in the final index. In this sense, the proposed AQI matches better with real data since it provides different levels for the air quality assessment, while

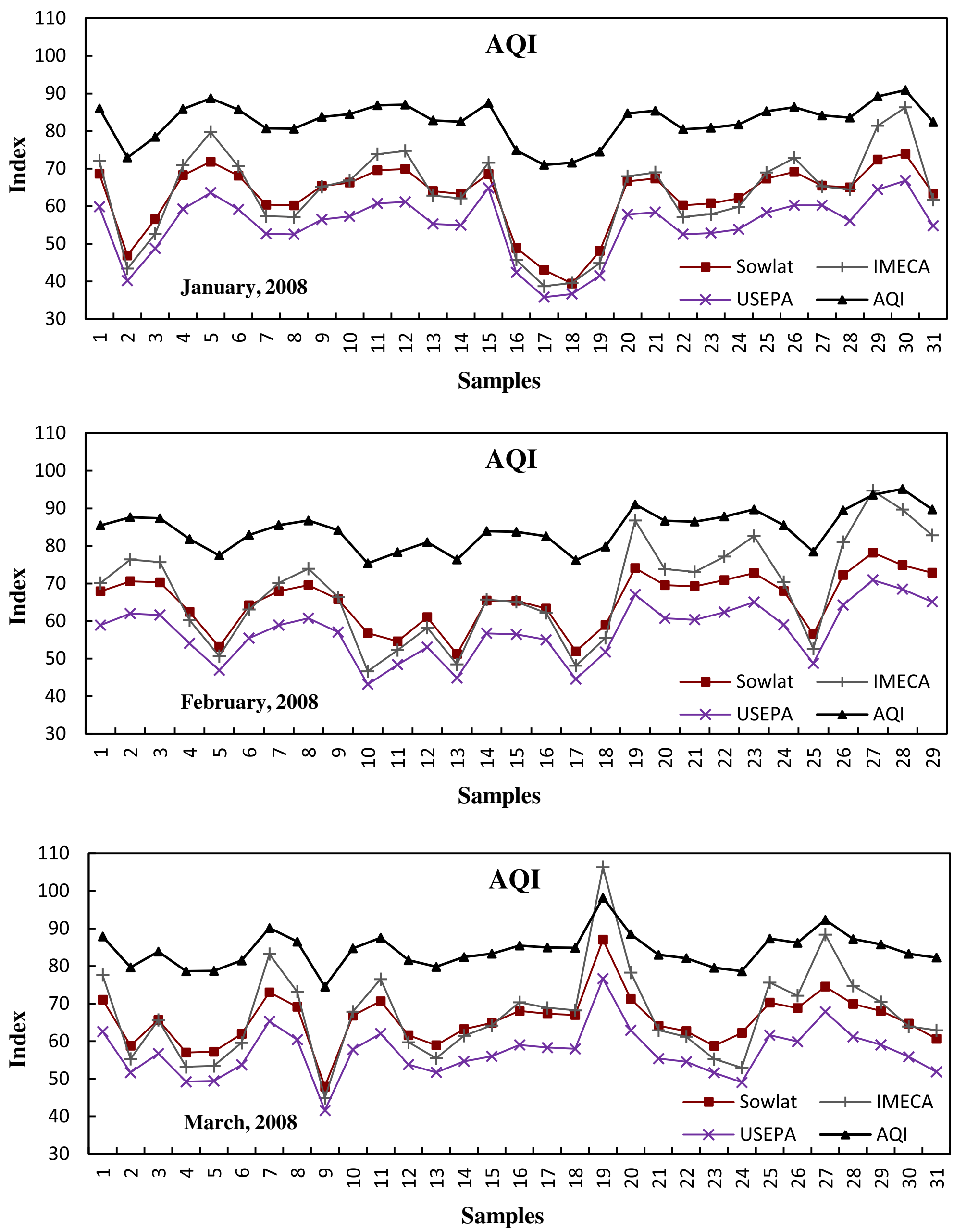

Fig. 8. Comparison between AQI, USEPA and IMECA from January to March, 2008. The fuzzy environment in the AQI affects the final score giving a more penalized assessment, due to the index integrates all affectations of minimal toxicity levels in all air pollutants.

Table 4
Comparison between AQI, IMECA, USEPA and the model proposed by Sowlat et al. (2011).

O_3 (ppm)	NO_2 (ppm)	SO_2 (ppm)	CO (ppm)	PM_{10} ($\mu g\ m^{-3}$)	AQI	IMECA	USEPA	Sowlat	Observations
0.0143	0.0336	0.0091	2.2	52.8	76.7	49.1	45.4	53	According to IMECA the air is excellent ¿Can a 49 value in a 0–50 range really be considered as an excellent index?
0.0121	0.0378	0.0091	2.31	55.2					
0.0077	0.0357	0.0091	2.31	60					
0.0077	0.0357	0.0091	2.42	68.4					
0.033	0.0126	0.0104	1.98	63.6	78.9	53.9	49.9	60.1	According to USEPA the air is good ¿Can a 49.9 value in a 0–50 range really be considered as a good index?
0.0374	0.0084	0.0091	1.87	62.4					
0.0374	0.0063	0.0091	1.65	62.4					
0.0385	0.0042	0.0091	1.43	61.2					
0.0187	0.0252	0.0026	0.66	112.8	95.8	99.4	73.2	81.5	According to IMECA the air is good. However, it is close to be in a regular level.
0.0198	0.0231	0.0026	0.66	115.2					
0.0231	0.021	0.0026	0.77	117.6					
0.0275	0.0189	0.0026	0.77	120					

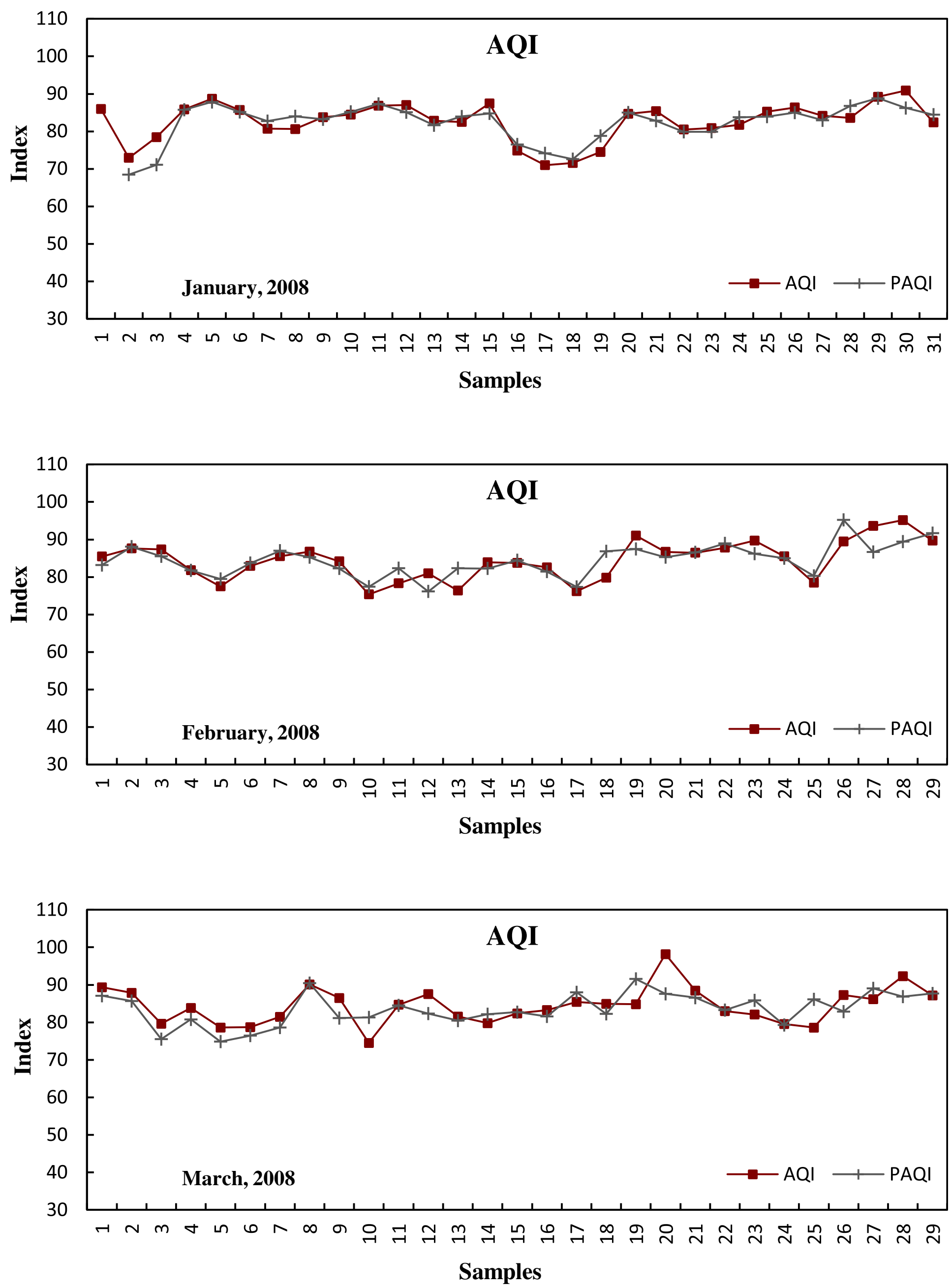

Fig. 9. Results of the prediction and assessment of air condition.

IMECA and USEPA only provide a true or false value and they change their values only when one air quality parameter exceeds its limits regardless other parameters. Table 4 describes the comparison among indexes where a set of measurements is analyzed. In the first evaluation IMECA and USEPA indexes report scores of 49.1 and 45.5 because PM_{10} presents values close to a regular situation, while other parameters are in excellent levels. However, the values of IMECA and USEPA are classified as *excellent* despite their proximity to *good*. In this case AQI evaluates sample deviations from a desired level and processes this information using the FIS, integrating all negative effects of toxicity levels giving a score of 76.7, which means a good air quality. The second analysis shows IMECA and USEPA values of 53.9 and 49.9 which means *good* and *excellent* levels respectively. However, it is hard understand why these values are too close to *good* and *excellent* levels. Another aspect to consider is that those values were considered only for PM_{10} values. The AQI solves this problem, since the analysis and integration of all situations generated by the air quality parameter set produces a score of 78.9. The index proposed by Sowlat et al. (2011) also solves those problems using a fuzzy inference system. However, in our proposed model the treatment of the inputs using the Sigma operator has a better performance.

5.4. Air quality prediction results

It is important to remark that air quality prediction is based in two steps: in the first step the air quality parameters are predicted, and in the second step the fuzzy inference system assesses predicted values having as a result a predicted air quality index. In our experiments, prediction tests were made using one day of information and one day of information was predicted (24 values); three months of measurements were extracted from data base (January, February and March, 2008). In other words, for one day of information, the next 24 h can be predicted using the AR model, and in the second step the predicted values are processed by the fuzzy inference system, calculating the predicted AQI (P-AQI).

The P-AQI performances were evaluated using correlation coefficient (R), mean error (ME), root mean square error (RMSE) and normalized root mean square error (NRMSE). In brief, the AQI predictions were optimum if R, ME, and RMSE were found to be close to 1, 0, and 0, respectively. Also R, RMSE, and ME were used to measure the prediction performance of P-AQI on the validation data set.

A comparison between assessed (AQI) and predicted (P-AQI) air quality are showed in Fig. 9. Scores of predicted air quality showed a good performance. In Fig. 10, statistical results about the prediction process are given, where the relationship between estimated and real values confirms the good predictability of the 24 h analysis. The points nearest to the diagonal line represent more accurate predictions. The values predicted by our model were too close to those measured values (see the line of exact fit); moreover ME showed low error rates (2.49, 1.95 and 3.26 from a [0–250] range).

6. Discussion

The methods proposed for assessing air quality provides a good approach in the air management field. The AQI assesses air quality using past information and giving a [0, 250] range for an air quality level. Currently, methodologies for air pollution evaluation do not consider the potentiality of reasoning processes in air quality assessment and the deviation of failed tests, which are very important for detection of potential harmful situations.

Traditional reports on air quality tend to be too technical and detailed, presenting monitoring data on individual substances, without providing a complete and interpreted description of air quality. To solve this gap, several air quality indexes have been developed to integrate air quality parameters. Traditional methodologies evaluate air quality in a rigorous sense, where certain levels of concentrations are classified in a strict level. The proposed AQI, follows a soft approach where the measured concentrations are processed giving an indication of the air pollution degree, integrating all compounds.

Although IMECA and USEPA solve the problem of air pollution assessment, the reasoning process of harmful situations in AQI provides a more accurate evaluation. In addition, the proposed model integrates all air quality parameter evaluations providing

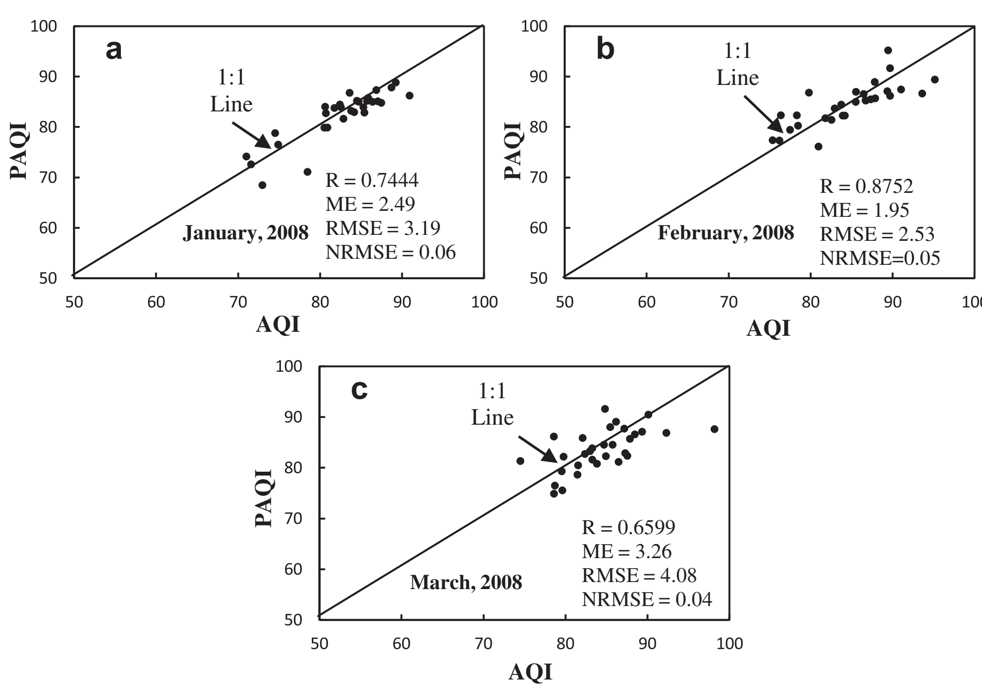

Fig. 10. Comparison between assessment and prediction, where the line represents the accuracy of the results.

a complete air quality index. Sowlat et al. (2011) propose a fuzzy inference system for air quality assessment; nevertheless, our model makes a treatment to the input of the system that significantly increases the performance of the air pollution assessment.

On the other hand, the prediction of air quality is an important factor in high polluted areas, where a crisis in air pollution can lead to serious health problems. Predicting bad air quality conditions would help to implement more effective preventive programs for controlling harmful pollutant emissions and to detect future dangerous concentrations. In this sense the IMECA, USEPA and international air quality indexes have not introduced methodologies for predicting air quality conditions using past values. The preventive programs are implemented after a high polluted concentration exceeds a critical limit.

7. Conclusions and future work

In this paper, a new model based on fuzzy inference systems has been introduced to assess air quality status. The proposed AQI works in two steps: first, the toxicity of a set of measured concentrations is classified by levels (Sigma operator); second, the effects in the ecosystem are evaluated in order to determine air quality status (fuzzy inference system). A comparison between models shows that a fuzzy environment directly affects the results providing a more accurate index dealing with real data. Experimental results showed that the proposed algorithm is an efficient way to monitor air pollution in urban areas.

A model to predict air quality was also developed and it works in two phases: first, a parameter prediction was done using one day of information; then predicted signals were assessed using the AQI in order to predict an air pollution level. A comparison of the assessment and the predicted air quality showed a good system performance. Therefore the proposed model in this research is a powerful tool in decision support for monitoring future air pollution environmental problems.

As future work we will assess and analyze specific polluted areas of Mexico City using the proposed AQI. We consider that the best way to validate an index performance is comparing it with impact indicators for air pollution. In this sense, more comparisons between air quality indices in the literature are going to be implemented in order to assess the effectiveness of the proposed system. Additionally, correlations between diseases and the proposed AQI scores will be studied in order to understand diseases appearing with bad air quality conditions.

Acknowledgments

The authors of the present paper would like to thank the following institutions for their support to develop this work: National Polytechnic Institute, Mexico, SIMAT, National Institute of Astrophysics, Optics and Electronics and CONACyT.

Appendix A

A.1 USEPA

USEPA measures the daily pollution index of the compounds for which the USEPA has established National Ambient Air Quality Standards (NAAQS). The index for a pollutant can be calculated as follows:

$$I_p = \frac{I_{\mathrm{Hi}} - I_{\mathrm{LO}}}{\mathrm{BP_{Hi}} - \mathrm{BP_{LO}}}\left(C_p - \mathrm{BP_{LO}}\right) + I_{\mathrm{LO}}$$

where, I_P is the index value for the pollutant, P; C_P is the pollutant truncated concentration; $\mathrm{BP_{Hi}}$ is the breakpoint that is greater or equal to C_P; $\mathrm{BP_{LO}}$ is the breakpoint that is lesser or equal to C_P; I_{Hi} is the AQI value corresponding to $\mathrm{BP_{Hi}}$, I_{LO} is the air quality value corresponding to $\mathrm{BP_{LO}}$. USEPA index is determined by considering the maximum index value (I_p) of a single pollutant (Bishoi et al., 2009; USEPA, 2009).

References

Alhanafy, T., Zaghlool, F., Saad, A., 2010. Neuro fuzzy modeling scheme for the prediction of air pollution. Journal of American Science 6 (12), 605–616.

Arreola, J., González, G., 1999. Análisis espectral del viento y partículas menores de 10 micrómetros (PM_{10}) en el área metropolitana de Monterrey, México. Revista Internacional de Contaminación Ambiental 15, 95–102.

Bartra, J., Mullol, J., del Cuvillo, A., Dávila, I., Ferrer, M., Jáuregui, I., Montoro, J., Sastre, J., Valero, A., 2007. Air pollution and allergens. Journal of Investigational Allergology and Clinical Immunology 17, 3–8.

Bell, M., McDermott, A., Zeger, A., Samet, J., Dominici, F., 2004. Ozone and short-term mortality in 95 US urban communities, 1987–2000. American Medical Association 292, 2372–2378.

Bishoi, B., Prakash, A., Jain, V., 2009. A comparative study of air quality index based on factor analysis and USEPA methods for an urban environment. Aerosol and Air Quality Research 9 (1), 1–17.

Borja, V., Loomis, D., Bangdiwala, S., Shy, C., Rascon, R., 1996. Ozone, suspended particulates, and daily mortality in Mexico City. American Journal of Epidemiology 145, 258–268.

Brockwell, P., Davis, R., 1996. Introduction to Time Series and Forecasting. Springer, New York.

Chapra, S., Canale, R., 1999. Numerical Methods for Engineers. McGraw-Hill, México.

Chatfield, C., 2004. The Analysis of Time Series: an Introduction, sixth ed. Chapman & Hall/CRC.

De la Fuente, D., García, D., 1988. Modelado de series temporales con métodos en bloque y recursivos. Desarrollo de estimadores y predictores adaptativos. Questiió 12, 281–313.

Dijkhof, W., Wensik, E., 2000. Small sample statistics of the Yule-Walker method for autoregressive parameter estimation. In: Proceedings of European Signal Processing Conference.

Dovilé, L., 2008. Nitrogen dioxide concentrations and their relation with meteorological conditions and some environmental factors in Kaunas. Environmental Research, Engineering and Management 1, 21–27.

Emmanuel, C., 1993. Digital Signal Processing: a Practical Approach. Addison-Wesley. [INEGI] National Institute for Statistics and Geography (Instituto Nacional de Estadística y Geografía, in Spanish). Avaliable at: http//www.inegi.org.mx (accessed August, 2009).

Kenney, J., Keeping, E., 1962. Mathematics of Statistics, third ed. Van Nostrand, Princeton, NJ.

Liu, K., Liang, H., Yeh, K., Chen, C., 2009. A qualitative decision support for environmental impact assessment using fuzzy logic. Journal of Environmental Informatics 13 (2), 93–103.

Mo-Yuen, C., 1997. Methodologies of Using Neural Network and Fuzzy Logic Technologies for Motor Incipient Fault Detection. World Scientific, Singapore.

NADF-009-AIRE, 2006. Environmental Standard for the Federal District (norma ambiental para el Distrito Federal). Gaceta Oficial del Distrito Federal (in Spanish), XVI epoch.

NOM-020-SSA1, 1993. Official Mexican Standard for the Environmental Health; Criteria for Ozone Assessment in Air Quality. Official Gazette for the Federal District.

NOM-021-SSA1, 1993. Official Mexican Standard for the Environmental Health; Criteria for Carbon Monoxide Assessment in Air Quality. Official Gazette for the Federal District.

NOM-022-SSA1, 1993. Official Mexican Standard for the Environmental Health; Criteria for Sulfur Dioxide Assessment in Air Quality. official gazette for the Federal District.

NOM-023-SSA1, 1993. Official Mexican Standard for the Environmental Health; Criteria for Nitrogen Dioxide Assessment in Air Quality. Official Gazette for the Federal District.

NOM-025-SSA1, 1993. Official Mexican Standard for the Environmental Health; Criteria for Particulate Matter Assessment in Air Quality. Official Gazette for the Federal District.

Ocampo, W., Ferré, N., Domingo, J., Schuhmacher, M., 2006. Assessing water quality in rivers with fuzzy inference systems: a case study. Environment International. Elsevier, 32, 733–742.

[PAHO] Pan American Health Organization. Available at: http://www.paho.org (accessed August 2009).

[RAMA] Automatic Atmospheric Monitoring Network. Available at: http://www.calidadaire.df.gob.mx (accessed August, 2009).

[REDMET] Manual Network for Atmospheric Monitoring. Available at: http://www.calidadaire.df.gob.mx (accessed August, 2009).

Ross, T., 2004. Fuzzy Logic with Engineering Applications. John Wiley & Sons.

Salazar, E., 2007. Development and comparative analysis of tropospheric ozone prediction models using linear and artificial intelligence – based models in Mexicali, Baja California (Mexico) and Calexico, California (US). Environmental Modeling and Software 23 (8), 1056–1069.

Shumway, R., Stoffer, D., 2000. Time Series Analysis and Its Applications. Springer-Verlag, New York.
[SIMAT] Atmospheric Monitoring System (Sistema de Monitoreo Ambiental, in Spanish). Available at: http://www.calidadaire.df.gob.mx (accessed August 2009).
[SMA] Mexican Ministry of Environment (Secretaría del Medio Ambiente, in Spanish). Available at: http://www.sma.df.gob.mx (accessed August, 2009).
Soler, V., 2007. Lógica difusa aplicada a conjuntos imbalanceados: aplicación a la detección del síndrome de Down. Ph. D. thesis. Departament de Microelectrònica i Sistemes Electrònics, Universitat Autònoma de Barcelona.
Sowlat, M., Gharibi, H., Yunesian, M., Mahmoudi, T., Lotfi, S., 2011. A novel, fuzzy-based air quality index (FAQI) for air quality assessment. Atmospheric Environment 45, 2050–2059.
Upadhyaya, G., Dashore, N., 2010. Monitoring of air polution by using fuzzy logic. International Journal on Computer Science and Engineering 02 (07), 2282–2286.
[USEPA] United States Environmental Protection Agency, 2000. Air Quality Criteria for Carbon Monoxide.
[USEPA] United States Environmental Protection Agency, 2009. Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI).
Wang, W., Men, C., Lu, W., 2008. Online prediction model based on support vector machine. Neurocomputing 71 (4 – 6), 550–558.
Wang, L., Jang, C., Zhang, Y., et al., 2011. Assessment of air quality benefits from national air pollution control policies in China. Part II: evaluation of air quality predictions and air quality benefits assessment. Atmospheric Environment 44, 3449–3457.
[WHO], World Health Organization, 1987. Sulfur Dioxide and Particulate Matter. Air Quality Guidelines for Europe. WHO regional publications, European series, 23, pp. 338–360.
[WHO], World Health Organization, 2005. WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide; Summary of Risk Assessment.
Yañez, C., López, I., de la Luz, G., 2008. Analysis and Prediction of Air Quality Data with the Sigma Classifier, Lecture Notes in Computer Science. Springer Verlag, pp. 651–658.
Yong, L., Huaicheng, G., Guozhu, M., Pingjian, Y., 2008. A Bayesian hierarchical model for urban air quality prediction under uncertainty. Atmospheric Environment 42, 8464–8469.
Zadeh, L., 1965. Fuzzy sets. Information Control 8, 338–353.
Zadeh, L., 1978. Fuzzy sets as a basis for theory of possibility. Fuzzy Sets and Systems 1, 3–28.