

# MODELACIÓN ESTADÍSTICA DE LOS APORTES DE LAS VÍAS COMO FUENTES DE EMISION A LAS PARTÍCULAS TOTALES EN SUSPENSION SEGÚN MODELO MCF, ZONA CENTRO DE MEDELLIN-ANTIOQUIA-COLOMBIA, 2004

## STATISTICAL MODELING OF ROAD CONTRIBUTION AS EMISSION SOURCES TO TOTAL SUSPENDED PARTICLES (TSP) UNDER MCF MODEL DOWNTOWN MEDELLIN-ANTIOQUIA -COLOMBIA, 2004

MIRYAM GÓMEZ

*Directora de investigación y Posgrados, Politécnico Colombiano Jaime Isaza Cadavid, investigaf@elpoli.edu.co*

JULIO SALDARRIAGA

*Ingeniero Sanitario, MSc, Grupo de Ingeniería y Gestión Ambiental, Universidad de Antioquia*

MAURICIO CORREA

*Ingeniero Sanitario, MSc, Grupo de Ingeniería y Gestión Ambiental, Universidad de Antioquia*

ENRIQUE POSADA

*Ingeniero Mecánico, MSc, Indisa S.A.*

FRANCISCO JAVIER CASTRILLÓN M

*Matemático y Estadístico, MSc, Universidad Nacional de Colombia, Sede Medellín*

Recibido para revisar agosto 03 de 2006. Aceptado noviembre 14 de 2006. Versión final febrero 15 de 2007

**RESUMEN:** Entre las fuentes de emisión a la contaminación del aire en el Valle de Aburrá Zona Centro, se encuentran: canchas de arena, fuentes biológicas, calderas de carbón, vías y construcciones, entre otras. La forma funcional de los aportes de las Vías a las partículas totales en suspensión generados por el modelo fuente-receptor, MCF, Modelo de Correlación de Fuentes, creado para la estimación en porcentaje del grado de contribución de una fuente de contaminación específica, es aproximadamente la de una distribución Gamma. Aquí se utiliza la prueba de bondad de ajuste Chi-cuadrado para modelar estadísticamente los aportes de las Vías como fuentes de emisión a las partículas totales en suspensión durante un periodo determinado de tiempo. Esta prueba permite estimar también los parámetros de la distribución mediante el método de máxima verosimilitud. Como algoritmo de convergencia se ha utilizado el Algoritmo EM (Estimation Maximization Algorithm). La estimación de la media de los factores de contribución de las Vías según el modelo Fuente-Receptor MCF es inmediata a partir de estas estimaciones y valida el factor de contribución de las Vías a la contaminación atmosférica en la zona de estudio.

**PALABRAS CLAVE:** Estimación por máxima verosimilitud, Algoritmo.

**ABSTRACT:** Sand fields, constructions, carbon boilers, roads, and biologic sources are air-contaminant-constituent factors in Down Town Valle de Aburrá, among others. The distribution of road contribution data to total suspended particles according to the source receptor model MCF, source correlation modeling, is nearly a Gamma distribution. Chi-square goodness of fit is used to model statistically. This test for goodness of fit also allows to estimate the parameters of the distribution utilizing maximum likelihood method. As a convergency criteria, the Estimation Maximization Algorithm is used. The mean of road contribution data to total suspended particles according to the source receptor model MCF, is straightforward and validates the road contribution factor to the atmospheric pollution of the zone under study.

**KEY WORDS:** Maximum Likelihood Estimation, Expectation Maximization Algorithm, Pollutant Sources, Particulate Matter.

## 1. INTRODUCCIÓN

En las últimas décadas la problemática de la contaminación del aire ha sido de primordial interés, a nivel local, regional y global, desde el punto de vista ambiental y de la salud pública. Por lo tanto, el muestreo de la calidad del aire hoy en día, hace parte de una de las estrategias de desarrollo del plan de competitividad del Departamento y la Nación.

De acuerdo con los resultados obtenidos por la red de monitoreo de la calidad del aire del Valle de Aburrá (REDAIRE Boletines 2000 – 2005) y por otros estudios realizados en el medio (Echeverry et al., 2001), (Bedoya, J., 2002), (Contraloría General de Medellín, 2001), y (Programa de Protección y Control de la Calidad del Aire, 1998), se puede concluir que con frecuencia en diferentes puntos de la ciudad de Medellín se están presentando episodios de contaminación que exceden los estándares nacionales de calidad del aire para partículas suspendidas totales (PST) y  $PM_{10}$  (partículas con diámetro aerodinámico menor a  $10\mu$ ). Por consiguiente, conocer la composición y la contribución de las diferentes fuentes responsables de la presencia de material particulado en la atmósfera es de mucho interés en la planificación de estrategias de control para alcanzar los estándares de calidad del aire en la ciudad de Medellín y demás ciudades del país.

Con base en lo anterior, entre las diferentes fuentes estudiadas, las vías se constituyen en fuentes de difícil evaluación y control, dadas las variables que afectan la emisión como lo son el desprendimiento de las partículas por el desgaste de las vías mezcladas con partículas provenientes de la combustión y componentes del suelo transportadas por la acción del viento.

Para proveer la información sobre la composición química tanto de las partículas emitidas por las fuentes como de las muestras del PST recolectadas sobre filtros de fibra de

vidrio, se ensayaron, adaptaron y calibraron metodologías de análisis químico reportadas por la literatura con la técnica de absorción atómica (Harrison, *et al.* 1986, Stevens, *et al.* 1974, EPA. 2004).

Esta información se empleó para la alimentación del modelo receptor MCF, diseñado por los grupos de investigación GIGA y GHYGAM de la Universidad de Antioquia y el Politécnico Jaime Isaza Cadavid. El modelo utilizado en este trabajo fue desarrollado en Visual Basic con bases de datos en Microsoft Access y salida de datos en Excel para determinar la influencia de un conjunto de fuentes contaminantes sobre una zona sujeta al efecto de dichas fuentes. Los factores de influencia se determinan con base en una serie de especies químicas, cuyas concentraciones se conocen tanto en las fuentes de emisión como en el aire de la zona de estudio.

Para el caso en particular -modulación de las vías de la zona centro de la ciudad de Medellín- las concentraciones modeladas se comparan con las concentraciones reales (Ec. 1.1) y se seleccionan aquellas soluciones (combinaciones de factores de influencia) que den errores combinados menores a un valor prefijado. A juicio del usuario se puede escoger una de las soluciones que tengan errores combinados menores a los estipulados o se pueden promediar estas soluciones para obtener una única solución.

$$\%E = \frac{\% \text{ valor real} - \% \text{ valor modelado}}{\% \text{ valor real}} * 100$$

(Ec. 1.1)

Una serie de ecuaciones formuladas y resueltas para una zona centro del Valle de Aburrá determina el impacto de las vías y de las demás fuentes, sobre ésta, partiendo de la suposición de que las masas de partículas se conservan sin cambio químico desde el momento en el cual son emitidas hasta que son capturadas en el filtro de muestreo. La ecuación del balance de

masa planteada para una especie química en particular, en términos de porcentaje es la siguiente:

$$\rho_{kj} = \sum_{i=1}^n F_{ik} P_{ij} \quad (\text{Ec. 1.2})$$

En la cual  $\rho_{kj}$  es la composición porcentual del componente  $j$  en la muestra recolectada en la zona  $k$ ;  $P_{ij}$  es la composición porcentual del componente  $j$  en una muestra de la fuente  $i$ ; y  $F_{ik}$  es el factor de influencia de la fuente  $i$  sobre la contaminación existente en la zona  $k$ , equivalente a la razón de contribución de la masa de la fuente  $i$  para el total de la masa colectada en la zona  $k$ .

En este artículo se hace una prueba de bondad de ajuste de los datos correspondientes a los aportes de las Vías como fuentes de emisión a las partículas totales en suspensión a una distribución gamma, se da la estimación de los parámetros de esta vía máxima verosimilitud y se da también la estimación de la media poblacional de estos aportes.

## 2. FUNCIÓN DE VEROSIMILITUD

La función de verosimilitud se define como una función de probabilidad basada en una muestra de datos; el procedimiento de máxima verosimilitud consiste en encontrar los puntos críticos que maximiza esta función.

Asumamos que  $X_1, X_2, \dots, X_n$  es una muestra aleatoria de una distribución  $f(x; \theta)$  en la que  $\theta$  es el vector de parámetros. La función de verosimilitud de la muestra es:

$$\begin{aligned} L(X_1, X_2, \dots, X_n; \theta) &= L(\theta | X_1, X_2, \dots, X_n) \\ &= f(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (\text{Ec. 2.1}) \end{aligned}$$

La cual es una función del parámetro desconocido  $\theta$ . Asumimos por la naturaleza de

los datos (Fig. 1), que hay  $k$  ceros solamente, lo que conoce como *ceros censurados*, y que el resto de los mismos es positivo. Así, la muestra es  $X_1 = 0, X_2 = 0, \dots, X_k = 0, X_{k+1}, X_{k+2}, \dots, X_n$ .

Suponemos también que el programa generado para determinar los errores de la modulación asigna *cero* a las diferencias menores de 0.2%; esto es, menores a 0.002 en términos absolutos. De este modo, la función de verosimilitud toma la forma:

$$\begin{aligned} L(\theta | X_1, X_2, \dots, X_n) &= P(X_1 < 0.2, X_2 < 0.2, \dots, X_k < 0.2, | \theta) \cdot \prod_{j=k+1}^n f(x_j; \theta) \\ L(\theta | X_1, X_2, \dots, X_n) &= [P(X_i < 0.2)]_{i=1,2,\dots,k} \cdot \prod_{j=k+1}^n f(x_j | \theta) \\ L(\theta | X_1, X_2, \dots, X_n) &= [P(X_i < 0.2)]^k \cdot \prod_{j=k+1}^n f(x_j | \theta) \\ L(\theta | X_1, X_2, \dots, X_n) &= \left[ \int_0^{0.2} f(x_1 | \theta) dx_1 \right]^k \cdot \prod_{j=k+1}^n f(x_j | \theta) \quad (\text{Ec. 2.2}) \end{aligned}$$

Para efectos de cálculo y por la propiedad que tiene el logaritmo, la función sobre la cual se hallan los puntos críticos que la maximizan es:

$$\begin{aligned} l(\theta | X_1, X_2, \dots, X_n) &= \log L(\theta | X_1, X_2, \dots, X_n) \\ &= k \log \left[ \int_0^{0.2} f(x_1 | \theta) dx_1 \right] + \sum_{j=k+1}^n \log f(x_j | \theta) \quad (\text{Ec. 2.3}) \end{aligned}$$

## 3. ESTIMACIÓN Y ANÁLISIS DE RESULTADOS

En el caso particular de una distribución Gamma con parámetro  $\theta = (\alpha, \beta)$ , la forma funcional es:

$$f(x | \theta) = f(x; \theta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, x > 0 \quad (\text{Ec. 3.1})$$

Para  $\alpha > 0, \beta > 0$ ,  $\alpha$  es conocido como parámetro de forma y  $\beta$  parámetro de escala.

Esta función se reemplaza en la Ecuación 2.2 para obtener la función de verosimilitud que se va a maximizar. Los  $X_i$ 's son los datos obtenidos en la toma de muestras durante 36 días (Gómez, *et al.* 2005).

La Tabla 1 presenta los valores de los aportes de las Vías como fuentes de emisión a las partículas totales en suspensión según el Modelo MCF en la Zona Centro del Valle de Aburrá, así como los errores arrojados por este modelo, para los 36 días modelados.

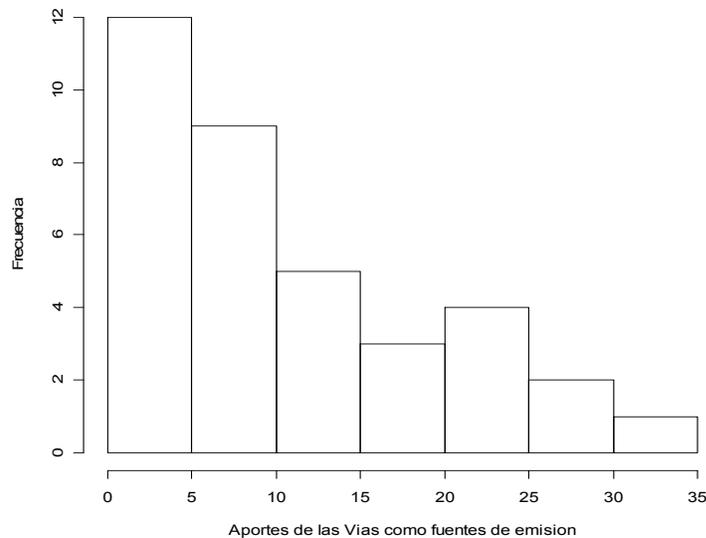
En realidad lo que interesa no es el valor máximo de la verosimilitud sino los puntos críticos que la maximizan. Estos puntos críticos son los estimadores de los parámetros; para obtenerlos se utiliza el software estadístico "R" en su versión 2.3.1.(2006). Con el fin de lograr cierta convergencia en las estimaciones de los parámetros  $\alpha$  y  $\beta$ , se utiliza el Algoritmo EM (Dempster et al.,1977) como criterio de convergencia.

En términos generales, el algoritmo hace una estimación inicial de los parámetros de la distribución Gamma a partir de los valores no nulos de un vector, en este caso, del vector *Aporte de las Vías*; luego utiliza esta distribución

para generar una población de tamaño igual a la longitud del vector original e incorporar al vector de datos no nulos tantos elementos como le fueron sacados al vector original. Con este último vector hace un nuevo ajuste de la distribución Gamma para obtener otras estimaciones de los parámetros repitiendo el proceso hasta notar alguna convergencia en los estimadores. Una vez obtenidos estos estimadores, se procede con ellos de acuerdo con la teoría estadística conocida para la distribución Gamma.

Ahora bien, en la distribución Gamma la media es igual al producto de sus parámetros; es decir,  $E[X] = \alpha / \beta$  y los resultados del programa arrojan  $\hat{\alpha} = 0.715$  y  $\hat{\beta} = 0.068$ , de modo que la media de la distribución es 10.5. En la Fig. 1 se muestra el histograma de los aportes (%) de las Vías a las partículas totales en suspensión; su forma motivó el estudio que nos ocupa.

La Tabla 2., muestra los resultados del Algoritmo EM para los estimadores de la distribución Gamma sugerida por aportes de las Vías como fuentes de emisión a las partículas totales en suspensión según el Modelo MCF en la Zona Centro del Valle de Aburrá para los 36 días modelados.



**Figura 1.** Distribución de frecuencias de contribución de las vías  
**Figure 1.** Road Contribution frequencies distribution

**Tabla 1.** Aportes de las Vías como fuentes de emisión a las partículas totales en suspensión según el Modelo (MCF) en la Zona Centro del Valle de Aburrá para los 36 días modelados (%)

**Table 1.** Road contribution as emission sources to total suspended particles (TSP) under MCF model downtown Valle de Aburrá during the 36-day-modeling (%)

Día	1	2	3	4	5	6
Factor (%)	7.9	7.0	13.0	14.0	0.5	22.0
Error modelado	3.37	3.61	2.46	1.86	2.25	3.60

Día	7	8	9	10	11	12
Factor (%)	0.5	9.0	6.7	34.0	1.0	6.5
Error modelado	0.29	3.73	0.73	3.44	0.54	4.52

Día	13	14	15	16	17	18
Factor (%)	12.0	10.9	0.5	15.7	9.3	27.0
Error modelado	3.46	6.79	3.29	2.25	6.68	2.26

Día	19	20	21	22	23	24
Factor (%)	27.1	0.0	24.0	0.0	1.00	0.0
Error modelado	2.89	2.83	2.55	4.53	0.44	3.78

Día	25	26	27	28	29	30
Factor (%)	0.0	8.0	10.3	0.2	2.6	17.3
Error modelado	2.23	2.80	2.18	2.46	2.46	1.53

Día	31	32	33	34	35	36
Factor (%)	6.6	24.0	24.0	0.0	15.9	6.3
Error modelado	1.02	1.91	2.56	2.29	2.75	2.29

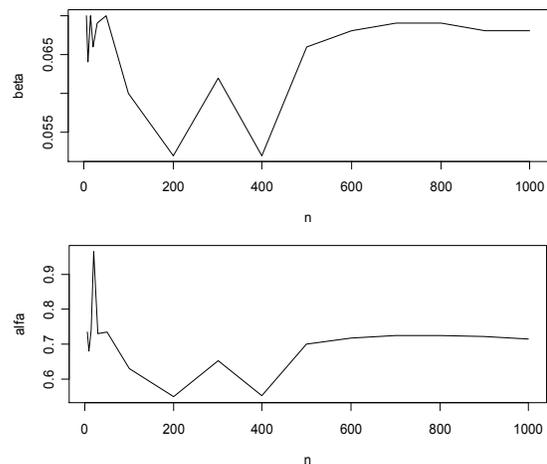
**Tabla 2.** Resultados del algoritmo EM

**Table 2.** Results of EM algorithm

$n$	5	10	15	20	30	50	100	200
$\hat{\alpha}$	0.735	0.678	0.739	0.694	0.728	0.733	0.629	0.549
$\hat{\beta}$	0.070	0.064	0.070	0.066	0.069	0.070	0.060	0.052

$n$	300	400	500	600	700	800	900	1000
$\hat{\alpha}$	0.651	0.551	0.698	0.717	0.725	0.725	0.721	0.714
$\hat{\beta}$	0.062	0.052	0.066	0.068	0.069	0.069	0.068	0.068

Estos resultados muestran que para tamaños muestrales grandes ( $n \geq 600$ ),  $\hat{\alpha}$  tiende a estar entre 0.71 y 0.72; mientras que  $\hat{\beta}$  tiende a estar entre 0.068 y 0.069.



**Figura 2.** Tendencia de los estimadores alfa y beta por el algoritmo EM

**Figure 2.** Alfa and Beta estimator pattern under EM algorithm

La Fig. 2 muestra que hay cierta tendencia de los estimadores a partir de  $n = 600$ .

Se tiene, entonces, que la media de los aportes de las Vías a las partículas totales en suspensión según el Modelo MCF en la Zona Centro del Valle de Aburrá es aproximadamente de 10.5%.

#### 4. CONCLUSIONES

Con base en la prueba de Bondad de Ajuste Chi-cuadrado con un nivel de significancia

$\alpha = 0.05$ , los resultados permiten concluir que los aportes (en porcentaje) de las Vías como fuentes de emisión a las partículas totales en suspensión, según el Modelo MCF para la Zona Centro de Medellín en el periodo de estudio, se pueden ajustar a una distribución  $Gamma(\alpha, \beta)$ . Para esta distribución los estimadores de máxima verosimilitud  $\hat{\alpha}$  y  $\hat{\beta}$  son 0.715 y 0.068 respectivamente.

Según los parámetros estimados en el punto anterior, la media y la desviación estándar de los aportes de las Vías como fuentes de emisión a las partículas totales en suspensión son 10.5% y 12.4%, respectivamente; lo que indica que desde un punto de vista práctico, las Vías aportan significativamente en términos porcentuales a la emisión de partículas totales en suspensión.

Los resultados estadísticos arrojados por la prueba de bondad de ajuste con el uso de la estimación de los parámetros vía máxima verosimilitud aplicada a los aportes de las vías como fuentes de emisión de las partículas totales en suspensión, son comparables con los producidos con las técnicas utilizadas por el Modelo de Fuente Receptor, MCF; la primera produjo una medida del factor de contribución de 10.5, mientras que la segunda arrojó una medida de 9.8.

Las metodologías estadísticas descritas en la conclusión anterior, abren el camino a la estimación de los promedios de otras fuentes que por su naturaleza y complejidad, dificultan la aplicación de métodos tradicionales de monitoreo y medición.

## REFERENCIAS

- [1] A.P., DEMPSTER, N.M., LAIRD, AND D.B., RUBIN. (1977). Maximum Likelihood From Incomplete Data Via the EM Algorithm. JRSS Series B. Vol. 39, pp. 1-38.
- [2] BEDOYA, V. JULIÁN (2002). La Gestión de la Calidad del Aire Urbano: Los Inventarios. Boletín Redaire, No. 9.
- [3] Contraloría General de Medellín. Informe Estado de los Recursos Naturales y del Medio Ambiente de Municipio de Medellín. 2001
- [4] ECHEVERRY, CARLOS et al. (2001). Evaluación de Material Particulado en Suspensión en el Valle de Aburrá. Boletín Redaire. No. 8.
- [5] GÓMEZ, M., CORREA, M. SALDARRIAGA J. C., CASTRILLÓN, F. Y POSADA, E. Software Modelo de Correlación de Fuentes de Material Particulado y Muestras Tomadas en el Ambiente, Grupo GIGA y Grupo de Desarrollo Sostenible, U. de A., Grupo de Investigación GHYGAM, Politécnico JIC, Año 2005
- [6] HARRISON, M. ROY, AND PERRY ROGER (1986). Handbook of Air Pollution Analysis. Chapman and Hall. Second Ed
- [7] Programa de Protección y Control de la Calidad del Aire. Área Metropolitana del Valle de Aburrá en Asocio con la Universidad de Antioquia, Universidad Pontificia Bolivariana, Universidad Nacional y Politécnico Colombiano Jaime Isaza Cadavid. 1998.
- [8] R Development Core Team. (2006). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna Austria
- [9] STEVENS, ROBERT K., HERGET, WILLIAM F. (1977). Analytical Method Applied to Air Pollution Measurements. USA: Ann Arbor Science
- [10] www.epa.gov. (2004). Method 3010A Acid Digestion of Aqueous Samples and Extract for Total Metals for Analysis by GFAA Spectroscopy
- [11] www.epa.gov. (2004). Method 3020A Acid Digestion of Aqueous Samples and Extract for Total Metals for Analysis by Spectroscopy.