

Pronóstico de la Concentración de Ozono en Guadalajara-México usando Redes Neuronales Artificiales

Ignacio García¹, Alonso Marbán¹, Yenisse M. Tenorio² y José G. Rodríguez²

(1) Instituto Politécnico Nacional, Centro Mexicano para la Producción Más Limpia, Departamento de Posgrado, Av. Acueducto s/n, Barrio La Laguna, Col. Ticomán, Del. Gustavo A. Madero, 07340 México, D.F.-México (e-mail: igarcias@ipn.mx; amarban@ipn.mx)

(2) Instituto Politécnico Nacional, Escuela Superior de Ingeniería y Arquitectura, Departamento de Investigación y Posgrado, U. Prof. "Adolfo López Mateos", Zacatenco, Edif. 12, Del. Gustavo A. Madero, 07738 México, D.F.-México (e-mail: ytenoriom0700@ipn.mx; adore81@prodigy.net.mx)

Resumen

Se desarrolló un modelo de pronóstico de la concentración de ozono máxima para el día de interés para la Zona Metropolitana de Guadalajara-México. Se usó una Red Neuronal Artificial alimentada con seis variables meteorológicas y tres químicas. Los nodos de la capa oculta se variaron entre doce y quince, las funciones de transferencia de la capa interna fueron log-sigmoidal, y de la de salida fue lineal. Para entrenar la red se empleó un algoritmo de Levenberg-Marquardt con datos históricos de 1999 a 2004. Datos del año 2005 fueron usados para verificar la capacidad predictiva de la red ya entrenada, y evaluar la calidad del aire en tres niveles: buena, moderada y mala. El modelo tuvo eficiencias globales de alrededor del 50%, llegando a 65% para concentraciones elevadas de ozono.

Palabras clave: red neuronal artificial, calidad de aire, algoritmo, pronóstico, ozono

Ozone Concentration Forecast in Guadalajara-Mexico using Artificial Neuronal Networks

Abstract

A forecasting model to predict the maximum ozone concentration in a specific day in the Metropolitan Area of Guadalajara-Mexico was developed. An Artificial Neuronal Network fed with six meteorological variables and three chemicals was used. Nodes in the hidden layer were varying in a number among 12 and 15. The transfer functions were log-sigmoid for the hidden layer and linear for the output layer. For the network training the Levenberg-Marquardt algorithm with historical dates from 1999 to 2004. Data for 2005 were used to evaluate the predictive capabilities of the trained network, evaluating the quality of the air at three levels: good, moderate, and unhealthy. The model presented global efficiencies of around 50%, reaching and 65% for high ozone concentrations.

Keywords: artificial neuronal network, air quality, algorithm, forecasting, ozone

INTRODUCCIÓN

El avance en los modelos matemáticos que describen la emisión, transporte, formación y desaparición de los contaminantes atmosféricos ha motivado un mayor entendimiento de la dinámica de esos contaminantes. Sin embargo, entre más complejos son los modelos, más información requieren para ser aplicados con la suficiente certidumbre de que los resultados obtenidos tendrán algún valor técnico o científico (Russell y Dennis, 2000). Estos modelos determinísticos requieren gran cantidad de información que no siempre es posible conseguir y los datos con los que se cuenta no auguran siempre buenos resultados de la aplicación del modelo (Roth, 1999), o la inversión necesaria en información para obtener resultados confiables puede ser prohibitiva (Pun y Louis, 2000). Existen otros métodos que requieren menor información y que pueden utilizarse para estudiar la contaminación atmosférica de ciertas zonas. Estos métodos, en general, hacen uso de técnicas estadísticas de regresión o ajuste de datos mediante otras técnicas numéricas para establecer las condiciones químicas de la atmósfera, basándose en datos históricos de diversos parámetros medidos rutinariamente. Los objetivos principales de estos métodos incluyen: investigar y estimar tendencias de la calidad del aire, obtener pronósticos ambientales e incrementar el conocimiento científico de los mecanismos que gobiernan la calidad del aire (Thompson *et al.*, 2001).

Entre las técnicas que continúan siendo estudiadas para relacionar la calidad del aire en una zona con ciertos parámetros físicos y químicos medidos, existen tres que han sido utilizadas con mayor frecuencia: i) regresión multivariable (Hubbard y Cobourn, 1998; Comrie y Diem, 1999; Davis y Speckman, 1999; Draxler, 2000; Gardner y Dorling, 2000), ii) Redes Neuronales Artificiales (RNA) (Al-Alawi *et al.*, 2007; Gardner y Dorling, 1999; Guardani *et al.*, 1999; Cannon y Lord, 2000; Hadjiiski y Hopke, 2000; Melas *et al.*, 2000; Pérez *et al.*, 2000; Kolehmainen *et al.*, 2001), y iii) series de tiempo y análisis espectral (Raga y Moyne, 1996; Chen *et al.*, 1998; Milanichus *et al.*, 1998; Salcedo *et al.*, 1999; Sebald *et al.*, 2000). Algunos autores han combinado modelos que describen de una manera simplista el transporte de contaminantes con modelos estadísticos (Cobourn y Hubbard, 1999).

Este artículo presenta un estudio donde se emplearon RNA para obtener modelos que permitan el pronóstico de concentraciones máximas de ozono en la Zona Metropolitana de Guadalajara (ZMG), México; cabe señalar que el estudio se centra en el pronóstico de la concentración de ozono troposférico, ya que en México éste es uno de los 5 principales contaminantes que generan efectos en la salud de la población. Los modelos se estructuraron utilizando mediciones rutinarias de ciertos parámetros meteorológicos y químicos que afectan de forma directa las reacciones en la química atmosférica, para estimar las condiciones de calidad del aire (niveles de ozono). Las RNA se entrenaron con datos históricos medidos en Guadalajara entre 1999 y 2004, y se evaluó su desempeño usando datos del 2005. Los modelos de pronóstico generados cuentan con la capacidad de pronosticar los niveles de contaminación por ozono basados en monitoreos ambientales del día previo y pronósticos de parámetros meteorológicos para el día de interés.

El modelo obtenido realiza pronósticos de concentración de ozono en tres niveles: bajo <0.06ppm, medio 0.06-0.11ppm y alto >0.11ppm, considerando la norma NOM-020-SSA1-1993 (2002) que rige en México los niveles máximos permisibles de concentración de ozono.

METODOLOGÍA

Existen varios tipos de RNA. En el área de las ciencias atmosféricas, las RNA de retropropagación en multicapas han sido las que han demostrado ser las más útiles (Gardner y Dorling, 1998), y son las que se utilizan en este trabajo.

Para el desarrollo de los modelos de pronóstico probados en este trabajo, se empleó el software MATLAB v.6.0 Release 12, y en particular el módulo de herramientas de redes neuronales (Demuth y Beale, 2000) (Neural NetworkToolbox). Todas las redes consistieron de estructuras de tres capas (alimentación, transferencia y salida) con nueve señales de entrada y una de salida. Las nueve señales de entrada corresponden a las nueve variables preseleccionadas, y la única señal de salida es el estimado de concentración máxima de ozono.

Las nueve variables de entrada son: Velocidad del viento promedio (km/h), Temperatura máxima (°C), Humedad Relativa promedio (%), Presión Barométrica promedio (mbar), Precipitación promedio (mm), Radiación Solar máxima (W/m²), concentración de NO₂ máxima en el rango de 15 a 23 hr del día anterior, concentración de NO_x (NO + NO₂) máxima en el rango de 15 a 23 hr del día anterior, y ozono máximo del día anterior. La selección de las variables usadas fue basada 1) en regresiones individuales entre la variable de interés (ozono máximo) y los diversos parámetros (temperatura, humedad, etc.), escogiendo sólo aquellos con índices altos de correlación (R² de 0.3 en adelante) (García, 2003). 2) así como la t – Ratio (t), con la que se obtiene el grado de importancia de cada una de las variables con respecto a la variable dependiente, es decir el ozono (ver tabla 1), este análisis indica, que entre mayor sea el número (considerando el valor absoluto), más necesaria es esta variable (Miller *et al.*, 1992) en el modelo de RNA. Todos los valores de las variables meteorológicas corresponden a los que se presentan en el día en el que se está generando el pronóstico de ozono, por lo que si se pretenden emplear los modelos en tiempo real se requerirán pronósticos de estas variables. Los valores de las variables químicas corresponden a datos del día anterior al de interés.

Tabla 1: Valores de t-Ratio para cada año.

	Modelo 1999	Modelo 2000	Modelo 2001	Modelo 2002	Modelo 2003	Modelo 2004
Constante	-2.344	-.242	-10.033	-.591	-1.295	-.795
VELOCIDAD DEL VIENTO	-.049	-.365	1.859	-.719	-2.617	-3.490
TEMPERATURA	7.847	4.637	2.519	10.951	7.031	5.036
HUMEDAD	1.340	1.583	-3.777	-.437	-5.512	-3.978
PRESION	2.261	.197	10.228	.499	1.358	.806
PRECIPITACION	-.071	-.905	.635	-1.721	-1.692	-1.092
RADIACION SOLAR	-2.959	-.778	3.132	-2.973	-2.699	.316
NO ₂ DEL DIA ANTERIOR	.243	-.875	1.293	.105	1.213	3.654
NO _x DEL DIA ANTERIOR	-1.143	1.223	-.950	-.296	.842	-.152

Se desarrollaron 2 diferentes estructuras de RNA: 9x12x1 (nueve señales de entrada, doce nodos en la capa de transferencia y un nodo en la capa de salida) (R-12) y 9x15x1 (R-15). El entrenamiento se desarrolló usando las tres diferentes funciones de desempeño que son la media del cuadrado del error (MSE), la media del cuadrado del error con regularización (MSEREG) y la suma del cuadrado del error (SSE). Como fuente de información para el proceso de entrenamiento de las RNA se empleó una base de datos con valores diarios máximos de Temperatura (°C), Radiación Solar (W/m²), Dióxidos de Nitrógeno (NO₂) del día anterior y Óxidos de nitrógeno (NO_x) del día anterior, estos dos últimos con un horario de las 15 – 23 hrs; y con valores promedio diarios de la Velocidad del viento (Km/h), Humedad (%), Presión (mbar) y Precipitación (mm) de los años 1999-2004. La base de datos contiene 2065 datos válidos para cada variable. La información provino de dos fuentes principales: Sistema Meteorológico Nacional (SMN), del cual se obtuvieron las variables meteorológicas estos datos fueron obtenidos de la estación de Chapala, que pertenece a las Estaciones de Monitoreo Automáticas (EMA's) y el Instituto Nacional de Ecología (INE) obteniendo de este las variables químicas sobre la estación de Miravalle. Una vez realizado el entrenamiento, se evaluó el desempeño con datos del 2005. Esta segunda base de datos cuenta con 173 valores válidos para cada parámetro, empleándose los datos en las mismas condiciones que la base de datos 1999-2004, correspondiendo a datos diarios de Enero a Junio del 2005.

Las funciones de transferencia seleccionadas para la capa oculta fue una de tipo log-sigmoidal, siendo una función diferenciable, tomando cualquier valor entre menos y más infinito, reduciéndolos a valores de salida entre 0 y 1.

$$f(x) = \left[\frac{1}{1 + e^{-x}} \right] \quad (1)$$

Donde $f(x)$ se refiere a la función de una variable y (x) será cualquier variable de las que se están manejando. Mientras que la función de transferencia para la capa de salida fue una lineal.

RESULTADOS Y DISCUSIÓN

Se presentan dos estructuras, ambas con una capa de alimentación (9 entradas, cada una correspondiente a las mismas variables usadas en el modelo de regresión), una capa oculta (con 12 y 15 nodos) y una capa de salida (con un único nodo) siendo la concentración estimada de O_3 . La función de transferencia seleccionada para la capa oculta fue log-sigmoidal y la de la capa de salida fue lineal. El algoritmo utilizado fue el de Levenberg-Marquardt (Demuth y Beale, 2000) para la retropropagación de los errores y ajuste de los pesos.

El modelo fue calibrado para un periodo de 1999 al 2004, y de acuerdo a la NOM-020-SSA1-1993 (2202), se clasificaron los datos de O_3 real y pronosticado dentro de 3 rangos de concentración, a partir de los cuales se cuantificó el número de datos que caían dentro de dichos rangos (porcentaje de aciertos) para así estimar el grado de certeza que ofrecían los modelos. Esto se estableció bajo el precepto de que resultaría muy difícil pretender obtener de los modelos estimados concentraciones puntuales, así se decidió que los rangos de concentración fueran los siguientes: bajo <0.06 ppm, medio $0.06-0.11$ ppm y alto >0.11 ppm (la notación "ppm" es una medida de concentración, y significa partes por millón, tomando como millón 10^6), este último rango correspondiente al límite permisible de la norma antes mencionada. Sobre estos intervalos, se revisó la cantidad de valores observados que estaban dentro del rango, para contar los valores que los modelos estimaron y que efectivamente correspondieron a las ocasiones que se presentó ese rango en la realidad, siendo ese la cantidad de veces que el modelo hizo una estimación acertada.

En la Tabla 2 se muestra el número y % de veces que se acertó en los rangos de concentración de ozono obtenidos por las redes respecto a los valores observados, en la fase de simulación, donde R-12 y R-15 se refiere a la red de 12 y 15 neuronas en la capa oculta respectivamente.

Tabla 2: Número y % de aciertos de los valores de O_3 estimados, respecto a los valores observados, en la fase de simulación.

	Bajo (<0.06 ppm)	Medio ($0.06-0.11$ ppm)	Alto (>0.11 ppm)
Observado (días)	79	77	17
R-12	41 (52%)	33 (43%)	0 (0%)
R-15	50 (63%)	48 (62%)	2 (12%)

La Figura 1 muestra el resultado del modelo de RNA correspondiente a 12 neuronas en la capa oculta (R-12), con los datos pertenecientes al año 2005 de Enero a Junio. El modelo desarrollado mostró un excelente desempeño para pronosticar concentraciones bajas, un desempeño aceptable para concentraciones medias, pero un nulo o escaso desempeño para concentraciones altas.

El desempeño obtenido es muy similar entre las diferentes redes. Sin embargo, resulta evidente cómo las diferentes redes difícilmente logran detectar concentraciones de ozono superiores a la norma, las cuales son las de importancia para este estudio.

A efecto de poder subsanar el problema de estimación de valores altos, se decidió escalar el valor estimado por la red. Así, el valor estimado final de ozono (\hat{O}_3) se calcula como:

$$\hat{O}_3 = \alpha \hat{O}_{3,m} \quad (2)$$

Donde α representa el factor de escalamiento y \hat{O}_3 es la concentración de ozono (ppm) estimada por el modelo de red neuronal. El valor de α se obtuvo mediante una búsqueda incremental de valores que al aplicar la ecuación se alcanzaba el número de ocasiones que sobrepasaban la norma las concentraciones observadas, sin saber (hasta ese momento) si las ocasiones que la concentración estimada marcó una excedencia correspondieron también a los días en que realmente se

presentaron. Este proceso arrojó un valor de α de 1.2. Así, se multiplicó el ozono estimado por cada red en 1.2, mostrando los resultados en la Tabla 3, de donde se observa que se pierde eficiencia en el rango inferior a costa de una ganancia en los rangos medio y alto (de mayor interés).

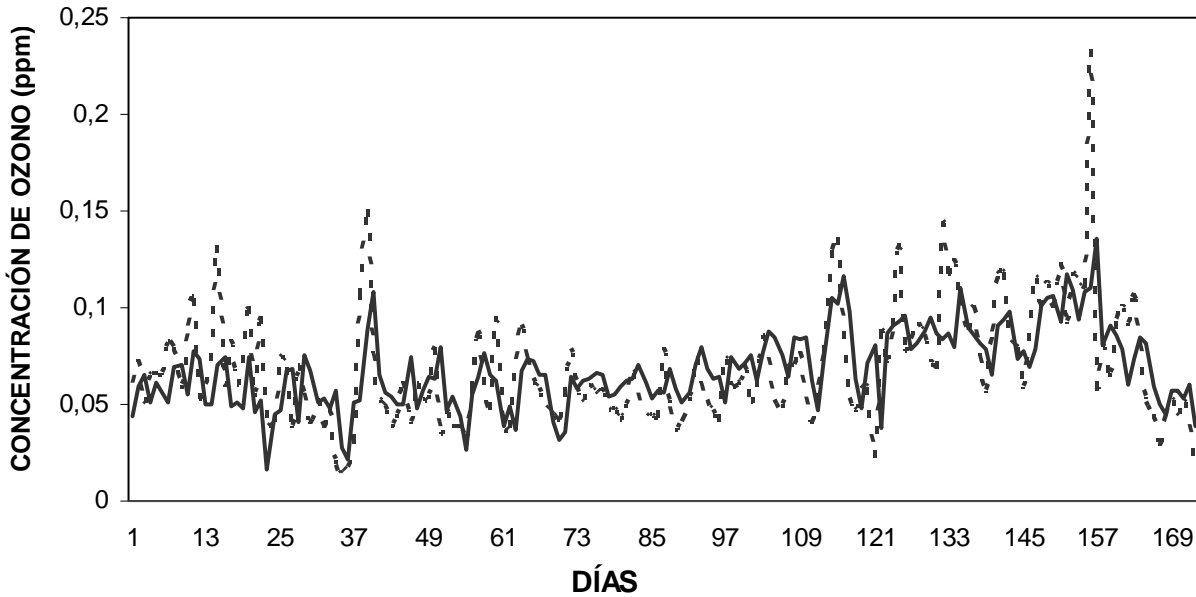


Fig. 1: Ozono Real vs. Ozono Pronóstico (sin factor). [..... Ozono Real — Ozono Pronóstico]

Tabla 3: Número y % de aciertos de los valores de O₃ estimados por las diferentes redes, respecto a los valores observados en la fase de simulación, aplicando el escalamiento correspondiente.

	Bajo (<0.06 ppm)	Medio (0,06-0.11 ppm)	Alto (>0.11 ppm)
Observado (días)	79	77	17
R-12 ($\alpha=1.216$)	18 (23%)	57 (74%)	11 (65%)
R-15 ($\alpha=1.208$)	23 (29%)	57 (74%)	8 (47%)

En la Tabla 3 se observa que el desempeño global de las redes resulta del 50%, y para la detección de concentraciones elevadas de ozono, del 65% con la estructura de R-12 y del 47% con la estructura R-15. La Figura 2 muestra el resultado del modelo al aplicar el factor de escalamiento.

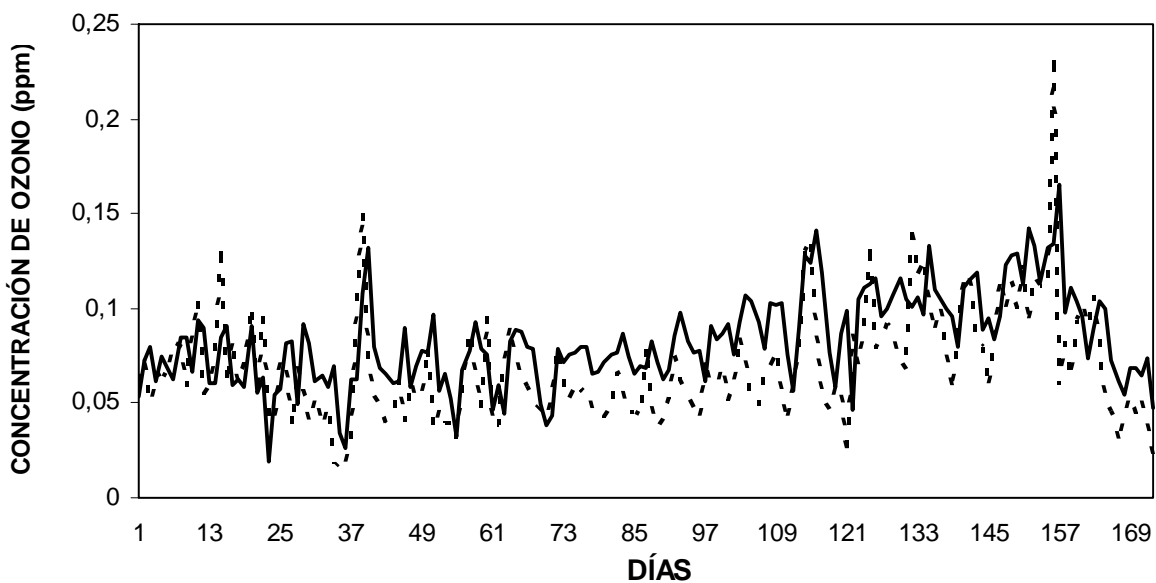


Fig. 2: Ozono Real vs. Ozono Pronóstico (con factor). [..... Ozono Real — Ozono Pronóstico]

CONCLUSIONES

El modelo RNA tiene un desempeño aceptable en cuanto al pronóstico de O₃ en el rango bajo y medio, sin embargo el objetivo de este estudio fue pronosticar concentraciones elevadas del contaminante en cuestión, por lo que fue necesario el uso del factor de escalamiento.

Con los resultados obtenidos por el modelo y las variables seleccionadas, se concluyó que para la ZMG las variables que tienen una mayor relevancia en la formación y dispersión de O₃ son: las variables meteorológicas que reducen la concentración de O₃ considerablemente en la troposfera como la velocidad del viento promedio, precipitación, y por otro lado las variables químicas meteorológicas que son importantes son: temperatura máxima, radiación solar máxima, O₃ del día anterior, y óxidos de nitrógeno (NO_x y NO₂) (Gómez *et al.*, 2006) ya que estas intervienen en el ciclo fotoquímico de formación de ozono (Wark y Warner, 2005) y estadísticas (R² y t - Ratio); aunando las características propias de la cuenca de Guadalajara, Jalisco (dirección de los vientos reinantes, topografía y ubicación de zonas industriales).

El modelo obtenido puede pronosticar concentraciones máximas de O₃ en ppm, para la ZMG, utilizando variables meteorológicas y químicas. Los modelos tuvieron eficiencias globales alrededor del 50%, y para la detección de concentraciones elevadas, del 65%.

Los modelos generados se caracterizan por una fácil implantación, requerimientos tecnológicos moderados y estructuras sencillas y de fácil entendimiento, lo que les confiere costos de operación mínimos. Estos modelos pueden ayudar a alertar a la población en escenarios no deseados de calidad de aire, tomando medidas precautorias para salvaguardar la salud de la comunidad.

REFERENCIAS

Al-Alawi, S.M., S.A. Abdul-Wahab y C.S. Bakheit; "Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone". *Environmental Modelling and Software*: 23 (4), 396-403, 2007

Cannon A.J. y E.R. Lord; "Forecasting summertime surface-level ozone concentrations in the lower fraser valley of British Columbia", *J. Air & Waste Manage. Assoc.*: 50, 322-339 (2000).

Chen J.-L., S. Islam y P. Biswas; "Nonlinear dynamics of hourly ozone concentrations: Nonparametric short term prediction", *Atmospheric Environment*: 32, 1839-1848 (1998).

Cobourn W.G. y M.C. Hubbard; "An enhanced ozone forecasting model using air mass trajectory analysis", *Atmospheric Environment*: 33, 4663-4674 (1999).

Comrie A.C. y J.E. Diem; "Climatology and forecast modeling of ambient carbon monoxide in Phoenix, Arizona", *Atmospheric Environment*: 33, 5023-5036 (1999).

Davis J.M. y P. Speckman; "A model for predicting maximum and 8 h average ozone in Houston", *Atmospheric Environment*: 33, 2487-2500 (1999).

Demuth H. y M. Beale; "Neural network toolbox, for use with Matlab", 6th Ed. The Mathworks, Inc. (2000).

Draxler R.R.; "Meteorological factors of ozone predictability at Houston, Texas", *J. Air & Waste Manage. Assoc.*: 50, 259-271 (2000).

García I.; "Aplicación de modelos semi-empíricos para el análisis y pronóstico de la calidad del aire en el Área Metropolitana de Monterrey, N.L.", Tesis de Maestría en Ciencias, Ingeniería Ambiental, ITESM, Monterrey, México (2003).

- Gardner M.W. y S.R. Dorling; "Artificial neural networks (the multilayer perceptron) – A review of applications in the Atmospheric Sciences," *Atmospheric Environment*. 32, 2627-2636 (1998).
- Gardner M.W. y S.R. Dorling; "Neural Network modeling and prediction of hourly NO_x and NO₂ concentrations in urban air in London", *Atmospheric Environment*. 33, 709-719 (1999).
- Gardner M.W. y S.R. Dorling; "Statistical surface ozone models: an improved methodology to account for nonlinear behaviour", *Atmospheric Environment*. 34, 21-34 (2000).
- Gomez, J. y otros cinco autores; "Neural networks for analysing the relevance of input variables in the prediction of tropospheric ozone concentration", *Atmospheric Environment*. 40 (32), 6173-6180, 2006.
- Guardani R., C.A.O. Nascimento y M.L.G. Guardani; "Study of atmospheric ozone formation by means of a neural network-based model," *J. Air & Waste Manage. Assoc.*: 49:316-323 (1999).
- Hadjiiski L. y P. Hopke; "Application of artificial neural networks to modeling and prediction of ambient ozone concentrations", *J. Air & Waste Manage. Assoc.*: 50, 894-901 (2000).
- Hubbard M. y W.G. Cobourn; "Development of a regression model to forecast ground-level ozone concentration in Louisville, KY", *Atmospheric Environment*. 32, 2637-2647 (1998).
- Kolehmainen M., H. Martikainen y J. Ruuskanen; "Neural networks and periodic components used in air quality forecasting", *Atmospheric Environment*. 35, 815-825 (2001).
- Melas D., I. Kioutsioukis e I.C. Ziomas; "Neural network model for predicting peak photochemical pollutant levels", *J. Air & Waste Manage. Assoc.*: 50, 495-501 (2000).
- Milanchus M.L., S.T. Rao e I.G. Zurbenko; "Evaluating the effectiveness of ozone management efforts in the presence of meteorological variability," *J. Air & Waste Manage. Assoc.*: 48, 201-215 (1998).
- Miller, I.R., J.E. Freund y R. Jonson; "Probabilidad y Estadística para Ingenieros", 4ta Ed. Prentice-Hall Hispanoamericana, pp. 356 – 358 (1992).
- NOM-020-SSA1-1993, Norma Oficial Mexicana sobre Criterio para evaluar el valor límite permisible para la concentración de ozono (O₃) de la calidad del aire ambiente. México, D. F. (2002).
- Pérez P., A. Trier y J. Reyes; "Prediction of PM_{2.5} concentrations several hours in advance using neural networks in Santiago, Chile", *Atmospheric Environment*. 34, 1189-1196 (2000).
- Pun B.K., J.F. Louis; "Ozone formation in California's San Joaquin Valley: A critical assessment of modeling and data needs", *J. Air & Waste Manage. Assoc.*: 50, 961-971 (2000).
- Raga G.B. y L. Le Moyne; "On the nature of air pollution dynamics in Mexico City – I. Nonlinear analysis", *Atmospheric Environment*. 30, 3987-3993 (1996).
- Roth P.M.; "A qualitative approach to evaluating the anticipated reliability of a photochemical air quality simulation model for a selected application", *J. Air & Waste Manage. Assoc.*: 49, 1050-1059 (1999).
- Russell A. y R. Dennis; "NARSTO critical review of photochemical models and modeling", *Atmospheric Environment*. 34, 2283-2324 (2000).
- Salcedo R.L.R., M.C.M. Alvim Ferraz, C.A. Alves y F.G. Martins; "Time-series analysis of air pollution data", *Atmospheric Environment*. 33, 2361-2372 (1999).

Sebald L., R. Treffeisen, E. Reimery T. Hies; "Spectral analysis of air pollutants. Part 2: ozone time series", *Atmospheric Environment*. 34, 3503-3509 (2000).

Thompson M.L. y otros cuatro autores; "A review of statistical methods for the meteorological adjustment of tropospheric ozone", *Atmospheric Environment*. 35, 617-630 (2001).

Wark, K. y C.F. Warner; "Contaminación del aire: Origen y control", México D. F.: Limusa: Noriega, pp. 510 – 530 (2005).