

## **Predicción de la concentración de contaminantes atmosféricos basada en un clasificador asociativo de patrones**

### **Prediction of air contaminant concentration based on an associative pattern classifier**

*Itzamá López Yáñez\**, *Cornelio Yáñez Márquez*, *Oscar Camacho Nieto*,  
*Amadeo José Argüelles Cruz*

Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN), Av. Juan de Dios Bátiz s/n casi esq. Miguel Othón de Mendizábal, Unidad Profesional “Adolfo López Mateos”, Edificio CIC. Col. Nueva Industrial Vallejo. C. P. 07738. México, D.F. México

(Recibido el 4 de marzo de 2011. Aceptado el 21 de mayo de 2011)

#### **Resumen**

Desde hace poco más de tres lustros, el Reconocimiento de Patrones ha incidido en el tratamiento de datos concernientes a la protección del medio ambiente; en especial, diversos grupos de investigadores han utilizado algoritmos genéticos y redes neuronales artificiales en la predicción de datos relacionados con las ciencias atmosféricas y el medio ambiente. En este artículo se presentan los resultados de aplicar el clasificador Gamma en la predicción de valores futuros de concentración de contaminantes atmosféricos, obteniendo resultados competitivos (RMSE de 0,556382 ppm para monóxido de carbono).

----- *Palabras clave:* Predicción de contaminantes atmosféricos, clasificación de patrones, clasificador Gamma

#### **Abstract**

From a little more than 15 years to this day, several methods and techniques taken from the area of Pattern Recognition have been employed on the treatment of data concerning environmental protection. In particular, diverse research groups have applied genetic algorithms and artificial neural networks to the prediction of data related to atmospheric sciences and the environment. In this paper, the authors present the results of applying the Gamma classifier

---

\* Autor de correspondencia: teléfono: + 52 + 5 + 55 + 729 60 00 - 5 65 84, fax: + 52 + 5+ 55 + 729 60 00 - 5 66 07, correo electrónico: ilopez@ipn.mx (I. López)

to the prediction of future values for air contaminants concentration, obtaining competitive results (RMSE of 0.556382 ppm for carbon monoxide).

----- *Keywords:* Air pollution prediction, pattern classification, Gamma classifier

## Introducción

El medio ambiente y temas relacionados como el calentamiento global del planeta Tierra se han posicionado, de súbito, en el pensamiento y acción de un buen sector de la sociedad mundial. En diferentes idiomas y a través de diversos medios, las organizaciones civiles lanzan campañas a fin de que la humanidad tome conciencia de la importancia que debe tener la protección del medio ambiente [1, 2] y han tenido una penetración tal, que los mismos gobiernos participan activamente en acciones de cuidado y conservación del medio ambiente [3-6].

La relevancia del uso de técnicas de inteligencia artificial y reconocimiento de patrones en las tareas de análisis de datos relacionados con el medio ambiente se ha evidenciado desde hace casi dos décadas [7-9], siendo notable el uso de las redes neuronales artificiales [10-12] y de los algoritmos genéticos [13].

Sin embargo, en este tipo de aplicaciones ha sido notoria la ausencia, hasta hace unos cuantos años, de una de las técnicas de reconocimiento de patrones más eficientes: las memorias asociativas [14]. En este sentido, una memoria asociativa es un modelo matemático cuyo propósito básico es recuperar correctamente patrones completos, partiendo de un patrón de entrada, mismo que puede haber sido modificado por alguna alteración.

En el año 2002 se crearon, en el Centro de Investigación en Computación del Instituto Politécnico Nacional de México, los modelos asociativos Alfa-Beta [14] cuya eficacia y eficiencia se ha mostrado mediante diferentes aplicaciones en bases de datos reales de distintas áreas del saber humano [15-27]. Las operaciones básicas de estas memorias, los operadores Alfa y Beta, han sido la base conceptual para un

nuevo algoritmo de clasificación de patrones: el clasificador Gamma [25-27]. En este artículo se presentan los resultados obtenidos al aplicar este clasificador a la predicción de la concentración de contaminantes atmosféricos, ligados al cálculo del Índice Metropolitano de la Calidad del Aire de la Ciudad de México (IMECA) [28].

El resto del artículo está organizado como sigue: primero se describen de manera concisa el Sistema de Monitoreo Atmosférico de la Ciudad de México (SIMAT) y el IMECA, para luego presentar el clasificador Gamma. Más adelante se introduce la propuesta principal de este trabajo, y posteriormente se discuten los resultados experimentales y se incluyen las conclusiones y trabajo futuro, para finalizar con las referencias.

## Materiales: SIMAT e IMECA

El Sistema de Monitoreo Atmosférico de la Ciudad de México (SIMAT) es el organismo gubernamental encargado de vigilar y monitorear la calidad del aire en la Ciudad de México. La información que se presenta a continuación está tomada de [28].

El SIMAT está comprometido a operar y mantener un sistema confiable para el monitoreo de la calidad del aire en la Ciudad de México, así como el análisis y la difusión de la información para dar cumplimiento con los requerimientos y normatividad vigente; está integrado por cuatro subsistemas especializados, una Unidad Móvil de Monitoreo Atmosférico y un Laboratorio de Transferencia de Estándares de Calibración.

El subsistema de interés para el presente trabajo es la Red Automática de Monitoreo Atmosférico (RAMA), misma que realiza mediciones continuas y permanentes de ozono (O<sub>3</sub>), dióxido de azufre (SO<sub>2</sub>), óxidos de nitrógeno (NO<sub>x</sub>),

monóxido de carbono (CO), partículas menores a 10 micrómetros (PM<sub>10</sub>) y partículas menores a 2,5 micrómetros (PM<sub>2,5</sub>). Cada muestra se toma de manera automática cada hora.

Por su parte, el Índice Metropolitano de la Calidad del Aire (IMECA) es un valor de referencia para que la población conozca los niveles de

contaminación prevalecientes en su zona de residencia, de manera precisa y oportuna, a fin de que tome las medidas pertinentes de protección. Cuando el IMECA de cualquier contaminante rebasa los 100 puntos, significa que sus niveles son perjudiciales para salud y en la medida en que aumenta el valor del IMECA se agudizan los síntomas (tabla 1).

**Tabla 1** El IMECA y sus implicaciones para la salud

<i><b>IMECA</b></i>	<i><b>Condición</b></i>	<i><b>Efectos a la Salud</b></i>
0-50: verde	Buena	Adecuada para llevar a cabo actividades al aire libre
51-100: amarillo	Regular	Posibles molestias en niños, adultos mayores y personas con enfermedades (conjuntivitis, irritación de nariz o garganta, problemas respiratorios, entre otras)
101-150: naranja	Mala	Causante de efectos adversos a la salud en la población, en particular los niños y los adultos mayores con enfermedades cardiovasculares o respiratorias como asma
151-200: rojo	Muy Mala	Causante de mayores efectos adversos a la salud en la población en general, en particular los niños y los adultos mayores con enfermedades cardiovasculares o respiratorias como asma
>200: morado	Extremadamente Mala	Causante de efectos adversos a la salud de la población en general. Se pueden presentar complicaciones graves en los niños y los adultos mayores con enfermedades cardiovasculares o respiratorias como asma

La generación del IMECA es una de las tareas primordiales del SIMAT; desde el primero de julio de 1998 se transmite las 24 horas del día a los diferentes medios de comunicación electrónicos e impresos.

En noviembre del año 2006 se publicó en la Gaceta Oficial del Distrito Federal la Norma Ambiental para el Distrito Federal NADF-009-AIRE-2006 [29], que establece los requisitos para elaborar el Índice Metropolitano de la Calidad del Aire (IMECA) de los contaminantes criterio, como son: O<sub>3</sub>, NO<sub>2</sub>, SO<sub>2</sub>, CO, PM<sub>10</sub> y PM<sub>2.5</sub>. Para cada uno de los contaminantes criterio la norma proporciona ecuaciones que permiten el cálculo del IMECA correspondiente, a partir de los datos de las concentraciones en partes por millón (ppm).

Entonces, para calcular el valor del IMECA se parte de la concentración atmosférica de cada contaminante criterio. A su vez, para determinar el nivel del IMECA, basta con ubicarlo en los rangos correspondientes (tabla 1). La principal fuente de información para estas tareas es el subsistema RAMA, mismo que muestrea, mide y monitorea la concentración de varios contaminantes de interés dentro de la Zona Metropolitana del Valle de México. Esta medición se lleva a cabo por una red de 34 estaciones de monitoreo (actualmente activas), mismas que toman una muestra y registran la correspondiente medición cada hora. Cabe mencionar que, para fines de informar a la población en general, el valor publicado del IMECA es el valor correspondiente al IMECA más alto entre los contaminantes criterio. Así, si

el IMECA de O<sub>3</sub>, es 86, mientras que el de NO<sub>2</sub> es el valor máximo con 105, entonces el IMECA general en ese momento será 105.

**Métodos: Clasificador Gamma**

Aunque de reciente introducción, este clasificador ha mostrado resultados experimentales competitivos al aplicarse a diversas tareas de clasificación de patrones y, específicamente, para tratar datos ambientales [26, 27]. Por otro lado, este clasificador toma su nombre del operador Gamma de similitud, que sirve de base para su funcionamiento. A su vez, este operador está basado en los operadores alfa (α), beta (β) y u<sub>β</sub> así como sus propiedades, sobre todo cuando trabajan con vectores binarios codificados usando el código Johnson-Möbius modificado. Es por ello que a continuación se presentan dichos componentes del clasificador Gamma: los operadores alfa, beta y u<sub>β</sub>, así como el código Johnson-Möbius modificado y el operador Gamma de similitud [14-19, 24-27], para terminar con el algoritmo del clasificador [24-27].

**Conceptos preliminares**

Los operadores alfa y beta fueron introducidos en [14-18] y son el fundamento de las memorias asociativas Alfa-Beta. Dichos operadores se definen de manera tabular como se muestra en la tabla 2, considerando que los conjuntos A y B están definidos como A = {0, 1} y B = {0, 1, 2}.

**Tabla 2** Definición de los operadores alfa y beta

$\alpha : A \times A \rightarrow B$			$\beta : B \times A \rightarrow A$		
<b>x</b>	<b>y</b>	<b>α(x, y)</b>	<b>x</b>	<b>y</b>	<b>β(x, y)</b>
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

Estos operadores han sentado las bases para el desarrollo de varios modelos matemáticos utilizados para el Reconocimiento de Patrones, tales como las memorias asociativas Alfa-Beta originales [14], las memorias asociativas bidireccionales Alfa-Beta [15-18], las máquinas asociativas Alfa-Beta con soporte vectorial [24] o el clasificador Gamma [25-27].

Por su parte, el operador u<sub>β</sub> recibe como entrada un vector binario n-dimensional x, entrega a la salida un entero no negativo y es calculado como se muestra a continuación:

$$u_{\beta}(\mathbf{x}) = \sum_{i=1}^n \beta(x_i, x_i) \tag{1}$$

Por otro lado, el código Johnson-Möbius modificado —propuesto en el grupo de investigación de los autores, mismo que es una variante del conocido código Johnson-Möbius [30]— permite la conversión de un conjunto de números reales a su representación binaria siguiendo estos pasos:

1. Restar el mínimo (del conjunto de reales) de cada número, dejando un conjunto de números reales no negativos, incluso si alguno era negativo antes de este paso.
2. Escalar los números (truncando los decimales restantes si es necesario) al multiplicar todos los números obtenidos en el paso 1 por alguna potencia de 10 apropiada; esto da como resultado un conjunto de números enteros no negativos.
3. Concatenar e<sub>m</sub>-e<sub>j</sub> ceros con e<sub>j</sub> unos, donde e<sub>m</sub> es el número entero máximo a ser codificado y e<sub>j</sub> es el número entero actual a ser codificado.

Por último, el operador Gamma de similitud generalizado γ<sub>g</sub>, que toma a la entrada dos vectores binarios x ∈ A<sup>n</sup> y y ∈ A<sup>m</sup> —con n, m ∈ ℤ<sup>+</sup>, n ≤ m— así como un entero no negativo θ; y entrega a la salida un número binario; se calcula como sigue:

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

donde mod 2 indica el operador módulo 2 usual. Esto es, el operador Gamma de similitud generalizado  $\gamma_g$  entrega como salida el valor 1 si ambos vectores se diferencian a lo más en  $\theta$  bits, y 0 en otro caso.

### Algoritmo del clasificador Gamma

Sean  $k, m, n, p \in \mathbb{Z}^+$ ;  $\{\mathbf{x}^\mu \mid \mu = 1, 2, \dots, p\}$  el conjunto fundamental de patrones (de aprendizaje o entrenamiento) con cardinalidad  $p$ , donde  $\forall \mu \mathbf{x}^\mu \in \mathbb{R}^n$ , y sea  $\mathbf{y} \in \mathbb{R}^n$  un patrón  $n$ -dimensional con valores reales a ser clasificado. Se asume que el conjunto fundamental está particionado en  $m$  clases diferentes (mutuamente excluyentes) en donde cada clase tiene una cardinalidad  $k_i$ ,  $i = 1, 2, \dots, m$ , haciendo que  $\sum k_i = p$ . Para clasificar el patrón y se siguen estos pasos:

1. Se codifican todos los patrones del conjunto fundamental usando el código Johnson-Möbius modificado, calculándose el criterio de paro  $r$ , como se muestra a continuación:

$$e_m(j) = \bigvee_{i=1}^p x_j^i \quad (3)$$

$$\rho = \bigwedge_{j=1}^n e_m(j) = \bigwedge_{j=1}^n \left[ \bigvee_{i=1}^p (x_j^i) \right] \quad (4)$$

donde  $\vee$  y  $\wedge$  representan los operadores usuales máximo y mínimo, respectivamente. Así,  $e_m(j)$  es el valor máximo entre los  $j$ -ésimos componentes de todos los patrones fundamentales, y  $r$  es el valor mínimo entre esos  $e_m(j)$ .

2. Se codifica el patrón a clasificar usando el código Johnson-Möbius modificado con los mismos parámetros utilizados para los patrones fundamentales.
3. Transformar los índices de los patrones fundamentales en dos índices: uno para

indicar su clase y otro para diferenciar los patrones de la misma clase (i.e.  $\mathbf{x}^\mu$  que pertenece a la clase  $i$  se convierte en  $\mathbf{x}^{i\omega}$ ).

4. Se inicializa  $\theta$  a 0.
5. Se compara el patrón a clasificar con todos los patrones fundamentales, utilizando el operador Gamma:  $\gamma_g(\mathbf{x}_j^{i\omega}, \mathbf{y}_j, \theta)$ .
6. Se calcula una suma ponderada por clase de los resultados de la comparación del paso anterior, de acuerdo con la siguiente ecuación:

$$c_i = \frac{\sum_{\omega=1}^{k_i} \sum_{j=1}^n \gamma_g(\mathbf{x}_j^{i\omega}, \mathbf{y}_j, \theta)}{k_i} \quad (5)$$

7. Si se tiene un máximo único entre las sumas (hay una clase que le gana a todas las demás), asignar el patrón a esa clase.

$$C_y = C_j \text{ tal que } \bigvee_{i=1}^m c_i = c_j \quad (6)$$

8. En caso de empate, incrementar  $\theta$  y repetir los pasos 5 y 6, hasta que haya un desempate o se alcance el criterio de paro:  $\theta \geq \rho$ .
9. Si se alcanzó el criterio de paro, asignar el patrón a clasificar a cualquiera de las clases empatadas con una suma máxima; por ejemplo, la primera de ellas.

Los primeros 4 pasos de este algoritmo se pueden ver como la fase de aprendizaje del clasificador, mientras que los pasos 5 a 9 (incluyendo las correspondientes iteraciones) conforman la fase de clasificación.

Aunque el clasificador Gamma no tiene mucho tiempo de haber sido propuesto, ya ha sido aplicado a diversos problemas: clasificación de la conocida base de datos Iris Plant, localización de estaciones móviles, estimación del esfuerzo en el desarrollo de programas pequeños y predicción de datos ambientales [25-27]. En estos problemas, algunos muy diferentes entre sí —algunos incluso que violan las premisas básicas

del clasificador no se cumplen—, el clasificador Gamma ha mostrado un desempeño competitivo en los resultados experimentales.

### Aplicación propuesta

En este artículo se propone la aplicación del clasificador Gamma en las bases de datos medioambientales que se derivan de los subsistemas del SIMAT. Para ello, se formarán patrones de datos, de acuerdo con las instancias de cada base de datos y las unidades de medida, con objeto de atacar un problema importante: la predicción automática de la concentración de contaminantes utilizados para calcular el IMECA, para un contaminante criterio en una fecha y una hora dadas.

Se experimenta específicamente con la base de datos RAMA (Red Automática de Monitoreo Atmosférico). Esta base de datos incluye las concentraciones registradas para cada contaminante criterio, muestreadas cada hora, en cada estación de monitoreo.

Para formar los patrones, se toman  $n$  valores de muestreo consecutivos y se concatenan para formar un patrón, tomando como su clase el siguiente valor. Así, el  $i$ -ésimo patrón incluye los valores  $i, i + 1, i + 2, \dots, i + (n - 1)$  y su clase sería el valor  $i + n$ , mientras que el patrón  $i + 1$  estaría formado por los valores  $i + 1, i + 2, \dots, i + n$  y su clase sería el valor  $i + (n + 1)$ .

### Resultados experimentales

En la fase de experimentación se utilizaron los datos de RAMA, en particular los datos de CO. Para llevar a cabo los experimentos, se tomó el conjunto de mediciones de monóxido de carbono (CO) en ppm obtenidas en la estación Iztacalco, muestreadas cada hora durante el año 2010. Con estos datos, se generaron patrones de 10 datos, enseñando este patrón como entrada y el dato siguiente como salida al clasificador Gamma.

El conjunto fundamental, a partir del cual aprendió el clasificador, quedó entonces integrado por 8749 asociaciones, y como conjunto de prueba se tomaron los datos obtenidos por la misma estación de monitoreo durante el mes de febrero de 2011, contándose así con un conjunto de prueba integrado por 661 asociaciones.

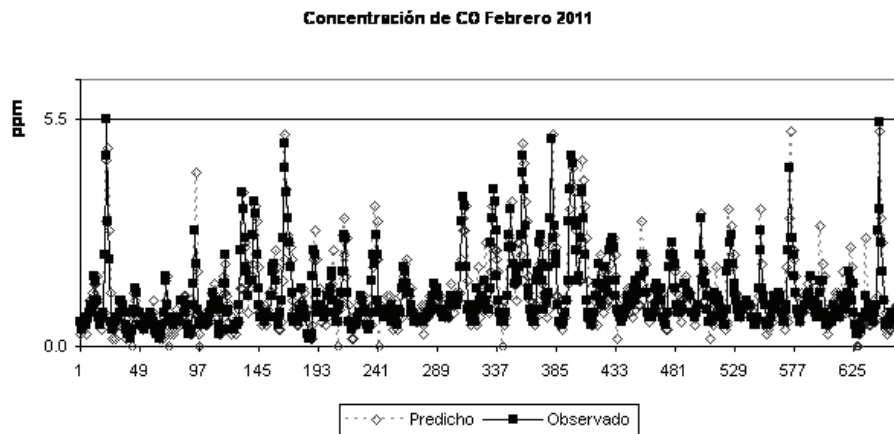
Ahora bien, los resultados experimentales presentados aquí predicen el valor del contaminante criterio CO; pero dado que estos valores sirven de base para calcular el valor del IMECA correspondiente, de acuerdo con lo anteriormente explicado con respecto al SIMAT, la predicción de la concentración del contaminante criterio es de suma importancia.

Veamos unos ejemplos. El día 2 de febrero, a las 19 horas, el sistema propuesto predijo una concentración de CO de 0,8 ppm, siendo el valor real de 0,8 ppm; como es de esperarse, los correspondientes valores del IMECA son iguales: 7 (condición Buena). Por otro lado, el 16 de febrero a las 13 horas, se obtuvieron los valores predicho y real de 1,7 ppm y 1,4 ppm, respectivamente, lo que da valores del IMECA de 15 y 13 (condición Buena para ambos casos). Como se puede ver, aunque la predicción no fue exacta en el valor de CO, el nivel de IMECA fue correctamente predicho: condición Buena. Finalmente, el día 23 de febrero a las 6 horas, el valor predicho fue de 0,6 ppm, mientras que el valor real fue de 1,8 ppm, y por ello los valores del IMECA de 5 y 16, respectivamente. A pesar de que la diferencia en los valores de CO es de 1,2, lo que genera una diferencia en los valores del IMECA de 11, el nivel del índice corresponde a una condición Buena para ambos casos.

Algunos resultados del mes de febrero de 2011 se pueden apreciar en la tabla 3, mientras que la comparación entre los resultados predichos y los reales (observados) se muestra en la figura 1.

**Tabla 3** Resultados experimentales; P significa “Predicho”, O significa “Observado”

Muestra	Concentración de CO (ppm)			Valor del IMECA			Nivel del IMECA		
	P	O	Error	P	O	Error	P	O	Error
2 de febrero 19:00	0,8	0,8	0,0	7	7	0	Buena	Buena	No
11 de febrero 14:00	1,0	0,8	0,2	9	7	2	Buena	Buena	No
16 de febrero 13:00	1,7	1,4	0,3	15	13	2	Buena	Buena	No
18 de febrero 11:00	3,4	2,2	1,2	31	20	11	Buena	Buena	No
23 de febrero 6:00	0,6	1,8	-1,2	5	16	-11	Buena	Buena	No
28 de febrero 9:00	5,2	3,3	1,9	47	30	17	Buena	Buena	No



**Figura 1** Comparación entre los datos predichos y reales para la concentración de CO tomados de la estación Instituto Mexicano del Petróleo durante marzo de 2001

Para determinar cuantitativamente qué tan buenos son estos resultados, se utilizaron dos medidas de desempeño. Por un lado el *Rooted Mean Square Error* (RMSE por sus siglas en inglés, raíz cuadrada del error cuadrático medio), que es una de las medidas de desempeño más utilizadas en redes neuronales y se calcula de acuerdo con la ecuación 7. Por otro lado, para describir qué tanto subestima o sobreestima la situación el algoritmo utilizado, se usó el sesgo (*bias*), que se calcula usando la ecuación 8. Para ambas ecuaciones,  $P_i$  es el  $i$ -ésimo valor predicho y  $O_i$  es el  $i$ -ésimo valor original (real, observado).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (7)$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (8)$$

Por otro lado, se utilizaron los mismos datos con los que aprendió y operó el método propuesto para probar otros métodos conocidos. Para ello, se utilizó la plataforma WEKA [31]. Los resultados obtenidos al predecir la concentración de CO para cada hora del mes de febrero de

2011, usando datos muestreados durante el año 2010 para aprender —tanto para el clasificador Gamma como para otros métodos— se presentan en la tabla 4.

Como se puede apreciar, con respecto a la concentración el clasificador Gamma presenta el segundo mejor resultado, siendo superado únicamente por SMOreg (regresión utilizando

SVMs) en RMSE (0,5564 ppm y 0,5546 ppm, respectivamente) y por LinearRegression en sesgo (-4,6 ppm y 0,3 ppm, respectivamente). En cuanto al valor del IMECA, la situación es mejor: de nueva cuenta el mejor método en RMSE es SMOreg y el método propuesto es el segundo mejor, con desempeños de 5,0640 puntos y 5,0682 puntos, respectivamente; mientras que el sesgo del clasificador Gamma es el mejor: 31 puntos.

**Tabla 4** Desempeño cuantitativo de la predicción

<i>Método</i>	<i>Concentración (ppm)</i>		<i>Valor IMECA (puntos)</i>	
	RMSE	Bias	RMSE	Bias
IBk 1 (1-NN)	0,631270	-20,1	5,784323	-192
RBFNetwork	0,587618	-17,0	5,367577	-156
MultilayerPerceptron	0,575232	-44,8	5,254347	-431
IBk 3 (3-NN)	0,561241	-13,7	5,110281	-142
ConjunctiveRule	0,559703	10,1	5,106135	54
<b>LinearRegression</b>	0,556912	0,3	5,077463	-31
M5P	0,556749	-15,5	5,100206	-156
<b>Método propuesto</b>	<b>0,556382</b>	-4,6	5,068218	-25
SMOreg	0,554639	-14,8	<b>5,064037</b>	-165

Cabe mencionar que, aunque SMOreg supera al clasificador Gamma en RMSE para la concentración y para el valor del IMECA, lo hace marginalmente: la diferencia entre ambos es de 0,001743 ppm para el primer caso y de 0,004181 puntos para el segundo. Por otro lado, SMOreg tiene un mayor sesgo en ambos casos, en particular con respecto al valor del IMECA: las diferencias son de 10,2 ppm y 140 puntos, respectivamente; estos últimos 140 puntos pueden representar una diferencia de casi dos niveles de IMECA, como se indican las fronteras entre niveles en la tabla 1. Además de lo anterior, el método SMOreg es particularmente lento, considerablemente más que los otros métodos probados en WEKA.

De manera similar, LinearRegression presenta un mejor sesgo que el método propuesto con

respecto a la concentración, con una diferencia de 4,9 ppm, a la vez que la diferencia en RMSE es de apenas 0,00053 ppm (menor incluso que la correspondiente diferencia entre SMOreg y el clasificador Gamma). En cuanto al valor de IMECA, LinearRegression tiene también un muy buen sesgo: el segundo mejor, superado por el clasificador Gamma por apenas 6 puntos.

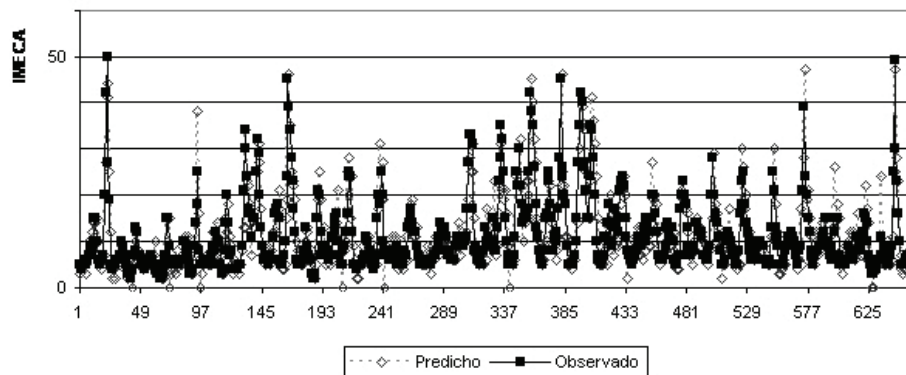
Asimismo, el sesgo exhibido por la mayoría de los métodos es negativo, lo que indica que es más común que se subestime la predicción a que se sobreestime.

Por otro lado, al observar la figura 1 se puede ver que los valores de concentración predichos son bastante cercanos a los observados. Asimismo, se puede ver que la predicción del valor del IMECA arrojado por la concentración de CO tiene un



desempeño muy similar al de la concentración (figura 2). La forma de ambas gráficas (figuras 1 y 2) es muy similar, además de que tanto el

RMSE como el Bias es bastante similar para ambos datos: la concentración de CO y su correspondiente valor de IMECA.



**Figura 2** Comparación entre los datos predichos y reales para el IMECA de CO tomados de la estación Instituto Mexicano del Petróleo durante marzo de 2001

### Conclusiones y trabajo futuro

En el presente trabajo se ha mostrado experimentalmente la utilidad del clasificador Gamma en la predicción de valores futuros de la concentración de contaminantes utilizados para el cálculo del Índice Metropolitano de la Calidad del Aire de la Ciudad de México (IMECA), específicamente en la predicción de valores horarios del contaminante criterio Monóxido de Carbono (CO) en la base de datos RAMA.

La utilidad de aplicar de manera sistemática herramientas de Reconocimiento de Patrones, en particular el clasificador Gamma, a esta tarea de predecir la calidad del aire en un futuro cercano, radica en cómo se puede soportar una mejor toma de decisiones.

Cabe mencionar que, como su nombre lo indica, el clasificador Gamma fue diseñado pensando en la tarea de clasificación de patrones, no en la tarea de predicción de datos. A pesar de ello, se puede ver que los resultados obtenidos son sumamente prometedores.

Una clara línea de investigación derivada del presente trabajo es la detección de mínimos y

máximos locales para mejorar el desempeño de la propuesta en esos casos particulares.

Por otro lado, están también presentes la posibilidad de aplicar este método a la predicción de otros contaminantes, tanto atmosféricos como no atmosféricos (acuáticos o de suelos, por ejemplo), así como la extensión del método para incluir más información además de la concentración conocida del contaminante de interés.

### Agradecimientos

Los autores agradecen el apoyo de las siguientes instituciones para la realización de esta obra: Secretaría de Investigación y Posgrado, Secretaría Académica, COFAA y CIC del Instituto Politécnico Nacional, Sistema de Monitoreo Atmosférico de la Ciudad de México (SIMAT), CONA-CyT, Sistema Nacional de Investigadores (SNI), así como al ICyTDF (proyectos PIUTE10-77 y PICSO10-85).

### Referencias

1. K. Toepfer. *Aliados Naturales: El Programa de las Naciones Unidas para el Medio Ambiente (PNUMA)*

- y la sociedad civil. Ed. UNEP-United Nations Foundation. Hertfordshire. UK. 2004. pp. 16-22.
2. L. Hisas (editora). *A Guide to the Global Environmental Facility (GEF) for NGOs*. Ed. UNEP-United Nations Foundation. Buenos Aires. Argentina. 2005. pp. 25-71.
  3. United Nations. *Rio Declaration on Environment and Development (Earth Summit)*. Ed. United Nations Conference on Environment and Development. Rio de Janeiro. Brazil. 1992. pp. 1-6.
  4. United Nations. *Kyoto Protocol to The United Nations Framework Convention on Climate Change*. Ed. United Nations Framework Convention on Climate Change. Kyoto. Japan. 1997. pp. 1-21.
  5. Secretaría de Comercio y Fomento Industrial. *Protección al Ambiente - Contaminación Atmosférica - Determinación de Neblina de Acido Fosfórico en los Gases que Fluyen por un Conducto*. Norma Mexicana NMX-AA-090-1986. México. 1986.
  6. Generalitat de Catalunya. *Web del Departamento de Medio Ambiente y Vivienda de la Generalitat de Catalunya*. 2007. Disponible en: <http://mediambient.gencat.net/>. Consultado el 16 de diciembre de 2010.
  7. J. Burnetti, R. Manley, W. Mitchell, D. Varnadore. "Visualizing environmental data for program decision support". *Proc. IEEE Conference on Visualization*. Vol. 1. 1991. pp. 398-404.
  8. B. Kompare, S. Dzeroski. "Two artificial intelligence methods for knowledge synthesis from environmental data". *Computer techniques in environmental studies V*. Vol. 2. 1994. pp. 265-272.
  9. J. Hart, I. Hunt, V. Shankaraman. "Environmental Management Systems - a Role for AI?." *Proc. Binding Environmental Sciences and Artificial Intelligence BESAI'98*. Ed. BESAI Working Group. Brighton. UK. 1998. pp. 1-9.
  10. M. W. Gardner, S. R. Dorling. "Artificial neural networks (the multilayer perceptron)--a review of applications in the atmospheric sciences". *Atmospheric Environment*. Vol. 32. 1998. pp. 2627-2636.
  11. G. Nunnari, A. F. M. Nucifora, C. Randieri. "The application of neural techniques to the modelling of time-series of atmospheric pollution data". *Ecological Modelling*. Vol. 111. 1998. pp. 187-205.
  12. G. Spellman. "An application of artificial neural networks to the prediction of surface ozone". *Applied Geography*. Vol. 19. 1999. pp. 123-136.
  13. E. Kalapanidas, N. Avouris. "Feature Selection Using a Genetic Algorithm Applied on an Air Quality Forecasting Problem". *Proc. Binding Environmental Sciences and Artificial Intelligence, BESAI 2002*. Ed. BESAI Working Group. 2002. pp. 9.1-9.5.
  14. C. Yáñez Márquez. *Memorias Asociativas basadas en Relaciones de Orden y Operadores Binarios*. Tesis de Doctorado. Instituto Politécnico Nacional. Centro de Investigación en Computación. México. 2002. pp. 53-62.
  15. M. E. Acevedo Mosqueda. *Memorias Asociativas Bidireccionales Alfa-Beta*. Tesis de Doctorado. Instituto Politécnico Nacional. Centro de Investigación en Computación. México. 2006. pp. 35-89.
  16. M. E. Acevedo Mosqueda, C. Yáñez Márquez, I. López Yáñez. "A New Model of BAM: Alpha-Beta Bidirectional Associative Memories". *Lecture Notes in Computer Science, LNCS 4263*. Ed. Springer-Verlag. Berlin. Heidelberg. 2006. pp. 286-295.
  17. M. E. Acevedo Mosqueda, C. Yáñez Márquez, I. López Yáñez. "Alpha-Beta Bidirectional Associative Memories Based Translator". *International Journal of Computer Science and Network Security*. Vol. 6. 2006. pp. 190-194.
  18. M. E. Acevedo Mosqueda, C. Yáñez Márquez, I. López Yáñez. "Alpha-Beta Bidirectional Associative Memories: Theory and Applications". *Neural Processing Letters*. Vol. 26. 2007. pp. 1-40.
  19. C. Yáñez-Márquez, E. M. Felipe-Riverón, I. López-Yáñez, R. Flores-Carapia. "A Novel Approach to Automatic Color Matching". *Lecture Notes in Computer Science. LNCS 4225*. Ed. Springer-Verlag. Berlin. Heidelberg. 2006. pp. 529-538.
  20. I. Román Godínez. *Aplicación de los modelos asociativos Alfa-Beta a la Bioinformática*. Tesis de Maestría. Instituto Politécnico Nacional. Centro de Investigación en Computación. México. 2007. pp. 39-73.
  21. I. Román Godínez, I. López Yáñez. "A New Classifier Based on Associative Memories". *Proc. 15th International Conference on Computing, CIC 2006*. Ed. IEEE Computer Society. Washington, D.C.(USA). 2006. pp. 55-59.
  22. E. Guzmán, O. Pogrebnyak, C. Yáñez Márquez. "Image Compression Algorithm Based on Morphological Associative Memories". *Lecture Notes in Computer Science. LNCS 4225*. Ed. Springer-Verlag. Berlin. Heidelberg. 2006. pp. 519-528.
  23. C. Yáñez Márquez, M. E. Cruz Meza, F. A. Sánchez Garfías, I. López Yáñez. "Using Alpha-Beta Associative Memories to Learn and Recall RGB Images". *Lecture Notes in Computer Science. LNCS 4493*. Ed. Springer-Verlag. Berlin. Heidelberg. 2007. pp. 828-833.

24. L. O. López Leyva, C. Yáñez Márquez, R. Flores Carapia, O. Camacho Nieto. "Handwritten Digit Classification Based on Alpha-Beta Associative Model". *Lecture Notes in Computer Science. LNCS 5197*. Ed. Springer-Verlag. Berlin. Heidelberg. 2008. pp. 437-444.
25. I. López Yáñez. *Clasificador Automático de Alto Desempeño*. Tesis de Maestría. Instituto Politécnico Nacional, Centro de Investigación en Computación. México. 2007. pp. 24-44.
26. C. Yáñez Márquez, I. López Yáñez, G. de la L. Sáenz Morales. "Analysis and Prediction of Air Quality Data with the Gamma Classifier". *Lecture Notes in Computer Science, LNCS 5197*. Ed. Springer-Verlag. Berlin. Heidelberg. 2008. pp. 651-658.
27. I. López Yáñez, C. Yáñez Márquez, V. M. Silva García. *Forecasting Air Quality Data with the Gamma Classifier*. Pattern Recognition. Y. Peng-Yeng (editor). Ed. INTECH. Croatia. 2009. pp. 499-512. Disponible en <http://sciyo.com/articles/show/title/forecasting-air-quality-data-with-the-gamma-classifier>.
28. Secretaría del Medio Ambiente. "Sistema de Monitoreo Atmosférico de la Ciudad de México". Disponible en: <http://www.sma.df.gob.mx/simat2/>. Consultado el 25 de febrero de 2011.
29. Gobierno del Distrito Federal. "Norma Ambiental para el Distrito Federal". Gaceta Oficial del Distrito Federal. Epoch XVI. 2006. Disponible en: [http://www.sma.df.gob.mx/simat/PCAA/noviembre06\\_29\\_141\\_imeca.pdf](http://www.sma.df.gob.mx/simat/PCAA/noviembre06_29_141_imeca.pdf). Consultado el 20 de febrero de 2011.
30. K. S Papadomanolakis, A. P. Kakarountas, N. Sklavos, C. E. Goutis. *A Fast Johnson-Mobius Encoding Scheme for Fault Secure Binary Counters*. Proc. Design, Automation and Test in Europe 2002. C.D. Kloos and J. da Franca(editors). Paris. France. 2002. pp. 1.
31. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten. "The WEKA Data Mining Software: An Update". *SIGKDD Explorations*. Vol. 11. 2009. pp. 10-18.