

---

## Aplicación del análisis discriminante a los datos de calidad del aire de la Zona Metropolitana de Guadalajara

A. Figueroa-Montaña<sup>1\*</sup>, C. Garibay-López<sup>2</sup>, P. Gutierrez-Gonzalez<sup>2</sup>

<sup>1</sup>Universidad de Guadalajara, Departamento de Física

<sup>2</sup>Departamento de Matemáticas, CUCEI, Av Revolución 1500, Guadalajara, Jalisco, 44420 México

---

*Application of discriminant analysis to air quality data of the Metropolitan Zone of Guadalajara*

### Abstract

The aim of this paper was to classify the eight monitoring stations of Guadalajara's atmospheric network, by means of discriminant analysis. Measured variables were hourly means of criterion pollutants (PM<sub>10</sub>, O<sub>3</sub>, CO, NO<sub>2</sub>, and SO<sub>2</sub>), as well as meteorological variables (temperature, relative humidity, and wind) for the period 1994-2005. The analysis was performed using the software STATGRAPHICS 5.1. Results suggested three discriminant functions to explain 82.36 % of total variability in the data. The first discriminant function explains 51.19 % of variability and distinguishes PM<sub>10</sub> as the variable that contributes the most to define this function. The second discriminant function explains 20.14 % of variability; within this function temperature and humidity as meteorological variables, as well as CO and NO<sub>2</sub> as criterion pollutants contribute the most to define this function. Finally, the third discriminant function explains 11.03 % of variability, and wind speed contributes the most to define it. The general classification table, as the result of the evaluation of the three discriminant functions, suggested an overall 27.47 % of observations correctly classified. Within this percentage, the stations best classified were Center, Miravalle and Atemajac with a 32.70, 32.16 % and 31.61% respectively. Such results do not agree with local environmental reports at which Miravalle highlights as the station with the poorest air quality and consequently the highest health impacts. Up on such findings, the authorities have set air pollution management policies for this particular area, thus failing to approach the problem in global manner. According to the results of this study, it is worth reconsidering such policies towards integrated management policies that involves the metropolitan area as a whole. With the data used and the results of the discriminant analysis it is not possible to state that the worst air pollution scenario and health impacts only occur at the southwest part of the city where Miravalle is located. Even though the overall classification of the stations was 27.47%, the stations best classified were Center, Miravalle and Atemajac. It is well identified that within this places economic and a brought range of services concentrate here, as well as industrial activities, and the valley region of the city. These stations best classification puts forward the strong relationship between meteorological-climatic conditions and pollution patterns within urban areas.

*Key words:* Discriminant analysis, criterion pollutants, Metropolitan Zone of Guadalajara.

### Resumen

El objetivo de este trabajo fue clasificar las ocho estaciones de la Red de Monitoreo Atmosférico de la Zona Metropolitana de Guadalajara, aplicando la técnica del análisis discriminante de la estadística multivariada. Las variables medidas fueron los promedios horarios de contaminantes criterio (PM<sub>10</sub>, O<sub>3</sub>, CO, NO<sub>2</sub>, y SO<sub>2</sub>) y las magnitudes meteorológicas (temperatura, humedad relativa, y viento) durante el periodo 1994-2005. El análisis se aplicó utilizando el paquete estadístico STATGRAPHICS 5.1. Los resultados del análisis proponen tres funciones discriminantes con las que se explica el 82.36 % de la variabilidad total de los datos. La primera función discriminante explica el 51.19% de la variabilidad y la variable que mayormente contribuye a definir la función corresponde a PM<sub>10</sub>. La segunda función discriminante explica el 20.14% de la

---

\*Autor de correspondencia

E-mail: arturo.figueroa@cucei.udg.mx; Tel: (33)13 78 59 00. Ext 7763; Fax: (33)36 19 82 92 y 36 19 8054.

variabilidad, siendo las variables que mayormente contribuyen a la definición de la función; las variables meteorológicas temperatura-humedad y los contaminantes CO-NO<sub>2</sub>. Finalmente, la tercera función discriminante explica el 11.03% de la variabilidad donde se distingue una mayor contribución de la velocidad del viento en la definición de la función. La tabla de clasificación general, producto de la evaluación de las tres funciones discriminantes, sugiere un total del 27.47% de casos correctamente clasificados. Dentro de este porcentaje, las estaciones mejor clasificadas fueron el Centro, Miravalle y Atemajac, con un porcentaje de clasificación del 32.70%, 32.16% y 31.61% respectivamente. Estos resultados no concuerdan con los reportes emitidos por las autoridades locales en materia ambiental, en donde se identifica de manera importante a la estación Miravalle como la estación con las peores condiciones de calidad del aire y por ende donde ocurren los mayores impactos a la salud de la población. Lo anterior ha hecho que las autoridades establezcan políticas de manejo y control de la calidad del aire en esta zona en particular, sacrificando el enfoque global del problema. Con base a los resultados de este estudio, es importante reorientar esas políticas hacia un plan integral de manejo de la calidad del aire en toda la zona metropolitana de Guadalajara, ya que en base a la clasificación de las estaciones no se observaron diferencias importantes en la variabilidad de los contaminantes y las magnitudes meteorológicas entre la estación Miravalle y el resto de las estaciones de la Red de Monitoreo de la Zona Metropolitana de Guadalajara. Con los datos analizados y los resultados del estudio no es posible identificar a la estación Miravalle o la porción suroeste de la ciudad donde se ubica dicha estación como la zona más contaminada de toda la ZMG. Aunque de manera general el porcentaje de clasificación de las estaciones fue únicamente del 27.47%, de manera importante las estaciones mejor clasificadas fueron la zona Centro, Miravalle y Atemajac, lugares donde se concentran las actividades económicas y de servicios, la actividad industrial y la zona de valles en la ciudad. La mejor clasificación para estas estaciones ponen de manifiesto la relación estrecha que existe entre las condiciones meteorológico-climáticas y los patrones de contaminación del aire en las zonas urbanas.

*Palabras clave:* Análisis discriminante, contaminantes criterio, Zona Metropolitana de Guadalajara.

---

### **Introducción**

La contaminación atmosférica en los ambientes urbanos en las grandes ciudades de México se percibe desde hace algún tiempo como un problema serio, llevando algunas de las metrópolis como la ciudad de México a implantar regulaciones de restricción de uso de vehículos automotores, y para la ZMG programas de afinación controlada y reestructuración de las paradas del transporte urbano a fin de reducir las emisiones. Las emisiones anuales de contaminantes en el país son superiores a 16 millones de toneladas, de las cuales el 75 % es de origen vehicular. De las emisiones anuales, el 23.6% se generan en la Ciudad de México, el 3.4% en Guadalajara y el 3 % en Monterrey. Los otros centros industriales del país generan el 70% restante (SIMA, 2001).

La contaminación atmosférica en la ZMG ha sido poco estudiada. Los primeros intentos por analizar los contaminantes en la atmósfera datan de los años 70s, cuando se instala una red manual. Sin embargo, los estudios más recientes de contaminación del aire inician a principios de los 90s, cuando la RAMA inicia sus actividades de

manera regular. Entre algunos de los trabajos, se encuentra la investigación de las elevadas concentraciones de ozono debido a la presencia de smog fotoquímico sobre la ZMG en Octubre de 1996 (Tereshchenko y Filonov, 1997), así como el estudio preliminar de los altos niveles de ozono y los efectos a la salud de la población (Tereshchenko *et al*, 1999). También se han realizado estudios de los niveles de contaminación por medio de modelos matemáticos (Davydova y Figuroa, 1999) y la dependencia de emisiones industriales con el viento, que han ayudado a definir el aporte de contaminantes a cada una de las zonas de la ZMG (Davydova y Filonov, 2001). Los estudios recientes abordan el análisis del ozono de tropósfera (Nájera, 2005), así como otros aspectos importantes de la contaminación urbana, como la lluvia ácida (García, 2001 y 2004).

El propósito principal de este trabajo es hacer una clasificación de las ocho estaciones de la red de monitoreo atmosférico de la Zona Metropolitana de Guadalajara, aplicando la técnica del análisis discriminante a los datos de contaminantes y variables meteorológicas. Así el conocimiento

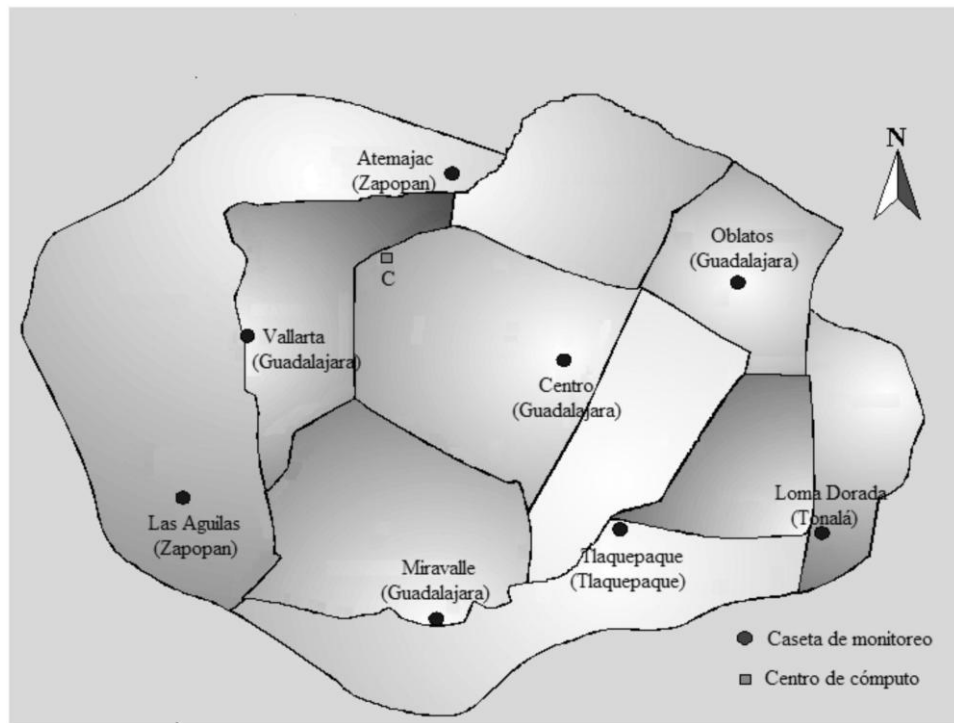
producto del presente estudio nos llevará a identificar diferencias de la variabilidad de los índices de contaminación del aire y las variables meteorológicas en la Zona Metropolitana de Guadalajara. Además de ser una valiosa herramienta de consulta para el establecimiento de planes y políticas de mejora de la calidad del aire más acertadas.

### Material y métodos

El material para la realización del trabajo corresponde a los promedios horarios de los datos de contaminación del aire ( $PM_{10}$ ,  $O_3$ ,  $CO$ ,  $NO_2$ , y  $SO_2$ ) y variables meteorológicas (temperatura, humedad relativa, y viento) para el periodo 1994-2005. Estos se obtuvieron de la red automática de monitoreo atmosférico de la Zona Metropolitana de Guadalajara. La red es operada de manera conjunta por la Secretaría del Medio Ambiente para el Desarrollo Sustentable del estado de Jalisco, y las Direcciones de Ecología de los Ayuntamientos de

Guadalajara, Zapopan, Tlaquepaque y Tonalá. Entre estos municipios se encuentran distribuidas un total de ocho estaciones automáticas (Fig. 1, Tabla 1).

La ZMG abarca una extensión territorial superior a los 350 km<sup>2</sup> donde habita una población de 3'665,739 habitantes (INEGI, 2002). Se ubica en la porción Suroeste de la altiplanicie central de México sobre el valle de Atemajac, a 20°39'54" de latitud N, 103°18'42" de longitud W; y una altitud de 1,551 msnm. Importantes complejos montañosos circundan la ZMG, los que constituyen parcialmente una barrera física natural importante para la dinámica local de los contaminantes. Un segundo factor importante son las inversiones térmicas y la ocurrencia de vientos débiles durante casi todo el año. En promedio los periodos de calma durante el día pueden tener una duración de hasta once horas. La clasificación de las estaciones se hizo por medio de la técnica de análisis discriminante con ayuda del paquete estadístico STATGRAPHICS 5.1.



**Figura 1.** Distribución de las estaciones de monitoreo atmosférico en la Zona Metropolitana de Guadalajara.

**Tabla 1. Estaciones de la red automática de monitoreo atmosférico y su ubicación geográfica.**

Estación Nombre/Número	Ubicación geográfica		
	Latitud Norte	Longitud Oeste	Altitud m.s.n.m*
Atemajac/1	20°43'10"	103°21'19"	1,563
Oblatos/2	20°42'01"	103°17'47"	1,608
Loma Dorada/3	20°37'45"	103°15'50"	1,645
Tlaquepaque/4	20°38'27"	103°18'45"	1,622
Miravalle/5	20°36'53"	103°20'36"	1,622
Las Águilas/6	20°37'51"	103°24'59"	1,633
Vallarta/7	20°40'49"	103°23'55"	1,640
Centro/8	20°40'26"	103°20'01"	1,582

\*Metros sobre el nivel del mar

Se trata de una técnica multivariada cuya finalidad es analizar si existen diferencias entre grupos de objetos en este caso las estaciones de monitoreo, respecto a un conjunto de variables medidas sobre los mismos, en este caso los contaminantes y las variables meteorológicas. De existir diferencias entre los objetos, la técnica explica en qué sentido se dan y proporciona procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido o de uno de los grupos analizados.

### Resultados y discusión

Los resultados del análisis proponen tres funciones discriminantes para explicar de manera acumulada

el 82.36 % de la variabilidad total en los datos (Tabla 2). La primera función discriminante explica el 51.19% de la variabilidad y las variables que mayormente contribuyen a definir las funciones se muestran en la tabla 3 a partir de sus coeficientes estandarizados. Así la variable que mayormente contribuye a definir la primera función corresponde a PM<sub>10</sub>. La segunda función discriminante explica el 20.14% de la variabilidad, siendo las variables que mayormente contribuyen a la definición de la función; las variables meteorológicas de temperatura y humedad, así como los contaminantes CO y NO<sub>2</sub>. Finalmente, la tercera función discriminante explica el 11.03% de la variabilidad donde se distingue una mayor contribución de la velocidad del viento en la

**Tabla 2. Funciones discriminantes y variabilidad explicada.**

Función discriminante	Valores propios	Porcentaje relativo	Correlación canónica
1	0.173047	51.19	0.38408
2	0.068094	20.14	0.25249
3	0.037283	11.03	0.18959
4	0.030767	9.10	0.17277
5	0.023824	7.05	0.15254
6	0.003430	1.01	0.05846
7	0.001578	0.47	0.03970

**Tabla 3. Contribución de cada una de las variables medidas en la definición de las funciones discriminantes.**

Variables medidas	Funciones discriminantes		
	1	2	3
Temperatura	0.248946	0.697947	0.002358
Humedad	0.304062	0.545673	-0.254736
Dirección viento	0.142081	0.245449	-0.221044
Velocidad viento	-0.224145	-0.287453	0.732719
PM <sub>10</sub>	1.10324	-0.0778987	0.252383
O <sub>3</sub>	-0.188597	0.105261	0.0100768
CO	-0.347229	0.611962	0.0688933
Humedad	0.248946	0.437803	0.348905

definición de la función. Siendo las expresiones matemáticas de las funciones:

$$D_1 = 0.25 * \text{Temperatura} + 0.34 * \text{Humedad} + 0.14 * \text{Dirección viento} - 0.22 * \text{Velocidad viento} + 1.10 * \text{PM}_{10} - 0.19 * \text{O}_3 - 0.34 * \text{CO} - 0.23 * \text{NO}_2 - 0.20 * \text{SO}_2.$$

$$D_2 = 0.70 * \text{Temperatura} + 0.54 * \text{Humedad} + 0.24 * \text{Dirección viento} - 0.28 * \text{Velocidad viento} - 0.08 * \text{PM}_{10} + 0.10 * \text{O}_3 + 0.61 * \text{CO} + 0.44 * \text{NO}_2 + 0.13 * \text{SO}_2.$$

$$D_3 = 0.002 * \text{Temperatura} - 0.25 * \text{Humedad} - 0.22 * \text{Dirección viento} + 0.73 * \text{Velocidad viento} + 0.25 * \text{PM}_{10} + 0.01 * \text{O}_3 + 0.07 * \text{CO} + 0.35 * \text{NO}_2 + 0.21 * \text{SO}_2.$$

A partir de la evaluación de las observaciones en las funciones discriminantes, se obtiene la tabla de clasificación general (Tabla 4). Dicha tabla muestra en paréntesis el porcentaje de casos correctamente clasificados para cada una de las ocho estaciones de monitoreo. De manera general el porcentaje de casos correctamente clasificados fue del 27.47%. Este valor se obtiene haciendo una relación del total de casos correctamente clasificados, esto es la suma de la diagonal principal, en relación al total de los 492 365 datos originales utilizados en el análisis. Dentro de este porcentaje, las estaciones mejor clasificadas fueron el Centro, Miravalle y Atemajac, con un porcentaje de clasificación del 32.70%, 32.16% y 31.61% respectivamente. La mejor clasificación de estas estaciones se puede relacionar

con la bien establecida actividad económica y la concentración de servicios para el caso de la estación Centro, así como la actividad industrial en la zona de Miravalle, y a mejores condiciones meteorológicas y de contaminación para el valle de la ciudad donde se localiza la estación de Atemajac.

### Discusión

Con la aplicación de la técnica de análisis discriminante a los datos de contaminación y variables meteorológicas no se obtuvo un porcentaje alto de casos correctamente clasificados, aunque la técnica resultó valiosa al encontrar la mejor clasificación para las estaciones de Centro, Miravalle y Atemajac. La mayor aportación que el estudio hace al entendimiento de la contaminación del aire en la Zona Metropolitana de Guadalajara es visualizar el problema de manera generalizada en toda la ZMG. Esto es los impactos de la mala calidad del aire son de igual forma importantes en cada una de las estaciones de la red de monitoreo y no solo en la estación de Miravalle como lo establecen las políticas de manejo de la calidad del aire recientes. La contribución importante que tienen las  $\text{PM}_{10}$  a definir la primera función discriminante y a explicar mejor los datos de contaminación para la estación Miravalle no únicamente se pueden relacionar a la actividad de la industria cementera y la fabricación de ladrillos que en la zona ocurren, sino que también los altos índices de partículas aquí registrados se deben a la deforestación y predominancia de suelo natural, así

**Tabla 4. Tabla de clasificación general mostrando el porcentaje de casos correctamente clasificados en paréntesis.**

Número Estación	Total observaciones	1	2	3	4	5	6	7	8
<b>1</b>	<b>75314</b>	23809	5403	4353	4644	2619	9196	12783	12508
		(31.61)	(7.17)	(5.78)	(6.17)	(3.48)	(12.21)	(16.97)	(16.61)
<b>2</b>	<b>39749</b>	4140	10059	3362	4295	4238	4298	6124	3231
		(10.42)	(25.31)	(8.46)	(10.81)	(10.66)	(10.81)	(15.41)	(8.13)
<b>3</b>	<b>58347</b>	8689	5909	11694	5843	11089	4787	3519	6817
		(14.89)	(10.13)	(20.04)	(10.01)	(19.01)	(8.20)	(6.03)	(11.68)
<b>4</b>	<b>54910</b>	4988	66.08	6758	11430	8341	4637	5929	6219
		(9.08)	(12.03)	(12.31)	(20.82)	(15.19)	(8.44)	(10.80)	(11.33)
<b>5</b>	<b>69001</b>	6686	7724	7176	8274	22192	6069	6486	4394
		(9.69)	(11.19)	(10.40)	(11.99)	(32.16)	(8.80)	(9.40)	(6.37)
<b>6</b>	<b>57177</b>	5507	10647	4134	3877	4393	14821	5257	8541
		(9.63)	(18.62)	(7.23)	(6.78)	(7.68)	(25.92)	(9.19)	(14.94)
<b>7</b>	<b>65885</b>	9455	10313	3535	5460	4703	8213	17715	6491
		(14.35)	(15.65)	(5.37)	(8.29)	(7.14)	(12.47)	(26.89)	(9.85)
<b>8</b>	<b>71984</b>	11922	3348	3520	6164	2490	10690	10314	23536
		(16.56)	(4.65)	(4.89)	(8.56)	(3.46)	(14.85)	(14.33)	(32.70)

como a los impactos que la actividad agrícola y el mal estado de las vialidades de igual forma generan en esta zona. Por otro lado el CO y NO<sub>2</sub> en la función discriminante dos, son indicadores de la contribución que hacen las emisiones vehiculares al estado de la calidad del aire en la ZMG, pues estudios recientes refieren que un 75 % de las emisiones contaminantes provienen de automotores y un 25 % de la actividad industrial (Molina and Molina, 2004, Figueroa, 2005). De igual manera en esta función las variables meteorológicas de temperatura y humedad juegan un papel importante en el comportamiento fisicoquímico de los contaminantes. Por ejemplo cielos despejados con intensa radiación solar promueven la foto-oxidación de los precursores de la formación de ozono troposférico, y gran contenido de humedad en la atmosfera forma conglomerados de partículas y aerosoles haciéndolos permanecer por menos tiempo en suspensión. Similarmente en la tercera función discriminante la velocidad del viento es responsable de los mecanismos de difusión y transporte de contaminantes en la cuenca atmosférica. Al respecto el patrón de vientos en la ZMG se distingue por vientos débiles con una magnitud promedio que apenas supera los 3.7 m s<sup>-1</sup> y periodos de calma de hasta 11 horas durante el día (Tereshchenko y Filonov, 2001).

### **Conclusiones**

Con el procedimiento de la técnica de análisis discriminante fue posible encontrar tres funciones discriminantes para explicar el 82.36 % de la variabilidad total de los datos. Las variables que mayormente contribuyen a definir cada una de las funciones discriminantes fueron PM<sub>10</sub> para la función uno, las variables meteorológicas de temperatura y humedad, y los contaminantes criterio de CO y NO<sub>2</sub> para la función dos. En la función tres la variable que mayormente contribuye a su definición es la variable velocidad del viento. La contribución que este grupo de variables hace a la definición de las funciones discriminantes nos ayuda a comprender que el problema de la calidad del aire en la ZMG es principalmente debido a las partículas, y que las variables meteorológicas de temperatura, humedad y velocidad del viento juegan un papel importante en los mecanismos de difusión y transporte de contaminantes en la metrópoli. Así pues las condiciones meteorológicas aminoran o

favorecen los eventos de contaminación. Por ejemplo la circulación del viento y la turbulencia atmosférica diluyen la concentración de contaminantes, pero por otro lado la radiación solar y las altas temperaturas aceleran la formación de contaminantes secundarios como el ozono debido a la foto-oxidación de NO<sub>x</sub> y otros gases producto de la combustión de hidrocarburos por tráfico vehicular, industria y plantas de generación de energía. Similarmente, las inversiones térmicas limitan el transporte vertical de los compuestos no deseados. De lo anterior se concluye que la calidad del aire de una cuenca atmosférica depende, en primera instancia, del volumen de contaminantes emitidos, del comportamiento fisicoquímico de estos y de los procesos dinámicos de la atmósfera que determinarán su dispersión, transformación y remoción.

### **Agradecimientos**

Los autores desean agradecer a la Secretaria del Medio Ambiente para el Desarrollo Sustentable por proporcionar los datos del monitoreo. De igual manera agradecen a la Universidad de Guadalajara, en particular a los Departamentos de Física y Matemáticas por el apoyo recibido para la realización del presente trabajo.

### **Bibliografía**

- Davydova V.B. 2001. Modelación matemática de los niveles de contaminación en la ciudad de Guadalajara, Jalisco México. Tesis Doctoral, Universidad Nacional Autónoma de México, México D.F.
- Davydova V.B. y Skiba Y.N. 1999. Estimación de los niveles de contaminación en la Zona Metropolitana de Guadalajara, Jalisco. México. IX Congreso Nacional de Meteorología. Universidad de Guadalajara, Guadalajara, Jalisco, pp. 140-143.
- Figueroa M.A. 2005. Investigación de los patrones meteorológico-climáticos y los patrones de contaminación atmosférica de la Zona Metropolitana de Guadalajara. Tesis Doctoral, Universidad de Guadalajara, Guadalajara, Jalisco.
- García G.M.E. 2004. Evaluación físico-química de la acidez de las precipitaciones y su dinámica en la Zona Metropolitana de Guadalajara, periodo 1994-2002. Tesis de Maestría, Universidad de Guadalajara, Guadalajara, Jalisco.
- García G.M.E. 2001. Estimación cuantitativa de la acidez de las lluvias en la Zona Metropolitana de Guadalajara, periodo 1994-1999. Tesis Ingeniería Química, Universidad de Guadalajara, Guadalajara Jalisco.
- INEGI. 2002. Síntesis de resultados de la Zona Metropolitana de Guadalajara. XII Censo General de Población y Vivienda 2000. INEGI, Guadalajara, Jalisco.
- Molina M.J. y Molina L.T. 2004. Megacities and atmospheric

- pollution, critical review. *Journal of the Air and Waste Management Association*, 54:644-680.
- SIMA. 2001. Calidad del Aire del Valle de México. (en [http://www.sima.com.mx/valle\\_de\\_mexico/](http://www.sima.com.mx/valle_de_mexico/)).
- Tereshchenko I.E. y Filonov A.E., 2001. Air temperature fluctuations in Guadalajara, México, from 1926 to 1994 in relation to urban growth. *International Journal of Climatology*, 21:483-494.
- Tereshchenko I.E., y Figuroa A.M. 1999. Estudio preliminar de las altas concentraciones de ozono y los efectos a la salud pública en la Zona Metropolitana de Guadalajara. IX Congreso Nacional de Meteorología. Universidad de Guadalajara, Guadalajara, Jalisco, pp. 136-139.
- Tereshchenko I.E. y Filonov A.E. 1997. Acerca de las causas de las elevadas concentraciones de ozono en la atmósfera de la Zona Metropolitana de Guadalajara, en octubre de 1996. *GEOS*, 17: 54-59.