

METODOLOGIA PARA LA APLICACIÓN DE MODELOS DE REGRESION DE USOS DEL SUELO EN LA ESTIMACIÓN LOCAL DE LA CONCENTRACION MENSUAL DE PM10 EN MEDELLIN – COLOMBIA

Libardo Antonio Londoño Ciro¹, Julio Eduardo Cañón Barriga²

¹ Doctorando en Ingeniería, Universidad de Antioquia.

² Doctor en Hidrología. Grupo de investigación en Gestión y Modelación Ambiental (GAIA). Facultad de Ingeniería, Universidad de Antioquia. Calle 67 N.º 53-108. AA. 1226 Medellín, Colombia. libaranto@gmail.com

RESUMEN

En este artículo se presenta una metodología que aplica modelos de regresión de usos del suelo (LUR) a datos de algunas posibles variables explicativas disponibles en enero de 2007 (usos del suelo, vías principales, milímetros de lluvia y decibeles de ruido) para estimar localmente la concentración promedio mensual del contaminante PM10 ($\mu\text{g}/\text{m}^3$) en 10 sitios de monitoreo en la ciudad de Medellín - Colombia. La metodología usa sistemas de información geográfica (SIG), el método de mínimo cuadrados ordinarios (OLS) y regresiones geográficamente ponderadas. De 11 modelos obtenidos, el mejor fue una función lineal de la distancia a las vías principales con un R^2 de 0,79. En la primera parte se presentan las características de los modelos LUR. En la metodología se implementan los modelos LUR; se obtienen los resultados, se hace su análisis y finalmente se presentan las conclusiones.

Palabras clave: Modelos de regresión de usos del suelo, contaminación del aire.

Recibido 29 de Julio de 2015. Aceptado 22 de Octubre de 2015

Received: July 29th, 2015. Accepted: October 22nd, 2015

METHODOLOGY FOR THE IMPLEMENTATION OF LAND USE REGRESSION MODELS TO MONTHLY LOCAL ESTIMATION OF THE PM10'S CONCENTRATION IN MEDELLIN – COLOMBIA

ABSTRACT

This article describes a methodology that applies the Land Use Regression models (LUR) to data of some potential explanatory variables available in January 2007 (land use, main roads, millimeters of rain and noise decibels) to estimate the local monthly average concentrations of PM10 ($\mu\text{g}/\text{m}^3$) at 10 monitoring sites of the air quality network (RedAire) of Medellín - Colombia. The methodology used geographic information systems (GIS), the method of ordinary least squares (OLS) and geographically weighted regression. From 11 models obtained, the best was a linear function of the distance to the main roads with a R^2 of 0.79. The first part introduces the LUR models. In the methodology the LUR models are implemented, the results are obtained, an analysis is carried out and finally, the conclusions are presented.

Keywords: Land Use Regression Models, Air Pollution.

Cómo citar este artículo: L. A. Londoño y J. E. Cañón, "Metodología para la aplicación de modelos de regresión de usos del suelo en la estimación local de la concentración mensual de PM10 en Medellín – Colombia" *Revista Politécnica*, vol. 11, no. 21, pp. 29-40, 2015.

1. INTRODUCCIÓN

En los centros urbanos densamente poblados se presentan problemas de calidad de aire por contaminación debida a distintos tipos de fuentes de emisión contaminantes que son mezclas complejas de sustancias y material particulado de diferentes diámetros [1]. El problema de la calidad del aire ha sido abordado mediante la implementación de sistemas de vigilancia de calidad del aire que realizan mediciones de la concentración de estas sustancias a través de redes de sitios de monitoreo. Sin embargo, no siempre es posible tener una configuración adecuada para la red que permita hacer mediciones exhaustivas en el espacio y tiempo por lo que se hace necesario el uso del modelamiento y la simulación para estimar el valor de un contaminante en algún punto del área de influencia a partir de los valores de otros contaminantes, del valor del mismo contaminante en otros puntos (modelos de interpolación espacial basados en el principio de autocorrelación espacial) y/o de variables meteorológicas, [2] [3] [4]. Existen modelos físicos de dispersión atmosférica a pequeña escala para simular la calidad del aire, modelos numéricos determinísticos gaussianos basados en principios físico-químicos que simulan las ecuaciones de conservación de masa y los efectos de procesos dinámicos que influyen en el equilibrio de masa y térmico en un volumen de control [5], modelos estocásticos basados en relaciones estadísticas entre las mediciones obtenidas y otras variables [6], modelos de redes neuronales [7] [8] y modelos de regresión de usos del suelo (land use regression models – LUR) que establecen correlaciones entre las mediciones de una variable dependiente y algunas posibles variables explicativas para construir modelos de regresión con datos espaciales [9] [10] [11].

Los modelos LUR buscan explicar la variabilidad local de una variable dependiente en función de variables independientes obtenidas mediante la intersección de buffers espaciales con un conjunto de posibles variables explicativas. Se han usado recientemente para estimar las concentraciones de un contaminante en un lugar determinado con base en variables explicativas tales como los usos del suelo que hay a su alrededor, la proximidad a vías existentes y densidades de tráfico entre otras. Utilizan SIG, el método de mínimos cuadrados ordinarios (Ordinary Mininum Squares -

OLS) y algunos el método de regresiones geográficamente ponderada (Geographically Weighted Regression - GWR®) [12] [13].

Muchas de los modelos LUR revisados en la literatura estiman la concentración de carbón negro y NO₂ y en menor proporción el material particulado [9] [14]. Las mayores aplicaciones LUR se han realizado en el contexto del proyecto “variación de la salud por contaminación del aire en áreas pequeñas” (small area variation in air pollution health - SAVIAH) en varias ciudades Europeas [10]. En general los resultados obtenidos varían con base en la cantidad de sitios de monitoreo, las variables explicativas utilizadas y la ciudad. Los valores de R² tienen una variación entre 0.66 y 0.87. Los modelos usan entre 3 y 6 variables explicativas con buffers de radio menor a 300m [9] [15] [16]. En algunas ciudades de Canadá también se han aplicado modelos LUR para estimar NO₂ usando hasta 4 variables explicativas y con valores de R² entre 0.54 y 0.76 [17] [18]. En todos estos estudios existen algunas características comunes:

- No necesariamente el aumento de variables explicativas mejoró el rendimiento del modelo.
- Las variables explicativas comunes en todos los modelos fueron densidad vehicular y vías.
- En general no se registra la inclusión de variables meteorológicas. Sin embargo, en el trabajo de Li et al., en la ciudad de Shangsha (China) para estimar NO₂ y PM₁₀ [19], se incluyeron buffers semicirculares para determinar indirectamente el efecto de la dirección del viento sobre la concentración de estas sustancias, pero dado que esta variable cambia constantemente en todos los sitios de monitoreo, medirla es una tarea costosa y requiere mucho tiempo. En la ciudad de Madrid los estudios de Pérez-Martínez y Miranda [20] consideraron los efectos de las variables meteorológicas y de tráfico sobre la concentración de PM₁₀, los resultados más significativos fueron:
 - La concentración de PM₁₀ disminuyó entre un 10.3% y un 12.4% por cada aumento de 1 m/s en la velocidad del viento.
 - La concentración de PM₁₀ aumentó en un 7.2% por cada aumento de 10°C en la temperatura.

- La concentración de PM10 aumentó en un 3.3% por cada aumento de 1 millón de vehículos por km de vía principal.
- La concentración de PM10 disminuyó en un 3.2% por cada disminución del 1% en la cantidad de vehículos pesados sobre la vía principal.
- La concentración de PM10 aumentó en un 0.73% por cada aumento de 1 KM/h en la velocidad de los vehículos pesados sobre la vía principal.

La implementación de un modelo LUR contempla las siguientes fases metodológicas:

- Construir buffers con distintos valores de radio con centro en las coordenadas de los sitios de monitoreo de la variable dependiente. En general se usa una geometría circular para el buffer aunque también se han usado geometrías anulares.
- Determinar para cada tamaño de buffer la intersección espacial con las posibles variables explicativas. Si para un tamaño de buffer dado, una posible variable explicativa tiene un valor de cero para más del 90% de los sitios de monitoreo, ésta variable es omitida por no ser significativa dentro del modelo LUR.
- Determinar el modelo LUR mediante la correcta especificación de un modelo de regresión para cada tamaño de buffer analizando cuales variables explicativas y/o sus posibles combinaciones cumplen con los criterios de especificación del método OLS.
- Validar el modelo LUR mediante la ejecución del mismo y el cálculo de los residuales y estadísticas de error (valor observado menos valor estimado) en cada sitio de monitoreo.
- Usar el modelo LUR para estimar localmente el valor de la variable dependiente en otros sitios en donde se tenga información de las variables explicativas.

Los criterios de especificación del método OLS son:

- Criterios de rendimiento del modelo en la medida en que las variables explicativas dan cuenta de la variación de la variable

dependiente (R^2 , R^2 ajustado, criterio corregido de información de Akaike - AICc).

- Criterio de consistencia del modelo en tanto los residuos se distribuyan normalmente, es decir, que no falten variables explicativas claves, que las relaciones sean lineales en un modelo lineal, que no hayan efectos de valores outliers o fuerte heteroscedasticidad (estadístico de Jarque-Bera - JB).
- Criterio de ausencia de redundancia o multicolinealidad entre las variables explicativas (factor de inflación de la varianza VIF).
- Criterio de significancia para los coeficientes de las variables explicativas (estadístico de residuales de Koenker Breusch-Pagan - K(BP)).
- Criterio de ausencia de residuos en la autocorrelación espacial (criterio de autocorrelación espacial residual mediante el índice global de Moran - SA).

En este trabajo se propone un modelo LUR para estimar la concentración promedio mensual en microgramos por metro cúbico $\mu\text{gm}/\text{m}^3$ del contaminante PM10 (material particulado de 10 micras de diámetro) a partir de datos disponibles en enero de 2007. Como posibles variables explicativas se pudo acceder a datos de mediciones de pluviometría (mm), de ruido (db), un mapa de usos del suelo y uno de vías principales.

2. METODOLOGÍA

2.1. Análisis de variables explicativas

Los datos de PM10 corresponden con mediciones hechas en 10 sitios de monitoreo de la red de calidad del aire (RedAire) de la ciudad de Medellín, Colombia. En la Fig.1 y en la Tabla 1 se muestran las coordenadas y los nombres de los sitios (Geographic Coordinate System: GCS MAGNA. Projected Coordinate System: MAGNA Colombia Bogotá, Projection Transverse Mercator).

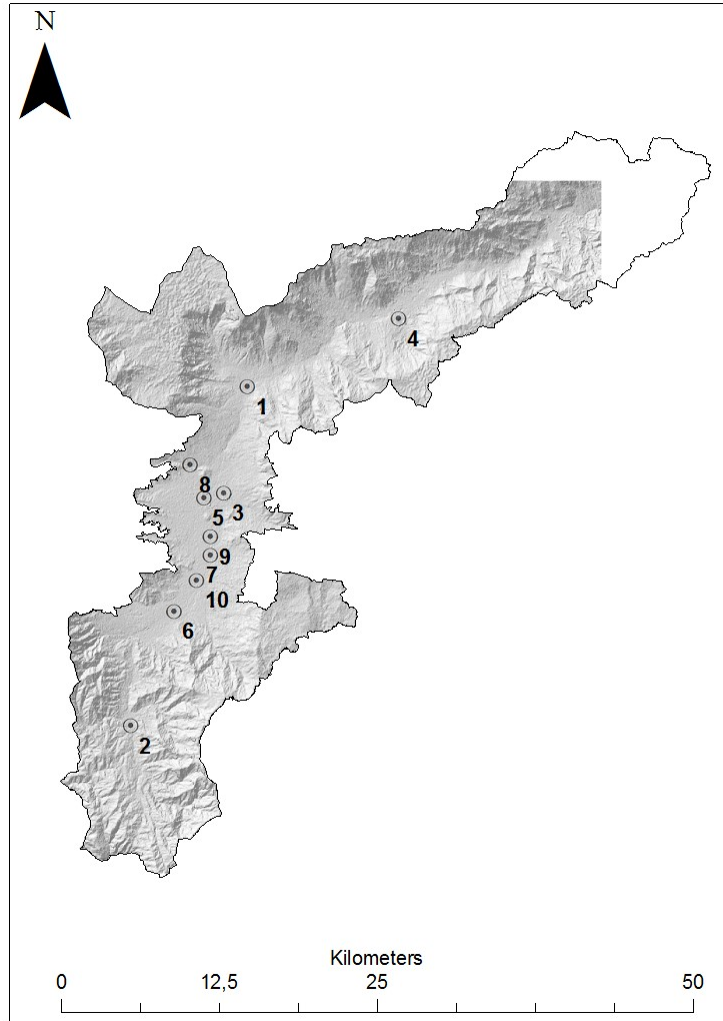


Fig 1. Ubicación sitios de monitoreo.

Tabla 1. Descripción sitios de monitoreo.

ID del sitio de monitoreo	X	Y	Nombre sitio de monitoreo	Descripción	Dato Observado PM10 ($\mu\text{g}/\text{m}^3$)
1	836785,42	1191924,48	BELLO	Municipio de Bello	70.00
2	827264,12	1165348,22	CALD	Municipio de Caldas	59.29
3	834995,21	1183470,75	AGUI	Edificio Miguel de Aguinaga	76.90
4	849226,63	1197678,07	GIRARDOTA	Municipio de Girardota	51.21
5	833423,81	1182995,98	CORA	Corporación Autónoma Regional de Antioquia	70.09
6	826710,56	1172556,94	ITAGUÍ	Municipio de Itagüí Colegio el Rosario	72.58
7	833901,20	1178597,43	POLITECN	Politécnico Colombiano Jaime Isaza Cadavid	63.25
8	832276,76	1185725,08	UNAL	Universidad Nacional de Colombia	88.21
9	833931,85	1179845,29	RIO	Corredor vial del río	91.71
10	832774,03	1176575,16	ITSFDO	Planta de Tratamiento de Agua San Fernando	69.54

Se pudo tener acceso a 11 posibles variables explicativas, clasificadas en las siguientes categorías:

- Usos del suelo: capa temática con polígonos e información sobre usos del suelo en el área de estudio. Se toma como dato el área en m² del uso. Los usos disponibles fueron uso residencial (AreaUresid), uso comercial (AreaUcomerc), uso industrial (AreaUind), uso verde (AreaUverde), uso infraestructura pública institucional (AreaUinst) y uso minero (AreaUmin).
- Vías: capa temática con líneas e información sobre las vías principales. Se hizo un buffer de 25 m sobre cada vía para tomar el área de la vía principal en m² (AreaVP) y se calculó el número de vías principales por área del buffer (#Vp/AreaB) en 1/m².
- Distancia: se calculó la distancia perpendicular en m de cada sitio de monitoreo a la vía principal más cercana (DistVPnear).
- Lluvia: capa temática con información sobre los milímetros (mm) de lluvia promedio mensual en los sitios de monitoreo (PluvAllocat).
- Ruido: capa temática con información sobre los decibeles (db) promedio mensual en el día en la vía principal cercana al sitio de monitoreo (RuidoNearVp).

Mediante sistemas de información geográfica se calcularon buffers circulares con centro en los sitios de monitoreo (Fig. 2). Para efectos exploratorios se construyeron buffers con radios de 50m, 100m, 150m, 300m, 500m y 1000m.

Se realizó un proceso de intersección espacial entre cada buffer y las posibles variables explicativas. Si el buffer se intercepta con alguna variable explicativa, se toma el valor de la misma, en caso contrario se asigna un valor de cero. Si para un tamaño de buffer dado, una posible variable explicativa tiene un valor de cero para más del 90% de los sitios de monitoreo, ésta variable es omitida por no ser significativa dentro del modelo LUR. Con base en lo anterior, no fueron significativas las variables AreaUcomerc, AreaUind, AreaUverde, AreaUinst y AreaUmin; y

en consecuencia el conjunto de posibles variables explicativas se redujo de 11 a 6. En la Tabla 2 se muestran las posibles variables explicativas que se obtuvieron para los distintos tamaños de buffers.

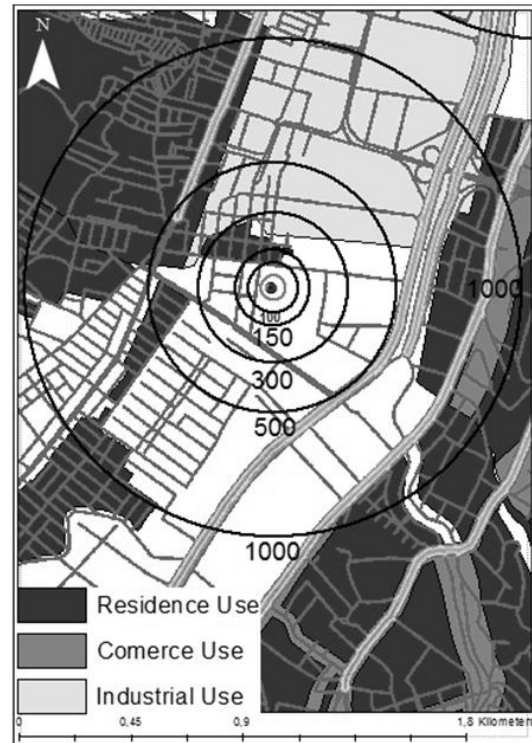


Fig 2. Buffers circulares y algunas variables explicativas.

Tabla 2. Posibles variables explicativas del modelo LUR.

Buffer	Número de sitios de monitoreo	ID del sitio de monitoreo (Tabla 1)	Posibles variables explicativas
50	4	3, 4, 6, y 8	
100	6	3, 4, 6, 7, 8 y 10	AreaVP, AreaUresid, DistVPnear, PluvAllocat, RuidoNearVP, #VP/AreaB
150			
300	9	1, 3, 4, 5, 6, 7, 8, 9 y 10	
500			
1000			

2.2. Criterios de especificación del modelo

Para implementar el modelo LUR se especificó un modelo de regresión para cada tamaño de buffer. Para saber cuáles variables explicativas de la Tabla 2 cumplen con los criterios de especificación del método OLS se usó la herramienta Exploratory Regression® de ArcGis10.2®. Esta es una herramienta de minería de datos que analiza todas las posibles combinaciones de variables explicativas para determinar el mejor modelo. Es similar a las herramientas de regresión Stepwise existentes en muchos software estadísticos y permite definir un número mínimo y un número máximo de variables explicativas en el análisis [21] [22] [23]. Los criterios de especificación usados fueron [13] [22]:

- Criterios de rendimiento del modelo para determinar si las variables explicativas dan cuenta de la variación de la variable dependiente. Para ello se usan el coeficiente de determinación (R^2), el coeficiente de determinación ajustado ($AdjR^2$) y el criterio corregido de información de Akaike (AICc). La Eq 1 define el valor mínimo aceptable para $AdjR^2$ por lo tanto el mejor modelo será aquel con mayor $AdjR^2$. Según la Eq 2, el mejor modelo será aquel con menor AICc.

$$AdjR^2 > 0.5 \quad Eq (1)$$

$$Min (AICc) \quad Eq (2)$$

- Criterio de consistencia del modelo para determinar si los residuos del modelo se distribuyen normalmente. Con ello se controla que no falten variables explicativas claves en el modelo, que las relaciones entre las variables explicativas sean lineales en un modelo lineal, que no haya efectos de valores extremos o fuerte heteroscedasticidad. Para ello se usa el p-valor del estadístico de Jarque-Bera (JB). La Eq 3 define el valor mínimo aceptable para el p-value de JB. En consecuencia, el mejor modelo será aquel con mayor JB.

$$(JB)_{p-value} > 0.10 \quad Eq (3)$$

- Criterio de ausencia de redundancia o multicolinealidad entre las variables

explicativas para determinar si éstas son independientes entre sí, sino lo fueran entonces existe multicolinealidad entre ellas y alguna de ellas debería ser removida. Para ello se usa el factor de inflación de la varianza (VIF). El mejor modelo será aquel con menor VIF. El valor mínimo aceptable para el VIF viene dado por la Eq 4.

$$VIF < 7.50 \quad Eq (4)$$

- Criterio de significancia para las variables explicativas que permite determinar si el coeficiente de la variable es significativo. Para ello se usa el p-valor de los residuales de Koenker Breusch-Pagan K(BP). El mejor modelo será aquel con menor K(BP). El valor mínimo aceptable para el K(BP) está dado por la Eq 5.

$$K(BP)_{p-value} < 0.05 \quad Eq (5)$$

- Criterio de ausencia de residuos en la autocorrelación espacial para determinar qué tan libre de residuos es la autocorrelación espacial. Para ello se usa el p-valor del estadístico de autocorrelación espacial residual mediante el índice global de Moran (SA). El mejor modelo será aquel con mayor SA. El valor mínimo aceptable para el p-valor del SA está dado por la Eq 6.

$$(SA)_{p-value} > 0.10 \quad Eq (6)$$

Con un mínimo de 1 y un máximo de 5 variables explicativas y sus posibles combinaciones se calculan los valores de los criterios de especificación con los datos de las Tablas 1 y 2. En la Tabla 3 se muestran los resultados de 11 modelos con las mejores combinaciones.

Evaluando los valores de los criterios de especificación de la tabla 3 y ordenando los modelos con base en el mejor cumplimiento de la mayor cantidad de criterios (mayor R^2 , mayor $AdjR^2$, menor AICs, mayor JB, menor K(BP), menor VIF, mayor SA), se obtienen los mejores 6 modelos que se muestran en la Tabla 4.

Tabla 3. Criterios de especificación de los mejores modelos.

Buffer	R ²	AdjR ²	AICc	JB	K(BP)	VIF	SA	Variables explicativas	
50	0.64	0.54	83.15	0.02	0.68	1.03	0.02	PluvAllocat, AreaVP	
	0.64	0.46	92.19	0.09	0.83	1.54	0.02	-DistVPnear,	RuidoNearVP,
100	0.61	0.51	58.66	0.99	0.26	1.00	0.48	AreaUResid	
	0.80	0.66	84.76	0.80	0.54	1.06	0.60	-DistVPnear	
	0.50	0.44	80.52	0.61	0.35	1.00	0.26	-DistVPnear, AreaUResid	
150	0.79	0.72	61.12	0.82	0.11	1.00	0.41	RuidoNearVP	
	0.50	0.44	80.52	0.61	0.35	1.00	0.26	-DistVPnear	
300	0.50	0.44	80.52	0.61	0.35	1.00	0.26	RuidoNearVP	
500	0.65	0.47	92.07	0.69	0.46	5.24	0.21	RuidoNearVP, -AreaVP	
	0.65	0.47	92.07	0.69	0.46	5.24	0.21	RuidoNearVP, -AreaVP	
1000	0.76	0.59	82.47	0.94	0.44	1.04	0.53	AreaUResid	
	0.77	0.61	82.17	0.81	0.14	1.23	0.26	-DistVPnear, AreaUResid	

Tabla 4. Mejores modelos según criterios de especificación del GWR.

Modelo	Buffer	R ²	AdjR ²	AICc	JB	K(BP)	VIF	SA	Variables explicativas
1	150	0.79	0.72	61.12	0.82	0.11	1.00	0.41	-DistVPnear
2	1000	0.77	0.61	82.17	0.81	0.14	1.23	0.26	-DistVPnear, AreaUResid
3		0.61	0.51	58.66	0.99	0.26	1.00	0.48	-DistVPnear
4	100	0.80	0.66	84.76	0.80	0.54	1.06	0.60	-DistVPnear, AreaUResid
5	150	0.50	0.44	80.52	0.61	0.35	1.00	0.26	RuidoNearVP
6	300	0.50	0.44	80.52	0.61	0.35	1.00	0.26	RuidoNearVP

3. RESULTADOS

De los 6 mejores modelos obtenidos en la tabla 4, el modelo 1 cumple con 7 de 7 criterios de especificación, seguido por el modelo 2 que cumple con 6 de 7. En consecuencia el modelo 1 es el mejor modelo y su valor de R^2 indica que explica un 79% de la variabilidad de PM10. Éste modelo estaría dado por la Eq 7.

$$PM10_{local} = Intercepto_{local} + Coeficiente_{local} * DistVPnear$$

Eq (7).

Para calcular los coeficientes de la Eq 7 y el valor estimado de PM10 en cada uno de los sitios de monitoreo de la Tabla 1, se utilizó la herramienta GWR® de ArcGis10.2®. En la Tabla 5 se muestra el resultado.

Tabla 5. PM10 estimado con la Eq 7.

ID del sitio de monitoreo	PM10 observado	PM10 estimado	R ² Local	Intercepto	Coeficiente de DistVPnear	Error estándar	Error estándar del Intercepto	Error estándar del coeficiente de DistVPnear
1	70.00	76.01	0.79	79.67	-0.02551	3.58	4.21	0.01
2	59.29	79.66	0.78	79.66	-0.02550	4.21	4.21	0.01
3	76.90	79.35	0.79	79.67	-0.02550	4.15	4.21	0.01
4	51.21	51.00	0.79	79.68	-0.02551	6.61	4.21	0.01
5	70.09	76.40	0.79	79.67	-0.02550	3.63	4.21	0.01
6	72.58	77.32	0.78	79.66	-0.02550	3.78	4.21	0.01
7	63.25	78.29	0.79	79.67	-0.02550	3.95	4.21	0.01
8	88.21	79.15	0.79	79.67	-0.02551	4.11	4.21	0.01
9	91.71	76.70	0.79	79.66	-0.02550	3.68	4.21	0.01
10	69.54	67.76	0.78	79.67	-0.02550	3.23	4.21	0.01

La Fig. 3 muestra los resultados de la estimación local de PM10 en cada sitio de monitoreo y el valor del coeficiente de determinación R² cuyo valor promedio fue de 0.78.

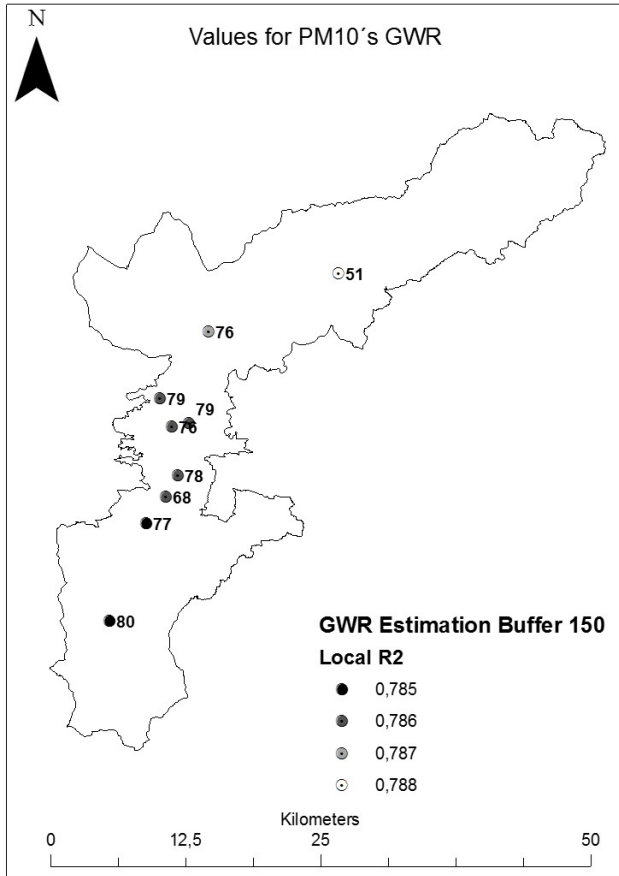


Fig. 3. PM10 estimado con la Eq 7.

Para validar los valores estimados con el modelo, se calculó en cada sitio de monitoreo el error absoluto relativo (*Absolut Relative Error – ARE*) y luego el error absoluto relativo promedio (*Mean Absolut Relative Error – MARE*) mediante las Eq 8 y 9.

$$ARE = \frac{|O_i - E_i|}{O_i} \quad \text{Eq (8)}$$

$$MARE = \frac{\sum_{i=1}^n |O_i - E_i|}{n} \quad \text{Eq (9)}$$

En donde n es la cantidad de sitios de monitoreo, E_i y O_i son los valores estimado y observado en cada sitio respectivamente. Los resultados se muestran en la tabla 6. El error absoluto relativo promedio obtenido fue de 0.12.

Tabla 6. ARE en cada sitio de monitoreo.

ID del sitio de monitoreo	O _i	E _i	Error Absoluto Relativo
1	70,00	76,01	0,09
2	59,29	79,66	0,34
3	76,90	79,35	0,03
4	51,21	51,00	0,00
5	70,09	76,40	0,09
6	72,58	77,32	0,07
7	63,25	78,29	0,24
8	88,21	79,15	0,10
9	91,71	76,70	0,16
10	69,54	67,76	0,03

4. ANÁLISIS DE RESULTADOS

En la realización de este trabajo se tuvieron limitaciones para el acceso a valores de variables para explicar la variabilidad de la concentración de PM10 en relación con trabajos similares que usan gran variedad de variables explicativas como la densidad vehicular (vehículos por día en vías adyacentes a sitios de monitoreo), la densidad poblacional (habitantes por km²), entre otras; y la información de la variable dependiente en gran cantidad de sitios de monitoreo (en promedio 40 sitios) [9]. En este caso la única información disponible fue de 11 variables asociadas con la pluviometría, el ruido, el área de vías principales, usos del suelo y el valor de PM10 en 10 sitios de monitoreo.

Se usaron y reportaron valores de 7 criterios de especificación para determinar la combinación más adecuada de las variables explicativas y proponer los mejores modelos de regresión. En trabajos similares solamente se reportan los valores de los criterios de rendimiento de los modelos y específicamente R² más no informan acerca de los demás criterios [9]. En la tabla 4 se obtuvieron 6 modelos con un R² variando entre 0.61 y 0.80 de los cuales 4 cumplen más adecuadamente con los

criterios de especificación. Dos de ellos están en función de 1 variable explicativa (distancia a vías principales) y los otros dos en función de 2 variables explicativas (distancia a vías principales y área del uso del suelo residencial). El valor de R^2 con una o dos variables explicativas prácticamente fue el mismo. Para el caso del modelo 1 con un buffer de 150m y una variable explicativa, R^2 fue de 0.79 y para modelo 4 con un buffer de 1000m y dos variables explicativas, R^2 fue de 0.80 (tabla 4).

El modelo 1 fue el que cumplió mejor los criterios de especificación y su valor de R^2 está dentro del rango de los estudios reportados en las referencias. Este modelo es similar a un modelo de proximidad a vías el cual es útil para hacer análisis de concentración e impacto de contaminantes por exposición directa por la cercanía a la fuente de emisión. La distancia a las vías principales y la densidad vehicular sobre la vía están directamente relacionadas ya que la cercanía a una vía densamente transitada incrementaría el impacto de esta fuente sobre la concentración del contaminante.

El error absoluto relativo del valor estimado en cada sitio de monitoreo fue en general menor o igual a 0.1 para distancias a vías principales entre 12 m y 143 m, es decir, distancias dentro del rango del tamaño de buffer de 150 m. La excepción se encontró en tres sitios de monitoreo (2, 7 y 9) ubicados cerca de vías principales pero que tienen una alta densidad de flujo vehicular que pudieron afectar el modelo.

Las variables de pluviometría y ruido no fueron estadísticamente significativas para explicar la variabilidad de PM10, esto pudo haber sido afectado por el hecho de que en general históricamente en el mes de enero se dan dos situaciones particulares en la ciudad: es un mes caluroso que está caracterizado por bajos niveles de lluvia y corresponde con la temporada vacacional con lo cual en la ciudad se disminuye la densidad de tráfico vehicular y por lo tanto los niveles de ruido asociados con ese factor. Los usos del suelo disponibles tampoco incidieron de manera importante en la explicación de la variación de PM10 debido a que durante el proceso de intersección espacial entre las variables explicativas y la ubicación de los sitios de monitoreo para distintos tamaños de buffer, la variable usos del suelo tuvo un valor de cero en más del 90% de los sitios de monitoreo.

5. CONCLUSIONES

Se presenta una metodología para la estimación local de la concentración de PM10 en la ciudad de Medellín – Colombia mediante modelos LUR. Es la primera vez que se utilizan estos modelos en la ciudad dado que no se tienen reportes de su aplicación. En general la metodología expuesta es sencilla en su implementación pero necesita tener acceso a datos de muchas variables que pudieran explicar la variabilidad local de la variable dependiente.

En los modelos explorados en la literatura las variables más comunes y significativas para estimar la concentración de PM10 a nivel urbano fueron densidad de tráfico y vías cercanas a los sitios de monitoreo. Las variables meteorológicas afectan de manera importante la concentración de PM10 pero por la dificultad de acceso a ellas y los altos costos de su medición; es común que no se incluyan estos tipos de modelos. En este trabajo no se tuvo acceso este tipo de variables.

De los modelos obtenidos, el mejor modelo fue un modelo lineal en donde la variabilidad local de PM10 está en función de la distancia a vías principales para un buffer de 150m con un R^2 de 0.79 y un error absoluto relativo promedio de estimación de 0.12. Este modelo de regresión cumplió todos los criterios de especificación del método OLS cuyos valores se registraron adecuadamente en la tabla 3. Tener en cuenta todos los criterios es muy importante para la correcta elección del modelo ya que no solamente el criterio del R^2 es totalmente válido. Si solamente se hubiese considerado este criterio en los resultados de los 11 modelos de la tabla 3, la elección del mejor modelo no habría sido correcta. En consecuencia se sugiere registrar los datos de todos los criterios de especificación en futuros estudios ya que estos serían unos indicadores muy útiles para los análisis de estados de arte y el perfeccionamiento de este tipo de modelos.

Los resultados obtenidos en este trabajo y otros similares, parecen indicar que las variables distancia a vías principales, densidad de tráfico vehicular y velocidad del viento; serían unas buenas candidatas para explicar la variabilidad local de la concentración de PM10 a nivel urbano. Se propone a futuro implementar en la ciudad modelos LUR que incluyan este tipo de información.

Estimar la variabilidad local de la concentración de PM10 con modelos LUR permitiría establecer relaciones con los efectos de la exposición directa de esta sustancia sobre la salud pública de la ciudad.

6. REFERENCIAS BIBLIOGRÁFICAS

- [1]. Laumbach. R.J., Kipen. H.M. Respiratory health effects of air pollution: update on biomass smoke and traffic pollution. *J Allergy Clin Immunol.* 129, 3–11, 2012.
- [2]. Deligiorgi, D., Philippopoulos, K. *Spatial Interpolation Methodologies in Urban Air Pollution Modeling: Application for the Greater Area of Metropolitan Athens, Greece*, Advanced Air Pollution. Dr. Farhad Nejadkoorki (Ed.). InTech. Disponible en: <http://www.intechopen.com/books/advanced-air-pollution/spatial-interpolation-methodologies-in-urban-air-pollution-modeling-application-for-the-greater-area> [consultado el 8 de agosto de 2015].
- [3]. Ramsay, J.O., Ramsay, T.O., Sangalli, L.M. *Spatial functional data analysis*. In: Ferraty, F. (Ed.). *Recent Advances in Functional Data Analysis and related topics*. Springer, 269-275, 2011.
- [4]. Sangalli, L.M., Ramsay, J.O., Ramsay, T.O. *Spatial splines regressions models*. *J. Roy. Statisc. Soc. Ser.* 75, 681-803, 2013.
- [5]. Dadvand. P., Rushton. S., Diggle. P., Goffe. L., Rankin. J., Pless-Mulloli. T. Using spatio-temporal modeling to predict long-term exposure to black smoke at fine spatial and temporal scale. *Atmospheric Environment*, 659 – 664, 2011.
- [6]. Şahin. Ü., Bayat. C., Uçan. U. Application of cellular neural network (CNN) to the prediction of missing air pollutant data. *Atmospheric Research*, 101, 314-326, 2011.
- [7]. Kang. P. Locally linear reconstruction based missing value imputation for supervised learning. *Neurocomputing*, 118, pp. 65-78, 2013.
- [8]. Eeftens. M., Beekhuizen. J., Beelen. R., Wang. M., Vermeulen. R., Brunekreef. B., Huss. A., Hoek. G. Quantifying urban street configuration for improvements in air pollution models. *Atmospheric Environment*, 72, 1-9, 2013.
- [9]. Dons. E., Van Poppe. M., Panis. L., De Prins. S., Berghmans. P., Koppen. G., Matheussen. C. Land use regression models as a tool for short, medium and long term exposure to traffic related air pollution. *Science of the Total Environment*, 476–477, 378-386, 2014.
- [10]. Jerrett M., Potoglou D., Arain A., Kanaroglou P., Sahuvaroglu T., Beckerman B, Morrison J., and Giovis C. A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, 15, 185–204, 2005.
- [11]. Gerard Hoek, Rob Beelen, Kees de Hoogh, Danielle Vienneau, John Gulliver, Paul Fischer, David Briggs. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*, 42, Issue 33, 7561-7578, 2008.
- [12]. Fotheringham. S., Brunsdon. C., Charlton. M. *Geographically Weighted Regression: the analysis of spatially varying relationships*, John Wiley & Sons, 2002.
- [13]. Fischer. M., Getis. A. *Handbook of Applied Spatial Analysis*, Springer, 2010.
- [14]. Wang M, Beelen R, Eeftens MR, Meliefste K, Hoek G, Brunekreef B. Systematic evaluation of land use regression models for NO2. *Environ Sci Tech*, 46, 4481–9, 2012.
- [15]. Beelen. R., Voogt. M., Duyzer. J., Zandveld. P., Hoek. G. Comparison of the performances of land use regression modelling and dispersion modelling in estimating small-scale variations in long-term air pollution concentrations in a Dutch urban area. *Atmospheric Environment*, 44, 4614 – 4621, 2010.
- [16]. Dons. E., Van Poppel. M., Kochan. B., Wets. G., Int Panis. L. Modeling temporal and spatial variability of traffic-related air pollution: hourly land use regression models for black carbon. *Atmospheric Environment*, 74, 237–246, 2013.
- [17]. N. Gilbert, M. Goldberg, B. Beckerman, J. Brook, M. Jerrett. Assessing Spatial Variability of Ambient Nitrogen Dioxide in Montreal, Canada, with a Land-Use Regression Model. *Journal of the Air & Waste Management Association*, 55, 1059–1063, 2005.
- [18]. M. Jerrett, B. Beckerman, J. Brook, M. Finkelstein, N. Gilbert. A Land Use Regression Model for Predicting Ambient Concentrations of Nitrogen Dioxide in Hamilton, Ontario, Canada. *Journal of the Air & Waste Management Association*, 56, 1059–1069, 2006.
- [19]. Xiaodong Li, Wu Liu, Zuo Chen c, Guangming Zeng, ChaoMing Hu, Tomas Leon, Jie Liang, Guohe Huang, Zhihua Gao, Zhenzhen Li,

Wenfeng Yan, Xiaoxiao He, Mingyong Lai, Yibin He. The application of semicircular-buffer-based land use regression models incorporating wind direction in predicting quarterly NO₂ and PM₁₀ concentrations. *Atmospheric Environment*, 103, 18-24. 2015.

- [20]. Perez-Martinez., P, Miranda. R, Temporal distribution of air quality related to meteorology and road traffic in Madrid. *Environmental Monitoring and Assessment*, 187, Issue 4, Disponible en: DOI: 10.1007/s10661-015-4452-3, [consultado el 8 de agosto de 2015].
- [21]. Burnham. K., Anderson. D. Model Selection and Multimodel Inference: a practical information-theoretic approach, 2nd Edition, New York, Springer, Section 1.5, 2002.
- [22]. Scott, L. Rosenshein, L. Janikas, M. Modeling Spatial Relationships using Regression Analysis. ESRI International User Conference, San Diego, CA. Technical Workshops, Disponible en: <http://www.arcgis.com/home/item.html?id=1824611e8e5c465ea492ef8124a15946>, [consultado el 8 de agosto de 2015]
- [23]. Raaschou-Nielsen. O., Andersen. ZJ., Beelen. R., Samoli. E., Stafoggia. M., Weinmayr G. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *Lancet Oncol*, 14, 813–22, 2013