

# EVALUACIÓN DE LA CALIDAD DEL AIRE MEDIANTE LA ESTIMACIÓN Y COMPARACIÓN DE CURVAS TÍPICAS NO PARAMÉTRICAS

## EVALUATION OF AIR QUALITY THROUGH ESTIMATION AND COMPARISON OF TYPICAL NON PARAMETRIC CURVES

DANIEL AYALA O.<sup>1</sup>, CÉSAR A. OJEDA E.<sup>1</sup>, JAVIER OLAYA O.<sup>1</sup>, LUZ A. PEREIRA H.<sup>1\*</sup>

Recibido 21-06-12, aceptado 17-07-12, versión final 26-07-12.

**RESUMEN:** Se estiman curvas típicas diarias para los contaminantes monóxido de carbono, dióxido de nitrógeno y ozono en los años 2003, 2004 y 2005 en la ciudad de Cali, mediante una técnica no paramétrica de suavización. Se estudia la evolución de las curvas comparándolas estadísticamente por medio de un análisis de covarianza no paramétrico, con el fin de identificar patrones y tendencias. Los resultados muestran que existe una reducción estadísticamente significativa de los niveles típicos de contaminación a través de los años. La información utilizada proviene de las estaciones de la red de monitoreo de calidad del aire en la ciudad.

**PALABRAS CLAVE:** Estadística no paramétrica, Métodos estadísticos, Métodos gráficos, Contaminación atmosférica.

**ABSTRACT:** We estimate typical daily curves for the pollutants carbon monoxide, nitrogen dioxide and ozone in the years 2003, 2004 and 2005 in Cali city, by means of a nonparametric smoothing technique. The curves evolution is studied through statistical comparison using a nonparametric analysis of covariance, hoping to find out patterns and trends. The results show the existence of a statistically significant reduction in the typical pollution levels throughout the years. The data comes from the stations of the Air Quality Monitoring Network.

**KEYWORDS:** Nonparametric statistics, Statistical methods, Graphical methods, Air pollution.

## 1 INTRODUCCIÓN

Publicaciones en el campo de la contaminación atmosférica evidencian los efectos nocivos que tienen sobre la salud de los seres humanos la concentración de los gases denominados contaminantes criterio (Folinsbee, 1993; Rosales et al., 2001; Glynn y Heynke, 1999; DAGMA, 2003). Dentro de este grupo se encuentran el monóxido de carbono (CO), el dióxido de nitrógeno (NO<sub>2</sub>) y el ozono (O<sub>3</sub>) (Suh et al., 2000). Folinsbee (1993) expone que la presencia de estos tres contaminantes a ciertas concentraciones genera efectos adversos en la salud humana, relacionados generalmente con afecciones de las vías respiratorias.

La ciudad de Santiago de Cali, como ciudad capital de continuo crecimiento, no es ajena a esta

<sup>1</sup> Escuela de Estadística, Universidad del Valle, Cali, Colombia, Calle 13 No 100-00

\* luz.pereira@correounivalle.edu.co

situación y se enfrenta a la problemática de salud pública que se genera alrededor de la contaminación del aire, por lo cual es de suma importancia para las autoridades ambientales locales contar con elementos de juicio que les permitan promover actividades de gestión para el control de la contaminación.

Para el estudio de esta problemática, es común la utilización de redes compuestas por estaciones fijas de monitoreo de la calidad del aire en las grandes ciudades alrededor del mundo. Por ejemplo, la zona metropolitana del Valle de México, cuenta con 47 estaciones operadas por el sistema de monitoreo atmosférico de la Ciudad de México; Beijing (China), cuenta con 28 estaciones automáticas de monitoreo de la calidad del aire; Londres (Inglaterra) estableció en 1993 la red de monitoreo ambiental (Gaitan et al., 2007). De la misma forma, Colombia cuenta con redes ubicadas en distintas ciudades, por ejemplo, la red de monitoreo de Bogotá (RMCAB) y la Red de Monitoreo de Calidad del Aire y Meteorología de Cali (RMCA) (DAGMA, 2003).

En este estudio se hace uso de los registros de la concentración en el aire de los tres gases mencionados: CO, NO<sub>2</sub> y O<sub>3</sub>, provenientes de la RMCA, con el objetivo de modelar el nivel de concentración típico diario en diferentes años. Se estimarán curvas típicas diarias de los contaminantes y posteriormente se compararán mediante un análisis de covarianza no paramétrico. Al menos un estudio previo con datos organizados de manera similar y cuyo propósito fue construir curvas típicas para efectos de pronósticos, demuestra que los modelos de series de tiempo no son muy adecuados y que en cambio los modelos de regresión no paramétrica resultan ser una mejor opción (Barrientos et al., 2007).

## 2 ESTIMACIÓN Y COMPARACIÓN DE CURVAS TÍPICAS DIARIAS

En el análisis de regresión no paramétrica se dispone de observaciones de las variables  $X$  y  $Y$   $((x_i, y_i), i = 1, 2, \dots, n)$ , acerca de las cuales se asume la existencia de cierta relación funcional  $m$  que no se ajusta a una forma paramétrica y que puede expresarse a través del modelo:

$$y_i = m(x_i) + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (1)$$

donde  $\epsilon_i$  son errores aleatorios con  $E(\epsilon_i) = 0$  y  $Var(\epsilon_i) = \sigma^2 < \infty$ .

Suponiendo que  $m$  es una función suave, es posible utilizar algún tipo de promedio de las respuestas  $y_i$  en los puntos  $x_i$  cercanos a  $x$  para obtener una estimación de  $m(x)$ . En la literatura sobre el tema están disponibles varios estimadores de  $m$  (Eubank, 1999), uno de los cuales, el estimador lineal local propuesto por Cleveland (1979), se utilizó en este proyecto, dada su eficiencia en la estimación de la función de regresión  $m$  en los valores extremos de  $X$ .

Para estimar el error estándar ( $se$ ) de  $\hat{m}(x)$ , y construir bandas de variabilidad similares a los intervalos de confianza en el ámbito paramétrico (Azzalini y Bowman, 1997), se utilizará el estimador de varianza propuesto por Rice (1984), para el cual la varianza de los  $\epsilon_i$  en el modelo (1) se basa en las diferencias sucesivas de la variable respuesta.

Una vez realizadas las estimaciones de las curvas típicas, se utiliza el análisis de covarianza no paramétrico para la comparación de estas. Esta técnica compara los valores de una variable respuesta a través de varios grupos en la presencia del efecto de una covariable. La forma general del modelo es:

$$y_{ij} = m_i(x_{ij}) + \epsilon_{ij}, \quad (2)$$

para  $i = 1, 2, \dots, p$  y  $j = 1, 2, \dots, n_i$ , en donde  $p$  es el número de grupos,  $n_i$  es el número de observaciones dentro del grupo  $j$  y los  $\epsilon_{ij}$  son errores aleatorios independientes  $N(0, \sigma^2)$ . El interés consiste en identificar si las funciones de regresión son diferentes a través de los grupos.

Las hipótesis se formulan como:  $H_0 : y_{ij} = m(x_{ij}) + \epsilon_{ij}$  (existe una única función de regresión  $m$ ) vs.  $H_1 : y_{ij} = m_i(x_{ij}) + \epsilon_{ij}$  (al menos un modelo  $m_i$  es diferente de  $m$ ). La distribución del estadístico de prueba ( $F$ ) en este caso se aproxima a través de una Chi-cuadrado con grados de libertad que dependen del segundo y tercer momento de  $F$  (Azzalini y Bowman, 1997).

Un procedimiento gráfico para evaluar la igualdad de dos curvas es el de banda de referencia, el cual crea en cada punto de interés  $x$  una banda centrada en el promedio de estas y se extiende a  $\pm 2 se\{\hat{m}_1(x) - \hat{m}_2(x)\}$  (Azzalini y Bowman, 1997). Por lo tanto, la banda es colocada alrededor del modelo de referencia  $y_{ij} = m(x_{ij}) + \epsilon_{ij}$  para indicar en dónde las curvas suavizadas de cada grupo están bajo la hipótesis nula de igualdad. Si las curvas suavizadas se encuentran por fuera de esta banda, se rechaza la hipótesis nula.

### 3 METODOLOGÍA

Los datos están conformados por los promedios horarios de la concentración de los contaminantes CO, NO<sub>2</sub> y O<sub>3</sub>, en los años 2003, 2004 y 2005. Estos promedios proceden de una muestra recolectada continuamente a lo largo de una hora en cada estación.

Según Morales y Montoya (2004), existe una forma o estructura general que puede definir un comportamiento típico diario para cada contaminante durante un año específico, del cual no se tiene conocimiento exacto, pero se puede construir una estimación a partir de sus observaciones mediante la suavización kernel. El modelo estadístico que describe este comportamiento puede ser expresado como (1), donde  $y$  son los niveles de concentración registrados de cada contaminante,  $x$  es la hora del día asociada a cada registro y  $m$  es la curva típica diaria de concentración atmosférica que se desea estimar.

Para la estimación de curvas típicas se tuvo en cuenta la estación de la que provienen los registros, el año del cual proceden y el tipo de día en el que fueron medidos (ordinarios y fin de semana). Para realizar la suavización kernel se utilizó el estimador lineal local. Aquí, la función ponderadora fue un kernel Gaussiano y el ancho de banda fue estimado por el método de validación cruzada (Stone, 1974). Una vez estimadas las curvas, se grafican junto con la norma que indica el nivel máximo permisible de contaminación en el año 2005 (IDEAM, 2005).

Para el proceso de comparación de curvas típicas diarias se utilizó la técnica de análisis de covarianza no paramétrico bajo el modelo (2). Esta técnica compara los niveles típicos de cada contaminante

( $y$ ) a través de los años estudiados ( $i$ ) en la presencia del efecto de una covariable  $x$ , que en este caso es la hora del día.

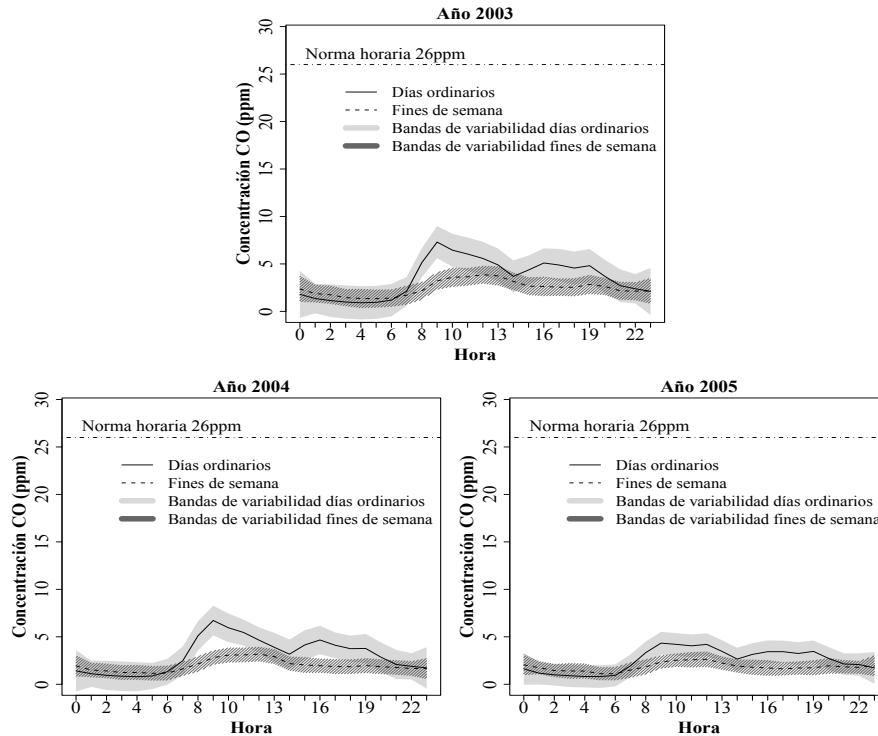


Figura 1: Curvas típicas diarias estimadas, bandas de variabilidad y norma de la concentración de CO en la estación CALLE15 para los tipos de día ordinario y fin de semana en los años 2003, 2004 y 2005.

## 4 RESULTADOS

Del ajuste de las curvas típicas diarias, se observa que el CO (estación CALLE15) en los años considerados presenta la más alta concentración entre las 8 y 9 a.m. para los días ordinarios, mostrando una reducción de los niveles en el transcurso de la tarde y un posterior pico de menor magnitud alrededor de las 7 p.m.. Para los días clasificados como fin de semana el comportamiento típico es bastante similar, con algunas diferencias en los niveles y las horas en las que se presentan los picos (Figura 1). También se observa que los niveles de concentración típicos se encuentran muy por debajo de la norma horaria (IDEAM, 2005), y al parecer existe una disminución a través de los años.

Para el dióxido de nitrógeno, en cada uno de los años, se presenta en general un pico de concentración a las 9 a.m. para ambos tipos de día, seguido por un descenso de los niveles hasta las 5 p.m., finalizando el día con un nuevo pico alrededor de las 9 p.m. Sin embargo, en la estación CALLE15, el comportamiento típico tiene forma acampanada, presentando el pico de concentración hacia el medio día. Cabe resaltar que para este contaminante, al igual que para el CO, las curvas típicas se mantienen por debajo de la norma horaria establecida (IDEAM, 2005). El comportamiento del ozono en general muestra una forma acampanada, en la cual el pico se encuentra alrededor de las

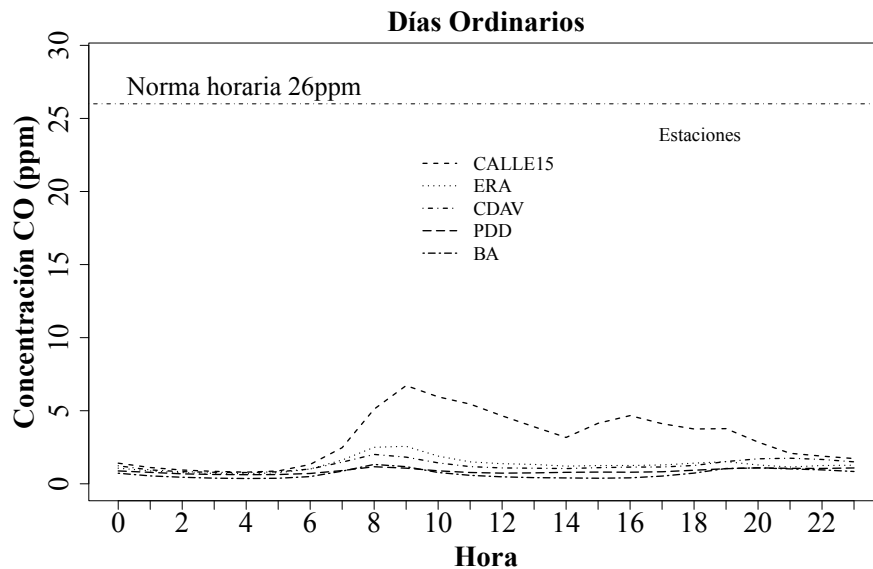


Figura 2: Curvas típicas diarias estimadas y norma para los contaminantes en las diferentes estaciones para el año 2004 (días ordinarios).

2 p.m. tanto para los días ordinarios, como para los fines de semana. Aquí, las curvas típicas son inferiores a la norma (IDEAM, 2005), pero su proximidad es muy notable en todas las estaciones. Al observar las curvas de los contaminantes a través de las estaciones de monitoreo (Figura 2), se observa que para el monóxido de carbono, la zona de la ciudad que presenta la situación más grave de contaminación es la zona centro, alcanzando concentraciones en la estación CALLE15 de más de 6 ppm en horas de la mañana en los días ordinarios, y de más de 3 ppm en los días fin de semana. Con respecto al dióxido de nitrógeno, nuevamente la zona centro presenta la situación más grave de contaminación, alcanzando concentraciones promedio de 60 ppb al medio día en los días ordinarios y de 40 ppb en los días fin de semana. Los niveles del ozono troposférico en la ciudad de Cali se mantienen aproximadamente constantes a través de las distintas zonas (urbana y rural), alcanzando concentraciones de 50 ppb para los dos tipos de día.

En cuanto a la comparación de las curvas típicas, se presenta el caso del  $O_3$  en la estación de monitoreo CDAV, donde se aprecia que existen diferencias significativas entre los años 2003 y 2005 según el tipo de día (Figura 3). En este caso las curvas estimadas para cada uno de los años se encuentran por encima y por debajo de la banda de referencia en todas las horas del día, evidenciando una disminución de los niveles típicos de concentración, el cual es soportado por los valores- $p$  asociados a las pruebas de comparación.

## 5 CONCLUSIONES

Según los resultados de este trabajo, las curvas típicas estimadas para los diferentes contaminantes se encuentran por debajo de los límites establecidos por la regulación ambiental local (IDEAM, 2005) en las distintas estaciones de monitoreo.

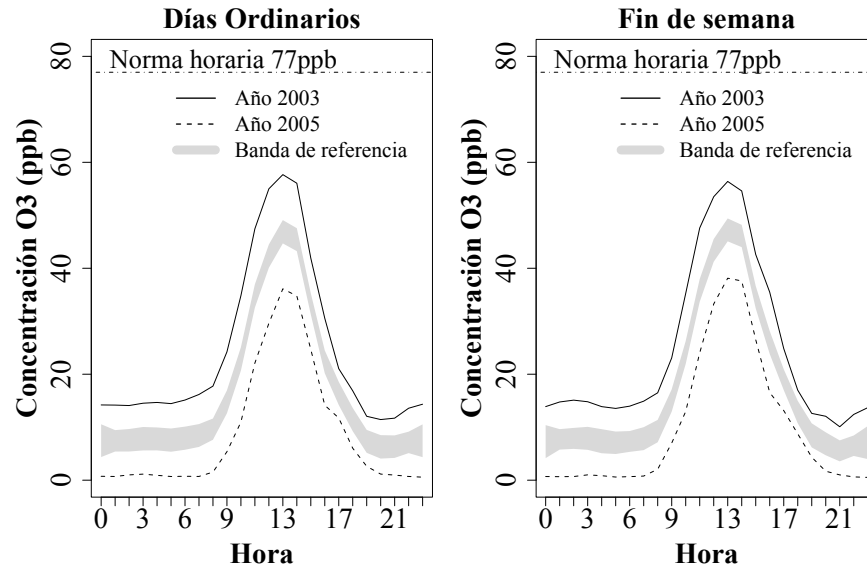


Figura 3: Comparación de las curvas típicas diarias y norma de la concentración de O<sub>3</sub> en la estación CDAV (Centro Diagnóstico Automotor del Valle) entre los años 2003 y 2005 según el tipo de día. Valor-*p* días ordinarios 0.0036, y 0.0037 fin de semana

Al observar las curvas típicas diarias para el monóxido de carbono y para el dióxido de nitrógeno (Figura 3), se encuentra que la zona centro de la ciudad es la que tiene mayores niveles de concentración de estos gases, sin embargo para el NO<sub>2</sub> se evidencia unos niveles medios sobresalientes en la estación ERA y CDAV, que corresponden la zona centro y norte de la ciudad. Los niveles del ozono troposférico en la ciudad de Cali se mantienen aproximadamente constantes a través de las distintas zonas (urbana y rural).

Los resultados del análisis de covarianza para la comparación de las curvas muestran que parece existir una reducción significativa a través del tiempo en los niveles de contaminación atmosférica por CO, NO<sub>2</sub> y O<sub>3</sub> en la ciudad de Cali. Este resultado soporta lo reportado por el IDEAM, en términos de que la tendencia nacional para NO<sub>2</sub>, O<sub>3</sub> y CO es a disminuir sus concentraciones para los años 2003 al 2006 y que en la ciudad de Cali esta tendencia se conserva. De igual manera cabe mencionar que para la fecha el DAGMA, había venido realizando programas de monitoreo y evaluación de la calidad del aire, control de fuentes fijas y móviles, así como también seguimiento de declaraciones ambientales, con el fin de ejercer el desarrollo de funciones de control y seguimiento de la contaminación atmosférica (IDEAM, 2007).

Como recomendaciones para futuras investigaciones, sería altamente informativo trabajar la metodología propuesta incluyendo factores tales como temperatura, radiación solar, dirección e intensidad del viento, entre otros, que ayuden a explicar de una forma más completa los fenómenos de contaminación atmosférica. Adicionalmente, sería interesante contar con un sistema de seguimiento en tiempo real de la contaminación atmosférica en la ciudad, en el cual se puedan

utilizar las herramientas implementadas en el presente estudio.

## Referencias

- Azzalini, A.; Bowman, A. (1997), *Applied smoothing techniques for data analysis: The Kernel approach with S-Plus illustrations*. New York: Oxford, Oxford University Press Inc.
- Barrientos, A.; González, V.; Olaya, J. (2007), Un modelo spline para el pronóstico de la demanda de energía eléctrica. *Revista Colombiana de Estadística*, **31**, 187-202.
- Cleveland, W. (1979), Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829-836.
- Departamento Administrativo de Gestión del Medio Ambiente (DAGMA). (2003), *Programa Integral para el Mejoramiento de la Calidad del Aire*. Boletín 1. Santiago de Cali: DAGMA, 3.
- Documento soporte norma de calidad del aire [en línea](2005), Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). [Consultada en mayo de 2010]. Disponible: [www.minambiente.gov.co/documentos/1748\\_Resumen\\_ejecutivo.pdf](http://www.minambiente.gov.co/documentos/1748_Resumen_ejecutivo.pdf)
- Eubank, R. (1999), *Nonparametric regression and spline smoothing 2a. Ed.*, New York: Marcel Dekker.
- Folinsbee, L. (1993), Human health effects of air pollution. *Environmental Health Perspectives*, **100**, 45-56.
- Gaitán, M.; Cancino, J.; Behrentz, E. (2007), Análisis del estado de la calidad del aire en Bogotá. *Revista de Ingeniería de la Universidad de los Andes*, **26**, 81-92.
- Glynn, J.; Heinke, G. (1999), *Ingeniería Ambiental 2a. Ed.*. México: Prentice Hall Hispanoamericana.
- Informe Anual sobre el Estado del Medio Ambiente y los Recursos Naturales Renovables en Colombia: Calidad del Aire. (2007) [en línea], Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM).[Consultada en mayo de 2010]. Disponible: <https://documentacion.ideam.gov.co/openbiblio/Bvirtual/publicaciones/publicaciones2.html>
- Morales, A. ; Montoya, M. (2004), *Estimación de curvas típicas diarias de los contaminantes CO, NO<sub>2</sub> y SO<sub>2</sub> y cálculo del índice de calidad del aire para Santiago de Cali*. Trabajo de grado. Universidad del Valle. Facultad de Ingenierías. Escuela de Ingeniería Industrial y Estadística. Santiago de Cali.
- Rice, J. (1984), Bandwidth choice for nonparametric regression, *The Annals of Statistics*, **4**, 1215-1230.
- Rosales, J.; Torres, V.; Olaiz, G; Borja, V. (2001), Los efectos agudos de la contaminación del aire en la salud de la población: evidencias de estudios epidemiológicos. *Revista Salud Pública de México*, **43**, 544-555.

- Stone, M. (1974), Cross-validators choice and assessment of statistical predictions. *J. Roy. Statist. Soc., Ser.B (Methodological)*, **36**, 111-147.
- Suh, H.; Bahadori, T.; Villarino, J. (2000), Criteria air pollutants and toxic air pollutants. *Environmental Health Perspectives*, **108**, 625-633.