

Sensing and Forecasting of Pollution Data in Mexico City



Elías-J. Ventura-Molina, Raúl Jiménez-Cruz, Adolfo Rangel-Díaz-de-la-Vega

Abstract: In this paper we present the characteristics of sensors used to monitor the pollution levels in Mexico City, namely sulfur dioxide (SO₂), nitrogen oxides (NO_x), ozone (O₃), and carbon monoxide (CO). A novel algorithm to predict contamination levels is presented: the Gamma classifier. Also, a new coding technique is introduced, allowing the conversion from a series of values taken from SIMAT databases into a set of patterns, which in turn are useful for the task of pollutant forecasting. Experimental results show a competitive performance by the Gamma classifier as a predictor, when compared to other methods.

Keywords : Associative Memory, Pattern Classifiers, Pollutant Forecasting; Pollutant Sensing.

I. INTRODUCTION

Environmental problems are common in Latin American countries [1,2]. In particular, atmospheric pollution in Mexico City is measured using specific sensors for each pollutant of interest by analyzing the concentration of said pollutants in small air samples. This data is gathered and managed by the Mexico City Atmospheric Monitoring System [3].

It is possible to find in the state of the art several algorithms to predict air pollutants, in order to give better and more comprehensive information to policy makers and the population: linear regression [4], neural networks [5-8], support vector machines [9], and pollution models such as Caline and IITLS [10]. In the current work, atmospheric pollution data from Mexico City were used, employing a method of recent proposal: the Gamma classifier [11-13]. This method exhibits competitive performance, as shown in the experimental results. The rest of the paper is organized as follows. In Section 2 the SIMAT (Mexico City Atmospheric Monitoring System) is presented and described. Then, section 3 presents the sensors used to sample and measure each pollutant of interest, according to the requirements indicated in section 2. Then, Sections 4 and 5 contain the description of the Gamma classifier and the experimental results, respectively. Finally, some conclusions and lines of future work are drawn in section 6.

II. SIMAT

SIMAT is dedicated to monitoring and predicting pollution conditions in Mexico City, as well as preventing the population and the authorities about health risks [3]. SIMAT consist of four subsystems:

- RAMA: it is responsible for measuring hourly levels of: ozone (O₃), sulfur dioxide (SO₂), nitrous oxides (NO_x), carbon monoxide (CO), PM10 and PM2.5)
- REDMA (Manual Atmospheric Monitoring Network, Red Manual de Monitoreo Atmosférico in Spanish) monitors particulate matter suspended in the air (PM10, PM2.5 and total suspended particulate matter PST), as well as their concentration and composition; each measurement is taken manually every six days.
- REDMET monitors meteorological parameters —such as wind direction and speed— and solar radiation, in order to elaborate meteorological forecasting and dispersion models; it also records and monitors the UV index.
- REDDA measures both dry and wet deposit, whose analysis allows the study of rain properties.

The air quality in Mexico City is measured with IMECA points. If they exceed 100, the health of the population is put at risk, as specified in table I (NADF-009-AIRE-2006 [14]).

Table- I: IMECA and its implications for health

IMECA	Condition ^a	Effects on Health
0-50: green	Good	Suitable for conducting outdoor activities
51-100: yellow	Regular	Possible discomfort in children, the elderly and people with illnesses
101-150: orange	Bad	Cause of adverse health effects on the population.
151-200: red	Very Bad	Cause of greater adverse health effects on the population.
>200: purple	Extremely Bad	Cause of adverse health effects in the general population. Serious effects may present in children and older adults with cardiovascular and/or respiratory illnesses such as asthma

With the equations specified in this norm, the IMECA value and IMECA level (condition) can be computed, in parts per million (ppm). Thus, in order to compute the IMECA value, the starting point is the atmospheric concentration of each of the criteria pollutants. The main source of information for this task is the RAMA subsystem. For this study, historical and current data about the concentration of specific pollutants was used.

III. POLLUTANTS SENSING

This section describes the sensors to measure these contaminants: CO, O₃, SO₂, and nitrogen oxides (NO, NO₂, and NO_x). In each case, the physical phenomenon of principle used by

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Elías-J. Ventura-Molina*, CIC, Instituto Politécnico Nacional, CDMX, México.

Raúl Jiménez-Cruz, CIC, Instituto Politécnico Nacional, CDMX, México.

Adolfo Rangel-Díaz-de-la-Vega, CIC, Instituto Politécnico Nacional, CDMX, México

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

the sensors is presented, as well as the specifications of operations; also, a brief explanation of the harmful effects of exposure to a high concentration and what constitutes a high concentration for each pollutant is included.

A. Carbon monoxide CO

The measurement of carbon monoxide is taken by a Filter 300E [15]. The model 300E is based on the absorption of infrared radiation (IR) by CO molecules at a wavelength of 4.7 micrometers (μm), computing the concentration through the use of the Beer-Lambert Law:

$$I = I_0 e^{-\alpha l C} \quad (1)$$

where

- I is the intensity of light after being absorbed by CO.
- I_0 is the intensity of light before absorption.
- α is the coefficient of absorption of CO.
- l represents the distance that the light follows while absorbed by the CO.
- C represents the CO concentration.

Fig. 1 shows how model 300E processes a sample in order to measure how much CO it contains. Part A indicates a source of IR, whose light will be made to pass through a rotating disc containing two gases: one compartment is filled with nitrogen (N_2), while the other is filled with CO (see parts B and C). While the disc is rotating, the light will go alternately through the N_2 and the CO cells of the disc.

When the light passes through the N_2 cell (measurement), no IR in the frequency of interest is absorbed; while by passing through the CO cell (reference), some light is absorbed. Thus, an alternating signal is generated, carrying two different intensities of light for the frequency of interest: one for sampling purposes and one for reference purposes. The difference between these intensities once the light has been presented to the gas sample will represent the absorption of

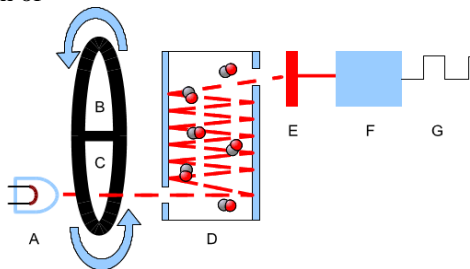


Fig. 1. Diagram of the CO measurement unit (model 300E)

light by said sample. This is accomplished in the absorption cell (D). Later, the light is passed through a filter (E), which only allows passage of the light in the frequency of interest: that having a wavelength around 4.7 μm . Finally, the light is presented to a photodetector (F), which generates the signal (G) with the two levels of light intensity mentioned above.

The 300E model thus exhibits a range of measurement from 0-1 ppm (parts per million) to 0-1000 ppm.

The Mexican Official Norm NOM-021-SSA1-1993 [16] states the limit of exposure for CO at an average of 11.0 ppm during 8 hours, for once a year at most. According to [3], exposure to concentrations of CO of 100ppm for more than one hour can result in headaches, while concentrations of

1600ppm during two hours may cause death

B. Ozone O_3

Meanwhile, ozone is measured with a UV Absorption O_3 Analyzer model 400E, also by Teledyne Instruments [17]. This sensor takes advantage of a particular property of ozone in order to determine its concentration in an air sample: absorption of ultraviolet radiation (UV) with a wavelength centered around 254 nanometers (nm). The concentration is calculated using the Beer-Lambert Law (see equation 1), where:

- I is the intensity of light after being absorbed by O_3 .
- I_0 is the intensity of light before absorption.
- α is the coefficient of absorption of O_3 .
- l represents the distance that the light follows while absorbed by the O_3 .
- C represents the O_3 concentration.

Fig. 2 depicts how model 400E works. First, a mercury lamp (part A) is used to generate a UV radiation beam, which is then passed through the absorption cell (part C). Said cell has two connections to it, one for the influx of the reference and measure samples (part B), and other for the exit of the samples (part D). Finally, a photodetector is used to generate an electrical signal which represents the concentration of ozone in the sample (parts E and F, respectively).

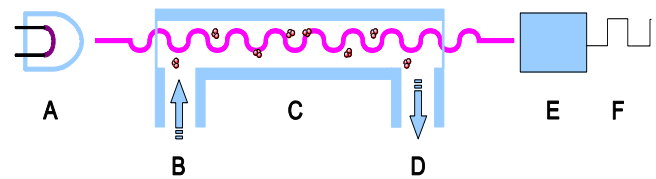


Fig. 2. Diagram of the O_3 measurement unit (model 400E)

In this manner, the 400E model has a measurement range from 100 ppb to 10 ppm, with a resolution of 0.5 ppb. Also, it can present results in several units: ppm, ppb, $\mu\text{g}/\text{m}^3$, and mg/m^3 .

The Mexican Official Norm NOM-020-SSA1-1993 [18] establishes the limit of exposure to O_3 at an average of 0.110 ppm during one hour, which should not be exceeded (not even once a year). Given that ozone is such an aggressive oxidizer, its effect on health are mainly on the respiratory system: difficulty to breath, degeneration of lung cells, depression of the immune system, and increased susceptibility to allergenic agents and diseases [3].

C. Sulfur dioxide SO_2

For the measurement of sulfur dioxide, a different phenomenon is taken advantage of. In this case, the Teledyne Instruments UV Fluorescence SO_2 Analyzer model 100E, based on the UV fluorescence principle, is used [19]. Here, the sample is irradiated with UV light of 214 nm in wavelength, exciting the SO_2 molecules. The amount of excitation generated in the sample depends on the average intensity of the applied UV light, according equation 2:

$$I_a = I_0 (1 - e^{-\alpha l C}) \quad (2)$$

where

- I is the average intensity of light.
- I_0 is average intensity of UV light applied.
- α is the coefficient of absorption of SO_2 at 214nm.
- l represents the distance that the light follows while absorbed by the SO_2 .
- C represents the SO_2 concentration.

However, SO_2 will not stay excited very long: it will go back to its more stable state of lower energy, emitting a UV photon of 330 nm in wavelength. The amount of light produced by this decaying of SO_2 depends on the speed of the reaction, as expressed by:

$$F = k(SO_2^*) \quad (3)$$

where

- F is the amount of light emitted.
- k is the speed of SO_2^* decaying into SO_2 .
- SO_2^* is the concentration of excited SO_2 .

The speed of the latter reaction is dependant on the temperature of the gas: at higher temperatures, decay of SO_2^* into SO_2 happens at a faster rate. Thus, when the light trajectory inside the reaction cell is short, the applied UV light intensity and gas temperature are both known, and there is no interference, the amount of light emitted by fluorescence is directly proportional to the SO_2 concentration inside the cell.

In order to avoid interference by other fluorescent compounds (such as nitrogen oxide) or by the applied UV radiation, several measures are taken. These measures include the following: the sample is filtered, removing some of the interfering compounds; also, the photodetector used to sense the amount of light emitted by fluorescence is located at a right angle from the photomultiplier tube used to feed the reaction; on the other hand, a filter is used to isolate the 214 nm UV light; and finally, the insides of the reaction cell are covered with a layer of Teflon in order to absorb stray light.

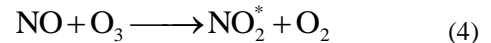
The 100E model has a range of measurement of 50 ppb to 20000 ppb, with a precision of 0.5% of the reading above 50 ppb. The measurement can be expressed in either ppb, ppm, $\mu g/m^3$, or mg/m^3 .

The limits of exposure before SO_2 are set at an average 0.13 ppm over 24 hours, once a year, and an annual arithmetic mean of 0.03 ppm, as published in the norm NOM-022-SSA1-1993 [20]. Among the effects that sulfur dioxide has on health are: respiratory diseases such as bronchoconstriction, bronchitis, tracheitis, and even death [3].

D. Nitrogen oxides NO_x

The sensor used to measure the nitrogen oxides concentration, which is the Teledyne Instruments Chemiluminescence $NO/NO_2/NO_x$ Analyzer model 200E [21], bases its operation on the principle of chemiluminescence. First, the sample is mixed with O_3 ,

which causes a reaction between nitrogen monoxide and the ozone, obtaining NO_2 and oxygen, as expressed below.



This reaction charges the NO_2 , which then frees the extra energy by emitting a photon with a wavelength between 600 nm and 3000 nm, with a maximum centered on 1200 nm. Given that the amount of light emitted is related to the amount on NO_2 present at the time of the reaction, it is possible to derive the concentration from the light intensity.

Notice, however, that this procedure only measures the concentration of NO . To measure NO_2 , it is reduced to NO . This is done by passing the sample through another cell where molybdenum at 315° celsius is used as a catalyst.



Now, the sample is again subjected to the chemiluminescence reaction. Thus, two readings are obtained from the light intensity: the concentration of NO , and that of NO_x , whose difference gives the NO_2 concentration. In order to avoid a noisy measure due to the presence of other gasses which may present a similar reaction to O_3 , an optical filter is used in order to suppress light with different wavelengths.

The model 200E specifications of operation include the following: ranges of measurement from 0-50 ppb to 0-20000 ppb, with independent ranges for NO , NO_2 , and NO_x ; lower detectable unit of 0.4 ppb; precision of 0.5% of reading; and the result presented in ppb, ppm, $\mu g/m^3$, and mg/m^3 .

The effects of nitrogen oxides vary, from the almost inert NO to the quite aggressive NO_2 ; however, they are known to cause respiratory affections similar to those caused by ozone, such as increased susceptible to respiratory infections, lung irritation and potentially lung damage similar to emphysema. The limit of exposure to nitrogen dioxide set by the norm NOM-023-SSA1-1993 [22] is an average of 0.21 ppm maximum one hour a year [3].

The preceding subsections present the different sensors used by the RAMA subsystem of SIMAT to automatically sample and measure each hour the concentration of several criteria pollutants. With these data, the IMECA is computed and periodically communicated to the population and the authorities. Although the criteria pollutants are given this quality in most countries, the sampling, measurement and monitoring of pollutants is done differently through the world. For instance, some countries use diverse units of measurement for the same pollutants: some use ppm while others use ppb, and yet others use mg/m^3 , for the same pollutant (i.e. ozone). In similar fashion, several methods have been applied to different databases around the planet, in order to predict future values for the concentration of a particular pollutant, once a set of past and current values is known. Some of these methods are taken from statistics, such as linear regression [4], while several methods from pattern classification and machine learning have been applied to this task, such as artificial neural networks [5-8] and support vector machines [9].

However, let us not forget the methods based on pollution modeling, such as Caline and IITLS [10]. These methods have been applied to data taken from diverse databases, whose records have been sampled and measured in a similar fashion to those of SIMAT, although not in quite the same manner. In the present paper, the Gamma classifier was applied to historical and current data regarding the hourly concentration of four pollutants —taken from the RAMA database— with the purpose of predicting the value of the next sample. However, in order to present the series of values to the classifier in an appropriate manner, the data was coded into patterns by using a novel technique. The Gamma classifier is presented in the following section, while the way in which the values delivered by the sensors were converted into patterns is further discussed in section 5.

IV. THE GAMMA CLASSIFIER

The Gamma classifier [11-13] is based on several concepts, namely: the gamma operator, alpha, beta, and u_β operators, and the Johnson-Möbius code.

Table- II: Alpha and Beta operators

$\alpha : A \times A \rightarrow B$			$\beta : B \times A \rightarrow A$		
x	y	$\alpha(x, y)$	x	y	$\beta(x, y)$
0	0	1	0	0	0
0	1	0	0	1	0
1	0	2	1	0	0
1	1	1	1	1	1
			2	0	1
			2	1	1

A. Preliminaries

Table II shows the Alpha and Beta operators, which were introduced in [23, 24]. They are the basis for the development of Alpha-Beta memories, the Gamma classifier, passing through models such as the Alpha-Beta-BAM [24], the Alpha-Beta support vector machines [25], and many other models [26-35].

The input to the operator u_β is a binary vector \mathbf{x} of dimension n , and at the output it throws a positive or zero integer:

$$u_\beta = \sum_{i=1}^n \beta(x_i, x_i) \quad (6)$$

On the other hand, the Johnson-Möbius code converts a collection of real numbers into binaries:

1. The minimum number is subtracted from each real number in the collection.
2. Perform a scaling so that only positive or zero integers remain.
3. Concatenate e_j ones to $e_m - e_j$ zeros. Here e_m is the greatest positive integer number (or zero) to be coded, and e_j is the current non-negative integer.

The generalized gamma operator γ_g . Two binary patterns form the input: \mathbf{x} and \mathbf{y} of dimensions n and $m - n$, m positive integers — in addition to a positive or zero integer value θ . The output is:

$$\gamma_g(\mathbf{x}, \mathbf{y}, \theta) = \begin{cases} 1 & \text{if } m - u_\beta[\alpha(\mathbf{x}, \mathbf{y}) \bmod 2] \leq \theta \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

A more extensive and detailed discussion on the former components of the gamma classifier can be found in [11].

B. The Gamma classifier algorithm

Consider p patterns \mathbf{x} with n real components. Let \mathbf{y} be a pattern with real n components that will be classified. There are m mutually exclusive classes with k_i patterns each. Thus,

it is fulfilled that: $\sum_{i=1}^m k_i = p$.

1. Apply the Johnson-Möbius code, to obtain a value e_m for each number (component) in the patterns:

$$e_m = \vee_{i=1}^p x_j^i \quad (8)$$

where \vee represents the usual maximum operator.

2. The stop parameter is calculated like this:

$$\rho = \bigwedge_{j=1}^n e_m(j) \quad (9)$$

3. Apply Johnson-Möbius code to \mathbf{y} .
4. Convert the indices of the patterns into two: that of the case and the place it occupies within that class..
5. Reset θ to 0.
6. Apply $\gamma_g(\mathbf{x}_j^{i\omega}, \mathbf{y}_j, \theta)$ (eq. 7).
7. Compute a weighted sum c_i for each class, according to this equation:

$$c_i = \frac{\sum_{\omega=1}^{k_i} \sum_{j=1}^n \gamma_g(\mathbf{x}_j^{i\omega}, \mathbf{y}_j, \theta)}{k_i} \quad (10)$$

8. If several maximums appear, add 1 to θ and repeat steps 6 and 7 until the maximum is unique, or else the following condition is true: $\theta \geq \rho$.
9. If the maximum is only one, do this:

$$C_y = C_j \text{ such that } \bigvee_{i=1}^m c_i = c_j \quad (11)$$

10. If something else happens, assign to \mathbf{y} the class of the first maximum that appears.

V. POLLUTANTS FORECASTING

In this section, the experiments done and the results obtained are presented and discussed.



A. Pattern coding

A positive n number is set and patterns are formed with n consecutive values of the contaminant under study. The pollutant value $n+1$ -th will be the class of the current pattern. Here we have chosen the value of $n = 10$.

B. Experiments

The types of experiments that are carried out in this work are shown in Table III. The experimental design is exemplified by experiment 1; data from the *Instituto Mexicano del Petróleo* (IMP) station feed experiment 1. Learning data: the whole 2000 year (8750 patterns); test (734): March 2001.

In this work, data from 4 stations were used: *Instituto Mexicano del Petróleo* (IMP) for experiments 1, 2, and 3; *Cerro de la Estrella* (CES) for experiments 4 and 5; *Tultitlán* (TLI) for experiments 6, 7, 10, 11, and 12; finally, data for the experiments 8 and 9 were taken from *Tacuba* (TAC) station.

Table- III: Experimental design

Exp.	Pollutant	Station	Set Period	
			Fund.	Test
1	CO	IMP	2000	March 2001
2	CO	IMP	2006	February 2007
3	CO	IMP	2006	May 2007
4	O ₃	CES	2006	February 2007
5	O ₃	CES	2006	May 2007
6	SO ₂	TLI	2006	February 2007
7	SO ₂	TLI	2006	May 2007
8	SO ₂	TAC	2001	February 2002
9	NO _x	TAC	2002	June 2003
10	NO	TLI	2009	May 2010
11	NO ₂	TLI	2009	May 2010
12	NO _x	TLI	2009	May 2010

C. Performance measure

In equations 12 and 13, P_i represents the i -th predicted (future or unknown) level, while O_i is the i -th original (real, actual, or known) datum.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (12)$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (13)$$

In general, small values are preferred for RMSE and bias.

VI. EXPERIMENTAL RESULTS

Table IV shows the experimental results of applying the Gamma classifier in the contamination data in the 12 designed experiments.

Table- IV: Experimental results (ppm)

Exp.	Pollutant	Bias	RMSE
------	-----------	------	------

1	CO	-0.028513	0.779154
2	CO	0.012042	0.726013
3	CO	0.062183	0.611769
4	O ₃	0.000918	0.012302
5	O ₃	0.000417	0.014443
6	SO ₂	0.000676	0.012096
7	SO ₂	0.000795	0.010487
8	SO ₂	0.000408	0.009218
9	NO _x	0.001543	0.026103
10	NO	0.000001	0.000037
11	NO ₂	0	0
12	NO _x	0	0

Note that the RMSE yields very similar values with the experiments of the same contaminant. With CO, for instance, the mean values are 0.705645 (exps. 1, 2, and 3) vs. 0.010600 for SO₂ (exps. 7 and 8).

Comparison with other methods

There are very few experimental pollutant prediction works that use SIMAT data. Table V shows the comparative results and there clearly shows the superiority of the Gamma classifier. The error values shown by the range in the experiments 4 and 5 are tiny compared to the errors reported in the state of the art, when the contaminant is O₃.

Table- V: Comparison of results (SIMAT O₃)

Algorithm	Performance (Abs. Mean Error)
Bayesian network [5]	0.221000
C4.5 [5]	0.176400
Neural network [5]	0.160000
Gamma exp. 4	0.000918
Gamma exp. 5	0.000417

In the specialized literature, research papers on contaminant prediction can be found. But the comparison of the experimental results cannot be straightforward, since the data comes from different sources and from different contaminants.

Table VI shows some indirect comparisons with works of the state of the art.

Table- VI: Comparison of results

Algorithm	Pollutant	Performance (RMSE)
IITLS [10]	NO _x (µg/m ³)	19.99
Neural network [8]	O ₃ (µg/m ³)	15
Neural network [7]	O ₃ (ppb)	13.79, 9.43
Online SVM [9]	SO ₂ (µg/m ³)	12.96, 10.90
Neural network [6]	CO, NO ₂ , SO ₂ , and O ₃ (ERPI)	5.852, 1.365
Gamma exp. 3 (current work & [13])	CO (ppm)	0.611769

Multivariate linear regression [4]	BaP (ng/m ³)	0.449
Gamma experiments 4,8,10,11,12 in ppm		
Gamma exp. 4	O ₃	0.012302
Gamma exp. 8	SO ₂	0.009218
Gamma exp. 10	NO	0.000037
Gamma exp. 11	NO ₂	0.000000
Gamma exp. 12	NO _x	0.000000

It is pertinent to note the great variability of the pollutants considered in the state of the art experiments.

The authors work with pollutants as disoluble as CO, NO₂, SO₂, and O₃, even a pollutant not very usual as it is benzo(a)pyrene (BaP) [4].

The problem of the comparison of results is complicated when the great variability of mediated units used by the different authors is taken into account: ppm, ppb, µg/m³, ng/m³, and ERPI (European Regional Pollution Index) [6].

To top it off, the methods used are also very different. However, a look at the values can give you an idea that the Gamma results are not bad, given the small values that throws the Gamma classifier given in Table VI.

VII. CONCLUSION

Atmospheric pollution in Mexico City is a grave problem at many different levels, both for Latin America and even the whole world, since such pollution favors global warming, with the consequences this entails. Different methods have been used to analyze air quality data and predict concentration levels for several pollutants which may endanger human health and lives, deteriorating the environment. In the present work, the data generated by several air quality sensors has been combined with a recently proposed pattern classifier, the Gamma classifier, and an emerging coding technique in order to forecast pollutant levels in Mexico City.

As a work in the future, the authors of this research suggest applying the results presented here, in the approaches contained in the article [36].

It is also suggested to apply this methodology in other pollutants or other combinations of several of them.

REFERENCES

1. D. Moreira and M. Vilhena, *Air Pollution and Turbulence: Modeling and Applications*. Boca Raton, FL, USA: CRC Press, 2010; pp. ix-x.
2. O. Bekir and G. Surhid, *Vehicular Air Pollution: Experiences From Seven Latin American Urban Centers*. Washington, D.C, USA: World Bank: 1997; pp. xiii-xiv.
3. Sistema de Monitoreo Atmosférico de la Ciudad de México. SIMAT (in Spanish), 2010. Available at <http://www.sma.df.gob.mx/simat2> (accessed September 12, 2010).
4. M.S. Callén, J.M. López, and A.M. Mastral, "Seasonal variation of benzo(a)pyrene in the Spanish airborne PM10. Multivariate linear regression model applied to estimate BaP concentrations," *Journal of Hazardous Materials*, vol. 180, 2010, pp. 1-3.
5. L.E. Sucar, J. Pérez-Brito, J.C. Ruiz-Suárez, and E. Morales, "Learning Structure from Data and Its Application to Ozone Prediction," *Applied Intelligence*, vol. 7(4), 1997, pp. 327-338.
6. K.P. Moustris, I.C. Ziomas, and A.G. Paliatsos, "3-Day-Ahead Forecasting of Regional Pollution Index for the Pollutants NO₂, CO,

- SO₂, and O₃ Using Artificial Neural Networks in Athens, Greece," *Water, Air, & Soil Pollution*, vol. 2019(1), 2010, pp. 29-43.
7. E. Salazar-Ruiz, et al., "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 Modelling and Software*, vol. 23(8), 2008, pp. 1056-1069.
8. A. Dutot, J. Rynkiewicz, F.E. Steiner, and J. Rude, "A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions," *Environmental Modelling and Software*, vol.22(9), 2007, pp. 1261-1269.
9. W. Wang, C. Men, and W. Lu, "Online prediction model based on support vector machine," *Neurocomputing*, vol. 71, 2008, pp. 4-6.
10. P. Goyal, N. Jaiswal, A. Kumar, J.K. Dadoo, and M. Dwarakanath, "Air quality impact assessment of NO_x and PM due to diesel vehicles in Delhi," *Transportation Research Part D: Transport and Environment*, vol. 15(5), 2010, pp. 298-303
11. Y. Villuendas-Rey, C. Yáñez-Márquez, J.A. Antón-Vargas, and I. López-Yáñez, "An extension of the Gamma associative classifier for dealing with hybrid data," *IEEE Access*, vol. 7(1), 2019, pp. 64198-64205.
12. A.V. Uriarte-Arcia, C. Yáñez-Márquez, J. Gama, I. López-Yáñez, and O. Camacho-Nieto, "Data Stream Classification Based on The Gamma Classifier," *Mathematical Problems in Engineering*, vol. 2015, 2015, Article 939175, 17 pages.
13. I. López-Yáñez, C. Yáñez -Márquez, C., and V.M. Silva-García, "Forecasting Air Quality Data with the Gamma Classifier," in *Pattern Recognition*, Peng-Yeng Yin, Ed. Croatia: INTECH, 2009; pp. 499-512.
14. Gobierno del Distrito Federal. Norma Ambiental para el Distrito Federal (in Spanish). Gaceta Oficial del Distrito Federal 2006, XVI Epoch.
15. Teledyne Advanced Pollution Instrumentation. Model 300E Family Carbon Monoxide Analyzers Technical Manual, Revision C. San Diego, USA, January 2009.
16. F. Pérez-Duarte, "Criterio para evaluar la calidad del aire ambiente con respecto al monóxido de carbono (CO)," Norma Oficial Mexicana NOM-021-SSA1-1993, Secretaría de Salud: México, D. F., México, 1994.
17. Teledyne Advanced Pollution Instrumentation. Model 400E Photometric Ozone Analyzer Technical Manual, Revision E. San Diego, USA, June 2009.
18. E. Enriquez-Rubio, "Criterios para evaluar la calidad del aire ambiente con respecto al ozono (O₃)," Modificación a la Norma Oficial Mexicana NOM-021-SSA1-1993, Secretaría de Salud: México, D. F., México, 2002.
19. Teledyne Advanced Pollution Instrumentation. Model 100E UV Fluorescence SO₂ Analyzer Instructions Manual. Revision B19, San Diego, USA, March 2009.
20. F. Pérez-Duarte, "Criterio para evaluar la calidad del aire ambiente con respecto al bióxido de azufre (SO₂)," Norma Oficial Mexicana NOM-022-SSA1-1993, Secretaría de Salud: México, D. F., México, 1994.
21. Teledyne Advanced Pollution Instrumentation. Model 200E Nitrogen Oxides Analyzer Technical Manual. Revision C2, San Diego, USA, July 2009.
22. F. Pérez-Duarte, "Criterio para evaluar la calidad del aire ambiente con respecto al bióxido de nitrógeno (NO₂)," Norma Oficial Mexicana NOM-023-SSA1-1993, Secretaría de Salud: México, D. F., México, 1994.
23. C. Yáñez-Márquez, I. López-Yáñez, M. Aldape-Pérez, O. Camacho-Nieto, A.J. Argüelles-Cruz and Y. Villuendas-Rey, "Theoretical Foundations for the Alpha-Beta Associative Memories: 10 Years of Derived Extensions, Models, and Applications," *Neural Processing Letters*, vol. 48, Issue 2, 2018, pp. 811-847.
24. M.E. Acevedo-Mosqueda, C. Yáñez -Márquez, and I. López-Yáñez, "Alpha-Beta Bidirectional Associative Memories: Theory and Applications," *Neural Processing Letters*, vol. 26(1), 2007, pp. 1-40.
25. L.O. López-Leyva, C. Yáñez-Márquez, R. Flores-Carapia, and O. Camacho-Nieto, "Handwritten Digit Classification Based on Alpha-Beta Associative Model," *Lecture Notes in Computer Science*, vol. LNCS 5197, 2008, pp. 437-444.
26. C. Lopez-Martin, "Applying a general regression neural network for predicting development effort of short-scale programs," *Neural Computing and Applications*, vol 20, 3, 2011, pp. 389-401.
27. A. Ferreira-Santiago, C. Lopez-Martin, C. Yáñez-Márquez, "Metaheuristic optimization of multivariate adaptive regression splines for predicting the schedule of software projects," *Neural Computing and Applications*, vol. 27, no 8, 2016, pp. 2229-2240.

28. I. López-Yáñez, C. Yáñez-Márquez, O. Camacho-Nieto, M. Aldape-Pérez, and A.J. Argüelles-Cruz, "Collaborative learning in postgraduate level courses," *Computers in Human Behavior*, vol. 51, Part B, 2015, pp. 938-944.
29. I. Kalichanins-Balich, C. Lopez-Martin, "Applying a feedforward neural network for predicting software development effort of short-scale projects," in *2010 Eighth ACIS International Conference on Software Engineering Research, Management and Applications*. IEEE, 2010. pp. 269-275.
30. M. Aldape-Pérez, C. Yáñez-Márquez, O. Camacho-Nieto, I. López-Yáñez, and I. & A.J. Argüelles-Cruz, "Collaborative learning based on associative models: Application to pattern classification in medical datasets," *Computers in Human Behavior*, vol. 51, Part B, 2015, pp. 771-779.
31. I. López-Yáñez, L. Sheremetov, and C. Yáñez-Márquez, "A Novel Associative Model for Time Series Data Mining," *Pattern Recognition Letters*, vol. 41, 2014, pp. 23-33.
32. A. García-Florian, C. López-Martín, C. Yáñez-Márquez, and A. Abran, "Support Vector Regression for Predicting Software Enhancement Effort," *Information and Software Technology*, vol. 97, 2018, pp. 90-109.
33. Y.O. Serrano-Silva, Y. Villuendas-Rey, and C. Yáñez-Márquez, "Automatic feature weighting for improving financial Decision Support Systems," *Decision Support Systems*, vol. 107, 2018, pp. 78-87.
34. A. Chavoya, C. Lopez-Martin, M.E. Meda-Campa, "Applying genetic programming for estimating software development effort of short-scale projects," in *2011 Eighth International Conference on Information Technology: New Generations*. IEEE, 2011. pp. 174-179.
35. Y. Villuendas-Rey, C.F. Rey-Benguría, A. Ferreira-Santiago, O. Camacho-Nieto, and C. Yáñez-Márquez, "The Naïve Associative Classifier (NAC): a novel, simple, transparent, and accurate classification model evaluated on financial data," *Neurocomputing*, vol. 265, 2017, pp. 105-115.
36. P. Quevauviller, O. Thomas, and A. van der Beken, *Wastewater Quality Monitoring and Treatment*, West Sussex, England: John Wiley & Sons, 2006.

AUTHORS PROFILE



Elías-J. Ventura-Molina. is last-semester computer science Ph.D. candidate at Instituto Politécnico Nacional. In 2007, he received a bachelor's degree in computer systems engineering at Technological University of Mexico and a master's degree in computer science from the Computing Research Center in 2014. In the gap between his Master and Ph.D. studies, he worked for over a year in the computational genomic department at the Mexican National Institute of Genomic Medicine. Additionally, he has realized a couple a research internship at the University of Windsor, Canada and University of California, Irvine in USA. His research and publication interests include Bioinformatics and Machine Learning with particular emphasis on: Deep Learning, Neural Networks, Support Vector Machines, Reinforcement Learning, Mobile Computing, and their Theoretical Foundations and Applications.



Raúl Jiménez-Cruz. is currently studying a PhD in computer science at the Center for Computer Research in Mexico. The professional career of Raul and his ability to handle different programming languages demonstrate his ability, interest and desire to help in the advancement of science. The mission of Raul is to provide knowledge in practice and theory, as well as to extend it in order to contribute in the scientific field and use it in practice.



Adolfo Rangel-Díaz-de-la-Vega. an experienced financial risk analyst with a 30 year career in the mexican mortgage market, obtained his Bachelor degree (1993) on Actuarial Science from the Science Faculty of the National Autonomous University of Mexico, his MSc degree (2015) in Quantitative Finance at Anahuac University and his MSc degree from the Center for Computing Research, where he is currently pursuing a PHD degree. His research interests include mobile computing, quantitative and computational finance, and financial machine learning.