

Metodología para Caracterizar Espacio-Temporalmente la Concentración de Material Particulado en Valles Intramontanos con Información Escasa

Tesis para optar al título de Doctor en Ingeniería



Por

Ing. Libardo Antonio Londoño Ciro

Director

Ing. Julio Eduardo Cañón Barriga, MSc., Ph.D.

Doctorado en Ingeniería

Facultad de Ingeniería

Universidad de Antioquia

Febrero de 2018

CONTENIDO

DEDICATORIA	5
AGRADECIMIENTOS	6
INTRODUCCIÓN.....	7
1 ESTADO DEL ARTE	10
2 MARCO TEÓRICO NECESARIO PARA EL DESARROLLO DE LA METODOLOGÍA.....	25
2.1 Algoritmos de interpolación espacial	25
2.2 Algoritmo de regresión espacial	37
2.3 Algoritmos econométricos	39
2.4 Validación de datos estimados.....	43
2.5 Gradientes espaciales	45
3 PROBLEMA Y OBJETIVOS	47
3.1 Problema	47
3.2 Objetivos.....	48
4 ÁREA Y DATOS DEL CASO DE ESTUDIO	49
4.1 Datos de PM2.5 y variables auxiliares.....	51
4.2 Datos de PM10 y variables auxiliares.....	54
5 CARACTERIZACIÓN ESPACIAL DE DATOS DE CONCENTRACIÓN DE MATERIAL PARTICULADO CON ALGORITMOS DE INTERPOLACIÓN O REGRESIÓN ESPACIAL.....	57
5.1 Caracterización espacial univariada de PM10	63
5.2 Caracterización espacial univariada de PM2.5	66
5.3 Caracterización espacial multivariada de PM10.....	68
5.3.1 Caracterización espacial multivariada de PM10 con el algoritmo geostatístico multivariado (Cokriging)	68
5.3.2 Caracterización espacial multivariada de PM10 con el algoritmo de regresión espacial	69
5.4 Caracterización espacial multivariada de PM2.5.....	71
5.4.1 Caracterización espacial multivariada de PM2.5 con el algoritmo geostatístico multivariado (Cokriging)	71
5.4.2 Caracterización espacial multivariada de PM2.5 con el algoritmo de regresión espacial	72

6	CARACTERIZACIÓN ESPACIO-TEMPORAL DE DATOS DE CONCENTRACIÓN DE MATERIAL PARTICULADO CON ALGORITMOS ECONOMÉTRICOS	74
6.1	Caracterización espacio-temporal de PM10.....	76
6.2	Caracterización espacio – temporal de PM2.5.....	79
7	MODELO DE PROXIMIDAD ESPACIAL (MPE)	83
8	ANÁLISIS DE RESULTADOS.....	86
9	CONCLUSIONES.....	114
	REFERENCIAS	120

Listado de Figuras

Figura 1. Problema de la Interpolación espacial. Fuente propia.	26
Figura 2. Metodología para tratamiento de Skewness y valores extremos.	31
Figura 3. Metodología para el cálculo del semivariograma empírico. Fuente propia.	32
Figura 4. Semivariograma teórico. Fuente propia.	33
Figura 5. Forma de la Matriz Γ . Fuente propia.	34
Figura 6. Forma del vector g . Fuente propia.	35
Figura 7. Cálculos de pesos ponderantes mediante Cokriging. Fuente propia.	36
Figura 8. Modelo AR-RD propuesto para estimar simular concentración de contaminantes del aire en el espacio-tiempo. Fuente propia.	42
Figura 9. Vecindad de una celda en una estructura raster. Fuente propia.	46
Figura 10. Mapa satelital de la Ciudad de Medellín y su Área Metropolitana. Google maps.	50
<i>Figura 11. Municipios del Área Metropolitana de la Ciudad de Medellín. Google maps. ...</i>	<i>50</i>
Figura 12. Sitios de monitoreo de material particulado. En la parte a PM10 y en la parte b PM2.5.	51
Figura 13. Densificación de datos mediante Algoritmos de Interpolación Espacial. Fuente propia.	59
Figura 14. Metodología para la caracterización espacial univariada de contaminantes del aire. Fuente propia.	60
Figura 15. Metodología para la caracterización espacial Multivariada de contaminantes del aire. Fuente propia.	62
Figura 16. Metodología propuesta para la caracterización espacio-temporal de material particulado. Fuente propia.	75
Figura 17. PM10 (O) en azul vs. PM10 (E) con AR-RD con dominio espacial GWR en rojo.	78
Figura 18. PM2.5 (O) en azul Vs PM2.5 (E) con AR – RD GWR en rojo.	81
Figura 19. Modelo de proximidad espacial. Fuente propia.	84
Figura 20. Insumos de entrada al Modelo de Proximidad Espacial.	85
Figura 21. Dato promedio mensual observado.	89
Figura 22. Dato promedio mensual estimado.	89
Figura 23. Valor promedio mensual de T. Figura 24. Valor promedio mensual de ∇T	91
Figura 25. Valor promedio mensual de V_v . Figura 26. Valor promedio mensual de ∇V_v	91
Figura 27. Foto panorámica de la ciudad de Medellín en marzo de 2013(a) y 2017(b).	94
Figura 28. Mapas de caracterización espacial de Pm2.5 período más contaminado.	96
Figura 29. Mapas de caracterización espacial de las variables explicativas de PM2.5 período más contaminado.	97
Figura 30. Mapas de caracterización espacial de PM2.5 período menos contaminado.	98
Figura 31. Mapas de caracterización espacial de las variables explicativas de PM2.5 período menos contaminado.	99
Figura 32. Mapas de caracterización espacial de PM2.5 período contaminación de promedio.	101
Figura 33. Mapas de caracterización espacial de las variables explicativas de PM2.5 período contaminación promedio.	102

Figura 34. Mapas de caracterización espacial de PM2.5 con ajuste de proximidad a fuentes.	102
Figura 35. Mapas mensuales de caracterización espacial de PM2.5.	103
Figura 36. Mapas mensuales de caracterización espacial de PM2.5.	104
Figura 37. Error estándar de estimación de PM2.5 primer cuatrimestre del año.	105
Figura 38. Error estándar de estimación de PM2.5 segundo cuatrimestre del año.....	106
Figura 39. Error estándar de estimación de PM2.5 tercer cuatrimestre del año.....	107
Figura 40. Propuesta de red de monitoreo de concentración de material particulado en Medellín (Londoño et al., 2017).....	109
Figura 41. Confiabilidad de la Red de monitoreo de concentración de material particulado propuesta para Medellín.	110
Figura 42. Metodología Final. Fuente propia.	115
Figura 43. Criterios espaciales para la ubicación de la red de sitios de monitoreo propuesta. Fuente propia.	119

Listado de Tablas

Tabla 1. Clasificación de los contaminantes por su estado de agregación (Laumbach & Kipen, 2012).	10
Tabla 2. Niveles máximos permisibles para contaminantes criterio (MinAmbiente, 2017).	12
Tabla 3. Objetivos y escalas de medición.	14
Tabla 4. Clasificación de los Algoritmos de Interpolación Espacial.	19
Tabla 5. Características de los Algoritmos de Interpolación Espacial.	20
Tabla 6. Algoritmos de interpolación espacial	27
Tabla 7. Descripción sitios de monitoreo de PM2.5.	52
Tabla 8. Datos observados promedio de PM2.5 (O) para cada mes del año ($\mu\text{g}/\text{m}^3$).	52
Tabla 9. Datos de temperatura promedio para cada mes del año ($^{\circ}\text{C}$).	52
Tabla 10. Datos de velocidad del viento promedio para cada mes del año (m/s).	53
Tabla 11. Datos de ∇T para cada mes del año (%).	53
Tabla 12. Datos de ∇V_v para cada mes del año (%).	53
Tabla 13. Descripción sitios de monitoreo PM10.	54
Tabla 14. Datos observados promedio de PM10 (O) para cada mes del año ($\mu\text{g}/\text{m}^3$).	54
Tabla 15. Datos de temperatura promedio para cada mes del año ($^{\circ}\text{C}$).	55
Tabla 16. Datos de velocidad del viento promedio para cada mes del año (m/s).	55
Tabla 17. Datos de ∇T para cada mes del año (%).	55
Tabla 18. Datos de ∇V_v para cada mes del año (%).	56
Tabla 19. Calculo de RMSE para cada interpolador	63
Tabla 20. Frecuencias de los modelos de semivariograma teórico para PM10.	63
Tabla 21. Datos estimados de PM10 (E) para cada mes del año con JB ($\mu\text{g}/\text{m}^3$).	64
Tabla 22. Estadísticos de Error Datos Estimados (E) de PM10 con JB.	65
Tabla 23. ARE y MARE Datos Estimados (E) de PM10 con JB.	65
Tabla 24. Calculo de RMSE para cada interpolador	66
Tabla 25. Frecuencias de los modelos de semivariograma teórico para PM2.5.	66
Tabla 26. Estadísticos de error de estimación promedio	66
Tabla 27. Datos Estimados de PM2.5 (E) para cada mes del año con HE ($\mu\text{g}/\text{m}^3$).	67
Tabla 28. Estadísticos de Error Datos Estimados (E) de PM2.5 con HE.	67
Tabla 29. ARE y MARE Datos Estimados (E) de PM2.5 con HE.	68
Tabla 30: Coeficientes de Correlación PM10.	69
Tabla 31. Criterios de especificación de los mejores modelos OLS en cada mes.	69
Tabla 32. Datos Estimados (E) de PM10 para cada mes del año con GWR ($\mu\text{g}/\text{m}^3$).	70
Tabla 33. Estadísticos de error datos estimados (E) de PM10 con GWR.	70
Tabla 34. ARE y MARE datos estimados (E) de PM10 con GWR.	70
Tabla 35. Coeficientes de correlación PM2.5.	72
Tabla 36. Criterios de especificación de los mejores modelos OLS en cada mes.	72
Tabla 37. Datos Estimados (E) de PM2.5 para cada mes del año mediante GWR ($\mu\text{g}/\text{m}^3$).	73
Tabla 38. Estadísticos de Error Datos Estimados (E) de PM2.5 con GWR.	73
Tabla 41. ARE y MARE Datos Estimados (E) de PM2.5 con GWR.	73
Tabla 40. Resultados pruebas de estacionariedad y causalidad para las variables explicativas de PM10.	76
Tabla 41. Mejores resultados de combinaciones de las variables explicativas para PM10.	77

Tabla 42. PM10 (E) para cada mes del año con modelo AR–RD de componente espacial de rezago distribuido GRW ($\mu\text{g}/\text{m}^3$).....	78
Tabla 43. Estadísticos de Error Datos Estimados de PM10 (E) con AR-RD GWR.....	79
Tabla 44. ARE y MARE Datos Estimados de PM10 (E) con AR-RD GWR.	79
Tabla 45. Estacionariedad y causalidad de las variables explicativas PM2.5.	80
Tabla 46. Mejores combinaciones de las variables explicativas para PM2.5.....	80
Tabla 47. PM2.5 (E) para cada mes del año con modelo AR RD de componente espacial de rezago distribuido GWR ($\mu\text{g}/\text{m}^3$).....	81
Tabla 48. Estadísticos de Error Datos Estimados por mes de PM2.5 (E) con AR RD GWR.	82
Tabla 49. ARE y MARE Datos Estimados por mes de PM2.5 (E) con AR RD GWR.....	82
Tabla 50: Resultados Caracterización Espacial y Espacio - Temporal de PM10.....	87
Tabla 51. Resultados Caracterización Espacial y Espacio - Temporal de PM2.5.....	87
Tabla 52. Resultado global de las estimaciones por mes.	87
Tabla 53. Estadísticos de error de estimación de PM2.5 por sitio de monitoreo.	90
Tabla 54. Resultado global de las estimaciones por sitio de Monitoreo.	90
Tabla 55. Resumen de los valores estimados de PM2.5 por sitio.	90
Tabla 56. Comportamiento intermensual de las variables T, ∇T y ∇Vv	92
Tabla 57. Tipología de los sitios de monitoreo de la Red de la Figura 40.	111
Tabla 58. Intervalos de confianza de los valores estimados en los sitios de monitoreo de la Red de la Figura 40 para el primer semestre.	112
Tabla 59. Intervalos de confianza de los valores estimados en los sitios de monitoreo de la Red de la Figura 40 para el segundo semestre.....	113

DEDICATORIA

A mí amada esposa e hijos por su solidaridad y ternura

AGRADECIMIENTOS

El autor agradece a los estudiantes de la Maestría en Geoinformática de la Universidad de San Buenaventura (Cohorte 1), de la Especialización en Sistemas de Información Geográfica de la Universidad de San Buenaventura (Cohortes 19 y 20) y de la Especialización en Medio Ambiente y Geoinformática de la Universidad de Antioquia (Cohorte 16) por el apoyo recibido en el procesamiento previo de los datos y la aplicación de la metodología.

INTRODUCCIÓN

En esta tesis doctoral se propone una metodología para caracterizar espacio-temporalmente la concentración de contaminantes particulados del aire en áreas urbanas intramontanas, tomando como caso de estudio el Valle de Aburrá (Antioquia, Colombia). La metodología integra aspectos de análisis espacial, análisis estadístico, análisis geoestadístico, econometría y sistemas de información geográfica, logrando resultados que son complementarios a los modelos físicamente basados para la comprensión y el modelamiento de los fenómenos de dispersión de contaminantes del aire. La tesis está organizada en ocho capítulos.

En el capítulo 1 se hace una revisión del estado del arte de los trabajos realizados en torno a la aplicación de algoritmos computacionales para la estimación de valores de la concentración de contaminantes del aire, haciendo especial énfasis en el material particulado, ya que este contaminante se usa como caso de estudio para la implementación de la metodología propuesta.

En el capítulo 2 se presenta el marco teórico necesario para el desarrollo de la metodología. Se trata el tema de los algoritmos de interpolación espacial, los algoritmos de regresión espacial, los algoritmos econométricos, la validación de datos estimados y los gradientes espaciales. Se hace una descripción de las consideraciones conceptuales y de los procedimientos operativos.

En el capítulo 3 se enuncia el problema a resolver y los objetivos de la tesis.

En el capítulo 4 se describe el área y los datos del caso de estudio y los datos que se usaron para la implementación de la metodología propuesta.

En el capítulo 5 se describe la metodología para la caracterización espacial de datos de concentración de contaminantes del aire usando algoritmos de interpolación y de regresión espacial. En este capítulo se busca describir las características espaciales de la concentración del material particulado de la región de estudio. Para el caso de los algoritmos de

interpolación espacial se hace énfasis en la descripción univariada y multivariada de la concentración del contaminante. Para el caso de los algoritmos geoestadísticos multivariados, se usan como variables explicativas para estimar la concentración del contaminante, la temperatura, la velocidad del viento y se calculan también los gradientes espaciales de éstas variables. Para el caso de los algoritmos de regresión espacial se hace énfasis en las consideraciones para la adecuada configuración de modelos de regresión y su solución con el método de mínimos cuadrados ordinarios y la técnica de regresión geográficamente ponderada. En cada sección se muestra esquemáticamente el componente metodológico a través de diagramas de flujo que ilustran de manera global los procedimientos para su desarrollo, se muestra la implementación de la metodología con datos de casos de estudio y se validan los resultados de los datos estimados de la concentración del material particulado con validación cruzada y el uso de diferentes estadísticos de error y de ajuste. Algunos aspectos fundamentales de la metodología de caracterización espacial han sido publicados en artículos en la Revista de la Facultad de Ingenierías de la Universidad de Antioquia en donde se usan interpoladores espaciales tipo Spline para la imputación de datos faltantes de material particulado de 10 micras (PM10) en la ciudad de Medellín (Londoño & Cañón, 2015a), en la Revista de Ingenierías de la Universidad de San Buenaventura Medellín en donde se hace una caracterización espacial de PM10 con algoritmos geoestadísticos en la ciudad de Medellín (Londoño et al., 2015), y en la Revista Politécnica del Politécnico Colombiano Jaime Isaza Cadavid en donde se ilustra la metodología para la aplicación de algoritmos de regresión espacial de usos del suelo en la estimación local de la concentración mensual de PM10 en Medellín (Londoño & Cañón, 2015b).

En el capítulo 6 se describe la metodología para la caracterización espacio - temporal de datos de concentración del material particulado usando algoritmos econométricos. En este capítulo se busca integrar la caracterización espacial y la temporal de los datos de concentración de material particulado de la región de estudio. En cada sección se muestra esquemáticamente el componente metodológico a través de diagramas de flujo que ilustran de manera global los procedimientos para su desarrollo, se muestra la implementación de la metodología con los datos del caso de estudio y se validan los resultados de datos estimados

de concentración con validación cruzada y el uso de diferentes estadísticos de error y de ajuste. Un aspecto importante en este capítulo es que se propone un modelo econométrico en el cual una de las variables explicativas es el dominio espacial de la concentración del material particulado obtenido de la Caracterización espacial con algoritmos geoestadísticos multivariados y con regresiones espaciales, siendo la de mejor desempeño ésta última.

En el capítulo 7 se hace un análisis de los resultados obtenidos en los dos capítulos anteriores y con el uso de los SIG se calculan los mapas de caracterización espacio-temporal de la concentración del material particulado. Los resultados se discuten en el contexto del comportamiento de las variables explicativas y se incluye la variable *distancia a fuentes de emisión y de mitigación del contaminante* en los mapas, con la aplicación de un modelo de proximidad a fuentes propuesto en esta tesis y publicado en la Revista de la Facultad Nacional de Salud Pública de la Universidad de Antioquia (Londoño et al., 2017).

En el capítulo 8 se hacen las consideraciones finales en donde se muestra que la concentración del material particulado es un proceso estacional que posee tres episodios en el año: en febrero-marzo en donde se dan los máximos valores de concentración, junio-julio en donde se dan los valores mínimos de concentración y en septiembre-octubre en donde la concentración aumenta nuevamente hacia el valor promedio anual.

1 ESTADO DEL ARTE

Los contaminantes del aire son mezclas complejas de gases y material particulado que está compuesto por partículas sólidas o líquidas suspendidas en el aire con diámetros menores a 10 micras (PM10) que se forman a partir de contaminantes primarios y secundarios. Los primarios son aquellos que permanecen en la atmósfera tal y como fueron emitidos. Dentro de ellos se encuentran: las partículas sólidas o líquidas, los gases de monóxido de carbono (CO), el dióxido de carbono (CO₂), los óxidos de nitrógeno (NO_x), los óxidos de azufre (SO_x) y los hidrocarburos. Los secundarios son aquellos primarios que han sufrido cambios químicos o son el producto de reacciones químicas entre dos o más contaminantes primarios, tales como el ácido sulfúrico (H₂SO₄), el ácido nítrico (HNO₃), el ozono (O₃), el smog fotoquímico y los compuestos orgánicos volátiles (COV_s). Como se muestra en la Tabla 1, los contaminantes también se clasifican por su estado de agregación. (Laumbach & Kipen, 2012).

Tabla 1. Clasificación de los contaminantes por su estado de agregación (Laumbach & Kipen, 2012).

Estado de Agregación	Designación	Ejemplos
Gaseoso	Gases	Óxidos de azufre Dióxido de carbono Monóxido de carbono Óxidos de nitrógeno Ozono
	Vapores	Alcoholes Aldehídos, cetonas Derivados halogenados Esteres Hidrocarburos alifáticos Hidrocarburos aromáticos
Líquido	Nieblas	Aceite mineral Ácidos clorhídrico, crómico, sulfúrico Hidróxido sódico
Sólido	Polvos	Asbestos, carbón, caolín Madera, óxidos metálicos, sílice
	Humos Humos metálicos	Asfalto, hidrocarburos policíclicos Aluminio, cadmio, cobre, cromo Estaño, hierro, manganeso Níquel, plomo, silicio, berilio

Los gases y las partículas poseen distintos niveles de concentración en la atmósfera debido principalmente a las condiciones meteorológicas, la topografía, los usos del suelo en zonas urbanas y la densidad del flujo vehicular (Singh et al., 2017; Rooney et al., 2012). En cuanto

a la meteorología, las variables que más inciden sobre la concentración son la velocidad del viento, la temperatura y la precipitación (Pérez et al., 2011). La topografía produce efectos de circulaciones cíclicas de vientos generadas por desfases entre los ángulos de los vectores gradientes de presión y temperatura con lo cual se pueden generar “microcuencas” en donde se aumenta la concentración por baja circulación del viento. En los usos del suelo asociados con la producción de la gran industria, la producción en consolidación y el uso industrial se presentan condiciones de alta concentración de contaminantes mientras que en las zonas verdes se atenúa (Marcon et al., 2014) y en las vías con alta densidad de flujo vehicular se aumenta la concentración de los contaminantes (Eeftens et al., 2013). La contaminación atmosférica es un tema de relevante preocupación a nivel mundial en los centros urbanos densamente poblados. Diversos estudios han demostrado que los contaminantes del aire tienen un efecto significativo sobre la salud (mortalidad y morbilidad) de los habitantes de las grandes ciudades, y en especial el material particulado de diámetro menor a 2.5 micras que constituye la denominada fracción inhalable de partículas y es capaz de llegar a la zona bronquio-traqueal (Pope & Dockery, 2006; Chow et al., 2006; Sanhueza et al., 2009). Éste material está compuesto por una variedad de sustancias sólidas y líquidas derivadas de fuentes naturales y actividades humanas que forman una mezcla compleja de carbono orgánico e inorgánico, metales, nitratos, sulfatos y fosfatos (Brunelli et al., 2007).

En Colombia, según el Ministerio de Ambiente y Desarrollo Sostenible (MinAmbiente), la contaminación atmosférica es el problema ambiental que más preocupa a sus habitantes. Al respecto, se han adoptado una serie de normas que buscan mantener una buena calidad del aire en el territorio nacional y en especial en los centros urbanos (MinAmbiente, 2017):

- Decreto 948 de 1995 en el que se define “*el marco de las acciones y los mecanismos administrativos de que disponen las autoridades ambientales para mejorar y preservar la calidad del aire, y evitar y reducir el deterioro del medio ambiente, los recursos naturales renovables y la salud humana ocasionados por la emisión de contaminantes químicos y físicos al aire; a fin de mejorar la calidad de vida de la población y procurar su bienestar bajo el principio del desarrollo sostenible*”.
- Decreto 979 de 2006 que modifica algunos artículos del Decreto 948 de 1995: artículo 7 (*clases de normas de calidad del aire y los distintos niveles periódicos de*

inmisión), 10 (niveles de prevención, alerta y emergencia por contaminación del aire), 93 (medidas para la atención de episodios), 94 (planes de contingencia por contaminación atmosférica) y 108 (clasificación de “áreas - fuente” de contaminación).

- Resolución 601 de 2006, "por la cual se establece la Norma de Calidad del Aire o Nivel de Inmisión, para todo el territorio nacional en condiciones de referencia".
- Resolución 650 de 2010, en la cual se define el Protocolo para el Monitoreo y Seguimiento de la Calidad del Aire.
- Resolución 2154 de 2010, mediante la cual se modifica el Protocolo para el Monitoreo y Seguimiento de Calidad del Aire, adoptado a través de la Resolución 650 de 2010.
- Resolución 651 de 2010, mediante la cual se crea “el Subsistema de Información sobre Calidad del Aire – SISAIRES, que hace parte del Sistema de Información Ambiental para Colombia SIAC, en lo referente a la información para el diseño, evaluación y ajuste de la política y las estrategias para la prevención y control de la contaminación del aire”.
- Resolución 2254 de 2017 que define los elementos fundamentales para la calidad del aire en Colombia y define los niveles máximos permisibles para contaminantes criterio como se muestra en la Tabla 2.

Tabla 2. Niveles máximos permisibles para contaminantes criterio (MinAmbiente, 2017).

Contaminante Criterio	Nivel máximo permisible ($\mu\text{g}/\text{m}^3$)	Tiempo de exposición
PST (Partículas Suspendidas Totales)	100	Anual
	300	24 horas
PM10	50	Anual
	100	24 horas
PM2.5	25	Anual
	50	24 horas
SO ₂	80	Anual
	250	24 horas
	750	3 horas
NO ₂	100	Anual
	150	24 horas

	200	1 hora
O ₃	80	8 horas
	120	1 hora
CO	10000	8 horas
	40000	1 hora

Los contaminantes criterio son monitoreados por las autoridades ambientales colombianas que se encuentran organizadas en 16 Corporaciones Autónomas Regionales (CAR) y 3 Autoridades Ambientales Urbanas (AAU), ubicadas principalmente en las ciudades de Bogotá, Medellín, Cali, Cartagena, Bucaramanga y en zonas de explotación minera de carbón del Cesar y de la Guajira, entre otros. El monitoreo se hace mediante 19 Sistemas de Vigilancia de la Calidad del Aire (SVCA) que poseen un total de 112 sitios de monitoreo diseñados y operados conforme a los criterios definidos por MinAmbiente. Esto corresponde a un monitoreo en 15 de los 32 departamentos y en 57 municipios y centros urbanos de los 1122 existentes en el país. Los contaminantes monitoreados prioritariamente por los SVCA son el material particulado PM10 y PM2.5 por sus altos efectos de corto y largo plazo sobre la salud (MinAmbiente, 2017).

Los SVCA evalúan la calidad del aire en una región determinada y para ello es necesario implementar redes de sitios de monitoreo que se diseñan para muestrear, analizar y procesar en forma continua durante un período, las concentraciones de los contaminantes presentes en el aire. Para la definición de una red de sitios de monitoreo, éstos deben estar adecuadamente localizados en el espacio y para ello se debe tener un conocimiento de las variables que afectan la dispersión de los contaminantes, de la normativa ambiental, de la distancia a las fuentes de emisión (Jerrett et al., 2005), de las condiciones y criterios de microlocalización (MinAmbiente, 2007) y de cómo se distribuyen las concentraciones del contaminante en el espacio a través de la implementación de campañas de mediciones preliminares (Nejadkoorki et al., 2011). Según la agencia de contaminación ambiental estadounidense EPA (EPA, 1994), para el diseño de una red de sitios de monitoreo deben tenerse en cuenta los objetivos y el área de influencia o la escala de medición como se muestra en la Tabla 3.

Tabla 3. Objetivos y escalas de medición.

Objetivos	Escala de medición
Estudio de zonas expuestas a altas concentraciones puntuales	Micro escala (1m – 100m)
Evaluación de los efectos de los contaminantes sobre la población	Meso escala (100m – 10km)
Determinación de la contaminación causada por varias fuentes	Micro escala con proximidad a los focos de emisión. (1m – 100m)
Establecer concentraciones de fondo y de transporte regional de contaminantes	Macro escala (más de 10km)

Según el Subsistema de Información sobre Calidad del Aire del Ministerio de Ambiente Vivienda y Desarrollo Sostenible de Colombia (SISAIRE) (MinAmbiente, 2007), los sitios de monitoreo se clasifican en tres niveles:

- Nivel 1. Según el área en donde se encuentra el sitio de monitoreo, el área hace referencia a la distribución o densidad de las edificaciones, se clasifican en urbanas (área totalmente urbanizada), suburbana (área parcialmente urbanizada mezclada con áreas no urbanizadas con usos agrícolas, bosque, lagos, etc.) y rurales (que no son de ninguno de los tipos anteriores).
- Nivel 2. Según el tiempo de monitoreo, se clasifican en fijos (permanece en un punto fijo por más de un año) e indicativos (permanece en un punto fijo por períodos menores a un año).
- Nivel 3. Según la influencia de los distintos tipos de fuentes emisoras, se clasifican en: de tráfico (la fuente emisora es el tráfico rodado cercano), de punto crítico (la fuente emisora es el tráfico rodado cercano, pero están ubicadas a nivel del suelo, muy usadas en estudios epidemiológicos), industriales (la fuente emisora son las zonas industriales) y de fondo (la fuente emisora no es directa si no la dispersión del contaminante por el régimen de vientos). Estos sitios de monitoreo a su vez pueden ser del nivel 1 y 2.

Adicionalmente a la clasificación anterior, en todos los sitios de monitoreo deberá indicarse la coordenada espacial, los objetivos de la medición, la tecnología de medición y algunos parámetros adicionales tales como:

- Para sitios de monitoreo urbanos y suburbanos: la población.

- Para sitios de monitoreo de tráfico: distancia (m) al bordillo, anchura de la calle o carretera entre bordillos, la longitud (m) entre fachadas de edificios, la altura (m) de las fachadas de los edificios (promedio de ambos lados de la calle en el área próxima a la estación), la intensidad media de tráfico (vehículos/día) en ambas direcciones expresada en promedio anual diario, la velocidad del tráfico (km/h) típico indicando franja horaria, la fracción de vehículos pesados (%) expresada en promedio diario a lo largo del año.
- Para sitios de monitoreo industriales: el tipo de industria, la distancia a las fuentes (km), la dirección a las fuentes (en grados).
- Sitios de monitoreo de punto crítico: la altura sobre el suelo, el tipo de fuente evaluada, el tipo de calle (calle encajonada o de valle o calle libre).
- Para sitios de monitoreo indicativos: el tiempo de monitoreo y las condiciones meteorológicas del monitoreo.
- Para sitios de monitoreo rurales de fondo: si son cercanos a ciudades o son regionales.

En general, las regulaciones normativas establecen marcos conceptuales y restrictivos aplicables a la problemática de la contaminación ambiental y la microubicación de los sitios de monitoreo (MinAmbiente, 2007) que para el caso de los centros urbanos definen como más adecuados los de tipo urbano, de monitoreo de tráfico, de monitoreo industrial, de monitoreo de punto crítico, de monitoreo de fondo y monitoreo fijo y/o indicativo. Excepto el monitoreo de fondo, los sitios de monitoreo se encuentran ubicados típicamente para detectar máximas concentraciones y aunque esta configuración es adecuada para identificar el potencial máximo de exposición y el riesgo, en muchos casos falla en la descripción de la caracterización espacial de la contaminación atmosférica debido a que éste es un fenómeno continuo en el espacio y estacional en el tiempo (Baklanov et al., 2007).

La configuración de las redes de monitoreo ha motivado a nivel mundial una serie de investigaciones que tienen como factor común la problemática asociada con la insuficiencia en la cantidad de datos en el espacio y en el tiempo para hacer una adecuada caracterización espacial y temporal del fenómeno de la contaminación del aire y hacer más efectiva la toma

de decisiones en cuanto a determinación de alertas tempranas que permitan la implementación de estrategias de mitigación de la contaminación (Diaz-Robles et al., 2008; Paschalidou et al., 2011). Actualmente se trabaja en el diseño y/o modificación de las redes de monitoreo de calidad de aire existentes para proponer metodologías que definan configuraciones óptimas para la micro y macro localización de los sitios de monitoreo que permitan hacer mediciones exhaustivas en el espacio y el tiempo y por tanto realizar mediciones espaciales de la concentración de contaminantes más confiables y rentables (Lozano et al., 2009; Mofarrah & Husain, 2010; Nejadkoorki et al., 2011). Al respecto, Li y Heap (Li & Heap, 2014), plantean que para la adecuada configuración de una red de sitios de monitoreo se deben tener en cuenta los siguientes factores:

- La densidad de muestreo, que está asociada con la cantidad de sitios de monitoreo por unidad de área ($D_m = \text{km}^2/\text{número de sitios de monitoreo}$).
- El diseño de muestreo, que está asociado con la distribución espacial de los sitios de monitoreo (regular, irregular, estratificado, por agregados, etc.), en función de la variabilidad espacial del contaminante.
- El espaciamiento entre los sitios de monitoreo que debe estar relacionado con la escala o las escalas de la variación del contaminante en el espacio (resolución espacial).
- La máxima distancia de autocorrelación espacial de la concentración del contaminante ya que una propiedad importante de la concentración es que sus valores son más similares entre las posiciones cercanas a un sitio de monitoreo que entre posiciones lejanas, es decir, el dato se correlaciona consigo mismo en función de la distancia (estructura espacial de los datos o rango).
- El tipo de superficie y paisajes, asociado con las heterogeneidades del medio físico que soporta el dato y que puede afectar la autocorrelación espacial de la concentración del contaminante (presencia de barreras físicas o cambios bruscos en la topografía).
- La correlación espacial entre la concentración del contaminante y otras variables, y su medición o no en el mismo sitio.
- La cantidad de sitios de monitoreo. Estudios al respecto han indicado que este es un factor importante para la adecuada definición de una red de monitoreo pero que debe

considerarse en interacción con los demás factores y no de manera aislada; y en especial en relación con la densidad de muestreo. Al respecto Li, y posteriormente Bennett y sus colaboradores (Li & Heap, 2011; Bennett et al., 2013), encontraron que en la medida en que se aumente la cantidad de sitios de monitoreo por unidad de área, los diferentes métodos de caracterización espacial del fenómeno de la contaminación tienden a tener el mismo desempeño. Esto es consistente con lo que Hengl planteó en 2007 indicando que, en la práctica, hay un umbral más allá del cual cualquier aumento en el tamaño de la muestra no mejoraría significativamente la caracterización espacial de las variables ambientales, sugiriendo en esa época que los estudios futuros deberían centrarse más en la calidad del muestreo y en la calidad de los predictores ambientales auxiliares, en lugar de hacer más observaciones (Hengl, 2007).

Según la EPA (EPA, 1998), la medición en un sitio de monitoreo se puede hacer:

- Mediante puntos de medición fija usando analizadores automáticos, captadores activos no automáticos o captadores difusivos.
- Mediante unidades móviles y tubos difusivos que se ubican en las zonas de máxima concentración o se distribuyen según un modelo de retícula.
- Mediante Sensores Remotos.
- Mediante Bioindicadores.

El problema de la insuficiencia en la cantidad de datos en el espacio para hacer una adecuada caracterización espacial del fenómeno de la contaminación del aire, ha sido abordado en los últimos 10 años usando algoritmos computacionales cuyo objetivo es estimar el valor de la concentración de un contaminante en algún sitio en donde no se tengan mediciones (Alam & McNabola, 2015). Los algoritmos computacionales han tenido tres enfoques: determinístico, estocástico e interpolación espacial (Shahraiyni & Sodoudi, 2016).

El enfoque determinístico se basa en el principio de la causalidad y para estimar el valor de una variable usa ecuaciones de conservación de masa y los efectos de los procesos dinámicos que influyen en el equilibrio de masa y térmico en un volumen de control como

es el caso de los modelos gaussianos (Daly & Zannetti, 2007), que implican numéricamente resolver un conjunto de ecuaciones diferenciales. Este enfoque requiere un conocimiento completo de las fuentes de contaminación, los inventarios de emisiones y su composición química y los procesos físicos-químicos de la dispersión de la concentración del contaminante en la atmósfera. Éste enfoque es muy bueno para estimar los eventos más frecuentes con una precisión razonable, pero no son capaces de estimar con exactitud los eventos extremos (Stern et al., 2008). Toda la información que requieren sobre el origen de los contaminantes y otros parámetros, con frecuencia no está disponible y estos parámetros deben ser estimados o simplemente ignorados debido a la falta de información lo que puede dar lugar a sesgos y errores significativos en las estimaciones de estos métodos (Borrego et al., 2008). Adicionalmente son matemáticamente complejos y requieren de alta capacidad computacional para su implementación.

El enfoque estocástico se basa en la descripción de relaciones estadísticas entre la concentración del contaminante y variables explicativas. Operativamente se estima el valor de una variable, implementando regresiones lineales o no lineales con el algoritmo de mínimos cuadrados ordinarios (Goyal et al., 2006) y para el caso de datos espaciales, su equivalente, mediante el algoritmo de Regresión Geográficamente Ponderada (Geographically Weighted Regression – GWR) (Fotheringham et al., 2002). Algunas variables explicativas usadas en este enfoque han sido: la concentración de otros contaminantes (Deligiorgi & Philippopoulos, 2011), variables meteorológicas (Ramsay et al., 2011; Dadvand et al., 2011), variables de uso del suelo, distancia a vías, distancia a fuentes de emisión, densidad de tráfico; entre otras (Modelos de Regresión de Usos del Suelo-Land Use Regression LUR) (Lee et al., 2017; Alam & McNabola, 2015; Londoño & Cañon, 2015b; Dons et al., 2014; Eeftens et al., 2013; Clougherty et al., 2013; Rooney et al., 2012; Gulliver et al., 2011; Beelen et al., 2010). Para el caso de la concentración del material particulado, los estudios han encontrado que las variables explicativas más importantes han sido en su orden, la velocidad del viento que tiene una relación directa con los procesos de dispersión atmosférica del material particulado y la temperatura del aire que afecta la movilidad de los sistemas de alta presión de la atmósfera y la turbulencia de energía cinética que influye en la altura de la capa de mezcla que afecta la concentración del material

particulado (Papanastasiou et al., 2007; Sayegh et al., 2014). Otras variables explicativas consideradas para la variación de corto plazo en zonas urbanas han sido la densidad de flujo vehicular y en caso de no tener información al respecto, algunos autores las han reemplazado por la concentración de SO₂ y NO_x (Chen L. et al., 2010). Adicionalmente se ha encontrado que la presencia de zonas verdes y cuerpos de agua afectan la concentración del material particulado en la medida en que impiden su resuspensión (Taheri Shahraini et al., 2015). En general los algoritmos del enfoque estocástico permiten realizar estimaciones locales, tienen buen desempeño y no requieren mucha información; pero no consideran la física detrás de los datos y además el algoritmo aplicado para un área de estudio no siempre es aplicable a otras.

El enfoque de la interpolación espacial se basa en el principio de autocorrelación espacial que exhiben las variables ambientales (en la cercanía a un sitio de monitoreo, el valor de la variable es similar y en la medida en que se aleja del sitio el valor difiere más). Operativamente, la mayoría de los algoritmos estiman el valor de la variable en un punto en donde no se tiene medición con base en una función que es el promedio ponderado de los valores medidos en otros puntos (Sangalli et al., 2013). Estudios recientes (Li & Heap, 2014), clasifican los algoritmos de interpolación espacial más usados en la estimación de variables ambientales en *algoritmos no geoestadísticos* y *algoritmos geoestadísticos*, según la forma en cómo se calculan los pesos ponderantes (ver Tabla 4).

Tabla 4. Clasificación de los Algoritmos de Interpolación Espacial.

No Geoestadísticos	Geoestadísticos	
	Univariados	Multivariados
Vecino más Cercano (Nearest neighbours - NN)	Kriging Simple (Simple Kriging - SK)	Kriging Universal (Universal Kriging - UK) Kriging con Deriva Externa (Kriging with an external drift - KED)
Interpolaciones relacionadas con Red de Triángulos Irregulares (Triangular Irregular Network - TIN)	Kriging Ordinario (Ordinary Kriging - OK)	CoKriging Ordinario (Ordinary CoKriging - OCK)
Vecinos Naturales (Natural Neighbours - NaN)	Kriging Indicador (Indicator Kriging - IK)	CoKriging Ordinario Estándarizado (Standardised OCK - SOCK)
Inverso de la Distancia (Inverse Distance Weighted - IDW)	Kriging Disyuntivo (Disjunctive Kriging - DK)	Kriging con Componentes Principales (Principal Component Kriging - PCK)
Splines y Superficies de Tendencias Locales (Splines and Local Trend Surfaces - LTS)		Kriging Probabilístico (Probability kriging - PK)

Los algoritmos de la Tabla 4 parten de suposiciones particulares y pueden ser univariados (U) o multivariados (M); determinísticos (D) o estocásticos (E); locales (L) si el valor

estimado se calcula con base en una vecindad de datos medidos o globales (G) si se calcula con todos los datos medidos; exacto (Ex) o inexacto (I); suave (S) o no suave (NS), según su nivel de desempeño y la forma de la superficie generada; convexo (C) o no convexo (NC), si genera resultados entre el valor mínimo y máximo del conjunto de datos. Además, tienen limitaciones, requerimientos de procesamiento computacional moderados (Spline y Kriging) a bajos y estructuras de datos de salida diferentes. En general el proceso de interpolación es rápido y puede operar a partir de datos dispersos en muestras regular o irregularmente espaciadas. En la Tabla 5 se muestran las características de los algoritmos de interpolación espacial más usados para la estimación de variables ambientales, siendo los de mejor desempeño el inverso de la distancia, el Spline y el kriging (Tuna & Buluç, 2015).

Tabla 5. Características de los Algoritmos de Interpolación Espacial.

Algoritmo	Suposición	U/ M	D/ E	L/ G	Ex/ I	S/ NS	C/ NC	Limitaciones	Salida
Vecino más Cercano	El mejor predictor local es el punto de datos más cercano	U	D	L	I	NS	C	El dato estimado por polígono es el mismo.	Polígonos o superficie con rejilla
Red de Triángulos Irregulares	El mejor predictor local son los puntos de datos en el triángulo circundante.	U	D	L	I	NS	C	El patrón de la red de triángulos depende de la distribución geométrica de los datos. Solo usa una vecindad de tres datos para hacer la estimación	Triángulos o Superficie con rejilla
Vecinos Naturales	Los mejores predictores locales son puntos de datos en los polígonos circundantes	U	D	L	I	NS	C	Solo usa como vecindad, los datos de los polígonos circundantes.	Superficie con rejilla
Inverso de la Distancia	Los mejores predictores locales son puntos de datos en una vecindad.	U	D	L	Ex	S	C	Los resultados dependen del tamaño de la ventana de búsqueda	Superficie con rejilla o contornos
Spline	El mejor predictor local son los puntos de datos más cercanos sobre la superficie de tendencia	U	D	L	Ex	S	NC	El dato estimado está en función de la superficie de tendencia	Superficie continua en cuadrícula
Kriging	Requiere que los datos medidos cumplan estacionariedad estadística.	U/ M	E	L	Ex	S	C	La estimación depende de la distribución de probabilidad de los datos.	Superficie en rejilla

Con relación al problema de la insuficiencia en la cantidad de datos en el tiempo para hacer una adecuada caracterización temporal del fenómeno de la contaminación del aire, en los últimos años se han desarrollado una serie de estudios que usan algoritmos computacionales para estimar y/o pronosticar el valor de la concentración de un contaminante en algún período en donde no se tengan mediciones (Sayegh et al., 2014). Éstos algoritmos

computacionales han tenido tres enfoques principales (Shahraiyni & Sodoudi, 2016): Regresión Lineal Multivariada (Multivariate Linear Regression – MLR), Redes Neuronales Artificiales (Artificial Neural Networks – ANN) y Econometría.

El enfoque MLR ha sido ampliamente utilizado para el pronóstico de material particulado en centros urbanos (Goyal et al., 2006). En un análisis hecho por Shahraiyni y Sodoudi (Shahraiyni & Sodoudi, 2016) de 36 aplicaciones del enfoque MLR en diferentes centros urbanos del mundo realizadas entre 1996 y 2014, la escala temporal de pronóstico más usada fue el día seguida de la hora y el mes, para series de tiempo entre 1 y 6 años de datos. En 18 ciudades se utilizó solamente 1 sitio de monitoreo y en las demás osciló entre 2 y 36 sitios de monitoreo, los resultados mostraron valores del coeficiente de correlación (ρ) entre 0,05 y 0,4 y valores de la raíz cuadrada del error cuadrático promedio (Root Mean Square – RMSE) entre 0,7 y 57. Las variables explicativas usadas fueron: distancia a vías, presión atmosférica, temperatura, velocidad y dirección del viento, precipitación, humedad relativa y densidad de tráfico. En general este enfoque muestra un buen desempeño, pero tiene un problema de precisión debido a la representación lineal de sistemas no lineales y muestra falencias para el pronóstico de los valores extremos de la variable (episodios). Adicionalmente se puede presentar multicolinealidad entre las posibles variables explicativas que se ha superado en algunos casos mediante análisis previos de componentes principales.

El enfoque ANN se ha utilizado para el pronóstico de la concentración de una amplia gama de contaminantes a diferentes escalas de tiempo con muy buenos resultados (Şahin et al., 2011). Permite incorporar un gran número de variables heterogéneas debido a su capacidad para simular problemas complejos con comportamiento no lineal. En comparación con el enfoque determinista, requiere menos datos de entrada y de tiempo de cálculo. Con respecto al enfoque estocástico no requiere presupuestos sobre la distribución de probabilidad de los datos. En relación con el enfoque MLR, tienen un mejor desempeño y permite hacer pronóstico de eventos extremos (episodios). El problema de este enfoque es que no es aplicable para otro período y lugar diferente al de los datos (Pilsung, 2013).

El enfoque econométrico ha sido ampliamente utilizado para el pronóstico de material particulado en centros urbanos. En este enfoque se pronostica el valor de la variable (llamada variable respuesta) a partir de otras variables explicativas (o regresoras). Si dentro de las variables explicativas no está la variable respuesta con un rezago en el tiempo, el enfoque se denomina de Rezago Distribuido (RD), si solo está la variable respuesta con un rezago en el tiempo el enfoque se denomina Autorregresivo (AR); y si está la variable respuesta con un rezago en el tiempo y otras variables explicativas, se denomina Autorregresivo con Rezago Distribuido (AR-RD) (Zhang, 2003). Operativamente éste enfoque usa los algoritmos ARMA (auto regresivo integrado de media móvil) y ARIMA (auto regresivo integrado de media móvil). En las aplicaciones hechas de estos algoritmos en los últimos años, se encuentra que la escala temporal de pronóstico más usada es el día seguida del mes y la hora, para series de tiempo entre uno y seis años de datos, usando entre uno y cuatro sitios de monitoreo y los resultados de los valores del coeficiente de determinación (R^2) están entre 0,5 y 0,9 y valores de la raíz cuadrada del error cuadrático promedio (Root Mean Square – RMSE) entre 8 y 29. Las variables explicativas más usadas han sido: velocidad del viento, temperatura, humedad relativa, variables de tráfico, CO, NO₂, SO₂ y la variable respuesta con un rezago en el tiempo (Pope & Wu, 2014; Shahraiyini & Sodoudi, 2016). Stadlober y otros mostraron que para el pronóstico del material particulado en centros urbanos, el rezago en el tiempo de la variable respuesta era la variable explicativa más significativa (Stadlober et al., 2008). Los algoritmos de éste enfoque, tienen un buen desempeño y permiten hacer pronóstico de eventos extremos (episodios). El problema de este enfoque es la representación lineal de sistemas no lineales y el requerimiento de datos históricos continuos.

En América del Sur se destacan algunos trabajos que han contribuido de forma significativa a la caracterización espacial de la concentración de contaminantes en centros urbanos:

- El modelo de Schmitz (2004), modelo CADM (The Chilean Air Pollution Dispersion Model) para evaluar el comportamiento espacial y temporal de las características físicas y químicas de los contaminantes del aire.

- El modelo de Dias de Freitas (Dias de Freitas et al., 2005), modelo RAMS (Regional Atmospheric Model System) aplicado para estimar la concentración de ozono y otros contaminantes en el área urbana de Sao Paulo.

A nivel de la ciudad de Medellín y su Área Metropolitana, se destaca la contribución de los siguientes trabajos:

- Jiménez & Mejía (2002), Toro & Marín (2006) y Bedoya & Martínez (2009) realizaron intentos aislados por estudiar los mecanismos de circulación de masas de aire, los cuales, en general, apuntan principalmente hacia la problemática de la calidad del aire en el área urbana asociada a procesos industriales y tráfico de automotores. En sus estudios se usan fuentes de datos georeferenciados.
- Gómez et al. (2007), hicieron una descripción de las fuentes de emisión de la contaminación del aire en la zona central de la ciudad y determinaron que 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. Utilizaron 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 período determinado. Con esta prueba estimaron los parámetros de la distribución Gamma mediante el método de máxima verosimilitud. Como algoritmo de convergencia usaron el algoritmo EMA (Estimation Maximization Algorithm). La estimación de la media de los factores de contribución de las vías según el modelo MCF fue inmediata a partir de estas estimaciones y validó el factor de contribución de las vías a la contaminación atmosférica en la zona de estudio.
- Los reportes presentados por la red de calidad del aire de la ciudad (RedAire) muestran que frecuentemente las concentraciones de los contaminantes exceden los estándares nacionales para partículas suspendidas totales (PST), PM10 (estándar de la EPA), monóxido de carbono (CO) y ozono (O₃). Por lo que existe una urgente preocupación por parte de las autoridades locales de revisar los estándares de calidad del aire e implementar estrategias de control y reducción de la contaminación.

- A partir del año 2005, se han desarrollado trabajos que buscan implementar modelos numéricos de simulación y predicción numérica de la concentración de contaminantes similares al modelo RAMS (Corantioquia, 2005; AMVA, 2007; AMVA, 2008).
- A partir del año 2011, en el Grupo de Investigación en Gestión Ambiental (GIGA) de la Universidad de Antioquia, se desarrollan una serie de investigaciones dentro de las cuales se destacan los trabajos inéditos de Hernández sobre la simulación de los campos de viento locales con alta resolución espacial para situaciones meteorológicas típicas, fundamentales para el estudio del fenómeno de dispersión de contaminantes y el trabajo de Correa sobre la modelación de la dinámica atmosférica y la dispersión de contaminantes (material particulado) en una región tropical con topografía compleja. Igualmente, los trabajos de Rendón que muestran mediante simulaciones numéricas que los cambios en el uso de la tierra causados por la urbanización, alteran la dinámica de las inversiones de temperatura, afectando así la calidad del aire en un valle urbano (Rendón et al., 2014). Jiménez demostró que los efectos urbanos alteran significativamente las condiciones atmosféricas en superficie y la altura de la capa límite de mezcla lo cual incide sobre la calidad del aire (Jiménez, 2016).

2 MARCO TEÓRICO NECESARIO PARA EL DESARROLLO DE LA METODOLOGÍA

Este capítulo tiene cinco secciones y busca desarrollar los aspectos conceptuales y procedimentales necesarios para el desarrollo de la metodología para la caracterización espacio-temporal de la concentración de material particulado en el aire en los siguientes capítulos. En la sección 2.1 se conceptúa sobre el dato espacial y el principio de autocorrelación espacial, se describen distintos tipos de algoritmos de Interpolación Espacial haciendo énfasis en el Spline, el Inverso de la Distancia (IDW) y en los aspectos metodológicos operativos de los algoritmos geoestadísticos univariados (Kriging) y multivariados (Cokriging). En la sección 2.2 se describen los criterios para la especificación adecuada de modelos de regresión y las características de los algoritmos de regresión espacial (Regresión Geográficamente Ponderada - GWR). En la sección 2.3 se hace una discusión en torno a los algoritmos econométricos: series de datos de panel, pruebas preliminares a series de tiempo, las características de los modelos Autorregresivos (AR), de Rezago Distribuido (RD), Autorregresivos con Rezago Distribuido (AR-RD); y un modelo propuesto en la tesis que incluye como variable explicativa los datos funcionales de la concentración del contaminante obtenidos a través de la metodología de caracterización espacial multivariada. En la sección 2.4 se describe el método de validación cruzada y los estadísticos de error usados para la validación de los datos estimados que resultan de la implementación de los aspectos de las secciones 2.2 a 2.3. Finalmente, en la sección 2.5 se definen las características de las funciones de gradiente espacial, útiles para el cálculo de la variación espacial como método para la obtención de variables explicativas complementarias.

2.1 Algoritmos de interpolación espacial

Un dato espacial $Z_i(x_i, y_i)$, es aquel que está definido en un dominio espacial $S(x, y)$ con x e y en algún sistema de proyección geográfico, contiene información sobre una variable objeto de estudio (value), una geometría asociada que puede ser del tipo punto, línea o polígono que representa un objeto en el espacio, un origen de coordenadas geográficas, una estructura

de almacenamiento de datos que puede ser del tipo matricial (*raster*) o vectorial y un código de colores (RGB) para representar la variación del dato en el espacio. Una propiedad importante del dato espacial es que cumple con el principio de autocorrelación espacial lo cual implica que los datos son más similares entre posiciones cercanas que entre posiciones lejanas, es decir, el dato se correlaciona consigo mismo en función de la distancia. A partir del principio de autocorrelación espacial, se puede estimar el valor de un dato espacial conociendo su valor en distintos sitios del espacio.

La interpolación espacial es un proceso matemático algorítmico que permite estimar el valor de un dato espacial de una variable Z definida en un dominio espacial $S(x, y)$ con x e y en algún sistema de proyección geográfico. En algunos puntos (x_i, y_i) se conoce el valor de Z , denominado $Z_i(x_i, y_i)$, pero en otros puntos (x_0, y_0) se desconoce el valor de Z , denominado $Z_0(x_0, y_0)$ (ver Figura 1). Una característica de Z en S es que entre más cercanos estén dos puntos entre sí, más similares serán sus valores; lo que se denomina principio de autocorrelación espacial (Tobler, 1970); por lo tanto una forma de estimar $Z_0(x_0, y_0)$ es a partir de una combinación lineal de $Z_i(x_i, y_i)$ mediante la Ecuación 1 en donde los λ_i son pesos ponderantes (Burrough & Mcdonnell, 1998).

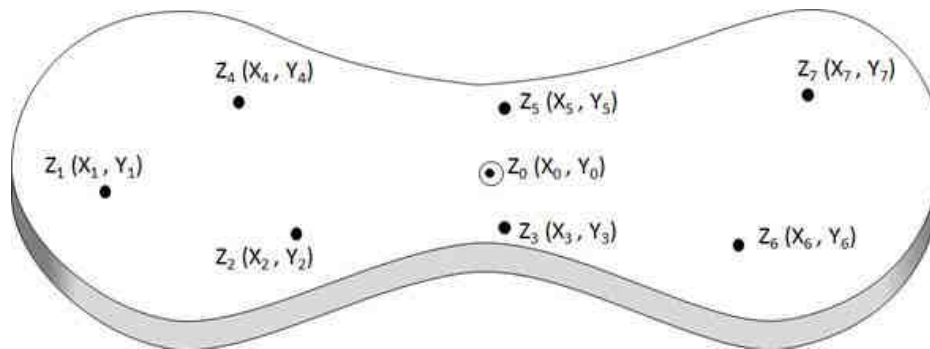


Figura 1. Problema de la Interpolación espacial. Fuente propia.

$$Z_0(x_0, y_0) = \sum_{i=1}^n \lambda_i Z_i(x_i, y_i) \quad \text{Ecuación 1}$$

Los algoritmos de interpolación espacial se clasifican con base en la forma cómo se calculen los pesos ponderantes, según la cantidad de datos usados para la estimación (globales si usa

todos los datos y locales si usa parte de los datos en una vecindad determinada) y según el valor del error del método (exacto e inexacto que depende de la naturaleza del fenómeno bajo estudio, la calidad y cantidad de los datos observados, la existencia o no de correlación espacial, el tipo de muestreo y la distribución espacial de los datos observados sobre el territorio) (Lam, 1983).

Otra forma de estimar $Z_0(x_0, y_0)$ es a partir de métodos de triangulación y ajustes de curvas con funciones de base radial como el Spline. Estos métodos no usan la Ecuación 1. En la Tabla 6 se muestra una clasificación de los métodos de interpolación espacial (Londoño & Valdés, 2015).

Tabla 6. Algoritmos de interpolación espacial

Algoritmo	Uso del estimador de la ecuación 1	Según cantidad de datos	Según valor del error
Polígonos de Thiessen	Si	Local.	Exacto para datos no conglomerados
Celdas Inverso de la distancia (IDW)	Si	Local	Exacto para datos no conglomerados Inexacto cuando hay valores extremos.
Geoestadístico	Si	Local, pero con semivariograma global.	Exacto pero dependen de la distribución de probabilidad de los datos y que éstos cumplan con el principio de estacionariedad.
Triangulación de Delanay	Si	Local	Exacto para distancias cortas.
Triangulación	No	Local	Exacto para distancias cortas.
Splines	No	Local	Exacto pero con problemas de acotamiento en la estimación de los valores mínimo y máximo del conjunto de datos.

En esta tesis se usaron los algoritmos IDW y Spline por ser los de mejor desempeño para variables ambientales según los reportes de la literatura (Tuna & Buluç, 2015).

El algoritmo del *Inverso de la distancia* (*Inverse Distance Weighted, IDW*) calcula el valor por estimar mediante la Ecuación 2.

$$Z_0 = \frac{\sum_{i=1}^n \frac{1}{(d_i)^p} * Z_i}{\sum_{i=1}^n \frac{1}{(d_i)^p}} \quad \text{Ecuación 2.}$$

Donde cada d_i , representa la distancia de cada una de las ubicaciones de los datos conocidos al punto por estimar. El valor usual del exponente p es 2.

El algoritmo *Spline* ajusta funciones polinómicas de base radial de manera similar a los procesos de interpolación global mediante modelos de regresión, pero de manera local. Con respecto a los demás algoritmos de la Tabla 3, presentan las siguientes ventajas:

- Operan considerando sólo los datos observados y su distribución espacial (Lennon & Turner, 1995).
- No dependen del tipo de distribución estadística de los datos observados.
- Los resultados obtenidos solo dependen de un parámetro de tensión similar al comportamiento de una membrana tensa (Spline tensión) o aflojada (Spline regularized) que pasa por los puntos de los datos observados.
- Los valores estimados pueden ser mayores y menores que los valores máximos y mínimos observados, por lo tanto, pueden ser más eficientes que los otros métodos en la estimación de valores en las fronteras de la región de estudio.
- No dependen de la forma de la red de muestreo como los métodos de Celdas y de polígonos de Thiessen, aunque lo ideal sería poder tener una red de muestreo con un diseño acorde con métodos de muestreo basados en el conocimiento previo de la estructura de variación de la variable a estimar.
- Requieren de pocos datos para realizar la estimación, aunque es claro que una mayor cantidad de datos mejora los resultados de cualquier método de interpolación espacial (Londoño & Cañón, 2015a).

Las desventajas de estos métodos son las siguientes:

- Su exactitud está en función de la resolución espacial de la estructura de datos que representa los valores observados, concretamente para el caso de estructuras matriciales en formato raster comunes en los SIG.
- La obtención del parámetro de tensión adecuado para la delimitación de los resultados del método entre el valor mínimo y máximo observado, es un proceso iterativo de ensayo y error. Los valores generados en la frontera de la región de estudio, tienden a desbordarse.

Un mapa de estimación con el algoritmo *Spline Tension* se obtiene mediante la Ecuación 3 (Burrough & McDonnell, 1998).

$$Z_0(x_0, y_0) = a_1 + \sum_1^n \lambda_i R(r_i)$$

Ecuación 3

$$R(r_i) = \frac{1}{2\pi\phi^2} \left[\ln\left(\frac{r_i}{2} \phi\right) + c + K_0(r_i \phi) \right]$$

Dónde:

$Z_0(x_0, y_0)$ es el dato estimado

n es la cantidad de puntos.

a_1 y λ_j son coeficientes determinados por la solución de un sistema de ecuaciones lineales.

r_i es la distancia del punto (x_0, y_0) al punto (x_i, y_i) .

ϕ^2 es un parámetro de peso.

K_0 es la función de Bessel modificada.

c es una constante igual a 0,577215.

En los algoritmos Inverso de la Distancia (IDW) y Spline, se evidenció que las estimaciones obtenidas por ambos están afectadas por la densidad de muestreo de los datos ($D_m = \text{km}^2/\text{número de sitios de monitoreo}$) del tal manera que en la medida en que la densidad de muestreo aumenta, se disminuye el desempeño de la estimación (Li & Heap, 2014).

El algoritmo ***Geoestadístico Univariado (Kriging)*** modela el cambio de la varianza de los datos con la distancia para calcular los pesos ponderantes de la Ecuación 1 y realizar la estimación de datos espaciales. Es extremadamente confiable pero su aplicación depende de la distribución de probabilidad de los datos, ya que estos deben cumplir con el principio de estacionariedad, es decir, se deben distribuir normalmente en el espacio, lo cual los hace altamente sensibles a la presencia de valores extremos y “*outliers*” en los datos. Adicionalmente dependiendo del nivel de escala del estudio (micro, meso o macro) se debe tener una baja densidad de muestreo ($\text{km}^2 / \text{Número de sitios de Monitoreo}$) (Li & Heap,

2014). Cuando la muestra de datos tiene pocos datos o alta densidad de muestreo, se sugiere usar previamente un algoritmo de interpolación espacial para densificar los datos en el área de estudio. El método Geoestadístico se denomina Kriging y por la forma en que se calcula los pesos ponderantes de la Ecuación 1, ésta se convierte en un estimador lineal sin sesgo y con mínima varianza (Best Linear Unbiased Estimation – BLUE) conocido también como filtro de Wiener. Los pasos generales del algoritmo son: un análisis exploratorio de los datos, un análisis estructural o variografía, la interpolación o estimación espacial (método del kriging) y finalmente la validación de la estimación (método de validación cruzada “*leave one out*”) (Pollice & Jona, 2009).

El análisis exploratorio de los datos permite determinar si los datos son estacionarios, es decir, si se distribuyen normalmente en el espacio lo cual implica que el valor de la media no varíe en el espacio (en diagramas de dispersión del dato contra las coordenadas no debe haber tendencias); para ello se realizan los siguientes procedimientos:

- Calcular las medidas de tendencia central (media, moda, mediana). En datos normalmente distribuidos éstas son similares.
- Hacer el histograma. Para datos normales, éste debe ser monomodal y simétrico con respecto a la media.
- Calcular el coeficiente de sesgo (skewness). Para datos normales el valor absoluto del skewness debe ser menor o igual a 0,5, si no, se usan transformaciones según los criterios de Webster y Oliver: si $0,5 < |Skewness| \leq 1$ se aplica una transformación raíz cuadrada al conjunto de datos; si no se aplica una transformación logarítmica (Webster & Oliver, 2001). Estas transformaciones se aplican de manera excluyente y no secuencial.
- Las medidas de tendencia central y el skewness, están afectadas por valores extremos, para los conjuntos de datos en los cuales las transformaciones anteriores no mejoran la estacionariedad, se realiza un análisis y detección de valores extremos. Éstos se retiran de los datos y se realiza nuevamente el análisis exploratorio, como se muestra en la parte inferior izquierda de la Figura 2, una vez determinado el conjunto de datos estacionarios, se hace el análisis estructural de los datos construyendo

primero el semivariograma experimental y luego ajustando un modelo de semivariograma teórico.

En la Figura 2, se propone una metodología para el análisis exploratorio en relación con el tratamiento del coeficiente de sesgo y de los valores extremos en el algoritmo geoestadístico realizada por Quiroz y Londoño (Quiroz et al., 2015).

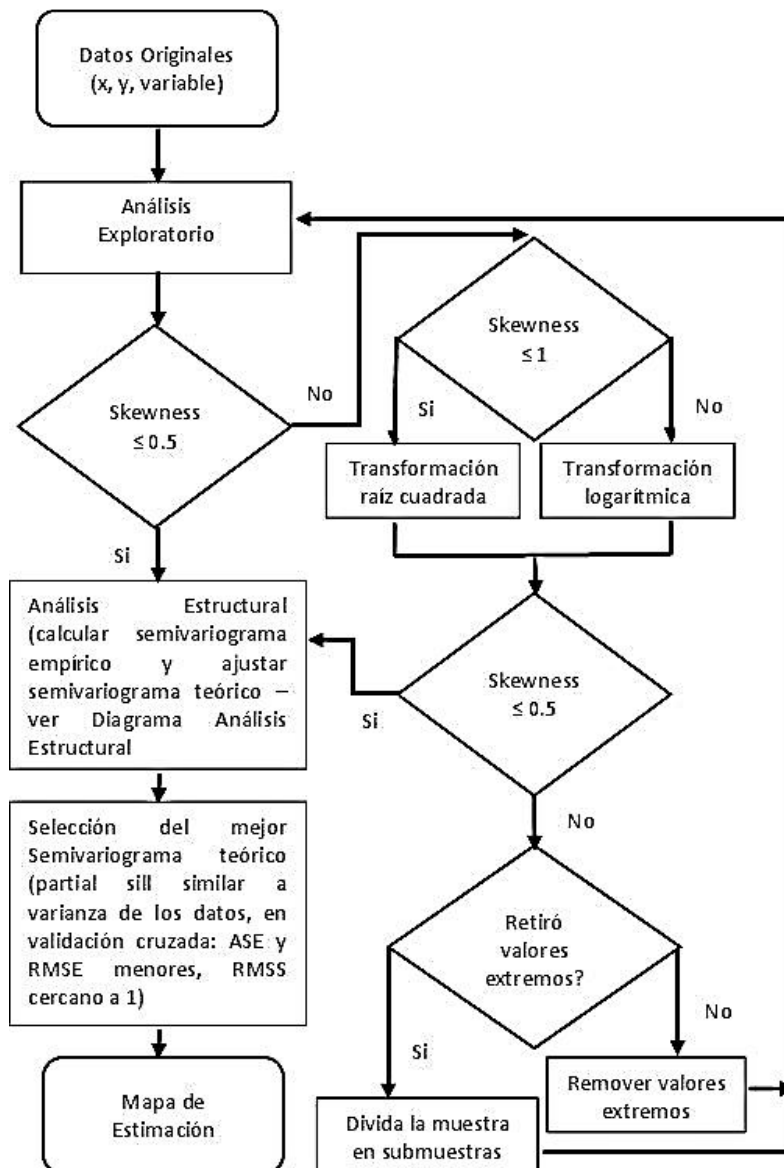


Figura 2. Metodología para tratamiento de Skewness y valores extremos.

ASE (Average Standard Error) es el error estándar promedio, RMSE (Root Mean Square error) es la raíz del error cuadrático promedio y RMSS (Root Mean Square error Standardized) es el RMSE estandarizado.

El análisis estructural, o variografía de los datos, busca determinar un modelo matemático que represente el cambio de la varianza de los datos con la distancia. Éste análisis se realiza en dos pasos:

- Se calcula el semivariograma empírico que es un gráfico de distancia euclidiana (h) contra semivarianza empírica (γ^*) que se calcula mediante la Ecuación 4 (Goovaerts, 1997).

$$\gamma^* = \frac{\Delta Z_i^2}{2} \quad \text{Ecuación 4}$$

En donde

$\Delta Z_i^2 = (Z_{i+h} - Z_i)^2$, con Z_i el valor del dato espacial Z en las coordenadas (x_i, y_i) y Z_{i+h} el valor del dato espacial Z en las coordenadas (x_{i+h}, y_{i+h}) . En la Figura 3 se propone una metodología para el cálculo del semivariograma empírico.

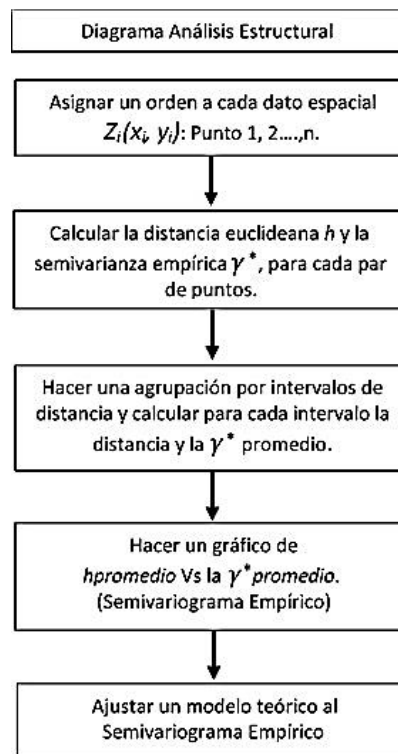


Figura 3. Metodología para el cálculo del semivariograma empírico. Fuente propia.

- Se ajusta un modelo teórico (γ) al semivariograma empírico (γ^*). En el aplicativo *Geostatistical Analyst*® de la Tecnología ArcGis® de ESRI se dispone de nueve modelos de semivariogramas teórico: *Spheric (SP)*, *Gaussian (GS)*, *Exponential (EX)*, *Tetraspheric (TS)*, *Pentaspheric (PS)*, *Rational Cuadratic (RC)*, *Hole Effect (HE)*, *J-Bessel (JB)* y *K-Bessel (KB)* (Johnston et al., 2003). En teoría un semivariograma es una función monótona creciente que alcanza un valor límite asintótico denominado meseta (*partial sill* - θ_s - sus unidades son las del dato al cuadrado) equivalente a la varianza de los datos (σ^2). La meseta se alcanza en un valor de distancia (h) llamado rango o alcance (*major range* - θ_r - con unidades de longitud). El rango determina una distancia de influencia en torno a un punto más allá de la cual la autocorrelación espacial es nula, pero no todos los semivariogramas alcanzan una meseta. El semivariograma en el origen es cero; pero puede mostrar un valor positivo (discontinuidad en el origen) dando lugar al denominado efecto pepita (nugget - θ_n - sus unidades son las del dato al cuadrado), cuando esto ocurre indica que la variable bajo estudio posee valores inferiores al valor mínimo de los datos de la muestra y mide simultáneamente las microregionalizaciones del fenómeno y los errores de mediciones de la variable. En la Figura 4 se muestra la forma típica y general de un semivariograma teórico.

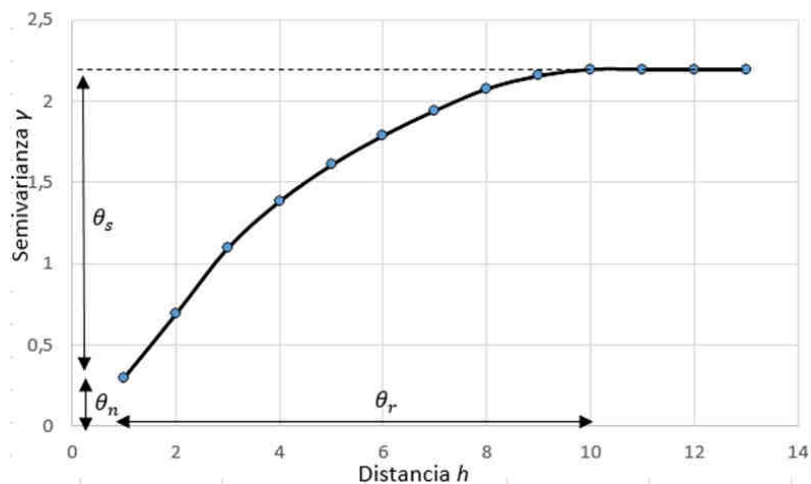


Figura 4. Semivariograma teórico. Fuente propia.

En general la función de semivariograma teórico (γ) tendrá la forma de la Ecuación 5.

$$\gamma = \theta_n + \gamma(\theta_s, \theta_r, h) \quad \text{Ecuación 5}$$

En donde $\gamma(\theta_s, \theta_r, h)$ tiene la forma matemática de alguno de los modelos disponibles en ArcGis®.

La estimación espacial se realiza con el algoritmo Kriging. Este algoritmo permite estimar datos desconocidos Z_0 en la coordenada (x_0, y_0) a partir de datos conocidos Z_i en las coordenadas (x_i, y_i) usando la Ecuación 1 y calculando los pesos ponderantes λ_i por medio de la función de semivariograma teórico (γ) mediante la Ecuación 6 (Johnston et al., 2003).

$$\lambda = \Gamma^{-1} * g \quad \text{Ecuación 6}$$

En donde

Γ^{-1} es la inversa de una matriz Γ cuyos elementos se calculan a partir de la función semivariograma teórico (γ). Los elementos de Γ representan el cambio de la semivarianza con la distancia entre cada dato conocido y los demás datos conocidos. La Figura 5 muestra la forma general de la matriz Γ , en donde P_i es un punto con coordenadas (x_i, y_i) . Los elementos $\gamma_{i,j}$ son simétricos con $\gamma_{j,i}$ y tienen la forma $\gamma = \theta_n + \gamma(\theta_s, \theta_r, h_{i,j})$, con $h_{i,j}$ la distancia euclídea entre el punto P_i y el punto P_j .

	P_1	P_2		P_3	P_n	Columna aumentada de 1
P_1	θ_n	$\gamma_{1,2}$		$\gamma_{1,3}$	$\gamma_{1,n}$	1
P_2	$\gamma_{2,1}$	θ_n		$\gamma_{2,3}$	$\gamma_{2,n}$	1
P_3	$\gamma_{3,1}$	$\gamma_{3,2}$		θ_n	$\gamma_{3,n}$	1
....	θ_n
P_n	$\gamma_{n,1}$	$\gamma_{n,2}$		$\gamma_{n,3}$	θ_n	1
Fila aumentada de 1	1	1		1	1	θ_n

Figura 5. Forma de la Matriz Γ . Fuente propia.

- g es un vector cuyos elementos se calculan también a partir de la función semivariograma teórico (γ). Estos elementos representan el cambio de la semivarianza con la distancia entre la coordenada de cada dato conocido y la coordenada (x_0, y_0) de dato desconocido. La Figura 6 muestra la forma general del vector g .

P_1	$\gamma_{1,0}$
P_2	$\gamma_{2,0}$
P_3	$\gamma_{3,0}$
...	...
P_n	$\gamma_{n,0}$
Fila aumentada de 1	1

Figura 6. Forma del vector g . Fuente propia.

Los elementos $\gamma_{i,0}$ tienen la forma $\gamma = \theta_n + \gamma(\theta_s, \theta_r, h_{i,0})$, con $h_{i,0}$ la distancia euclídea entre el punto P_i y el punto P_0 de coordenadas (x_0, y_0) .

Dependiendo de cómo sea la media de los datos en el área de estudio, existen algunas variantes del algoritmo Kriging a saber:

- Kriging Simple, cuando el valor de la media se conoce y es constante en toda el área de estudio.
- Kriging Ordinario, cuando el valor de la media se conoce y es constante zonalmente en el área de estudio. En el contexto de este trabajo se utilizó éste tipo de Kriging.
- Kriging Universal, cuando el valor de la media se desconoce y varía localmente en el área de estudio.

Algoritmo **Geoestadístico multivariado (Cokriging)**. Si en los puntos (x_i, y_i) se conoce el valor de una variable objeto de estudio $Z(x_i, y_i)$ y el valor de otras variables que están correlacionadas con ella $Z_{v1}(x_i, y_i), \dots, Z_{vr}(x_i, y_i)$; es posible usar la información adicional de las variables correlacionadas para calcular el valor de la variable estudio en otros puntos $Z(x_0, y_0)$ a partir del modelamiento del cambio de la covarianza con la distancia, es decir, a

partir de la implementación de un modelo de correlograma teórico. La estimación de $Z(x_0, y_0)$, se hace mediante la Ecuación 7 (Bogaert et al., 1995).

$$Z(x_0, y_0) = \sum_i^{n1} a_i Z_{v1}(x_i, y_i) + \sum_j^{n2} b_j Z_{v2}(x_j, y_j) + \dots + \sum_k^{nr} \alpha_k Z_{vr}(x_k, y_k) \quad \text{Ecuación 7}$$

Los pesos ponderantes de la Ecuación 7 ($a_i, b_j, \dots, \alpha_k$), se calculan mediante el modelo lineal de correogionalización que asume que todos los semivariogramas simples y cruzados (correlogramas) pueden expresarse como una suma de modelos básicos idénticos (exponencial, esférico, gaussiano, etc.) (Giraldo, 2003). Para el caso de dos variables se obtendrían a partir de la solución de un sistema como el que se muestra en la Figura 7.

$$\begin{bmatrix}
 \gamma_{v1}(1,1) & \dots & \gamma_{v1}(n1,1) & \gamma_{v1v2}(1,1) & \dots & \gamma_{v1v2}(n2,1) & 1 & 0 \\
 \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\
 \gamma_{v1}(1,n1) & \dots & \gamma_{v1}(n1,n1) & \gamma_{v1v2}(1,n1) & \dots & \gamma_{v1v2}(n2,n1) & 1 & 0 \\
 \gamma_{v1v2}(1,1) & \dots & \gamma_{v1v2}(n1,1) & \gamma_{v2}(1,1) & \dots & \gamma_{v2}(n2,1) & 0 & 1 \\
 \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\
 \gamma_{v1v2}(1,n2) & \dots & \gamma_{v1v2}(n1,n2) & \gamma_{v2}(1,n2) & \dots & \gamma_{v2}(n2,n2) & 0 & 1 \\
 \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\
 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 \\
 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0
 \end{bmatrix}
 \begin{bmatrix}
 a_1 \\
 \vdots \\
 a_{n1} \\
 b_1 \\
 \vdots \\
 b_{n2} \\
 \mu_1 \\
 \mu_2
 \end{bmatrix}
 =
 \begin{bmatrix}
 \gamma_{v1}(0,1) \\
 \vdots \\
 \gamma_{v1}(0,n1) \\
 \gamma_{v1v2}(0,1) \\
 \vdots \\
 \gamma_{v1v2}(0,n2) \\
 1 \\
 0
 \end{bmatrix}$$

Figura 7. Cálculos de pesos ponderantes mediante Cokriging. Fuente propia.

En la Figura 7, la matriz del lado izquierdo contiene los valores de las funciones de semivarianza (en recuadros rojo y azul para las variables Z_{v1} y Z_{v2} respectivamente) y de semivarianza cruzada (covarianza de las variables Z_{v1} y Z_{v2} en recuadro negro). Los términos de las dos últimas filas de las matrices corresponden a la restricción de no sesgamiento del predictor, con μ_1 y μ_2 los multiplicadores de Lagrange. El vector del lado derecho contiene los valores de las funciones de semivarianza y semivarianza cruzada evaluados para las distancias entre los sitios de monitoreo (de ambas variables) y el sitio donde se desea hacer la estimación (x_0, y_0) . La ventaja del método Cokriging frente al Kriging radica en el hecho de que cuando las variables $Z_{v1}(x_i, y_i), \dots, Z_{vr}(x_i, y_i)$ están fuertemente correlacionadas con la variable de estudio Z , se puede obtener una disminución en la varianza de estimación, no obstante dicha variable tenga menor densidad de muestreo (Bogaert et al., 1995).

2.2 Algoritmo de regresión espacial

El algoritmo de Regresión Espacial permite realizar estimaciones locales de una variable de estudio Z en las coordenadas (x_i, y_i) mediante la implementación del método de mínimos cuadrados ordinarios (Ordinary Least Squares – OLS) con datos espaciales de las variables explicativas $Z_j(x_i, y_i)$ (Dons et al., 2014) mediante la Ecuación 8.

$$Z(x_i, y_i) = \beta_0 + \sum_{j=1}^n \beta_j Z_j(x_i, y_i) \text{ Ecuación 8}$$

Inicialmente se debe especificar adecuadamente el modelo de regresión (cuantas y cuales variables explicativas $Z_j(x_i, y_i)$) usando los criterios de especificación del método OLS. En este trabajo se usó la herramienta Exploratory Regression® de ArcMap® que es una herramienta de minería de datos, similar a las herramientas de regresión Stepwise existentes en muchos softwares estadísticos, que analiza todas las posibles combinaciones de variables explicativas para determinar la mejor configuración de la Ecuación 8. Los criterios de especificación que se usan son (Fischer & Getis, 2010; Scott et al., 2011):

- **Criterio de rendimiento** 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 Ecuación 9 define el valor mínimo aceptable para $AdjR^2$ por lo tanto la mejor configuración de la Ecuación 8 será aquella con mayor $AdjR^2$ o según la Ecuación 10 menor AICc.

$$AdjR^2 > 0,5 \quad \text{Ecuación 9}$$

$$Min (AICc) \quad \text{Ecuación 10}$$

- **Criterio de consistencia** para determinar si los residuos se distribuyen normalmente. Con ello se controla que no falten variables explicativas claves, 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 Ecuación 11 define el valor mínimo aceptable para el p-value de JB. En consecuencia, la mejor configuración de la Ecuación 8 será aquella con mayor JB.

$$(JB)_{p\text{-value}} > 0,10 \quad \text{Ecuación 11}$$

- **Criterio de ausencia de redundancia o multicolinealidad** entre las variables explicativas para determinar si éstas son independientes entre sí, si no lo fueran entonces existe multicolinealidad entre ellas y alguna de ellas debería ser removidas. Para ello se usa el factor de inflación de la varianza (VIF). La mejor configuración de la Ecuación 8 será aquella con menor VIF. El valor mínimo aceptable para el VIF viene dado por la Ecuación 12.

$$VIF < 7,50 \quad \text{Ecuación 12}$$

- **Criterio de significancia para las variables explicativas** que permite determinar si el coeficiente de las variables explicativas es significativo. Para ello se usa el p-valor de los residuales de Koenker Breusch-Pagan K(BP). La mejor configuración de la Ecuación 8 será aquella con menor K(BP). El valor mínimo aceptable para el K(BP) está dado por la Ecuación 13.

$$K(BP)_{p\text{-value}} < 0,05 \quad \text{Ecuación 13}$$

- **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). La mejor configuración de la Ecuación 8 será aquella con mayor SA. El valor mínimo aceptable para el p-valor del SA está dado por la Ecuación 14.

$$(SA)_{p\text{-value}} > 0,10 \quad \text{Ecuación 14}$$

En este trabajo se usó la herramienta de regresiones geográficamente ponderadas (Geographically Weighted Regression - GWR®) del aplicativo ArcMap®, para calcular el término independiente, los coeficientes y el valor estimado de la Ecuación 8 en cada coordenada (x_i, y_i) (Fotheringham et al., 2002).

2.3 Algoritmos econométricos

Los algoritmos econométricos se basan en estructuras de regresión en las cuales es posible estimar el valor de una variable de respuesta Y a partir de variables explicativas o regresoras X_j , más un error. Para la variable de respuesta y las variables explicativas se deben tener series de datos para un tiempo dado, series de tiempo, o datos de panel (series de tiempo en diferentes sitios de monitoreo, donde cada sitio es un panel o unidad de corte transversal) (Woolridge, 2002). Los paneles están balanceados cuando cada variable tiene el mismo número de observaciones en cada panel y si los datos obtenidos provienen de una muestra mayor, el cálculo del modelo de regresión se hará mediante el método de efectos aleatorios. Dado que las variables ambientales se miden en el espacio en diferentes sitios de monitoreo, la estructura de datos de panel es más pertinente para la estimación espacio temporal de la variable de respuesta.

Para la implementación de la regresión se requiere que las series de datos sean estacionarias, es decir, que su media y varianza sean constantes en el tiempo y el valor de su covarianza entre dos períodos dependa sólo de la distancia o rezago entre estos dos períodos, y no del tiempo en el cual se calculó. La estacionariedad de una serie se puede validar mediante una prueba de raíces unitarias como la prueba de Dickey-Fuller aumentada (*Augmented Dickey-Fuller* – ADF) (Dickey & Fuller, 1979). Otra condición que debe cumplir las series de datos, es que se debe demostrar que X_i explica a Y y esto se puede hacer mediante una prueba de causalidad como la de Wiener - Granger (Wiener, 1956; Granger, 1969).

La estructura general de un modelo de regresión está dada por la Ecuación 15.

$$Y_{it} = \beta_1 + \sum_{i=2}^n (\beta_i X_{it}) + \text{error de estimación}_{it} \quad \text{Ecuación 15}$$

En donde β_1 es el valor esperado de Y cuando los β_i son 0, y los β se calculan mediante el método de mínimos cuadrados ordinarios. Con la Ecuación 15, se pueden definir varios modelos en la medida en que se tengan datos de varias variables explicativas X_{it} . Para determinar cuál es el modelo que mejor explica la variabilidad de Y se usan distintos indicadores y su valoración a partir de las siguientes consideraciones (Diebold, 1971):

- Menor valor de la suma de los cuadrados de los residuos (SCR) que es el mayor propósito de un buen modelo.
- Mayor valor del coeficiente de determinación (R^2) y del coeficiente de determinación ajustado $\text{Adj}R^2$ (Adjusted R-squared) de Theil (Theil, 1971). En general $\text{Adj}R^2 \leq R^2$.
- Significatividad individual, es decir, que las variables explicativas sean significativas a los niveles de significación habituales de 0,1, 0,05 o 0,01 (la probabilidad de que los coeficientes β ($\text{Prob}(\beta)$) sea menor que 0,1, 0,05 o 0,01; respectivamente).

Se tienen varios tipos de modelos de regresión dentro de los cuales se implementan en este trabajo los modelos de rezagos distribuidos (RD) y los modelos Autorregresivos (AR). En el modelo RD se incluyen valores actuales y valores rezagados (pasados) de las variables explicativas. La Ecuación 16 representa un modelo de rezagos distribuidos con un rezago finito de k períodos. El rezago k , es el tiempo hasta el cual se da la dependencia de Y con respecto a X .

$$Y_t = \alpha + \sum_{i=0}^k (\beta_i X_{t-i}) + \text{error de pronóstico}_t \quad \text{Ecuación 16}$$

El coeficiente β_0 se conoce como multiplicador de corto plazo o de impacto porque representa el cambio en el valor medio de Y que sigue a un cambio unitario en X en el mismo período; $\beta_0 + \beta_1$ es el cambio en el valor medio de Y en el período siguiente, $\beta_0 + \beta_1 + \beta_2$ en el período que sigue y así sucesivamente. $\beta_0 + \beta_1 + \beta_2 + \dots + \beta_k = \beta$, siempre que β exista, es el multiplicador de rezagos distribuidos total o de largo plazo. Resolver el modelo de la Ecuación 16 consiste en encontrar los valores de α y los β , y para ello existen dos métodos: el método Ad Hoc y el método de Alt (Alt, 1942). En el método Ad Hoc como la X_t no es estocástica, o por lo menos no está correlacionada con el error de pronóstico, entonces las X_{t-i} tampoco; por lo tanto, se puede aplicar el método de mínimos cuadrados ordinarios para

resolver la regresión de Y_t sobre X_t , luego la de Y_t sobre X_t y X_{t-1} , luego Y_t sobre X_t, X_{t-1}, X_{t-2} ; y así sucesivamente. El proceso se detiene cuando el coeficiente β de al menos una de las variables regresoras cambia de signo (de + a – o viceversa), entonces la iteración $i - 2$ será el mejor modelo de regresión y la i de X_{t-i} será el valor k del rezago.

En el modelo AR, se incluyen valores actuales y uno o más valores rezagados de la variable dependiente como variables explicativas. La Ecuación 17, representa un modelo AR(k) con un rezago finito de k períodos.

$$Y_t = \alpha_0 + \sum_{i=0}^k (\alpha_{i+1} Y_{t-i}) + \text{error de pronóstico}_t \quad \text{Ecuación 17}$$

Los modelos RD y AR se pueden combinar para construir un modelo autorregresivo con rezago distribuido (AR-RD). En éste modelo se incluyen como variables explicativas los valores actuales más uno o más valores rezagados de la variable dependiente y valores actuales y valores rezagados de las variables explicativas. La Ecuación 18 representa éste modelo con un rezago finito de k_1 y k_2 períodos.

$$Y_{jt} = \alpha_j + \sum_{j=1}^n \sum_{i=0}^{k_1} (\beta_{ji} X_{j(t-i)}) + \sum_{i=0}^{k_2} (\alpha_{j(i+1)} Y_{j(t-i)}) + \text{error}_{jt} \quad \text{Ecuación 18}$$

Los modelos AR y AR –RD, han mostrado un buen desempeño para la estimación del material particulado en zonas urbanas (Wang et al., 2017; Stadlober et al., 2008; Slini et al., 2006; Hooyberghs et al., 2005), con valores de R^2 que oscilan entre 0,38 y 0,98. En ellos se han incorporado como variables explicativas la velocidad del viento (Sayegh et al., 2014), la temperatura, la densidad vehicular (para efectos de largo plazo), la concentración de CO, NO_x y SO₂ (en reemplazo de variables de tráfico) (Papanastasiou et al., 2007), los usos del suelo, la presencia de zonas verdes (Taheri & Sodoudi, 2016), los cuerpos de agua, la densidad de población y variables meteorológicas (Zafra et al., 2017; Chen et al., 2010).

En la Figura 8 se muestra el esquema de un modelo AR – RD propuesto en este trabajo en donde una de las variables explicativas es un componente espacial de rezago distribuido que representa un dato funcional que proviene de la caracterización espacial univariada o

multivariada en un tiempo t de la variable respuesta mediante algoritmos de interpolación o regresión espacial que se complementan con el componente autorregresivo del modelo.

En la Ecuación 19 se muestra la estructura general del modelo AR–RD de la Figura 8.

$$S_{it} = \beta_0 + \sum_{r=0}^j \alpha_{r+1} S_{ri\text{spatial}} + \sum_{r=0}^k \beta_{r+1} O_{i_r} + \varepsilon_t \quad \text{Ecuación 19}$$

El subíndice i hace referencia a las coordenadas espaciales x_i, y_i ; con $i = 1, \dots, n$, en el Sistema de Coordenadas Geográficas MAGNA Bogotá - Colombia con Proyección Transversa de Mercator. Los términos $\beta_0, \alpha_{r+1}, \beta_{r+1}$; son los coeficientes de regresión y ε_t , es el error de la regresión. O_{i_r} es el dato observado en el espacio i en un tiempo r , así: O_{i_0} es el dato observado en el tiempo t , O_{i_1} en el $t - 1, \dots, O_{i_k}$ en el $t - k$. El término $S_{ri\text{spatial}}$ de la Ecuación 19 es el componente espacial de rezago distribuido que representa un dato funcional así: $S_{0i\text{spatial}}$ es el dato estimado en el tiempo t , $S_{1i\text{spatial}}$ el dato estimado en el tiempo $t - 1, \dots$; y $S_{ji\text{spatial}}$ es el dato estimado en el tiempo $t - j$.

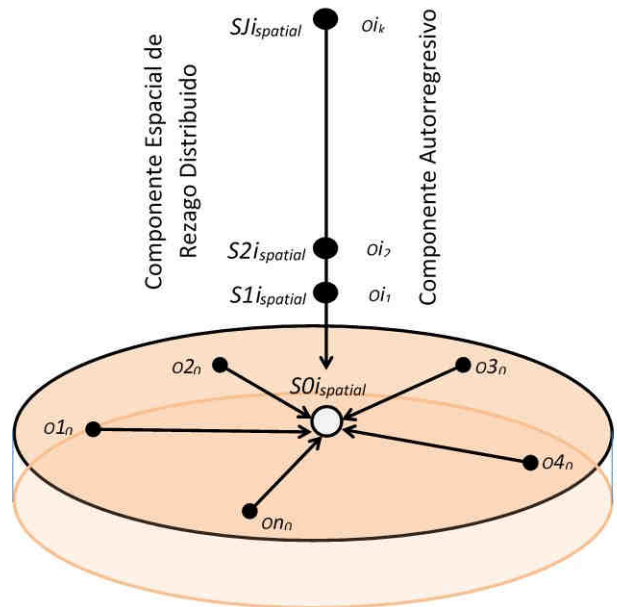


Figura 8. Modelo AR-RD propuesto para estimar simular concentración de contaminantes del aire en el espacio-tiempo.

Fuente propia.

Finalmente, el término S_{it} es el dato estimado de concentración de contaminante en el espacio-tiempo como variable dependiente de un componente espacial de rezago distribuido (dato funcional estimado en el espacio - $Sri_{spatial}$) y un componente autorregresivo (valor observado de la concentración en un sitio de monitoreo en el tiempo - Oi_t). Para calcular los modelos econométricos, en este trabajo se usó el aplicativo Eviews® validando la estacionariedad de las series de tiempo mediante pruebas de raíces unitarias y se demostró que las variables regresoras explican la variable de respuesta mediante la prueba de causalidad de Granger. Se calculó el valor del rezago con el método Ad Hoc obteniéndose un valor de 1. Adicionalmente se usó el método de efectos aleatorios para las secciones transversales (paneles) y el estimador Swamy y Arora para el componente de las varianzas.

2.4 Validación de datos estimados

Cada vez que se use algún algoritmo de estimación, se debe calcular su desempeño mediante validación cruzada y el uso de estadísticos de error. En el método de validación cruzada (*leave one out*) (Pollice & Jona, 2009), se retira un valor observado O_i en el sitio de monitoreo i y con algún algoritmo de estimación se calcula su valor estimado E_i , esto permite determinar el error de estimación $r_i = O_i - E_i$ o residual para cada uno de los datos observados. Con el residual se calculan los siguientes estadísticos de error (Kousa et al., 2001):

- ASE (*Average Standard Error*), error estándar promedio dado por la Ecuación 20.

$$ASE = \frac{\sum_1^n \sigma_{r_i}}{n} \quad \text{Ecuación 20.}$$

En donde:

$\sigma_{r_i} = \sqrt{(r_i - \mu_{r_i})^2}$, es el error estándar con μ_{r_i} la media de los residuales y n la cantidad de datos.

- RMSE (*Root Mean Square error*), la raíz del error cuadrático promedio dado por la Ecuación 21.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (r_i)^2}{n}} \quad \text{Ecuación 21}$$

- RMSS, (*Root Mean Square error Standarized*), la raíz del error cuadrático promedio estandarizado dado por la Ecuación 22, en donde σ es la desviación estándar de los r_i .

$$RMSS = \frac{RMSE}{\sigma} \quad \text{Ecuación 22}$$

Para el caso de los algoritmos geoestadísticos, se usan los anteriores estadísticos de error para determinar cuál es el mejor de los nueve modelos de semivariogramas teóricos de ESRI y como se muestra en la parte inferior izquierda de la Figura 2, el mejor modelo de estimación será aquel (Johnston et al., 2003):

- Cuyo θ_s sea más similar a la varianza de los datos (σ^2).
- Que tenga el menor ASE.
- Que tenga el menor RMSE.
- Que tenga el RMSS más cerca de 1.

Otros estadísticos de error utilizados son:

- El índice de ajuste (*Index of Agreement - IOA*) que se calcula mediante la Ecuación 23 (Willmott et al., 2012).

$$IOA = 1 - \frac{\sum_{i=1}^n (r_i)^2}{\sum_{i=1}^n (|Ei - \mu_o| + |Oi - \mu_o|)^2} \quad \text{Ecuación 23}$$

En donde μ_o es la media de los datos observados.

- El error absoluto relativo (*Absolut Relative Error - ARE*) y su valor promedio (*Mean Absolut Relative Error - MARE*) que se calculan mediante la Ecuación 24 (Li & Heap, 2014).

$$ARE_i = \frac{|r_i|}{o_i} \quad MARE = \frac{\sum_{i=1}^n ARE_i}{n} \quad \text{Ecuación 24}$$

La forma de interpretación de los estadísticos de error es la siguiente (Li & Heap, 2014):

- Valores de IOA y RMSS cercanos a 1, significan un buen ajuste entre el dato observado y el dato estimado.
- Valores pequeños de MARE, ASE y RMSE, son indicativos de buenos ajustes entre el dato observado y el dato estimado.

2.5 Gradientes espaciales

En este trabajo se usan los gradientes espaciales para obtener variables explicativas adicionales. El gradiente espacial se puede calcular mediante la Ecuación 25 (Burrough & Mcdonnell, 1998).

$$\nabla Z = \tan^{-1} \sqrt{\left(\frac{\nabla Z_{ij}}{\nabla x_{ij}}\right)^2 + \left(\frac{\nabla Z_{ij}}{\nabla y_{ij}}\right)^2} * \frac{180}{\pi} \quad \text{Ecuación 25}$$

Dónde:

∇Z es el gradiente del valor del dato con la distancia entre las posiciones (x_i, y_i) y (x_j, y_j) en unidades porcentuales.

∇Z_{ij} es el cambio del valor del dato entre las posiciones i e j .

∇x_{ij} es el cambio de la coordenada x entre las posiciones i y j en metros.

∇y_{ij} es el cambio de la coordenada y entre las posiciones i y j en metros.

Para calcular el cambio de la variable en la dirección x ($\frac{\nabla Z_{ij}}{\nabla x_{ij}}$) y la dirección y ($\frac{\nabla Z_{ij}}{\nabla y_{ij}}$), en una estructura de almacenamiento de datos matricial, se toman las ocho celdas vecinas a una celda de referencia. Como se muestra en la Figura 9, si la celda de referencia es la e entonces las ocho celdas vecinas serían a, b, c, d, f, g, h y la i . Si en estas celdas existe un valor de la

variable, entonces los cambios en las direcciones x e y, se calculan mediante las Ecuaciones 26 y 27 respectivamente.

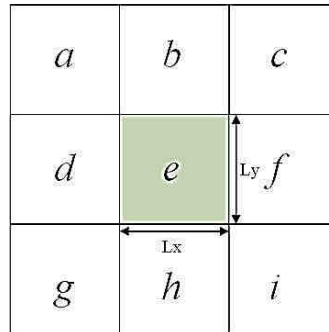


Figura 9. Vecindad de una celda en una estructura raster. Fuente propia.

$$\frac{\nabla Z_{ij}}{\nabla x_{ij}} = \frac{(c+2f+i)-(a+2d+g)}{8 \cdot \text{longitud X de la celda}} \text{ Ecuación 26}$$

$$\frac{\nabla Z_{ij}}{\nabla y_{ij}} = \frac{(g+2h+i)-(a+2b+c)}{8 \cdot \text{longitud Y de la celda}} \text{ Ecuación 27}$$

3 PROBLEMA Y OBJETIVOS

3.1 Problema

Los altos valores de concentración de los contaminantes del aire disminuyen la calidad del aire en los centros urbanos densamente poblados y afectan la salud de los habitantes. En los últimos años se ha prestado particular atención al material particulado de 10 y 2.5 micras de diámetro (PM10, PM2.5) por ser un factor importante en el aumento del número de casos de enfermedades respiratorias especialmente en niños menores de 6 años y adultos mayores de 60 años (Song et al., 2017; García et al., 2016; MinAmbiente, 2016).

Las problemáticas asociadas con la contaminación del aire hacen necesaria la implementación de sistemas de vigilancia de la calidad del aire para medir la concentración de los contaminantes a través de redes de sitios de monitoreo. La cantidad de mediciones no son suficientes para hacer una adecuada caracterización espacio-temporal del fenómeno de la contaminación del aire. El problema de la ausencia de datos, tanto en el espacio como en el tiempo, ha sido abordado a través del uso de algoritmos de estimación desde tres enfoques: determinista, estocástico e interpolación espacial. Desde el enfoque determinista se usan simulaciones numéricas, sin embargo, son costosas en la medida en que requieren datos de muchas variables para realizar estimaciones locales de la concentración. Desde el enfoque estocástico se han usado técnicas estadísticas tradicionales de imputación de datos (Urrutia et al., 2010; Borges et al., 2009) hasta regresiones espaciales lineales multivariadas. Finalmente, desde el enfoque de la interpolación espacial se han usados técnicas de análisis espacial implementadas mediante SIG como una alternativa complementaria, económica y eficiente.

Con base en lo anterior, en esta tesis doctoral se plantea como problema, el diseño de una metodología que integre aspectos de análisis espacial, estadístico, geoestadístico, econométrico y los SIG para caracterizar espacio-temporalmente la concentración mensual de contaminantes del aire en áreas urbanas intramontanas.

3.2 Objetivos

General

- Diseñar una metodología basada en análisis espacial, estadístico, geoestadístico, econométrico y de SIG para la caracterización espacio-temporal de la concentración de material particulado del aire en áreas urbanas intramontanas con información escasa.

Específicos

- Explorar las características fundamentales de los algoritmos de estimación de la concentración de material particulado.
- Analizar los aspectos conceptuales y procedimentales para la caracterización espacial univariada, multivariada y temporal de los algoritmos de estimación de la concentración de material particulado en áreas urbanas intramontanas con información escasa.
- Proponer una metodología para la caracterización espacio-temporal de la concentración de material particulado de áreas urbanas intramontanas con información escasa.
- Implementar la metodología propuesta con un caso de estudio y validar los resultados obtenidos.

4 ÁREA Y DATOS DEL CASO DE ESTUDIO

Como caso de estudio para la aplicación de la metodología, se toma el área metropolitana de la ciudad de Medellín que se encuentra en el centro geográfico del Valle de Aburrá sobre la cordillera central de los Andes entre $6,0^{\circ}$ - $6,5^{\circ}$ N y $75,5^{\circ}$ - $75,7^{\circ}$ W con una extensión de 328 km^2 y una población, a 2016, de 3'004.215 habitantes (ver Figuras 10 y 11). El Valle corresponde a la tipología de una región intramontana, es decir, es fruto de procesos de sedimentación y está enmarcado por una topografía irregular que varía entre los 1300-2800 msnm con pendientes entre 0% y 50% y dos cerros tutelares en su centro (Nutibara y El Volador); tiene una longitud aproximada de 60 km de largo y un ancho variable entre 10 y 20 km en su centro. Las estribaciones de la cordillera de los Andes que lo encierran, dan lugar a la configuración de diversos microclimas, aunque en general se establece un clima subtropical monzónico templado y húmedo con unos valores medios anuales de temperatura entre $16 - 28^{\circ}\text{C}$, 1673 mm de precipitación, 68% de humedad relativa y vientos suaves entre $5,4 - 11,3 \text{ km/h}$. La ciudad de Medellín es un centro urbano densamente poblado en donde se presentan problemas de calidad de aire y de salud pública por la contaminación debida a distintos tipos de fuentes de emisión de contaminantes (emisiones producidas por fuentes fijas asociadas con los procesos industriales y emisiones producidas por fuentes móviles debidas a la densidad del flujo vehicular).

Se toma como variable de estudio, la concentración promedio mensual del material particulado (PM10, PM2.5 en $\mu\text{g}/\text{m}^3$) disponibles de la Red de Calidad del Aire (RedAire) en algunos sitios de monitoreo en los años 2013 y 2014 (Geographic Coordinate System: GCS MAGNA. Projected Coordinate System: MAGNA Colombia Bogotá, Projection Transverse Mercator) (RedAire, 2015). Como variables explicativas, se pudo acceder a datos de temperatura (T en $^{\circ}\text{C}$) y velocidad del viento (Vv en m/s) en 17 sitios de monitoreo del Sistema de Alertas Tempranas del Departamento de Antioquia (SIATA, 2016). No todos estos sitios coinciden con las coordenadas de las mediciones hechas del material particulado. Estas variables son las que tiene mayor influencia en la concentración del material particulado en los centros urbanos (Papanastasiou et al, 2007; Sayegh et al., 2014).

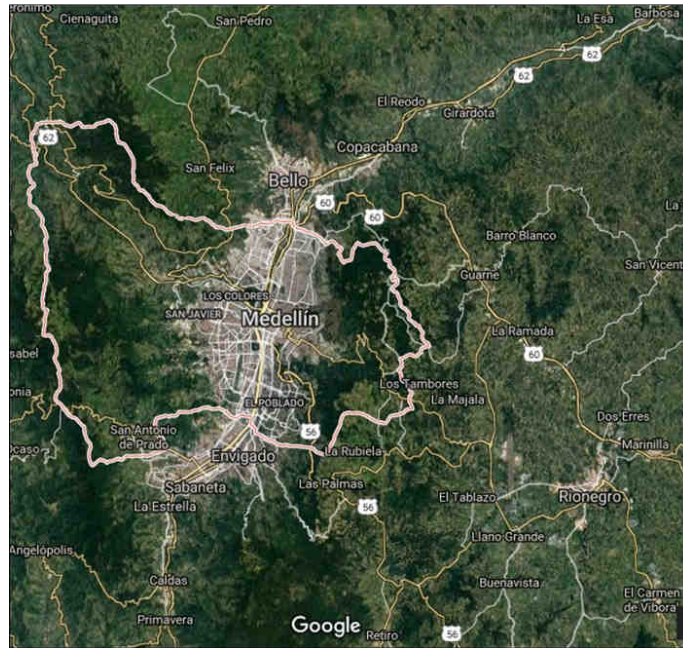


Figura 10. Mapa satelital de la Ciudad de Medellín y su Área Metropolitana. Google maps.

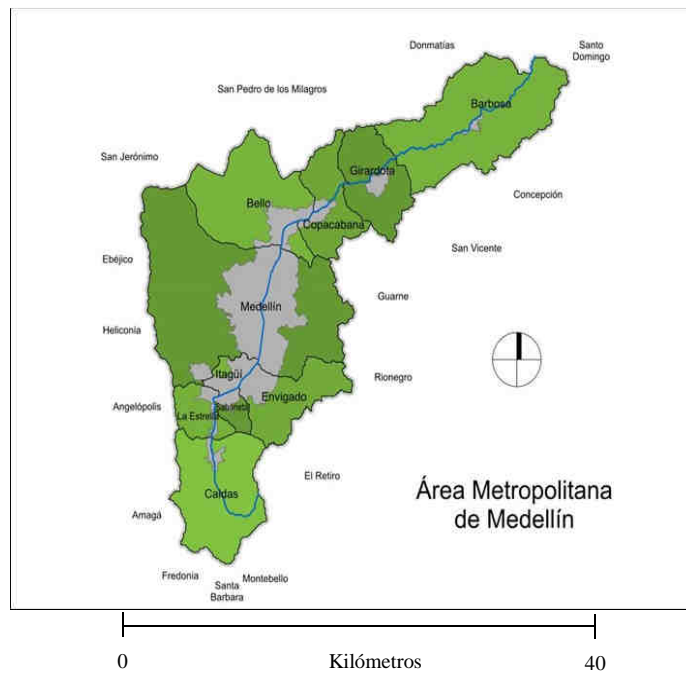


Figura 11. Municipios del Área Metropolitana de la Ciudad de Medellín. Google maps.

En la Figura 12 se muestra la ubicación de los sitios de monitoreo de la concentración del material particulado. En la Figura 12a (en los sitios identificados con los números 7, 11, 5, 6, 4, 3 y 1) se monitorean simultáneamente ambas variables. La metodología se implementa

con las herramientas *Geostatistical Analyst*® y *GWR Analyst*® del aplicativo *ArcMap*® de ESRI y con el software *SPSS*®.

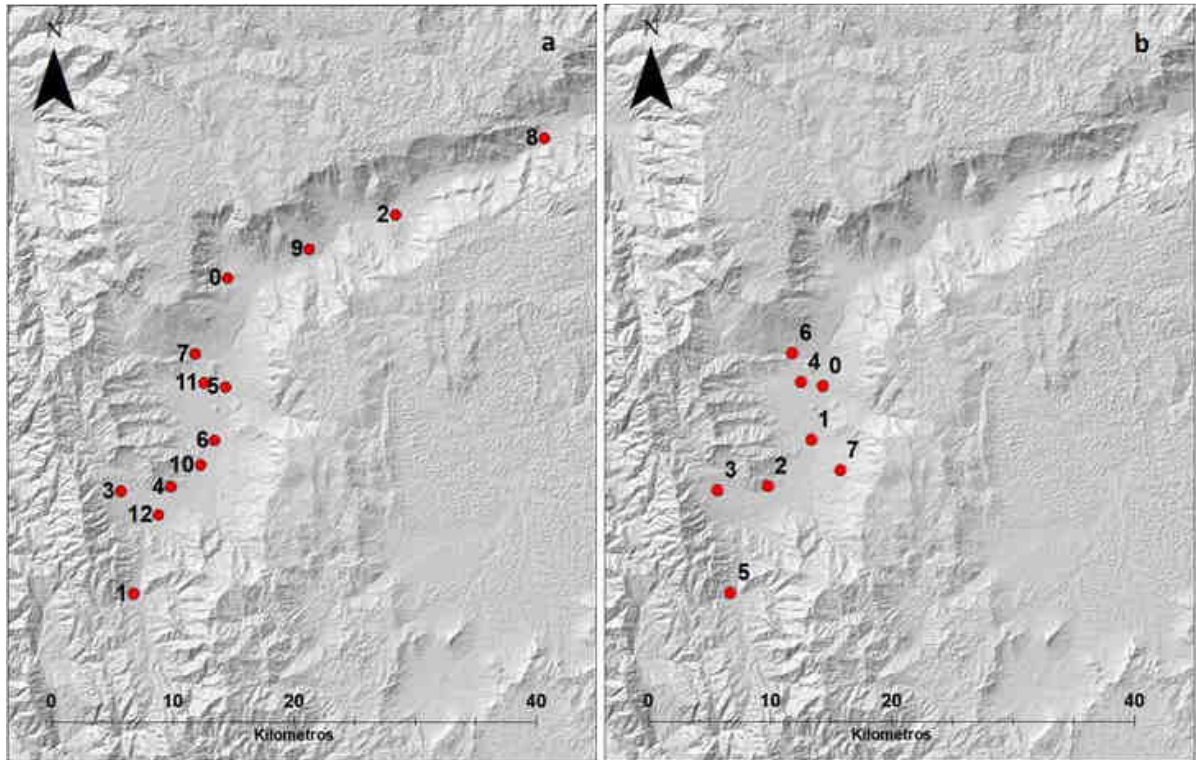


Figura 12. Sitios de monitoreo de material particulado. En la parte a PM10 y en la parte b PM2.5.

4.1 Datos de PM2.5 y variables auxiliares

Los datos disponibles de concentración promedio mensual de PM2.5 entre 2013 y 2014 se midieron en los sitios de monitoreo de la Tabla 7. La superficie de la región en donde se hicieron las mediciones es de 535,13 km², que corresponde con una densidad de muestreo de 66,89 km²/sitio de monitoreo. En la Tabla 8, se muestran los valores medidos.

Tabla 7. Descripción sitios de monitoreo de PM2.5.

ID del sitio de monitoreo	Xi	Yi	Descripción
0	834818,87	1183026,29	Museo de Antioquia
1	833901,20	1178597,43	Politécnico Colombiano Jaime Isaza Cadavid
2	830336,17	1174772,86	Casa de Justicia de Itagüí
3	826213,92	1174416,11	Concejo de Itagüí
4	833067,09	1183369,30	Universidad Nacional Núcleo Volador
5	827265,67	1165931,52	Corporación Universitaria Lasallista
6	832336,01	1185768,31	Universidad Nacional Facultad de Minas
7	836286,00	1176113,00	El Poblado

Tabla 8. Datos observados promedio de PM2.5 (O) para cada mes del año ($\mu\text{g}/\text{m}^3$)

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	39,1	44,5	42,8	36,4	33,1	28,1	28,3	30,5	33,8	36,6	40,2	40,7
1	24,3	32,0	29,9	26,3	23,5	20,9	18,4	19,8	20,1	19,5	18,8	18,2
2	26,8	36,7	38,9	32,5	26,5	20,4	20,9	25,8	30,8	33,5	31,1	32,3
3	26,9	35,0	33,4	28,3	25,0	21,1	19,4	20,0	23,0	24,0	22,3	25,9
4	33,3	37,1	39,7	30,2	24,3	19,0	25,2	26,1	28,0	28,8	28,0	28,5
5	28,0	34,4	32,5	25,2	17,1	19,1	22,5	21,2	24,5	22,8	21,2	27,3
6	25,3	31,7	33,6	24,2	27,2	22,7	22,3	24,9	23,3	27,5	21,7	19,3
7	27,0	33,7	34,2	27,5	24,3	17,8	16,0	18,7	20,3	23,8	22,3	27,8

En el mismo período se pudo tener acceso a los datos promedio mensual de temperatura (T en °C) y velocidad del viento (Vv en m/s) en 17 sitios de monitoreo, no todos los sitios de monitoreo coinciden con las coordenadas de la Tabla 7. Para estimar el valor de T y de Vv en las coordenadas de la Tabla 7, se usaron los algoritmos de interpolación espacial IDW y Spline, que son los de mejor desempeño para estas variables (Sayegh et al., 2014), y de los mapas resultantes, con la función *Extract* del aplicativo *Spatial Analyst®* de *ArcGis®* y el comando *extract values to points*, se extrajo el valor estimado de estas variables en dichas coordenadas.

Tabla 9. Datos de temperatura promedio para cada mes del año (°C).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	22,11	22,08	22,56	22,99	22,18	23,39	23,72	22,26	22,57	21,80	21,70	22,11
1	22,31	21,79	22,36	22,99	21,90	23,10	23,60	22,55	22,39	21,55	21,39	21,89
2	21,20	20,93	21,20	22,10	21,18	22,40	22,67	21,52	21,46	20,75	20,61	21,10
3	20,07	19,61	20,01	20,16	19,46	20,65	20,98	19,98	19,91	19,24	19,20	19,61
4	21,83	21,68	22,19	22,71	21,70	23,02	23,35	21,92	22,23	21,35	21,25	21,60
5	17,83	17,33	17,63	18,36	17,75	18,63	18,55	17,52	17,39	16,93	16,81	17,22
6	17,36	18,62	19,01	19,56	19,42	20,46	22,46	17,79	21,43	20,67	20,57	20,96
7	17,26	17,62	17,91	20,48	19,72	20,83	22,32	20,23	21,20	20,50	20,40	20,80
RMSE	7,27	6,56	6,77	6,12	5,29	5,60	3,03	4,80	2,92	2,83	2,75	2,87

El mejor interpolador para la variable T fue el Spline con un RMSE promedio de 4,73 y para la variable Vv el IDW con un RMSE promedio de 0,58. Los resultados se muestran en las Tablas 9 y 10 respectivamente.

Tabla 10. Datos de velocidad del viento promedio para cada mes del año (m/s).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	1,38	1,28	1,31	1,24	1,09	1,37	1,47	1,35	1,23	1,10	1,06	1,19
1	0,48	0,59	0,64	0,35	0,28	0,38	0,35	0,51	0,45	0,67	0,60	0,50
2	1,48	1,39	1,39	1,35	1,25	1,45	1,42	1,31	1,09	1,00	0,94	1,09
3	0,96	0,87	0,91	0,89	0,82	0,95	0,84	1,00	0,95	0,85	0,79	0,80
4	1,83	1,64	1,67	1,46	1,20	1,60	1,58	1,51	1,36	1,20	1,24	1,46
5	1,41	1,25	1,29	1,27	1,21	1,27	1,36	1,31	1,21	1,24	1,20	1,30
6	1,86	1,87	1,82	1,74	1,56	1,97	2,13	1,98	1,78	1,58	1,57	1,78
7	1,00	1,00	1,01	1,18	1,02	1,31	1,50	1,40	1,22	1,13	1,12	1,14
RMSE	0,71	0,63	0,61	0,61	0,57	0,71	0,70	0,59	0,51	0,41	0,41	0,46

Tabla 11. Datos de ∇T para cada mes del año (%).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	0,359	0,123	0,127	0,142	0,142	0,146	0,050	0,310	0,046	0,044	0,043	0,044
1	0,063	0,057	0,061	0,188	0,157	0,163	0,024	0,049	0,022	0,020	0,020	0,020
2	0,255	0,232	0,231	0,445	0,413	0,440	0,092	0,137	0,085	0,082	0,079	0,085
3	0,344	0,312	0,329	0,359	0,349	0,384	0,043	0,099	0,045	0,041	0,046	0,049
4	0,104	0,126	0,132	0,132	0,117	0,130	0,040	0,165	0,038	0,032	0,032	0,030
5	0,514	0,489	0,505	1,395	1,336	1,407	0,548	0,515	0,532	0,516	0,504	0,522
6	0,132	0,121	0,125	0,124	0,086	0,095	0,018	0,058	0,018	0,018	0,017	0,016
7	0,033	0,028	0,029	0,014	0,014	0,014	0,003	0,015	0,003	0,003	0,003	0,003

Tabla 12. Datos de ∇Vv para cada mes del año (%).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	0,0025	0,0070	0,0037	0,0047	0,0038	0,0053	0,0082	0,0075	0,0061	0,0053	0,0056	0,0061
1	0,0017	0,0014	0,0013	0,0026	0,0023	0,0029	0,0031	0,0024	0,0021	0,0013	0,0014	0,0018
2	0,0015	0,0012	0,0012	0,0004	0,0004	0,0004	0,0001	0,0001	0,0004	0,0004	0,0005	0,0001
3	0,0005	0,0006	0,0005	0,0002	0,0002	0,0002	0,0014	0,0008	0,0006	0,0006	0,0006	0,0008
4	0,0020	0,0020	0,0014	0,0009	0,0011	0,0011	0,0073	0,0057	0,0047	0,0041	0,0032	0,0026
5	0,0466	0,0341	0,0331	0,0304	0,0285	0,0306	0,0272	0,0339	0,0309	0,0335	0,0331	0,0339
6	0,0298	0,0173	0,0212	0,0083	0,0095	0,0112	0,0167	0,0155	0,0149	0,0127	0,0121	0,0155
7	0,0081	0,0079	0,0074	0,0058	0,0052	0,0066	0,0083	0,0064	0,0057	0,0035	0,0041	0,0049

Con base en los mapas interpolados de T y Vv, se calculan dos variables adicionales: el gradiente espacial de temperatura (∇T) (Tabla 11) y el gradiente espacial de la velocidad del viento (∇Vv) (Tabla 12). Con la función *Extract* se obtienen los valores de los gradientes espaciales en las coordenadas de la Tabla 7 para cada mes.

4.2 Datos de PM10 y variables auxiliares

Los datos disponibles de concentración promedio mensual de PM10 entre 2013 y 2014, se midieron en los sitios de monitoreo de la Tabla 13. La superficie de la región en donde se hicieron las mediciones es de 1046,49 km², lo cual corresponde con una densidad de muestreo de 80,50 km²/sitio de monitoreo. En la Tabla 13, se muestran los valores medidos.

Tabla 13. Descripción sitios de monitoreo PM10.

ID del sitio de monitoreo	Xi	Yi	Descripción
0	834998,29	1192029,45	Universidad de San Buenaventura – Bello
1	827265,67	1165931,52	Plaza de mercado – Caldas
2	848816,38	1197276,60	Institución Educativa Colombia - Girardota
3	826213,92	1174416,11	Liceo Concejo Itagií
4	830336,17	1174772,86	Colegio el Rosario – Itagií
5	834818,87	1183026,29	Éxito San Antonio – Medellín
6	833901,20	1178597,43	Politécnico Colombiano Jaime Isaza C.
7	832336,01	1185768,31	Universidad Nacional - Facultad de Minas
8	861066,40	1203605,81	Hospital San Vicente de Paúl – Barbosa
9	841707,31	1194438,13	Hospital Santa Margarita - Copacabana
10	832774,04	1176575,16	Planta de Tratamiento de Aguas San Fernando-
11	833067,09	1183369,30	Corantioquia – Medellín
12	829314,51	1172440,35	Centro Administrativo Municipal - Sabaneta

Tabla 14. Datos observados promedio de PM10 (O) para cada mes del año (µg/m3).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	35,70	48,41	48,75	39,30	34,83	32,39	34,71	36,67	40,38	43,29	39,02	40,96
1	65,43	78,26	76,73	59,79	55,53	58,17	57,94	54,80	59,34	54,04	48,28	58,12
2	50,72	54,48	55,57	49,11	38,32	38,92	42,38	40,07	38,20	42,11	44,62	37,87
3	48,52	58,59	56,00	51,17	48,49	42,41	42,73	43,13	51,24	49,45	44,05	51,81
4	62,84	69,29	68,95	62,30	61,82	53,07	52,92	49,31	54,97	54,69	52,66	59,68
5	56,85	66,07	64,22	59,59	56,14	48,85	52,45	55,16	57,93	63,94	64,36	66,73
6	48,62	67,23	64,51	62,11	53,13	37,66	36,95	40,26	42,49	47,01	42,92	47,62
7	57,48	66,72	67,09	62,42	57,82	51,80	54,22	53,19	62,85	62,28	63,47	55,35
8	32,47	37,48	42,00	32,56	24,89	31,71	39,92	25,31	30,55	24,32	21,03	25,14
9	43,63	44,85	39,71	37,93	33,78	32,64	38,90	30,44	32,08	30,77	28,15	31,51
10	36,31	42,05	41,50	39,33	34,30	34,46	36,46	33,82	42,41	39,57	38,85	40,12
11	36,29	40,09	44,67	40,89	41,95	33,82	34,27	34,38	35,20	35,31	39,85	38,22
12	44,19	47,97	46,45	43,38	38,11	35,10	36,44	35,77	40,36	38,84	38,89	42,10

En el mismo período se toman los datos promedio mensual de temperatura (T en °C) y velocidad del viento (Vv en m/s) en 17 sitios de monitoreo, aunque no todos los sitios de monitoreo coinciden con las coordenadas de la Tabla 12. Para estimar el valor de T y de Vv en las coordenadas de la Tabla 13, se usó el mismo proceso que para la variable PM2.5. El mejor interpolador para la variable T fue el Spline con un RMSE promedio de 4,64 y para la

variable Vv el IDW con un RMSE promedio de 0,57. Los resultados se muestran en las Tablas 15 y 16 respectivamente.

Tabla 15. Datos de temperatura promedio para cada mes del año (°C).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	18,99	19,22	19,67	20,24	20,40	21,40	22,26	20,38	21,33	20,57	20,49	20,83
1	18,30	17,98	18,28	19,17	18,48	19,51	20,75	19,17	19,64	19,01	18,93	19,31
2	13,15	13,37	13,60	14,19	14,41	14,91	22,22	20,79	21,53	20,83	20,70	21,06
3	17,73	17,93	18,24	19,65	18,91	19,96	21,61	19,57	20,50	19,82	19,72	20,12
4	19,54	19,47	19,76	20,44	19,63	20,75	21,98	20,45	20,82	20,13	20,02	20,46
5	19,13	18,73	19,09	19,42	18,74	19,85	20,87	19,68	19,79	19,14	19,08	19,48
6	17,29	18,32	18,70	19,39	19,38	20,42	22,45	18,48	21,42	20,67	20,56	20,95
7	20,38	20,23	20,70	21,89	20,94	22,09	23,11	21,52	21,94	21,15	21,02	21,48
8	13,42	13,74	14,03	14,57	20,76	21,70	22,36	20,86	21,50	20,78	20,63	20,92
9	19,83	19,93	20,22	20,68	20,71	21,50	22,45	21,16	21,71	21,08	20,91	21,22
10	20,38	21,05	21,50	22,11	21,31	22,47	23,28	20,77	22,17	21,40	21,31	21,70
11	16,83	16,44	16,71	17,10	16,53	17,37	18,12	16,90	17,02	16,53	16,45	16,82
12	20,30	20,70	21,16	21,68	20,80	22,00	23,02	20,48	21,93	21,10	21,01	21,37
RMSE	6,98	6,40	6,60	6,07	5,30	5,61	2,95	4,65	2,85	2,75	2,69	2,78

Tabla 16. Datos de velocidad del viento promedio para cada mes del año (m/s).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	1,65	1,54	1,57	1,46	1,29	1,51	1,56	1,42	1,36	1,23	1,30	1,38
1	1,41	1,25	1,29	1,26	1,20	1,25	1,36	1,31	1,21	1,24	1,20	1,30
2	1,84	1,84	1,91	1,71	1,59	1,65	1,49	1,70	1,58	1,51	1,70	1,68
3	0,96	0,87	0,91	0,89	0,82	0,95	0,84	1,00	0,95	0,85	0,79	0,80
4	1,48	1,39	1,39	1,35	1,25	1,45	1,42	1,31	1,09	1,00	0,94	1,09
5	1,38	1,28	1,31	1,24	1,09	1,36	1,47	1,35	1,23	1,10	1,06	1,19
6	0,48	0,59	0,64	0,35	0,28	0,38	0,35	0,51	0,45	0,67	0,60	0,50
7	1,86	1,87	1,82	1,73	1,55	1,95	2,13	1,98	1,78	1,58	1,57	1,78
8	0,75	0,75	0,80	0,82	0,77	0,79	1,69	1,68	1,65	1,49	1,56	1,54
9	1,00	1,10	1,12	1,02	1,90	2,12	2,01	2,11	1,92	1,74	1,79	1,86
10	1,02	1,02	1,04	1,02	0,90	1,12	1,25	1,20	1,05	1,03	0,99	1,02
11	1,83	1,64	1,67	1,46	1,20	1,60	1,58	1,51	1,36	1,20	1,24	1,46
12	1,26	1,17	1,20	1,17	1,07	1,26	1,31	1,26	1,11	1,02	0,97	1,06
RMSE	0,71	0,62	0,61	0,61	0,56	0,71	0,7	0,58	0,51	0,41	0,41	0,46

Con base en los mapas interpolados de T y Vv, se calculan el gradiente espacial de temperatura (∇T) y el gradiente espacial de la velocidad del viento (∇Vv). Con la función *Extract* se obtienen los valores de los gradientes espaciales en las coordenadas de las Tabla 13 para cada mes. Los resultados se muestran en las Tablas 17 y 18 respectivamente.

Tabla 17. Datos de ∇T para cada mes del año (%).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	0,13	0,09	0,10	0,08	0,03	0,03	0,01	0,06	0,01	0,01	0,01	0,01
1	0,08	0,05	0,05	0,04	0,04	0,04	0,03	0,08	0,03	0,03	0,03	0,03
2	0,25	0,27	0,27	0,31	0,35	0,37	0,02	0,05	0,01	0,01	0,01	0,00
3	0,01	0,01	0,01	0,02	0,02	0,02	0,02	0,01	0,02	0,02	0,02	0,02

4	0,15	0,14	0,13	0,10	0,09	0,10	0,05	0,09	0,05	0,05	0,04	0,05
5	0,30	0,25	0,27	0,20	0,19	0,22	0,03	0,12	0,03	0,03	0,03	0,04
6	0,13	0,12	0,12	0,11	0,05	0,05	0,02	0,04	0,02	0,02	0,02	0,01
7	0,03	0,03	0,03	0,05	0,05	0,05	0,02	0,03	0,01	0,01	0,01	0,01
8	0,12	0,12	0,12	0,14	0,02	0,02	0,00	0,03	0,00	0,00	0,00	0,00
9	0,05	0,05	0,04	0,04	0,01	0,01	0,00	0,01	0,00	0,00	0,00	0,00
10	0,19	0,08	0,08	0,05	0,05	0,05	0,03	0,16	0,03	0,03	0,03	0,03
11	0,24	0,23	0,23	0,26	0,24	0,26	0,28	0,26	0,28	0,27	0,26	0,27
12	0,07	0,09	0,09	0,08	0,07	0,07	0,03	0,08	0,03	0,02	0,02	0,02

Tabla 18. Datos de \overline{Wv} para cada mes del año (%).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	0,0003	0,0000	0,0001	0,0000	0,0002	0,0004	0,0006	0,0006	0,0004	0,0004	0,0003	0,0004
1	0,0466	0,0341	0,0331	0,0314	0,0294	0,0314	0,0272	0,0339	0,0309	0,0335	0,0331	0,0339
2	0,0001	0,0001	0,0001	0,0001	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000
3	0,0005	0,0006	0,0005	0,0006	0,0005	0,0006	0,0014	0,0008	0,0006	0,0006	0,0006	0,0008
4	0,0015	0,0012	0,0012	0,0010	0,0011	0,0010	0,0001	0,0001	0,0004	0,0004	0,0005	0,0001
5	0,0025	0,0070	0,0037	0,0067	0,0056	0,0074	0,0082	0,0075	0,0061	0,0053	0,0056	0,0061
6	0,0017	0,0014	0,0013	0,0022	0,0020	0,0025	0,0031	0,0024	0,0021	0,0013	0,0014	0,0018
7	0,0298	0,0173	0,0212	0,0129	0,0121	0,0157	0,0167	0,0155	0,0149	0,0127	0,0121	0,0155
8	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
9	0,0004	0,0003	0,0003	0,0003	0,0003	0,0003	0,0001	0,0002	0,0002	0,0002	0,0002	0,0002
10	0,0205	0,0165	0,0154	0,0181	0,0179	0,0189	0,0198	0,0152	0,0126	0,0067	0,0066	0,0118
11	0,0020	0,0020	0,0014	0,0032	0,0039	0,0041	0,0073	0,0057	0,0047	0,0041	0,0032	0,0026
12	0,0066	0,0110	0,0081	0,0094	0,0083	0,0107	0,0093	0,0064	0,0034	0,0027	0,0026	0,0048

5 CARACTERIZACIÓN ESPACIAL DE DATOS DE CONCENTRACIÓN DE MATERIAL PARTICULADO CON ALGORITMOS DE INTERPOLACIÓN O REGRESIÓN ESPACIAL

Los datos de las variables ambientales se caracterizan porque se miden en sitios de monitoreo ubicados en cierta zona de estudio con base en un sistema de referenciación espacial, es decir, son datos espaciales. La relación entre el valor del área de la zona de estudio y la cantidad de sitios de monitoreo es la densidad de muestreo $D_m = \text{km}^2/\text{número de sitios de monitoreo}$. Adicionalmente se miden periódicamente en cada sitio de monitoreo lo cual implica que se tiene una unidad de análisis temporal para los datos (para este caso se procesaron los datos disponibles de las mediciones del material particulado hechas durante dos años, encontrando que para un periodo de análisis mensual; se pudo observar estacionalidad en la concentración del material particulado).

En este capítulo se plantea una metodología que permite caracterizar espacialmente contaminantes del aire (variable de estudio o variable respuesta) de manera uni o multivariada. El primer caso se hace cuando solo se tienen datos disponibles de la variable respuesta y la caracterización se realiza mediante algoritmos de interpolación espacial. Se debe probar con diversos algoritmos (geoestadísticos o no) y determinar cual tiene mejor desempeño mediante pruebas de validación cruzada. Cuando se dispone de datos de otras variables que afectan la variable respuesta (variables explicativas), se puede hacer una caracterización espacial multivariada con algoritmos geoestadísticos multivariados o algoritmos de regresión espacial. La caracterización espacial uni o multivariada permite hacer estimaciones de la variable respuesta en el espacio a partir de funciones de análisis espacial. En consecuencia, los datos estimados son datos funcionales y permiten definir un dominio espacial para los procesos de caracterización espacio – temporales.

La metodología para la caracterización espacial univariada tiene dos propósitos, por un lado, permitiría hacer imputación espacial de datos faltantes y por otro lado analizar cómo se comporta en el espacio la variable. Tiene las siguientes fases:

- Si los datos no son estacionarios (no se distribuyen normalmente en el espacio), probar con algoritmos no geoestadísticos aplicando el proceso mostrado en la Figura 13.
- Para probar si los datos son estacionarios y poder aplicar algoritmos geoestadísticos, se hace un análisis exploratorio para los datos de la variable respuesta en cada uno de los meses mediante la metodología de la Figura 2. Con la metodología de la Figura 3, se hace un análisis variográfico para determinar el modelo de semivariograma teórico que tiene mejor desempeño usando pruebas de validación cruzada,
- Con los mejores semivariogramas teóricos obtenidos en cada mes, hacer un análisis de frecuencias para determinar el más frecuente.
- Con el semivariograma teórico más frecuente, hacer la estimación en cada uno de los meses y determinar los parámetros (θ_s, θ_r, h) y los estadísticos de error ASE, RMSE y RMSS.
- Validar si los parámetros calculados en el paso anterior son estadísticamente similares a su valor promedio en mes. Si la validación es satisfactoria se tiene el valor de cada parámetro para cada mes del año. Esto se hace por medio de pruebas ANOVA de una sola vía.
- Calcular el valor promedio de los datos de la variable respuesta para cada mes del año.
- Con los cálculos de los dos pasos anteriores, se hace la estimación para cada mes del año y se validan los valores estimados mediante pruebas de validación cruzada.

En la Figura 14 se muestra la metodología para la caracterización espacial univariada para variables estacionarias mediante algoritmos geoestadísticos.

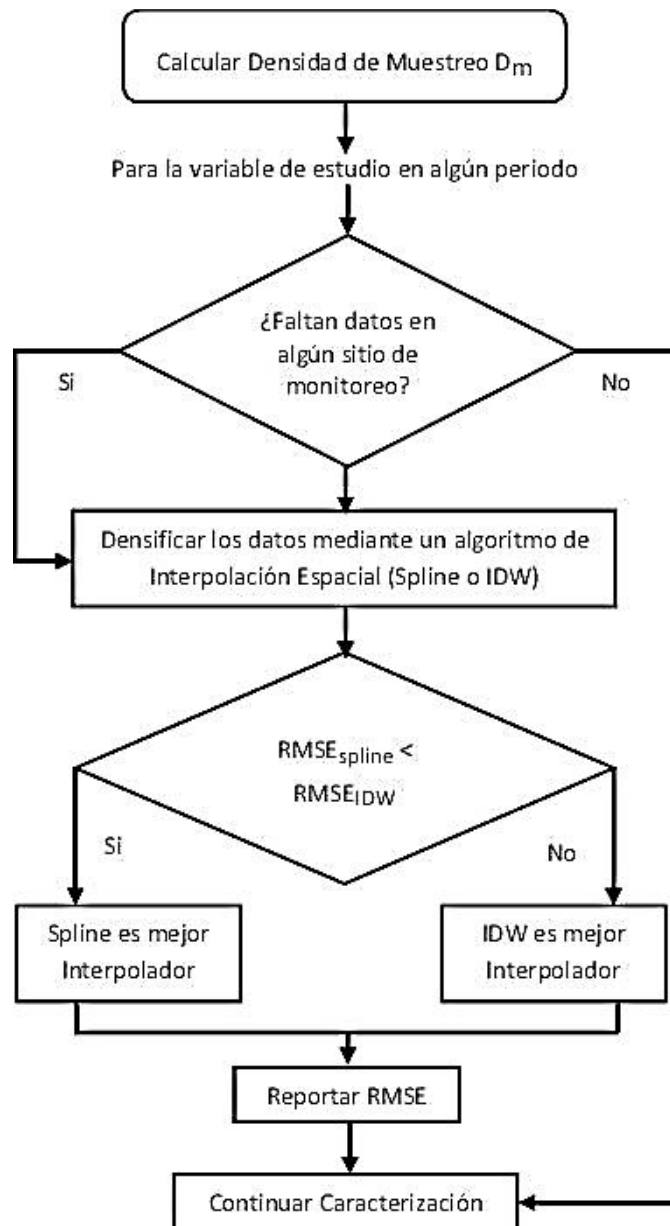


Figura 13. Densificación de datos mediante Algoritmos de Interpolación Espacial. Fuente propia.

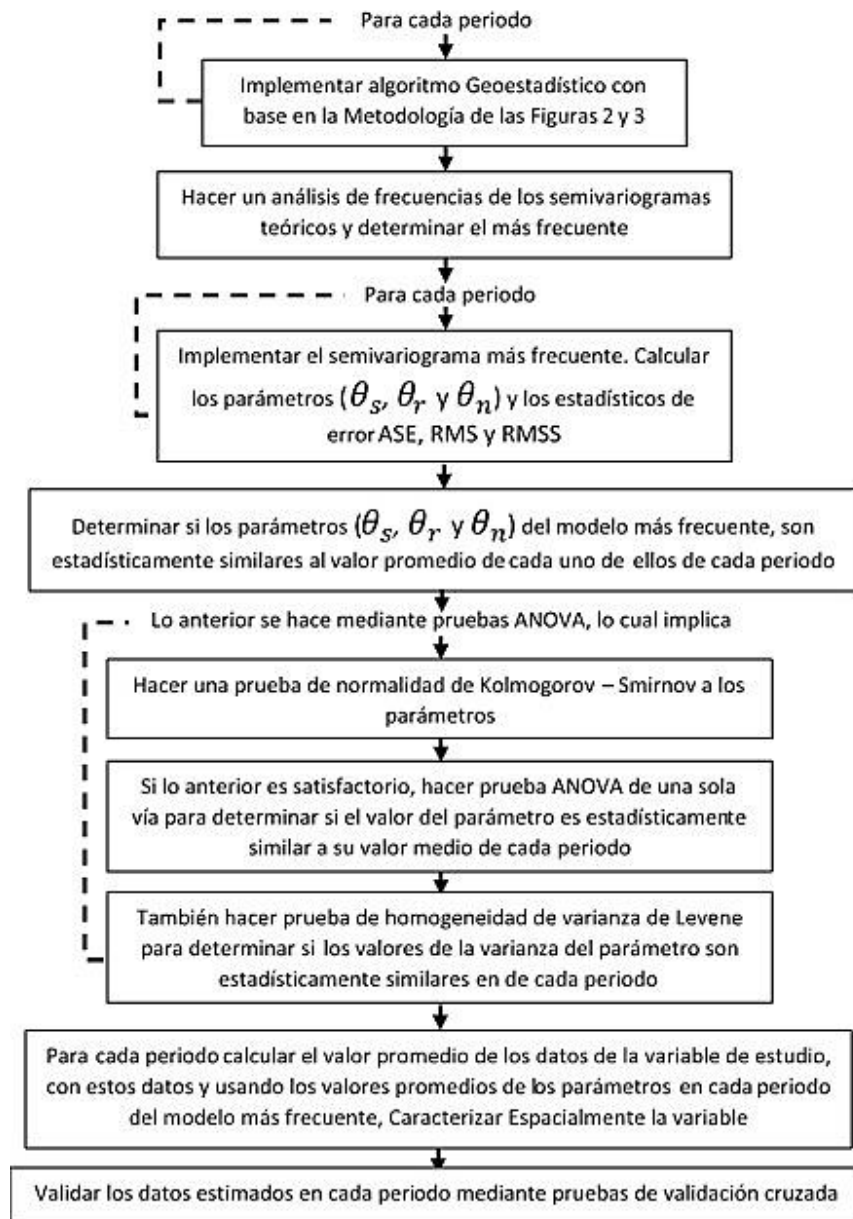


Figura 14. Metodología para la caracterización espacial univariada de contaminantes del aire. Fuente propia.

Si se tienen datos de otras variables que afectan la variable respuesta (variables explicativas), entonces se hace una caracterización espacial multivariada. La metodología tiene las siguientes fases:

- Si los datos de las variables explicativas están disponibles en sitios de monitoreo diferentes a los sitios de monitoreo de la variable respuesta, entonces se estiman los datos de las variables explicativas con algoritmos de interpolación espacial y

mediante funciones de análisis espacial se extraen los valores de las mismas en los sitios de monitoreo de la variable respuesta.

La caracterización espacial multivariada se puede hacer con algoritmos geoestadísticos multivariados (cokriging) o con algoritmos de regresión espacial (GWR®). La metodología para la caracterización espacial multivariada con algoritmos geoestadísticos multivariados tiene las siguientes fases:

- Previamente para la variable respuesta se debe hacer una caracterización espacial univariada con el fin de obtener el semivariograma teórico más frecuente.
- Se debe calcular el coeficiente de correlación entre la variable respuesta y cada una de las variables explicativas.
- Con las variables explicativas que tengan correlaciones fuertes con la variable respuesta y la información del primer paso, se implementa el cokriging.
- Los datos estimados se validan mediante validación cruzada.

La metodología para la caracterización espacial multivariada con algoritmos de regresión espacial tiene las siguientes fases:

- A partir de las posibles variables explicativas, se debe configurar un conjunto de ellas que cumplan adecuadamente con los criterios de especificación del método OLS. Se chequean exhaustivamente las combinaciones entre las posibles variables explicativas, y aquellas combinaciones que cumplan los criterios de especificación OLS ($\text{AdjR}^2 > 0,5$, el mínimo valor de AICc, (JB) $p\text{-value} > 0,1$, $\text{VIF} < 7,5$, K(BP) $p\text{-value} < 0,05$ y (SA) $p\text{-value} > 0,1$), serán las usadas en la Ecuación 8.
- Para calcular los términos de la Ecuación 8 con cada combinación de variables explicativas, se usa la herramienta GWR® del aplicativo ArcMap® de ESRI para datos espaciales. En cada período el mejor modelo de regresión será aquel con mayor valor de AdjR^2 y menor VIF. Esto se repite para cada mes.
- Los datos estimados se validan mediante validación cruzada.

En la Figura 15 se muestra la metodología para la caracterización espacial multivariada de contaminantes del aire.

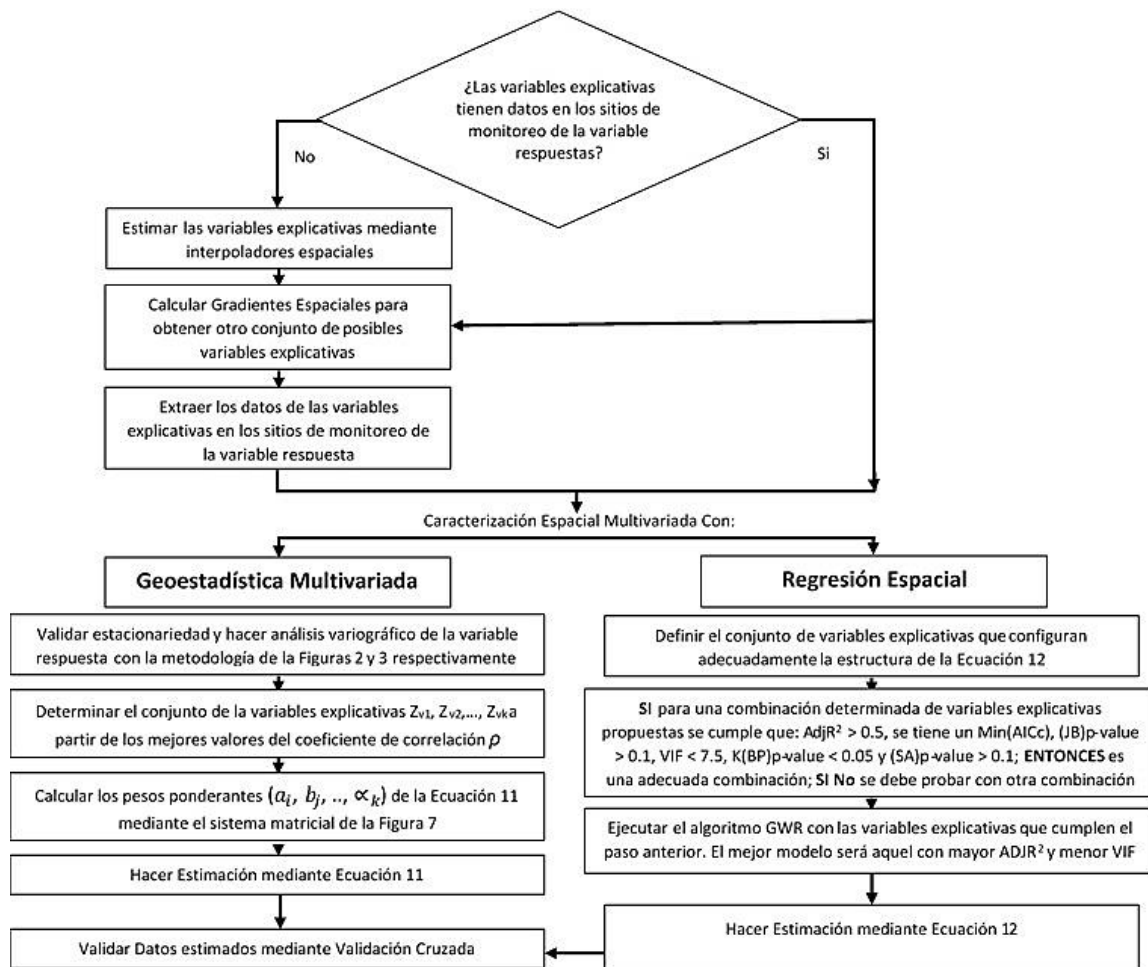


Figura 15. Metodología para la caracterización espacial Multivariada de contaminantes del aire. Fuente propia.

En las secciones 5.1 y 5.2, se aplica la metodología de caracterización espacial univariada para datos disponibles de PM10 y PM2.5 respectivamente en la zona de estudio. En las secciones 5.3 y 5.4, se aplica la metodología para la caracterización espacial multivariada para datos disponibles de las mismas variables. Como posibles variables explicativas se usan datos disponibles de Temperatura (T) y Velocidad del viento (Vv) con los que se calculan las variables Gradiente de Temperatura (∇T) y Gradiente de la Velocidad del viento (∇Vv).

5.1 Caracterización espacial univariada de PM10

En esta sección se muestran los cálculos para la caracterización espacial univariada de PM10. Con la metodología de la Figura 13 y usando los datos de la Tabla 14, en la Tabla 19, se muestran los valores de RMSE para el interpolador IDW y Spline.

Tabla 19. Cálculo de RMSE para cada interpolador

Algoritmo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc	Promedio	Des. Est
IDW	13,3	16,4	15,2	13,5	12,8	10,6	10,5	11,2	12,1	13,9	13,5	13,0	13,0	1,8
Spline	12,9	15,5	15,1	12,8	12,2	10,4	10,2	10,7	11,6	13,0	13,0	12,21	12,5	1,8

Seguidamente se usa la metodología de las Figuras 2 y 3 y se logra demostrar la estacionariedad de los datos y calcular los semivariogramas teóricos de cada mes. En la Tabla 20 se muestran las frecuencias de los modelos de semivariogramas teóricos obtenidos para los 24 meses.

Tabla 20. Frecuencias de los modelos de semivariograma teórico para PM10.

Modelo	Frecuencia Absoluta	Frecuencia Relativa
Hole Effect (HE)	8	0,33
J-Bessel (JB)	10	0,42
K-Bessel (KB)	1	0,04
Gauss (GS)	2	0,08
Exponential (EX)	2	0,08
Rational Cuadratic (RC)	1	0,04
Suma	24	1

Como el modelo J-Bessel es el más frecuente, se valida su desempeño en los 24 meses, obteniendo buenos resultados. Por tanto, éste es un buen modelo para la caracterización espacial de PM10. Usando la metodología de la Figura 14, y con pruebas ANOVA de una sola vía, se acepta que los valores medios de los parámetros en los diferentes grupos (meses) son estadísticamente similares para el modelo J_Bessel. La Ecuación 28 representa el modelo J-Bessel (Johnston et al., 2003).

$$\gamma(h, \theta) = \theta_n + \theta_s \left[1 - \frac{2^{\theta_d} \Gamma(\theta_d + 1)}{\phi} J_{\theta_d}(\phi) \right] \text{Ecuación 28}$$

En donde:

$$\phi = \Omega_{\theta_d} \frac{\|h\|}{\theta_r}$$

θ_n = parámetro *nugget*.

θ_s = parámetro *partial sill*.

θ_r = parámetro *major range*.

θ_d = parámetro calculado con la herramienta *Geostatistical Analyst*®.

$\Gamma(\theta_d + 1)$ = función gamma.

Ω_{θ_d} = valor calculado numéricamente sujeto a que $\theta_s \geq 0$, $\theta_r \geq 0$, $\theta_d \geq 0$; con $\gamma(\theta_r) = \theta_s$ y $\gamma'(\theta_r) < 0$.

Con base en los resultados de la prueba ANOVA, los valores de los parámetros del modelo J-Bessel (θ_s , θ_r y θ_n) de la Ecuación 28 se puede tomar como el valor promedio de los datos de cada mes y en este caso las unidades de θ_s y θ_n son $(\mu\text{g}/\text{m}^3)^2$ y las de θ_r son km. Usando los valores promedios de los parámetros del modelo JB de cada mes y la herramienta *Geostatistical Analyst*® del aplicativo *ArcMap*® de ESRI, se estima el valor de PM10 para cada mes del año. En la Tabla 21 se muestran los resultados.

Tabla 21. Datos estimados de PM10 (E) para cada mes del año con JB ($\mu\text{g}/\text{m}^3$).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	39,3	55,5	56,8	40,3	37,2	34,0	35,5	38,2	42,1	47,4	43,2	45,2
1	56,8	78,5	81,3	63,7	54,9	58,3	59,2	56,9	64,4	60,0	57,1	67,8
2	48,8	56,0	59,4	39,8	33,9	38,2	44,2	41,6	40,0	51,7	54,8	39,7
3	48,6	60,3	63,5	52,9	42,4	39,6	38,3	43,7	55,7	54,2	47,6	59,4
4	63,3	68,8	72,2	57,4	54,6	50,0	51,6	48,5	53,4	53,5	53,0	58,1
5	59,8	71,4	73,4	59,8	58,0	52,4	50,8	55,9	60,0	69,3	70,4	70,1
6	50,0	72,6	63,1	47,7	45,7	38,3	34,0	37,2	40,2	48,7	44,3	50,3
7	58,7	70,2	72,7	63,7	59,1	55,6	55,6	55,0	66,3	63,9	67,5	55,4
8	30,7	38,1	48,0	34,5	25,9	37,9	41,7	27,3	22,5	24,3	23,7	24,9
9	43,0	47,8	38,3	36,5	30,4	35,2	37,7	32,5	28,6	31,9	31,8	29,7
10	43,7	51,4	54,0	44,8	37,4	44,4	35,5	38,1	38,5	40,0	42,5	41,1
11	38,1	43,0	52,8	40,1	36,5	36,9	31,8	37,4	31,9	34,2	41,7	39,3
12	45,9	51,0	51,9	46,9	43,7	40,4	31,0	32,3	40,5	38,6	40,3	44,3

En las Tablas 22 y 23 se muestran los estadísticos de error para los datos estimados de PM10 mediante Kriging. El RMSE promedio de la estimación es de 4,2. Comparando los resultados de RMSE promedio de las Tablas 19 y 22, el mejor interpolador para el PM10 es el Kriging.

Tabla 22. Estadísticos de Error Datos Estimados (E) de PM10 con JB.

Estadísticos de Error	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
ASE	2,5	2,2	2,9	4,1	3,9	2,8	1,9	1,7	3,2	2,7	2,1	2,5
RMSE	3,6	4,3	6,6	5,5	4,5	4,2	2,5	2,3	3,7	4,0	4,8	4,0
RMSS	1,0	1,5	1,7	1,0	1,0	1,2	1,1	1,0	1,0	1,2	1,7	1,2

Tabla 23. ARE y MARE Datos Estimados (E) de PM10 con JB.

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc	MARE Sitio
0	0,10	0,15	0,16	0,03	0,07	0,05	0,02	0,04	0,04	0,10	0,11	0,10	0,08
1	0,13	0,00	0,06	0,07	0,01	0,00	0,02	0,04	0,08	0,11	0,18	0,17	0,07
2	0,04	0,03	0,07	0,19	0,12	0,02	0,04	0,04	0,05	0,23	0,23	0,05	0,09
3	0,00	0,03	0,13	0,03	0,13	0,07	0,10	0,01	0,09	0,10	0,08	0,15	0,08
4	0,01	0,01	0,05	0,08	0,12	0,06	0,03	0,02	0,03	0,02	0,01	0,03	0,04
5	0,05	0,08	0,14	0,00	0,03	0,07	0,03	0,01	0,04	0,08	0,09	0,05	0,06
6	0,03	0,08	0,02	0,23	0,14	0,02	0,08	0,08	0,05	0,03	0,03	0,06	0,07
7	0,02	0,05	0,08	0,02	0,02	0,07	0,03	0,03	0,05	0,03	0,06	0,00	0,04
8	0,06	0,02	0,14	0,06	0,04	0,19	0,04	0,08	0,26	0,00	0,13	0,01	0,09
9	0,01	0,07	0,04	0,04	0,10	0,08	0,03	0,07	0,11	0,04	0,13	0,06	0,06
10	0,20	0,22	0,30	0,14	0,09	0,29	0,03	0,13	0,09	0,01	0,09	0,02	0,13
11	0,05	0,07	0,18	0,02	0,13	0,09	0,07	0,09	0,09	0,03	0,05	0,03	0,08
12	0,04	0,06	0,12	0,08	0,15	0,15	0,15	0,10	0,00	0,01	0,04	0,05	0,08
MARE Mes	0,06	0,07	0,12	0,08	0,09	0,09	0,05	0,06	0,08	0,06	0,09	0,06	

De la Tabla 23, el valor promedio del MARE por sitio y por mes, es de 0,07 con una desviación estándar de 0,02. Por tanto, se infiere que en general el modelo JB hace una buena estimación de los datos por mes y por sitio de monitoreo. El MARE por mes varía entre 0,05 (julio) – 0,12 (marzo). El MARE por estación varía entre 0,04 - 0,13 en los sitios de monitoreo con ID 4 (Colegio el Rosario en Itagüí) y 10 (Planta de Tratamiento de Aguas San Fernando).

Lo anterior permite concluir que con los modelos J-Bessel se puede caracterizar espacialmente la concentración de PM10 en cada sitio de monitoreo y en cada mes del año, es decir, en cada mes se tiene una función de datos espaciales (datos funcionales - dominio espacial del fenómeno).

5.2 Caracterización espacial univariada de PM2.5

En esta sección se muestran los cálculos para la caracterización espacial univariada de PM2.5. Con la metodología de la Figura 13 y usando los datos de la Tabla 8, en la Tabla 24 se muestran los valores de RMSE para los interpoladores IDW y Spline.

Tabla 24. Cálculo de RMSE para cada interpolador

Algoritmo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc	Promedio	Desv. Est
IDW	4,7	4,5	4,6	4,6	4,9	4,0	3,0	3,7	5,3	6,4	7,9	8,5	5,2	1,6
Spline	4,5	4,4	4,6	4,6	4,8	3,7	3,0	3,8	5,4	6,3	7,8	8,3	5,1	1,6

De igual manera se usa la metodología de las Figuras 2 y 3 y se logra demostrar la estacionariedad de los datos y calcular los semivariogramas teóricos de cada mes. En la Tabla 25 se muestran las frecuencias de los modelos de semivariogramas teóricos obtenidos para los 24 meses.

Tabla 25. Frecuencias de los modelos de semivariograma teórico para PM2.5.

Modelo	Frecuencia Absoluta	Frecuencia Relativa
Hole Effect (HE)	9	0,38
J-Bessel (JB)	7	0,29
K-Bessel (KB)	5	0,21
Gauss (GS)	2	0,08
Rational Cuadratic (RC)	1	0,04
Suma	24	1

En la Tabla 26 se muestran los estadísticos de error de estimación promedio para los 24 meses modelados con *Hole Effect* de donde se infiere que se puede caracterizar espacialmente la concentración de PM2.5 en cada sitio de monitoreo y en cada mes del año con este modelo, es decir, en cada mes se tiene una función de datos espaciales (datos funcionales - dominio espacial del fenómeno).

Tabla 26. Estadísticos de error de estimación promedio

	ASE	RMSE	RMSS
Hole Effect	0,5	0,6	2,6

La Ecuación 29 representa el modelo *Hole Effect* (Johnston et al., 2003).

$$\gamma(h, \theta) = \theta_n + \theta_s \left[\frac{1 - \text{sen}\left(\frac{2\pi h}{\theta_r}\right)}{\text{sen}\left(\frac{2\pi h}{\theta_r}\right)} \right] \text{Ecuación 29}$$

En donde:

θ_n = parámetro *nugget*.

θ_s = parámetro *partial sill*.

θ_r = parámetro *major range*.

h = magnitud del vector de distancia.

Con base en los resultados de la prueba ANOVA, los valores de los parámetros del modelo *Hole Effect* (θ_s , θ_r y θ_n) de la Ecuación 29 se pueden tomar como el valor promedio de los datos de cada mes y en este caso las unidades de θ_s y θ_n son $(\mu\text{g}/\text{m}^3)^2$ y las de θ_r son km. En la Tabla 27 se muestran los valores promedio estimados de PM2.5 con el modelo *Hole Effect* para cada mes del año.

Tabla 27. Datos Estimados de PM2.5 (E) para cada mes del año con HE ($\mu\text{g}/\text{m}^3$).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	36,4	41,7	39,8	36,6	29,5	24,9	25,7	33,7	30,7	39,0	37,7	39,3
1	24,1	33,2	31,3	25,6	24,9	22,4	21,5	21,7	19,6	18,8	16,1	17,7
2	24,7	33,2	40,4	31,1	27,3	23,5	22,9	24,7	29,5	35,8	34,5	35,6
3	25,9	31,6	31,3	31,4	25,3	22,5	22,7	20,8	21,8	27,8	19,6	29,5
4	30,3	38,1	39,5	32,2	22,0	21,6	28,5	26,9	25,3	31,1	28,7	24,6
5	26,7	31,7	30,0	23,2	18,3	17,2	24,3	23,0	21,4	20,4	23,9	28,5
6	27,7	31,4	35,5	21,7	26,4	19,1	21,9	27,4	26,0	27,0	19,3	20,8
7	28,1	35,2	35,6	25,7	21,1	19,0	17,9	19,9	22,9	21,2	20,5	30,8

En las Tablas 28 y 29 se muestran los estadísticos de error para los datos estimados de PM2.5 mediante Kriging. El RMSE promedio de la estimación es de 5,4.

Tabla 28. Estadísticos de Error Datos Estimados (E) de PM2.5 con HE.

Estadísticos de Error	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
ASE	5,1	4,6	4,5	4,5	4,9	3,6	3,5	4,2	5,3	6,3	7,8	7,8
RMSE	4,2	5,6	4,0	5,0	5,4	4,7	4,2	3,4	5,5	6,7	7,9	7,8
RMSS	0,9	1,2	0,9	1,1	1,3	1,7	0,9	0,9	1,1	1,1	1,1	1,1

Tabla 29. ARE y MARE Datos Estimados (E) de PM2.5 con HE.

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc	MARE Sitio
0	0,07	0,06	0,07	0,01	0,11	0,12	0,09	0,10	0,09	0,07	0,06	0,03	0,07
1	0,01	0,04	0,05	0,03	0,06	0,07	0,17	0,10	0,03	0,03	0,14	0,03	0,06
2	0,08	0,10	0,04	0,04	0,03	0,15	0,09	0,04	0,04	0,07	0,11	0,10	0,07
3	0,04	0,10	0,06	0,11	0,01	0,06	0,17	0,04	0,05	0,15	0,12	0,14	0,09
4	0,09	0,03	0,01	0,06	0,09	0,14	0,13	0,03	0,10	0,08	0,03	0,14	0,08
5	0,05	0,08	0,08	0,08	0,07	0,10	0,08	0,08	0,13	0,11	0,13	0,04	0,09
6	0,10	0,01	0,06	0,11	0,03	0,16	0,02	0,10	0,12	0,02	0,11	0,08	0,07
7	0,04	0,04	0,04	0,07	0,13	0,07	0,12	0,06	0,13	0,11	0,08	0,11	0,08
MARE Mes	0,06	0,06	0,05	0,06	0,07	0,11	0,11	0,07	0,08	0,08	0,10	0,08	

De la Tabla 29, el valor promedio del MARE Sitio es de 0,08 con una desviación estándar de 0,02 y el valor promedio del MARE Mes es de 0,08 con una desviación estándar de 0,01.

5.3 Caracterización espacial multivariada de PM10

En esta sección se muestran los cálculos para la caracterización espacial multivariada de PM10 mediante la metodología propuesta en la Figura 15. Con PM10 como variable respuesta y T, ∇T , V_v y ∇V_v como posibles variables explicativas, es posible caracterizar espacialmente PM10 por medio del algoritmo geoestadístico multivariado (Cokriging) o con el algoritmo de regresión espacial (GWR). Para el cálculo se tomaron los datos de las Tablas 14 a 18.

5.3.1 Caracterización espacial multivariada de PM10 con el algoritmo geoestadístico multivariado (Cokriging)

Para implementar este algoritmo es necesario previamente hacer una caracterización espacial univariada de la variable respuesta con la metodología planteada en la Figura 14. En la aplicación de esta metodología en la sección 5.1, se obtienen datos funcionales para PM10 con el modelo J-Bessel. Este modelo será usado para el cálculo de las matrices de semivarianzas y covarianzas de cada una de las variables en el algoritmo Cokriging. Adicionalmente como se indica en la metodología de la Figura 15, se debe calcular el coeficiente de correlación entre la variable respuesta y las posibles variables explicativas. En la Tabla 30 se muestra el cálculo del coeficiente de correlación (ρ) de PM10 para cada mes con relación a T (ρ_{PM10-T}), ∇T ($\rho_{PM10-\nabla T}$), V_v (ρ_{PM10-V_v}) y ∇V_v ($\rho_{PM10-\nabla V_v}$).

Tabla 30: Coeficientes de Correlación PM10.

Mes	ρ_{PM10-T}	$\rho_{PM10-\nabla T}$	$\rho_{PM10-Vv}$	$\rho_{PM10-\nabla Vv}$
En	0,24	-0,07	0,25	0,52
Feb	0,19	-0,11	0,09	0,51
Mz	0,06	-0,02	0,16	0,53
Ab	0,24	-0,05	0,08	0,33
My	-0,13	0,07	-0,07	0,32
Jn	-0,08	0,02	0,11	0,58
Jl	-0,04	-0,22	0,22	0,48
Ag	-0,06	-0,04	-0,14	0,56
Sp	-0,07	-0,16	-0,20	0,63
Oc	-0,10	-0,14	-0,28	0,46
Nv	-0,17	0,02	-0,24	0,40
Dc	-0,19	-0,03	-0,36	0,52

De la Tabla 30 se infiere que en general las correlaciones de PM10 con T, ∇T , Vv y ∇Vv son débiles y por tanto no sería posible aplicar el algoritmo del Cokriging.

5.3.2 Caracterización espacial multivariada de PM10 con el algoritmo de regresión espacial

Como se indica en la metodología de la Figura 15, para implementar este algoritmo es necesario previamente hacer una adecuada configuración del modelo OLS, es decir, determinar la mejor combinación de las variables explicativas en cada mes para definir la mejor estructura de la Ecuación 8 y calcular sus coeficientes. Para esto se aplican los criterios de especificación del modelo OLS usando el aplicativo *Exploratory Regression*® de *ArcGis*®. En la Tabla 31 se muestran los valores de los criterios de especificación del mejor modelo OLS en cada mes.

Tabla 31. Criterios de especificación de los mejores modelos OLS en cada mes.

Criterios de Especificación	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
AdjR ²	0,15	0,14	0,16	0,12	0,09	0,22	0,11	0,20	0,27	0,09	0,03	0,11
AICc	103,15	108,35	106,20	110,17	108,12	97,61	96,90	99,43	101,12	106,25	108,57	106,59
JB	0,87	0,65	0,72	0,46	0,71	0,79	0,72	0,75	0,51	0,80	0,71	0,74
K(BP)	0,63	0,90	0,94	0,07	0,85	0,72	0,23	0,71	0,61	0,74	0,17	0,36
VIF	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,84	0,93	0,96	0,89	0,03	0,94	0,53	0,59	0,17	0,24	0,87	0,03
Variables explicativas del modelo	T, ∇Vv	T, ∇Vv	T, ∇Vv	T, ∇Vv	T, ∇Vv	T, ∇Vv	∇T , ∇Vv	T, ∇Vv	∇T , ∇Vv	∇T , ∇Vv	T, ∇Vv	T, ∇Vv

Con base en los valores de $AdjR^2$, los modelos implementados explican entre el 3% - 27% de la variabilidad de PM10 con un valor promedio de 0,14. Los valores estimados de PM10 con el algoritmo de regresión espacial se muestran en la Tabla 32.

Tabla 32. Datos Estimados (E) de PM10 para cada mes del año con GWR ($\mu g/m^3$).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	31,4	51,4	51,4	44,0	38,4	37,4	40,4	37,9	41,4	43,0	44,4	43,3
1	70,0	75,2	72,0	55,3	51,6	53,0	53,2	54,3	60,3	56,1	46,7	57,0
2	51,4	51,4	51,4	45,1	41,2	37,3	40,2	37,4	39,6	40,2	39,4	39,7
3	51,6	53,7	51,6	48,8	47,9	37,6	40,7	38,2	46,4	45,3	45,2	46,6
4	58,0	65,1	65,0	57,9	56,6	47,7	48,2	43,8	50,1	51,0	48,6	55,1
5	53,4	63,5	59,4	54,5	52,1	45,0	47,5	51,4	53,0	58,4	59,4	61,3
6	52,0	62,2	59,0	57,2	49,4	38,5	41,4	38,9	43,7	44,9	45,9	46,0
7	61,3	61,5	63,3	58,4	52,2	47,1	50,0	48,4	58,1	57,3	58,1	51,6
8	35,4	41,5	44,4	35,2	27,3	37,3	40,2	27,2	33,9	26,9	23,2	27,9
9	45,5	47,6	41,5	41,3	38,9	37,4	40,2	37,6	37,6	34,8	31,3	35,7
10	40,1	44,0	47,1	41,8	37,4	36,8	40,2	35,2	45,3	44,4	43,6	46,0
11	40,1	42,5	48,1	44,6	45,8	39,3	37,1	38,5	37,2	38,0	45,2	42,8
12	47,9	52,8	48,9	46,8	44,3	40,6	39,9	40,9	45,0	42,1	41,7	45,9

En las Tablas 33 y 34 se muestran los estadísticos de error para los datos estimados de PM10 con GWR®. El RMSE promedio de la estimación es de 3,90.

Tabla 33. Estadísticos de error datos estimados (E) de PM10 con GWR.

Estadísticos de Error	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
ASE	2,7	3,6	3,7	3,8	3,8	4,2	3,4	3,4	3,2	3,2	3,7	3,6
RMSE	3,6	3,8	4,0	3,9	4,1	4,5	3,8	4,0	3,7	3,5	4,0	3,9
RMSS	1,1	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0

Tabla 34. ARE y MARE datos estimados (E) de PM10 con GWR.

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc	MARE Sitio
0	0,12	0,06	0,05	0,12	0,10	0,16	0,16	0,03	0,03	0,01	0,14	0,06	0,09
1	0,07	0,04	0,06	0,08	0,07	0,09	0,08	0,01	0,02	0,04	0,03	0,02	0,05
2	0,01	0,06	0,08	0,08	0,07	0,04	0,05	0,07	0,04	0,05	0,12	0,05	0,06
3	0,06	0,08	0,08	0,05	0,01	0,11	0,05	0,11	0,09	0,08	0,03	0,10	0,07
4	0,08	0,06	0,06	0,07	0,08	0,10	0,09	0,11	0,09	0,07	0,08	0,08	0,08
5	0,06	0,04	0,07	0,09	0,07	0,08	0,09	0,07	0,09	0,09	0,08	0,08	0,08
6	0,07	0,07	0,08	0,08	0,07	0,02	0,12	0,03	0,03	0,04	0,07	0,04	0,06
7	0,07	0,08	0,06	0,06	0,10	0,09	0,08	0,09	0,08	0,08	0,08	0,07	0,08
8	0,09	0,11	0,06	0,08	0,10	0,18	0,01	0,08	0,11	0,11	0,10	0,11	0,09
9	0,04	0,06	0,04	0,09	0,15	0,14	0,03	0,24	0,17	0,13	0,11	0,13	0,11
10	0,10	0,05	0,13	0,06	0,09	0,07	0,10	0,04	0,07	0,12	0,12	0,15	0,09
11	0,10	0,06	0,08	0,09	0,09	0,16	0,08	0,12	0,06	0,08	0,13	0,12	0,10
12	0,08	0,10	0,05	0,08	0,16	0,16	0,10	0,14	0,12	0,08	0,07	0,09	0,10
MARE Mes	0,07	0,07	0,07	0,08	0,09	0,11	0,08	0,09	0,07	0,07	0,09	0,08	

De la Tabla 34, el valor promedio del MARE sitio es de 0,08 con una desviación estándar de 0,02; y el valor promedio del MARE Mes es de 0,08 con una desviación estándar de 0,01.

5.4 Caracterización espacial multivariada de PM2.5

En esta sección se muestran los cálculos para la caracterización espacial multivariada de PM2.5 mediante la metodología propuesta en la Figura 15. Con PM2.5 como variable respuesta y T, ∇T , V_v y ∇V_v como posibles variables explicativas, es posible caracterizar espacialmente PM2.5 por medio del algoritmo geoestadístico multivariado (Cokriging) o con el algoritmo de regresión espacial (GWR). Para el cálculo se tomaron los datos de las Tablas 8 a 12.

5.4.1 Caracterización espacial multivariada de PM2.5 con el algoritmo geoestadístico multivariado (Cokriging)

Al igual que para PM10, se hizo previamente una caracterización espacial univariada de PM2.5 con la metodología de la Figura 14. En la aplicación de esta metodología en la sección 5.2, se obtuvieron datos funcionales para PM2.5 con el modelo *Hole Effect*. Este modelo se usa para el cálculo de las matrices de semivarianzas y covarianzas de cada una de las variables en el algoritmo Cokriging. Adicionalmente como se indica en la metodología de la Figura 15 se calcula el coeficiente de correlación entre la variable respuesta y las posibles variables explicativas. En la Tabla 35 se muestra el cálculo del coeficiente de correlación (ρ) de PM2.5 para cada mes con relación a T ($\rho_{PM2.5-T}$), ∇T ($\rho_{PM2.5-\nabla T}$), V_v ($\rho_{PM2.5-V_v}$) y ∇V_v ($\rho_{PM2.5-\nabla V_v}$).

De la Tabla 35 se infiere que en general las correlaciones de PM2.5 con T, ∇T , V_v y ∇V_v son débiles pero más altas que las calculadas para PM10. Por lo tanto, no sería posible aplicar el algoritmo del Cokriging.

Tabla 35. Coeficientes de correlación PM2.5.

Mes	$\rho_{PM2.5-T}$	$\rho_{PM2.5-\nabla T}$	$\rho_{PM2.5-Vv}$	$\rho_{PM2.5-\nabla Vv}$
En	0,54	0,29	0,35	-0,23
Feb	0,53	0,00	0,14	-0,19
Mz	0,49	-0,15	0,50	-0,35
Ab	0,69	-0,27	0,07	-0,44
My	0,67	-0,69	0,13	-0,63
Jn	0,63	-0,24	0,10	-0,15
Jl	0,48	0,14	0,39	-0,23
Ag	0,55	0,29	0,45	-0,07
Sp	0,48	0,02	0,25	-0,11
Oc	0,47	-0,21	0,29	-0,22
Nv	0,54	-0,18	0,07	-0,24
Dc	0,46	0,06	0,10	-0,12

5.4.2 Caracterización espacial multivariada de PM2.5 con el algoritmo de regresión espacial

Al igual que para PM10 se aplica la metodología de la Figura 15, se hace previamente una adecuada configuración del modelo OLS determinando la mejor combinación de las variables explicativas en cada mes con los criterios de especificación del modelo OLS. Se usa el aplicativo *Exploratory Regression*® de *ArcGis*® y luego se calculan los coeficientes de la Ecuación 8 con el algoritmo *GWR*®. En la Tabla 36 se muestran los valores de los criterios de especificación del mejor modelo OLS en cada mes.

Tabla 36. Criterios de especificación de los mejores modelos OLS en cada mes.

Criterios de Especificación	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
AdjR ²	0,17	0,12	0,25	0,16	0,15	0,14	0,13	0,17	0,15	0,11	0,23	0,16
AICc	64,46	53,45	60,60	51,22	52,36	52,09	55,18	55,49	59,09	61,18	64,71	64,95
JB	0,68	0,93	0,83	0,95	0,44	0,27	0,84	0,69	0,75	0,88	0,48	0,86
K(BP)	0,49	0,10	0,55	0,07	0,42	0,78	0,27	0,07	0,25	0,27	0,52	0,80
VIF	1,04	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
SA	0,27	0,79	0,85	0,90	0,54	0,85	0,58	0,52	0,66	0,58	0,59	0,84
Variables explicativas del modelo	T, Vv	T	T, Vv	T	∇T	∇Vv	T	∇T	∇Vv	∇Vv	∇T	Vv

Con base en los valores de AdjR², los modelos implementados explican entre el 11% - 25% de la variabilidad de PM2.5 con un valor promedio de 0,16. Los valores estimados de PM2.5 con el algoritmo de Regresión Espacial se muestran en la Tabla 37.

Tabla 37. Datos Estimados (E) de PM2.5 para cada mes del año mediante GWR ($\mu\text{g}/\text{m}^3$).

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	34,7	40,5	38,5	31,7	29,4	26,3	25,3	25,7	30,6	32,4	36,1	37,6
1	27,6	35,5	34,3	29,7	26,3	23,4	22,2	22,4	23,8	17,8	20,3	22,3
2	31,1	35,3	37,4	30,3	24,5	21,5	22,8	23,1	25,9	29,9	27,8	27,4
3	29,3	33,1	30,2	27,3	29,0	24,5	21,1	22,8	25,9	27,9	26,1	27,9
4	30,6	39,9	44,2	31,2	26,6	21,4	22,1	23,9	25,7	27,5	26,2	25,0
5	26,8	30,4	30,7	27,5	13,9	20,1	20,1	24,2	27,2	24,1	22,5	23,8
6	28,4	33,0	37,4	26,3	25,9	21,0	20,7	22,6	25,1	26,5	26,3	23,6
7	24,1	32,2	31,4	25,8	27,4	21,2	19,7	22,0	24,6	27,5	26,4	24,5

En las Tablas 38 y 39 se muestran los estadísticos de error para los datos estimados de PM2.5 con GWR®. El RMSE promedio de la estimación es de 3,00.

Tabla 38. Estadísticos de Error Datos Estimados (E) de PM2.5 con GWR.

Estadísticos de Error	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
ASE	3,0	2,4	3,3	2,3	2,8	1,7	2,7	3,0	3,1	2,5	2,9	3,3
RMSE	3,2	2,8	3,5	2,6	2,9	2,3	2,8	3,1	3,4	2,9	3,3	3,7
RMSS	0,9	1,0	0,9	0,9	0,9	1,1	0,9	0,9	1,0	0,9	1,0	1,0

Tabla 39. ARE y MARE Datos Estimados (E) de PM2.5 con GWR.

Sitio de Monitoreo	En	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc	MARE Sitio
0	0,11	0,09	0,10	0,13	0,11	0,07	0,11	0,16	0,09	0,12	0,10	0,08	0,11
1	0,13	0,11	0,15	0,13	0,12	0,12	0,21	0,13	0,18	0,09	0,08	0,22	0,14
2	0,16	0,04	0,04	0,07	0,08	0,05	0,09	0,11	0,16	0,11	0,11	0,15	0,10
3	0,09	0,06	0,10	0,04	0,16	0,16	0,09	0,14	0,13	0,16	0,17	0,08	0,11
4	0,08	0,08	0,11	0,03	0,10	0,13	0,12	0,09	0,08	0,04	0,06	0,12	0,09
5	0,04	0,12	0,06	0,09	0,19	0,05	0,10	0,14	0,11	0,06	0,06	0,13	0,10
6	0,12	0,04	0,11	0,09	0,05	0,08	0,07	0,10	0,08	0,04	0,21	0,22	0,10
7	0,11	0,04	0,08	0,06	0,13	0,19	0,23	0,17	0,21	0,15	0,18	0,12	0,14
MARE Mes	0,11	0,07	0,09	0,08	0,12	0,11	0,13	0,13	0,13	0,10	0,12	0,14	

De la Tabla 39, el valor promedio del MARE sitio es de 0,11 con una desviación estándar de 0,02; y el valor promedio del MARE Mes es de 0,11 con una desviación estándar de 0,02.

6 CARACTERIZACIÓN ESPACIO-TEMPORAL DE DATOS DE CONCENTRACIÓN DE MATERIAL PARTICULADO CON ALGORITMOS ECONOMÉTRICOS

Como se indicó en el capítulo 5, los datos de las variables ambientales son espaciales, es decir, se miden simultáneamente en el espacio y en tiempo (en sitios de monitoreo y en algún período), por tanto, la estructura de las series de tiempo más adecuada para su análisis espacio-temporal es la de los datos de panel. En esta capítulo se plantea una metodología que permite caracterizar espacio-temporalmente la concentración de material particulado en el aire (variable de estudio o variable respuesta) y para ello se usan como posibles variables explicativas los datos funcionales estimados de la variable respuesta con la metodología de caracterización espacial multivariada de la Figura 15 (GWR) y los datos de T , V_v , ∇T y ∇V_v .

La metodología para la caracterización espacio-temporal de la concentración de PM10 y PM2.5 tiene las siguientes fases:

- Organizar las series de tiempo de los datos de la variable respuesta y de las posibles variables explicativas en estructuras de datos de panel.
- Para cada serie de tiempo, probar su estacionariedad mediante pruebas de raíces unitarias (PURT).
- Determinar cuáles posibles variables explicativas causan o efectivamente explican a la variable respuesta. Esto se hace a través de pruebas de causalidad para definir el conjunto real de las variables explicativas (Prueba de Causalidad de Granger – PCG).
- Calcular con el método Ad-Hoc el valor del rezago de las variables explicativas.
- Probar distintos modelos temporales y determinar cuál de ellos es el mejor. En este trabajo se explora con modelos RD, AR, AR-RD y con el modelo AR-RD propuesto en la Figura 8 y la Ecuación 19 usando como componente espacial de rezago distribuido del modelo ($SJ_i^{spatial}$) los datos funcionales calculados con GWR®.
- De los modelos explorados en el paso anterior, determinar el mejor de ellos con base en el menor valor de SCR y el mayor valor de AdjR².
- Validar los datos estimados por el mejor modelo con pruebas de validación cruzada.

En la Figura 16 se muestra la metodología desarrollada para la caracterización espacio - temporal del material particulado.

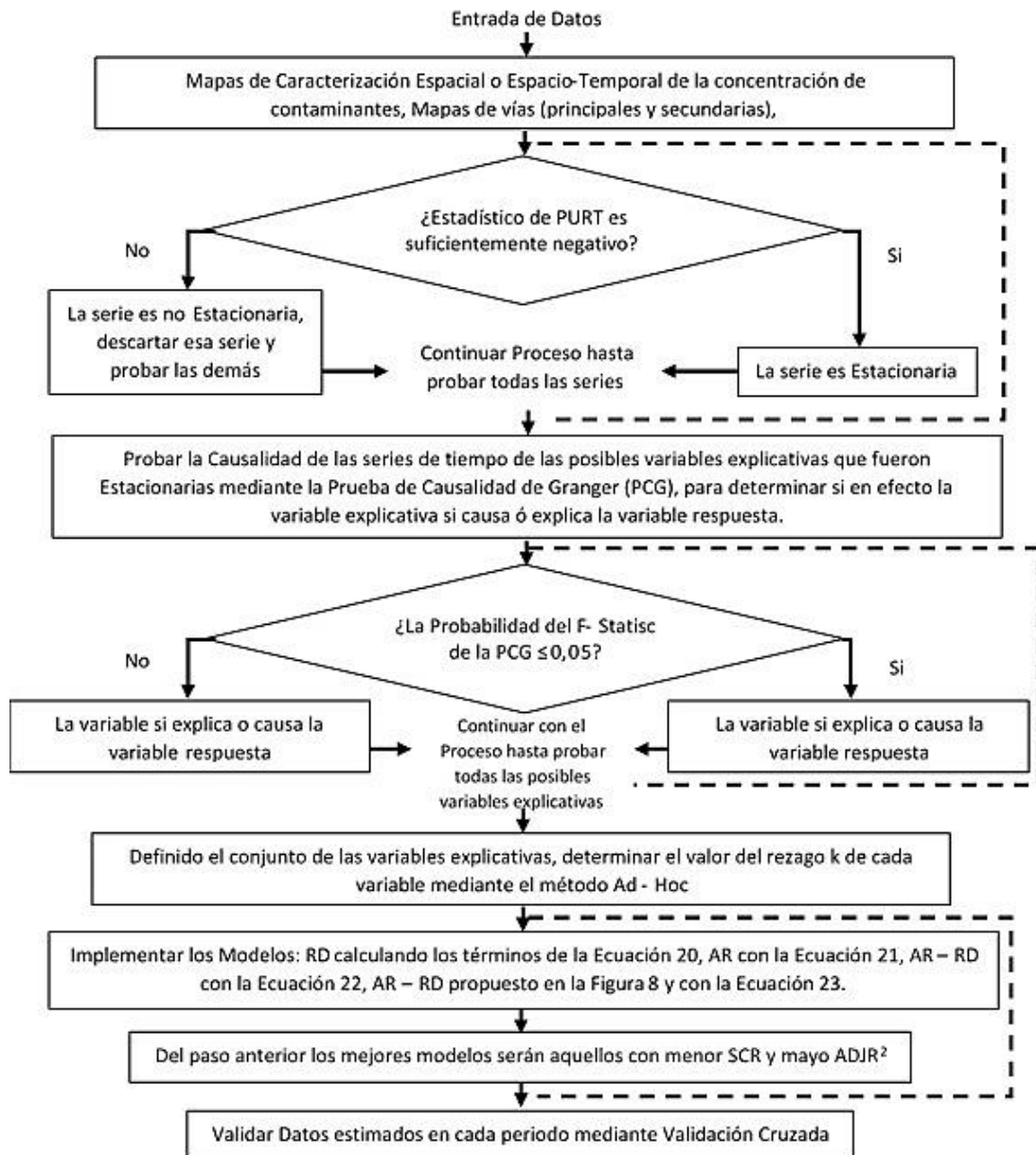


Figura 16. Metodología propuesta para la caracterización espacio-temporal de material particulado. Fuente propia.

6.1 Caracterización espacio-temporal de PM10

En esta sección se muestra la aplicación de la metodología propuesta en la Figura 16 para caracterizar espacio temporalmente la variable PM10. Se toman los datos disponibles del período 2013 a 2014 en los sitios de monitoreo de la Tabla 13 y como datos observados de PM10 (O_i) los datos mensuales de la Tabla 14. Las posibles variables explicativas son:

- La temperatura T de Tabla 15.
- La Velocidad del viento V_v de Tabla 16.
- El Gradiente Espacial de la Temperatura ∇T de Tabla 17.
- El Gradiente Espacial de la Velocidad del viento ∇V_v de Tabla 18.
- Para la caracterización espacio temporal usando el modelo AR-RD propuesto en la Ecuación 19 y la Figura 8, se usa como dominio espacial el de los datos funcionales de PM10 calculados con GWR de la Tabla 31. Estos datos serán el componente espacial de rezago distribuido del modelo ($SJ_{i \text{ spatial}}$).

En la Tabla 40 se muestran los resultados de las pruebas de estacionariedad (PUR) y de causalidad (PCG) de las posibles variables explicativas, la variable GWR son los datos funcionales de PM10 calculados con GWR. De la segunda columna de la Tabla 40 se infiere que las series de las posibles variables explicativas son estacionarias, pero la tercera columna muestra que solo las variables T, ∇V_v y GWR causan o explican a PM10. En consecuencia, se correrán los algoritmos econométricos con combinaciones de estas variables. El valor del rezago calculado con el método Ad-Hoc fue de 1 ($k = 1$).

Tabla 40. Resultados pruebas de estacionariedad y causalidad para las variables explicativas de PM10.

Variable	Resultado PUR	Resultado PCG (valor probabilidad F-Statistic)
T	-5,38	0,010
V_v	-4,12	0,250
∇T	-2,69	0,340
∇V_v	-3,95	0,040
GWR	-2,86	0,004

Los mejores resultados de combinaciones de las variables explicativas se muestran en la Tabla 41 para diferentes modelos econométricos.

Tabla 41. Mejores resultados de combinaciones de las variables explicativas para PM10.

Modelo	Variables Explicativas	ADJR ²	SCR
RD	T(-1) ∇Vv(-1)	0,56	2241
AR	PM10(-1)	0,88	2150
AR – RD	T (-1) ∇Vv(-1) PM10(-1)	0,88	2128
RD con GWR como variable explicativa	GWR(-1) T(-1) ∇Vv(-1)	0,55	2261
AR con GWR como variable explicativa	GWR(-1) PM10(-1)	0,88	2150
AR-RD con GWR como variable explicativa	GWR(-1) T(-1) ∇Vv(-1) PM10(-1)	0,88	2090

De la Tabla 41 se infiere que el modelo que explica mejor la variabilidad de PM10 en un 88% es un AR–RD como el propuesto en la Figura 8, con un componente espacial de rezago distribuido cuyo dominio espacial corresponde a los datos calculados con la metodología de caracterización espacial multivariada usando GWR con rezago 1. Igualmente son buenos modelos el AR–RD que incluye T, ∇Vv y PM10(-1) que explica también un 88% de la variabilidad de PM10 pero con mayor SCR, así como los modelos AR–RD y AR. En la Figura 17 se muestran, para cada panel (sitio de monitoreo) las series de tiempo observadas de PM10 (O) vs los datos estimados de PM10 (E) con el modelo AR-RD de componente espacial de rezago distribuido GWR.

Los resultados de Figura 17 se obtienen con la Ecuación 30.

$$PM10_t = 10,73 - 0,23GWR_{t-1} + 0,09T_{t-1} + 126,49∇Vv_{t-1} - 0,94PM10_{t-1} \quad \text{Ecuación 30.}$$

Los valores estimados de PM10 con la Ecuación 30 y algoritmos econométricos con un modelo AR–RD de componente espacial de rezago distribuido GWR se muestran en la Tabla 42.

En las Tablas 43 y 44 se muestran los estadísticos de error para los datos estimados de PM10 con el modelo AR-COK. El RMSE promedio de la estimación es de 3,40.

De la Tabla 44, el valor promedio del MARE sitio es de 0,07 con una desviación estándar de 0,02 y el valor promedio del MARE Mes es de 0,07 con una desviación estándar de 0,01.

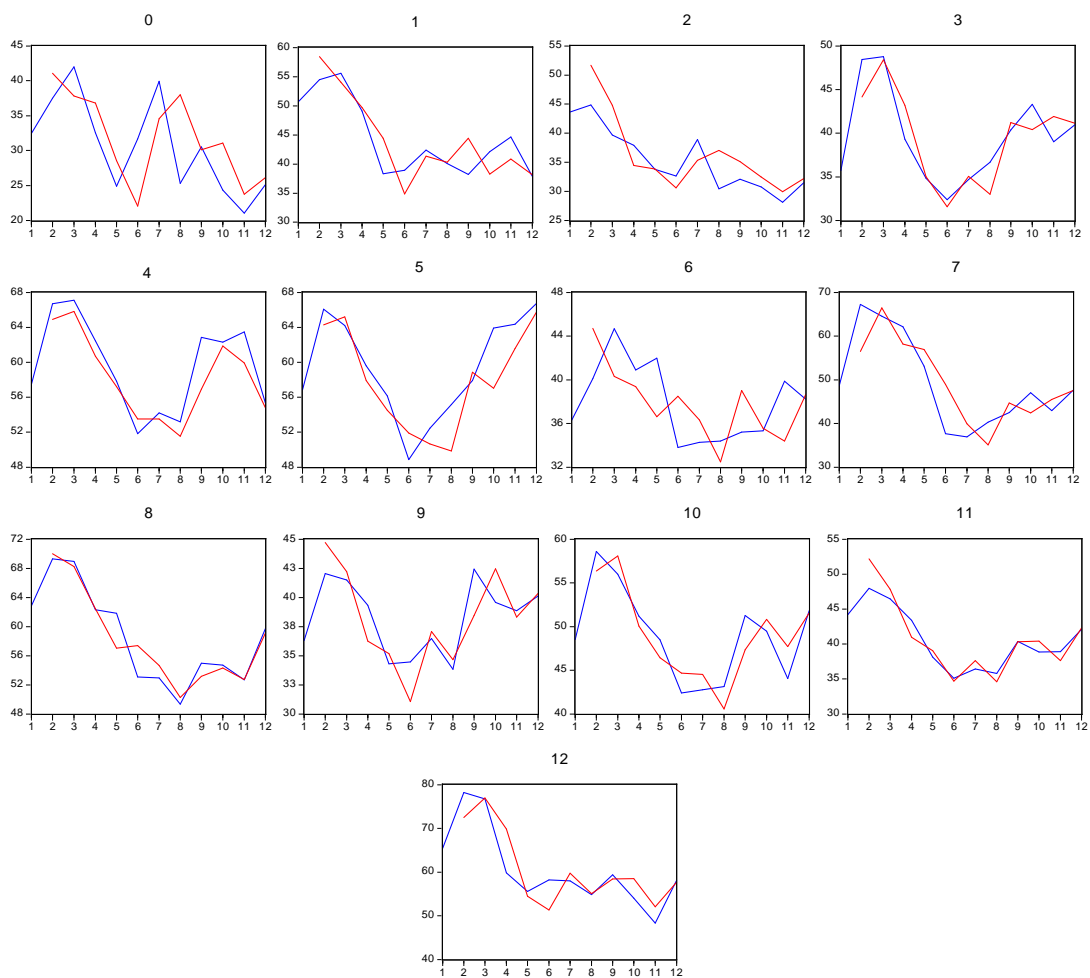


Figura 17. PM_{10} (O) en azul vs. PM_{10} (E) con AR-RD con dominio espacial GWR en rojo.

Tabla 42. PM_{10} (E) para cada mes del año con modelo AR-RD de componente espacial de rezago distribuido GRW ($\mu\text{g}/\text{m}^3$).

Sitio de Monitoreo	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	45,4	44,1	43,4	37,2	35,0	31,6	35,1	37,0	41,2	40,4	41,9
1	74,0	72,5	56,9	59,9	54,5	53,3	59,7	55,0	58,4	52,5	54,1
2	51,4	58,4	54,0	41,7	41,4	47,8	41,4	40,3	44,4	40,3	40,8
3	54,6	58,4	54,1	50,1	46,4	44,6	44,5	46,6	47,4	47,8	47,7
4	66,0	70,0	64,2	62,5	57,0	57,4	53,6	50,2	53,2	54,3	54,7
5	61,4	62,3	63,2	57,9	52,5	54,9	50,7	55,9	59,9	60,0	63,6
6	63,0	63,5	66,5	57,2	41,9	41,0	37,9	39,1	44,7	40,4	45,5
7	63,3	64,9	65,8	60,7	56,2	51,5	51,5	58,5	57,9	61,9	59,9
8	35,4	40,1	36,8	29,8	28,6	37,1	29,6	28,0	27,1	25,1	23,7
9	43,5	43,7	42,8	35,5	33,8	34,6	35,3	35,0	35,1	32,5	29,9
10	40,1	44,7	42,2	36,2	36,2	31,1	37,1	38,6	38,4	42,5	38,3
11	43,1	46,7	38,3	39,4	36,6	38,5	36,3	32,5	39,0	35,5	34,4
12	43,9	52,2	47,8	40,9	39,0	34,7	37,6	36,6	40,3	40,4	37,6

Tabla 43. Estadísticos de Error Datos Estimados de PM10 (E) con AR-RD GWR.

Estadísticos de Error	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
ASE	1,3	2,8	1,8	1,3	1,9	3,6	2,4	1,7	2,8	3,1	2,2
RMSE	3,4	3,2	3,7	2,9	3,4	3,8	3,3	3,6	3,0	3,4	3,3
RMSS	1,7	1,0	1,4	1,5	1,3	1,0	1,1	1,5	1,0	1,0	1,1

Tabla 44. ARE y MARE Datos Estimados de PM10 (E) con AR-RD GWR.

Sitio de Monitoreo	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc	MARE Sitio
0	0,06	0,09	0,10	0,07	0,08	0,09	0,04	0,08	0,05	0,04	0,02	0,07
1	0,05	0,06	0,05	0,08	0,06	0,08	0,09	0,07	0,08	0,09	0,07	0,07
2	0,06	0,05	0,10	0,09	0,06	0,13	0,03	0,05	0,05	0,10	0,08	0,07
3	0,07	0,04	0,06	0,03	0,09	0,04	0,03	0,09	0,04	0,09	0,08	0,06
4	0,05	0,02	0,03	0,01	0,07	0,08	0,09	0,09	0,03	0,03	0,08	0,05
5	0,07	0,03	0,06	0,03	0,07	0,05	0,08	0,04	0,06	0,07	0,05	0,06
6	0,06	0,02	0,07	0,08	0,11	0,11	0,06	0,08	0,05	0,06	0,04	0,07
7	0,05	0,03	0,05	0,05	0,08	0,05	0,03	0,07	0,07	0,03	0,08	0,05
8	0,06	0,05	0,13	0,20	0,10	0,07	0,17	0,08	0,11	0,19	0,06	0,11
9	0,03	0,10	0,13	0,05	0,04	0,11	0,16	0,09	0,14	0,15	0,05	0,10
10	0,05	0,08	0,07	0,06	0,05	0,15	0,10	0,09	0,03	0,09	0,05	0,07
11	0,08	0,05	0,06	0,06	0,08	0,12	0,06	0,08	0,10	0,11	0,10	0,08
12	0,08	0,12	0,10	0,07	0,11	0,05	0,05	0,09	0,04	0,04	0,11	0,08
MARE Mes	0,06	0,06	0,08	0,07	0,08	0,09	0,08	0,08	0,07	0,08	0,07	

6.2 Caracterización espacio – temporal de PM2.5

En esta sección se muestra la aplicación de la metodología propuesta en la Figura 15 para caracterizar espacio temporalmente la variable PM2.5. Se toman los datos disponibles del período 2013 y 2014 en los sitios de monitoreo de la Tabla 7 y como datos observados de PM2.5 (O_i) los datos mensuales de la Tabla 8. Las posibles variables explicativas son:

- La temperatura T de Tabla 9.
- La Velocidad del viento Vv de Tabla 10.
- El Gradiente Espacial de la Temperatura ∇T de Tabla 11.
- El Gradiente Espacial de la Velocidad del viento ∇V_v de Tabla 12.
- Para la caracterización espacio temporal usando el modelo AR – RD propuesto en la Ecuación 19 y la Figura 8, se usan como dominio espacial los datos funcionales de PM2.5 calculados con GWR de la Tabla 36; estos datos serán el componente espacial de rezago distribuido del modelo ($SJ_i^{spatial}$).

En la Tabla 45 se muestran los resultados de las pruebas de estacionariedad (PUR) y de causalidad (PCG) de las posibles variables explicativas. La variable GWR son los datos funcionales de PM2.5 calculados con GWR. De la segunda columna de la Tabla 45 se infiere que las series de las posibles variables explicativas son estacionarias, pero la tercera columna muestra que solo las variables T, ∇V_v y GWR causan o explican a PM2.5. En consecuencia, se correrán los algoritmos econométricos con combinaciones de estas variables. El valor del rezago calculado con el método Ad-Hoc fue de 1 ($k = 1$).

Tabla 45. Estacionariedad y causalidad de las variables explicativas PM2.5.

Variable	Resultado PUR	Resultado PCG (valor probabilidad F-Statistic)
T	-5,31	0,007
V _v	-3,69	0,720
∇T	-2,52	0,080
∇V_v	-4,03	0,030
GWR	-3,12	0,012

Los mejores resultados de combinaciones de las variables explicativas se muestran en la Tabla 46 para diferentes modelos econométricos.

Tabla 46. Mejores combinaciones de las variables explicativas para PM2.5.

Modelo	Variables Explicativas	ADJR ²	SCR
RD	T(-1), $\nabla V_v(-1)$	0,76	497
AR	PM2.5(-1)	0,86	448
AR – RD	T (-1), $\nabla V_v(-1)$, PM2.5(-1)	0,86	444
RD con GWR	GWR(-1), T(-1), $\nabla V_v(-1)$	0,75	504
AR con GWR	GWR(-1), PM2.5(-1)	0,86	431
AR-RD con GWR	GWR(-1), T(-1), $\nabla V_v(-1)$, PM2.5(-1)	0,87	404

De la Tabla 46 se infiere que el modelo que explica mejor la variabilidad de PM2.5 en un 87% es un modelo como el propuesto en la Figura 8, AR–RD más un componente espacial de rezago distribuido cuyo dominio espacial son los datos calculados con la metodología de caracterización espacial multivariada usando GWR con rezago 1. Igualmente son buenos un modelo AR más un componente espacial de rezago distribuido cuyo dominio espacial está dado por los datos calculados con la metodología de caracterización espacial multivariada usando GWR que explica un 86% de la variabilidad de PM2.5 y un modelo AR–RD que también explica el 86% de la variabilidad de PM2.5 pero con mayor SCR. En la Figura 18 se muestran, para cada panel (sitio de monitoreo), las series de tiempo observadas de PM2.5

(O) vs. los datos estimados de PM2.5 (E) con el modelo AR–RD de componente espacial de rezago distribuido GWR.

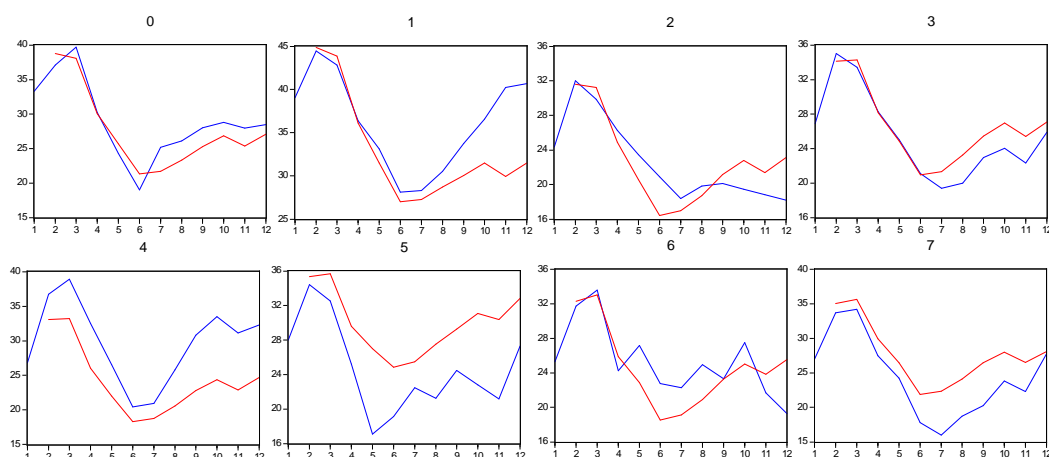


Figura 18. PM2.5 (O) en azul Vs PM2.5 (E) con AR – RD GWR en rojo.

Los resultados de Figura 18 se obtienen con la Ecuación 31.

$$PM2.5_t = 5,95 - 0,49GWR_{t-1} + 0,41 T_{t-1} - 6,53 \nabla v_{t-1} + 0,85 PM2.5_{t-1} \quad \text{Ecuación 31}$$

Los valores estimados de PM2.5 con la Ecuación 31 y algoritmos econométricos con un modelo AR RD con componente espacial de rezago distribuido GWR se muestran en la Tabla 47.

Tabla 47. PM2.5 (E) para cada mes del año con modelo AR RD de componente espacial de rezago distribuido GWR ($\mu\text{g}/\text{m}^3$).

Sitio de Monitoreo	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
0	44,89	43,89	36,25	31,57	27,38	28,26	30,02	31,32	33,21	32,03	33,91
1	31,28	32,15	24,02	21,09	16,06	16,39	18,42	19,96	22,12	21,39	23,20
2	33,15	33,86	26,69	22,81	19,62	20,40	22,38	25,12	26,75	25,22	27,17
3	34,64	34,83	29,17	25,83	21,56	21,52	23,08	25,29	26,28	24,16	25,45
4	38,39	37,88	29,54	25,31	21,17	22,14	24,13	26,51	28,47	27,28	29,20
5	35,08	35,31	29,58	26,07	20,69	20,88	22,71	25,38	26,99	25,50	27,04
6	33,01	33,01	27,88	23,63	19,78	20,89	23,02	26,47	28,45	27,20	29,14
7	34,72	34,12	27,52	24,68	23,01	22,55	23,38	23,65	24,29	22,94	24,82

En las Tablas 48 y 49 se muestran los estadísticos de error para los datos estimados de PM2.5 con el modelo AR RD GWR. El RMSE promedio de la estimación es de 3,17.

Tabla 48. Estadísticos de Error Datos Estimados por mes de PM2.5 (E) con AR RD GWR.

Estadísticos de Error	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc
ASE	1,18	1,88	2,23	2,81	2,34	2,17	2,31	2,44	2,61	3,71	3,90
RMSE	1,52	2,38	3,01	3,82	2,92	2,87	2,62	2,91	3,31	4,50	5,05
RMSS	0,94	0,94	0,94	0,94	0,94	0,94	0,94	0,94	0,94	0,94	0,94

Tabla 49. ARE y MARE Datos Estimados por mes de PM2.5 (E) con AR RD GWR.

Sitio de Monitoreo	Feb	Mz	Ab	My	Jn	Jl	Ag	Sp	Oc	Nv	Dc	MARE Sitio
0	0,01	0,03	0,00	0,05	0,03	0,00	0,02	0,07	0,09	0,20	0,17	0,06
1	0,02	0,08	0,09	0,10	0,23	0,11	0,07	0,01	0,14	0,14	0,27	0,11
2	0,10	0,13	0,18	0,14	0,04	0,03	0,13	0,18	0,20	0,19	0,16	0,13
3	0,01	0,04	0,03	0,03	0,02	0,11	0,16	0,10	0,09	0,08	0,02	0,06
4	0,04	0,05	0,02	0,04	0,11	0,12	0,08	0,05	0,01	0,02	0,03	0,05
5	0,02	0,09	0,17	0,52	0,08	0,07	0,07	0,04	0,18	0,20	0,01	0,13
6	0,04	0,02	0,15	0,13	0,13	0,06	0,08	0,14	0,03	0,25	0,51	0,14
7	0,03	0,00	0,00	0,02	0,29	0,41	0,25	0,17	0,02	0,03	0,11	0,12
MARE Mes	0,03	0,05	0,08	0,13	0,12	0,11	0,11	0,09	0,10	0,14	0,16	

De la Tabla 49, el valor promedio del MARE sitio es de 0,10 con una desviación estándar de 0,03; y el valor promedio del MARE Mes es de 0,10 con una desviación estándar de 0,04.

7 MODELO DE PROXIMIDAD ESPACIAL (MPE)

Derivado de esta tesis, los autores publicaron un artículo (Londoño et al., 2017) en donde se analiza el efecto que tiene la proximidad a las fuentes de emisión y de mitigación, sobre el valor de la concentración del material particulado en la ciudad de Medellín por medio de un Modelo de Proximidad Espacial (MPE), basado en la idea de que, cuando hay una aproximación a fuentes de emisión como las vías y las industrias, la concentración del contaminante aumenta y cuando hay una aproximación a las fuentes de mitigación como las zonas verdes la concentración disminuye. En el modelo se calcula un factor de proximidad espacial que contempla simultáneamente las cercanías a fuentes de emisión y de mitigación, el factor corrige el valor de la concentración del contaminante entre un 80% y un 220% y permite hacer un análisis más detallado de la caracterización espacial por la inclusión de la variable distancia euclidiana a las fuentes. Como insumos de entradas para el MPE, se toman los datos de la caracterización espacial o espacio-temporal de la concentración del contaminante, las vías principales y secundarias de la ciudad como fuentes de emisión móviles de material particulado, las zonas de producción industrial como fuentes de emisión fijas y como fuentes de mitigación las zonas verdes. El modelo se implementa mediante la herramienta *Spatial Analyst*® del aplicativo *ArcMap*® de ESRI. En la Figura 19, se muestra el diagrama de flujo del MPE.

El factor de efecto combinado FT supone que el eje central de las vías principales contribuye con un 100% en el aumento de la concentración del material particulado, las zonas industriales contribuyen con un aumento del 20% y las zonas verdes contribuyen con una atenuación del 20%. En la Figura 19, el factor F1, está dado por la Ecuación 32.

$$F_1 = \frac{(\lambda_1 O_{R1} + \lambda_2 O_{R2})}{[\lambda_1 Max(O_{R1}) + \lambda_2 Max(O_{R2})]} \text{Ecuación 32}$$

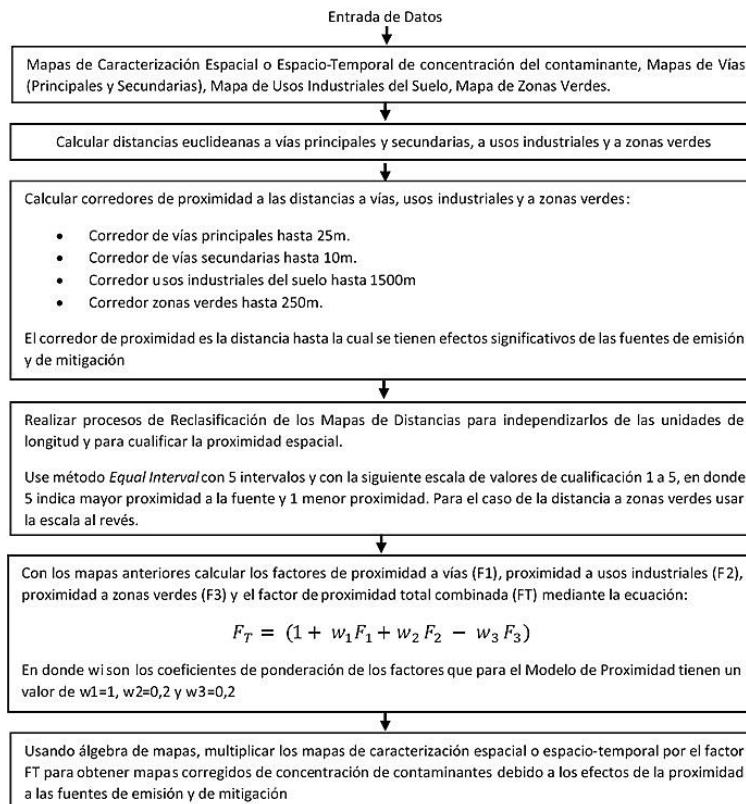


Figura 19. Modelo de proximidad espacial. Fuente propia.

En la Ecuación 32, λ_1 y λ_2 son los pesos ponderados del mapa de reclasificación de la distancia a vías principales (O_{R1}) y a vías secundarias (O_{R2}), calculados mediante el método de análisis multicriterio del proceso jerárquico analítico AHP de Saaty (Gass & Fu, 2013). Siendo O_{R2} más importante que O_{R1} , los pesos calculados con el algoritmo AHP fueron $\lambda_1=0,83$ y $\lambda_2 = 0,17$. La función $Max(O_{Rj})$, calcula el valor máximo de O_{Rj} tomando en cada celda una vecindad rectangular de 3×3 celdas mediante funciones de estadística de bloque. Los factores F_2 y F_3 normalizan los mapas reclasificados de distancia a usos industriales y zonas verdes respectivamente y se calculan mediante la Ecuación 33, con j igual a 2 y 3, en donde O_{R3} es el mapa reclasificado de la distancia a usos industriales y O_{R4} es el mapa reclasificado de la distancia a zonas verdes.

$$F_j = 0,2 * O_{Rk} \quad \text{Con } j=2 \text{ cuando } k =3 \text{ y } j=3 \text{ cuando } k = 4. \quad \text{Ecuación 33.}$$

En la Figura 20 se muestran los datos de entrada usados en el Modelo de Proximidad Espacial: vías principales, secundarias, zonas de usos industriales y zonas verdes.

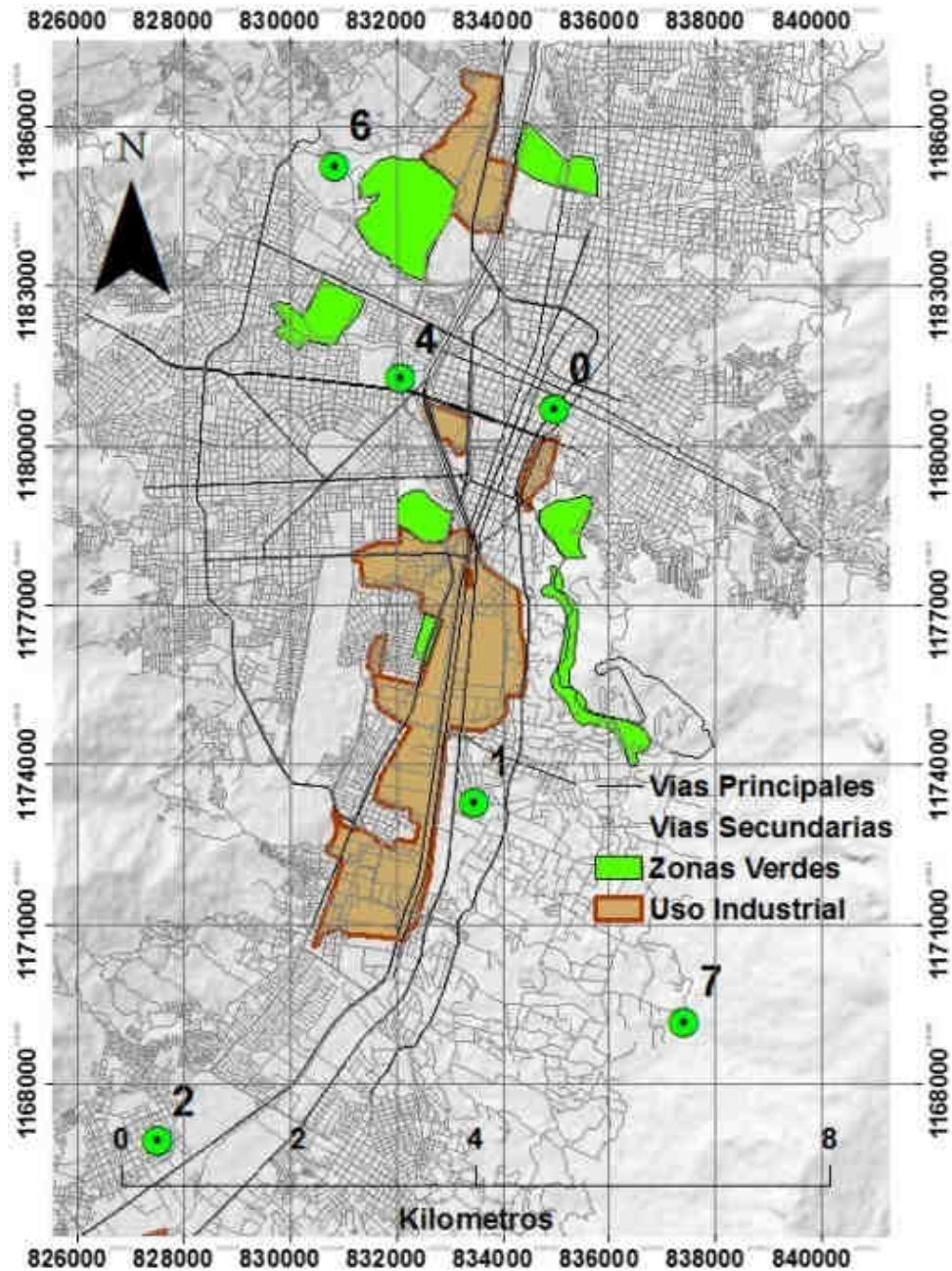


Figura 20. Insumos de entrada al Modelo de Proximidad Espacial.

8 ANÁLISIS DE RESULTADOS

En los capítulos anteriores se presentó una propuesta metodológica para caracterizar espacial y espacio-temporalmente la concentración de material particulado en el aire. Ésta se implementa a partir de la disponibilidad de datos de PM10 y PM2.5, de temperatura y velocidad del viento en el área metropolitana de la ciudad de Medellín en distintos períodos. Esta metodología tiene las siguientes características:

- Los aspectos procedimentales claves para su desarrollo se presentan esquemáticamente con diagramas de flujo (Figuras 2, 3, 13, 14, 15 y 16).
- Se ejemplifica su desarrollo en detalle con la aplicación de los algoritmos a las variables PM10 y PM2.5.
- Se validan los resultados de los valores estimados con cada algoritmo usando pruebas de validación cruzada mostrando en cada caso, 4 estadísticos de error.

El consolidado de los resultados obtenidos en la caracterización espacial y espacio-temporal de PM10 se muestran en la Tabla 50, de donde se ve que los modelos de caracterización espacial explican poco su variabilidad, la cual mejora significativamente al incluir el término temporal que implícitamente contiene elementos de la caracterización espacial multivariada en el componente de rezago distribuido del dato funcional. Otro aspecto importante se da en la inclusión de las variables gradiente de temperatura y gradiente de la velocidad del viento, que resultaron importantes para mejorar la descripción de la variabilidad de la concentración del material particulado. Para el caso de PM2.5, las variables explicativas temperatura y gradiente de la velocidad del viento fueron también relevantes en la descripción del fenómeno, como se muestra en el consolidado de la Tabla 51. Al comparar los estadísticos de error ambas Tablas, en general se obtuvieron mejores resultados en la estimación de la concentración de PM2.5 con relación al PM10.

Un resumen consolidado de los valores promedios mensuales de los estadísticos de error de cada contaminante, se muestran en la Tabla 52.

Tabla 50: Resultados Caracterización Espacial y Espacio - Temporal de PM10.

Algoritmo	Variables	AdjR ²	Estadísticos de Error por Mes			
Geoestadístico Univariado (J-Bessel)			ASE	RMSE	RMSS	MARE
Período 2013 – 2014						
Dm 64,3	PM10	Min	1,70	2,30	1,00	0,05
Área 836,4 km ²		Max	4,10	6,60	1,70	0,12
13 Sitios Monitoreo (156 datos)		Media	3,43	4,04	1,01	0,05
		Desviación Estándar	0,36	1,25	0,27	0,02
Regresión Espacial						
Multivariada (GWR)	T, VVv (En, Fb, Mz, Ab, My, Jn, Ag, Nv, Dc).		ASE	RMSE	RMSS	MARE
Período 2013 – 2014		0,14				
Dm 64,3	VT, VVv (Jl, Sp y Oc).	Min	2,70	3,50	1,00	0,07
Área 836,4 km ²		Max	4,20	4,50	1,10	0,11
13 Sitios Monitoreo (156 datos)		Media	3,53	3,90	1,01	0,08
		Desviación Estándar	0,38	0,26	0,03	0,01
Econométrico (AR – RD con componente espacial de rezago distribuido GWR)	GWR t-1		ASE	RMSE	RMSS	MARE
Período 2013 - 2014	T t-1	0,88				
Dm 64,3	VVv t-1	Min	1,30	2,90	1,00	0,06
Área 836,4 km ²	PM10 t-1	Max	3,60	3,80	1,70	0,09
13 Sitios Monitoreo (156 datos)		Media	2,26	3,36	1,24	0,07
		Desviación Estándar	0,75	0,27	0,25	0,01

Tabla 51. Resultados Caracterización Espacial y Espacio - Temporal de PM2.5.

Algoritmo	Variables	ADJR ²	Estadísticos de Error por Mes			
Geoestadístico Univariado (Hole – Effect))			ASE	RMSE	RMSS	MARE
Período 2013 - 2014						
Dm 67,1	PM2.5	Min	3,50	3,40	0,90	0,05
Área 536,7 km ²		Max	7,80	7,90	1,70	0,11
8 Sitios Monitoreo (96 datos)		Media	5,20	5,40	1,10	0,08
		Desviación Estándar	1,40	1,40	0,20	0,02
Regresión Espacial						
Multivariada (GWR)	T (Fb, Ab, Jl). T, Vv (En, Mz).		ASE	RMSE	RMSS	MARE
Período 2013 - 2014	VT (My, Ag, Nv).	0,16				
Dm 67,1	VVv (Jn, SP, Oc).	Min	1,70	2,30	0,90	0,07
Área 536,7 km ²	Vv (Dc).	Max	3,30	3,70	1,10	0,14
8 Sitios Monitoreo (96 datos)		Media	2,75	3,04	0,95	0,11
		Desviación Estándar	0,46	0,40	0,07	0,02
Econométrico (AR RD con componente espacial de rezago distribuido GWR)	GWR t-1		ASE	RMSE	RMSS	MARE
Período 2013 - 2014	T t-1	0,87				
Dm 67,1	VVv t-1	Min	1,18	1,52	0,94	0,03
Área 536,7 km ²	PM2.5 t-1	Max	3,90	5,05	0,94	0,16
8 Sitios Monitoreo (96 datos)		Media	2,51	3,17	0,94	0,10
		Desviación Estándar	0,77	0,98	0,00	0,04

Tabla 52. Resultado global de las estimaciones por mes.

	PM10				PM2.5			
	ASE	RMSE	RMSS	MARE	ASE	RMSE	RMSS	MARE
Min	1,98	2,61	0,95	0,05	1,32	2,06	0,91	0,06
Max	3,78	4,49	1,74	0,09	3,20	3,76	1,14	0,14
Media	2,87	3,59	1,11	0,07	2,42	2,89	0,97	0,10
Desviación Estándar	0,53	0,62	0,26	0,01	0,53	0,50	0,07	0,03

El PM10 y el PM2.5 han llamado la atención de los investigadores por los peligros potenciales sobre la salud de los habitantes de los centros urbanos densamente poblados. En especial el PM2.5, ya que sus altas concentraciones generan trastornos respiratorios en poblaciones vulnerables de niños menores de cinco años y adultos mayores de 60 años (Song et al., 2017; García et al., 2016; MinAmbiente, 2016). Por tanto, en adelante se hará un análisis de los resultados obtenidos de este contaminante.

Del modelo de estimación espacio - temporal de PM2.5 (Ecuación 31) se puede decir que la concentración mensual y espacial de éste contaminante depende de:

- Como se comporta su concentración en el espacio en el mes anterior. Este comportamiento se representa con la caracterización espacial multivariada hecha con el algoritmo GWR que incluye los efectos de T, ∇T , Vv y ∇Vv en cada mes (fila 2, columna 2 de la Tabla 51) y que define el componente espacial de rezago distribuido por medio de datos funcionales.
- El valor de T en el sitio en el mes anterior.
- El valor de ∇Vv en el sitio en el mes anterior.
- El valor de la concentración de PM2.5 en el sitio en el mes anterior. Con base en los resultados de la Tabla 46, ésta es la variable explicativa más importante en la variabilidad de PM2.5 en el tiempo ya que, comparando los mejores modelos con el modelo AR, las demás variables explicativas solo aportan en un 1% la mejora del desempeño de los modelos.

En la Figura 21 se muestran los datos mensuales observados de PM2.5. En amarillo el valor promedio, en verde el valor mínimo y en rojo el valor máximo. En ella se observa un comportamiento cíclico (estacional) en la concentración del contaminante (Liu et al., 2017). En la curva amarilla se ve un punto de máxima entre febrero y marzo con valores de concentración entre 29,9 y 44,5 $\mu\text{g}/\text{m}^3$, un punto de mínima entre junio y julio con valores de concentración entre 16,90 y 28,3 $\mu\text{g}/\text{m}^3$ y un punto promedio entre septiembre y octubre con valores de concentración entre 19,5 y 36,6 $\mu\text{g}/\text{m}^3$. De allí se infiere que el período más contaminado en la ciudad se da entre febrero-marzo, el menos contaminado en junio-julio y el de contaminación cercana al valor promedio anual en septiembre-octubre.

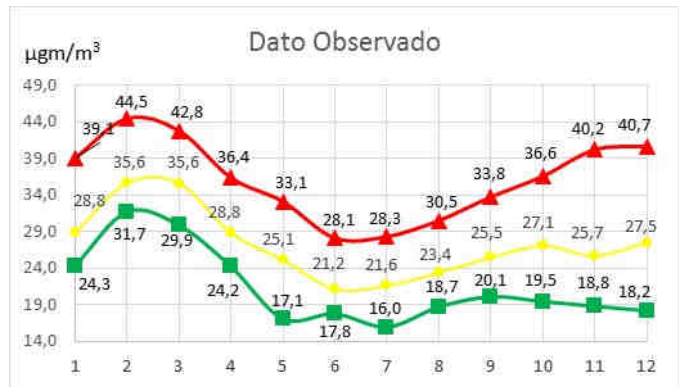


Figura 21. Dato promedio mensual observado.

Los valores mensuales estimados por cada uno de los algoritmos de caracterización espacial y espacio-temporal muestran el mismo patrón cíclico mensual de los datos observados (Figura 22).

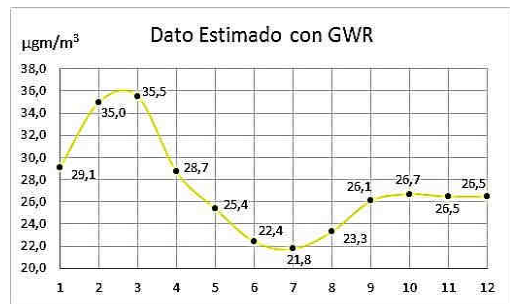
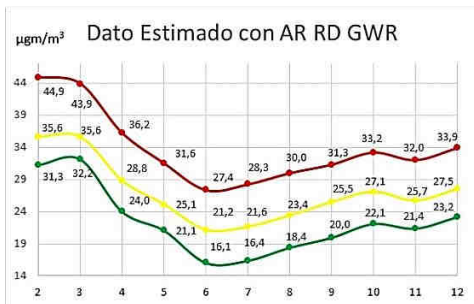


Figura 22. Dato promedio mensual estimado

Con relación a las estimaciones hechas en los sitios de monitoreo, en la Tabla 53 se muestran los estadísticos de error de estimación por sitio. Con base en ellos se infiere que la estimación de mejor desempeño es la del sitio con ID = 4 (Universidad Nacional Núcleo

Volador) y la de menor desempeño es la del sitio con ID = 6 (Universidad Nacional Facultad de Minas).

Tabla 53. Estadísticos de error de estimación de PM2.5 por sitio de monitoreo.

Estadísticos de Error	0	1	2	3	4	5	6	7
ASE	2,49	2,44	4,21	1,52	1,52	2,71	3,14	2,38
RMSE	3,58	2,81	4,62	1,77	1,72	3,57	3,98	3,19
RMSS	1,22	0,96	1,47	1,55	1,03	1,21	1,02	1,06
MARE	0,06	0,12	0,14	0,07	0,05	0,13	0,14	0,12

Un resumen de los estadísticos de error de la Tabla 53 se muestra en la Tabla 54.

Tabla 54. Resultado global de las estimaciones por sitio de Monitoreo.

	ASE	RMSE	RMSS	MARE
Min	1,52	1,72	0,96	0,05
Max	4,21	4,62	1,55	0,14
Media	2,55	3,16	1,19	0,10
Desviación Estándar	0,87	1,02	0,22	0,03

Comparando los valores de las desviaciones estándar promedio de las estimaciones de la parte derecha de la Tabla 52 con las de la Tabla 54, se infiere que hay una mayor desviación estándar en los estadísticos de error de los sitios de monitoreo que en el de los meses. Por tanto, hay menos dispersión en los resultados de la estimación mensual que en los resultados de la estimación por sitios. Analizando los valores de la concentración promedio del contaminante en los sitios de monitoreo (Tabla 55), se infiere que la zona más contaminada de la ciudad está entre los sitios de monitoreo con ID = 0 (Museo de Antioquia) e ID = 4 (Universidad Nacional Núcleo Volador), que la zona menos contaminada de la ciudad está entre los sitios de monitoreo con ID = 1 (Politécnico Colombiano Jaime Isaza Cadavid) e ID = 7 (El Poblado) y que la zona con una confiabilidad tendiente al valor promedio anual está en el sitio de monitoreo con ID = 5 (Corporación Universitaria Lasallista – Municipio de Caldas).

Tabla 55. Resumen de los valores estimados de PM2.5 por sitio.

Valores Sitio Monitoreo	0	1	2	3	4	5	6	7
Mínimo	27,4	16,1	19,6	21,5	21,2	20,7	19,8	22,6
Máximo	44,9	32,2	33,9	34,8	38,4	35,3	33,0	34,7
Promedio	33,9	22,4	25,7	26,5	28,2	26,8	26,6	26,0
Desviación estándar	5,8	5,3	4,6	4,6	5,6	4,9	4,4	4,4

De las consideraciones anteriores se concluye que en febrero-marzo, en la zona entre ID = 0 e ID = 4, la contaminación es crítica (esta zona corresponde con la parte central de la ciudad). En junio-julio, en la zona entre ID = 1 e ID = 7 la contaminación es la más baja (está zona corresponde al sector del barrio El Poblado de la ciudad). En septiembre-octubre la contaminación tiende al valor promedio anual en la zona con ID =5 (que corresponde con el sur del área metropolitana en el Municipio de Caldas).

El comportamiento mensual promedio de las variables explicativas del modelo espacio-temporal (T , ∇T , Vv y ∇Vv), se muestra en las Figuras 23 a 26.

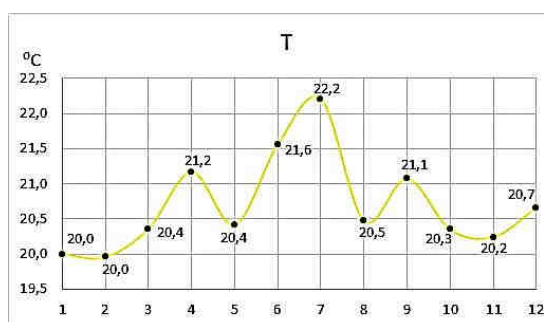


Figura 23. Valor promedio mensual de T.

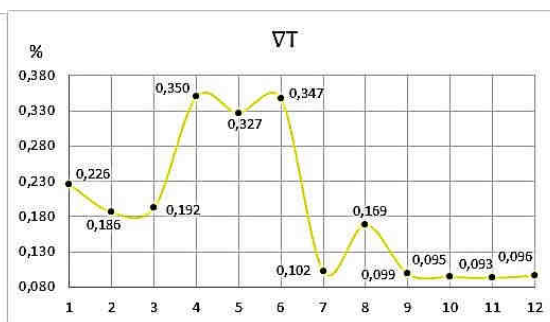


Figura 24. Valor promedio mensual de ∇T .

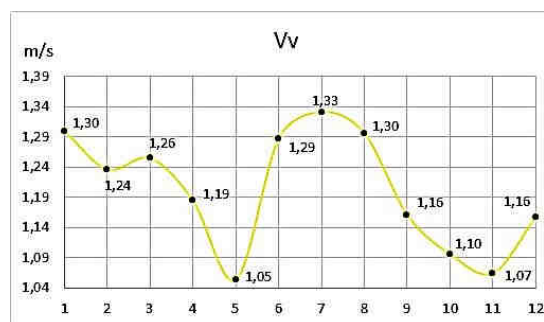


Figura 25. Valor promedio mensual de Vv .

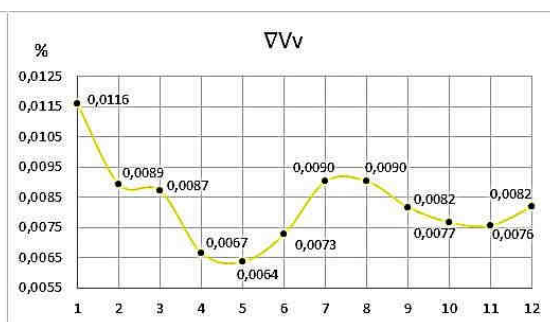


Figura 26. Valor promedio mensual de ∇Vv .

De las Figuras anteriores se ve que las variables explicativas tienen un comportamiento periódico. La T promedio mensual muestra un comportamiento simétrico, similar en el primer y el segundo semestre del año con un punto de máxima en el mes de julio. El ∇T promedio mensual no es tan simétrico como T, cae de enero a febrero en donde permanece casi estable hasta marzo, desde allí muestra un aumento brusco hasta su valor máximo en abril en donde alcanza una meseta en la que oscila levemente hasta junio y luego cae

abruptamente hasta casi su valor mínimo en julio, se restablece un poco en agosto y cae nuevamente a sus valores mínimos y allí permanece hasta diciembre. El valor promedio mensual de V_v no es tan simétrico como el de la temperatura, sin embargo, muestra una tendencia descendiente durante el primer semestre, entre febrero y marzo permanece casi estable y cae a su valor mínimo en mayo para subir rápidamente hasta julio a su valor máximo, luego desciende nuevamente hasta casi su valor mínimo en noviembre y se restablece un poco en diciembre. El ∇V_v promedio mensual muestra un patrón similar al de V_v , cae desde su valor máximo en enero hasta febrero y allí permanece estable hasta marzo, desciende a su valor mínimo en mayo y sube hasta julio en donde permanece nuevamente estable en agosto, vuelve a descender hasta noviembre y se restablece un poco en diciembre. Los comportamientos de las variables explicativas parecen tener una relación con el comportamiento de la concentración del contaminante. Con el propósito de describirla, se analizará el cambio intermensual de las variables explicativas T , ∇T y ∇V_v , en relación con el cambio mensual de la concentración del contaminante. Para ello se usará la siguiente convención:

- \uparrow : la variable aumenta entre ambos meses.
- \downarrow : la variable disminuye entre ambos meses.
- $\uparrow_{\text{Máximo}}$: la variable tiene el máximo aumento entre ambos meses.
- $\uparrow_{\text{Mínimo}}$: la variable tiene el mínimo aumento entre ambos meses.
- $\downarrow_{\text{Máximo}}$: la variable tiene la máxima disminución entre ambos meses.
- $\downarrow_{\text{Mínimo}}$: la variable tiene la mínima disminución entre ambos meses.
- \leftrightarrow : la variable no cambia entre ambos meses.

Los resultados de las consideraciones anteriores se sintetizan en la Tabla 56.

Tabla 56. Comportamiento intermensual de las variables T , ∇T y ∇V_v .

	En-Fb	Fb-Mz	Mz-Ab	Ab-My	My-Jn	Jn-Jl
∇V_v	$\downarrow_{\text{Máximo}}$	$\downarrow_{\text{Mínimo}}$	\downarrow	\downarrow	\uparrow	$\uparrow_{\text{Máximo}}$
∇T	\downarrow	\uparrow	$\uparrow_{\text{Máximo}}$	\downarrow	\uparrow	$\downarrow_{\text{Máximo}}$
T	\leftrightarrow	$\uparrow_{\text{Mínimo}}$	\uparrow	\downarrow	$\uparrow_{\text{Máximo}}$	\uparrow
	Jl-Ag	Ag-Sp	Sp-Oc	Oc-Nv	Nv-Dc	
∇V_v	\leftrightarrow	\downarrow	\downarrow	\leftrightarrow	$\uparrow_{\text{Mínimo}}$	
∇T	\uparrow	\downarrow	\downarrow	$\downarrow_{\text{Mínimo}}$	$\uparrow_{\text{Mínimo}}$	
T	$\downarrow_{\text{Máximo}}$	\uparrow	\downarrow	$\downarrow_{\text{Mínimo}}$	\uparrow	

Con base en los resultados de la Tabla 56, se encuentran los siguientes patrones:

- En el período más contaminado entre febrero y marzo, se da la mínima disminución de ∇Vv , el mínimo aumento de T y casi el mínimo aumento de ∇T , es decir, hay un cambio inverso de ∇Vv con T y con ∇T . Esto viene antecedido por la máxima disminución de ∇Vv y ningún cambio en T entre los meses de enero y febrero. Lo anterior genera condiciones favorables para que en el mes de marzo se den los más altos valores en la concentración de PM2.5 en la ciudad. Esto no ocurre, por ahora, en los otros meses del año.
- En el período menos contaminado entre junio y julio, se da el máximo aumento de ∇Vv y la máxima disminución de ∇T , es decir, cambian inversamente. Esto está antecedido por el máximo aumento de T y un aumento de ∇Vv entre los meses de mayo y junio. Lo anterior genera condiciones favorables en los meses de junio y julio para que se den los más bajos valores en la concentración de PM2.5 en la ciudad. Esto lamentablemente no ocurre nuevamente en los otros meses del año.
- Después del período menos contaminado, la contaminación empieza a aumentar hacia su valor promedio anual, entre septiembre y octubre. En este período el ∇Vv y el ∇T disminuyen, es decir, los gradientes cambian directamente. Esto está antecedido por la máxima disminución de T y ningún cambio de ∇Vv entre los meses de julio y agosto. Lo anterior genera condiciones favorables en los meses de septiembre y octubre para que se den valores en la concentración de PM2.5 cercanos al valor promedio anual. Esto no ocurre nuevamente, por ahora, en los otros meses del año, pero queda la duda de si en el futuro los niveles de contaminación subirán por encima del valor promedio anual.

Lo anterior plantea una situación que debe monitorearse a partir de la fecha, pues la realidad actual junto con los resultados de las estimaciones, indican que en marzo los valores de la concentración de PM2.5 están por encima de $31 \mu\text{g}/\text{m}^3$ en la ciudad de Medellín. Como ejemplo, las fotos de la Figura 27 a y b, tomadas en marzo 08 de 2016 a las 8:35 horas y en marzo 10 de 2017 a las 17:30 horas desde el sector centro oriental de la ciudad, muestran al fondo una nube de contaminación que se viene presentando desde el año 2015 en el mismo mes.



Figura 27. Foto panorámica de la ciudad de Medellín en marzo de 2013(a) y 2017(b).

Lo anterior es preocupante ya que el valor máximo promedio anual para la concentración de PM_{2.5} establecido en Colombia, según la normativa de la resolución 610 de 2010 del Ministerio de Ambiente, Vivienda y Desarrollo Territorial, es de 25 $\mu\text{g}/\text{m}^3$. Así las cosas, según el valor promedio mensual de PM_{2.5} (curva amarilla de la Figura 21), la ciudad está contaminada de PM_{2.5} desde el inicio del año, alcanzando su valor más crítico en marzo y disminuyendo en mayo a valores permisibles para continuar con la descontaminación en junio-julio y empezar nuevamente a subir hasta valores permisibles en septiembre para aumentar finalmente hasta el valor promedio anual medido de 27,1 $\mu\text{g}/\text{m}^3$ en diciembre.

Usando los resultados de la estimación espacio temporal de los datos de concentración de PM_{2.5} y la herramienta *Spatial Analyst*® del aplicativo *ArcMap*® de ESRI, se calculan y se muestran los mapas de caracterización espacio-temporal de PM_{2.5} en el período más contaminado junto con las variables explicativas asociadas del mes anterior, en el menos contaminado y en el mes de valor tendiente al promedio anual.

La caracterización espacio-temporal de los datos estimados en marzo (Figura 28a) muestra un patrón similar al dato observado en el mismo mes (Figura 28b). Comparando las Figuras 28 a y b y la Figura 28c, se concluye que la caracterización espacio-temporal en un mes es similar al comportamiento que se da en el mes anterior. Esto se puede explicar por la memoria de corto plazo dada por el valor del rezago $k = 1$ calculado en el algoritmo econométrico. Como se indicó en la Tabla 55, las zonas más contaminadas de la ciudad están en la vecindad de los sitios de monitoreo con ID 0 y 4 como se muestra en las Figuras

28 a-b-c. El sitio con ID 0 corresponde a la parte céntrica de la ciudad de Medellín en donde se da una alta densidad de flujo vehicular. Esta zona en el mes anterior muestra altos valores de temperatura (Figura 29a) y bajos valores en los gradientes de velocidad del viento (Figura 29b) y de temperatura (Figura 29c) lo cual favorece la alta contaminación. Los patrones de la temperatura (Figura 29a) y de ∇Vv (Figura 29b), muestran que la temperatura tiene una tendencia del sur-occidente hacia el nor-orienté y el ∇Vv una tendencia occidente-orienté con una variación de los valores mínimos en la parte central. Esto puede favorecer que la más alta contaminación tienda hacia el nor-orienté. Esta tendencia también se ve en el mapa estimado de PM_{2.5} mediante GWR que en el mes de febrero se calculó bajo el efecto de la temperatura. En el eje definido por los sitios con ID 1 e ID 7 se observa una disminución en la contaminación, ya que entre esos sitios hay cambios en los valores de la temperatura y en el valor de ∇Vv . En general en marzo los valores de la concentración de PM_{2.5} están por encima de 31 $\mu\text{g}/\text{m}^3$ en toda la ciudad.

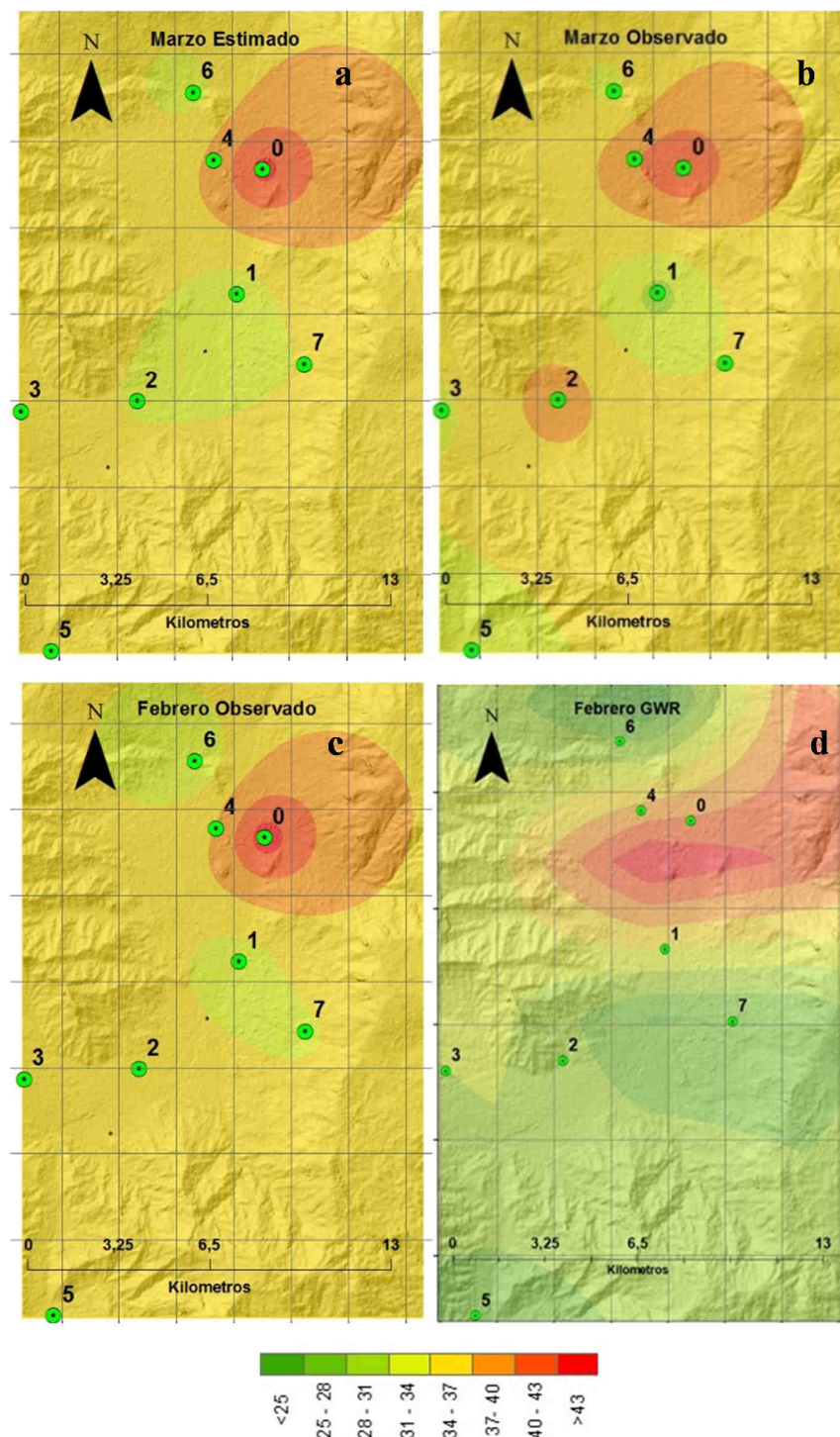


Figura 28. Mapas de caracterización espacial de Pm_{2.5} período más contaminado.

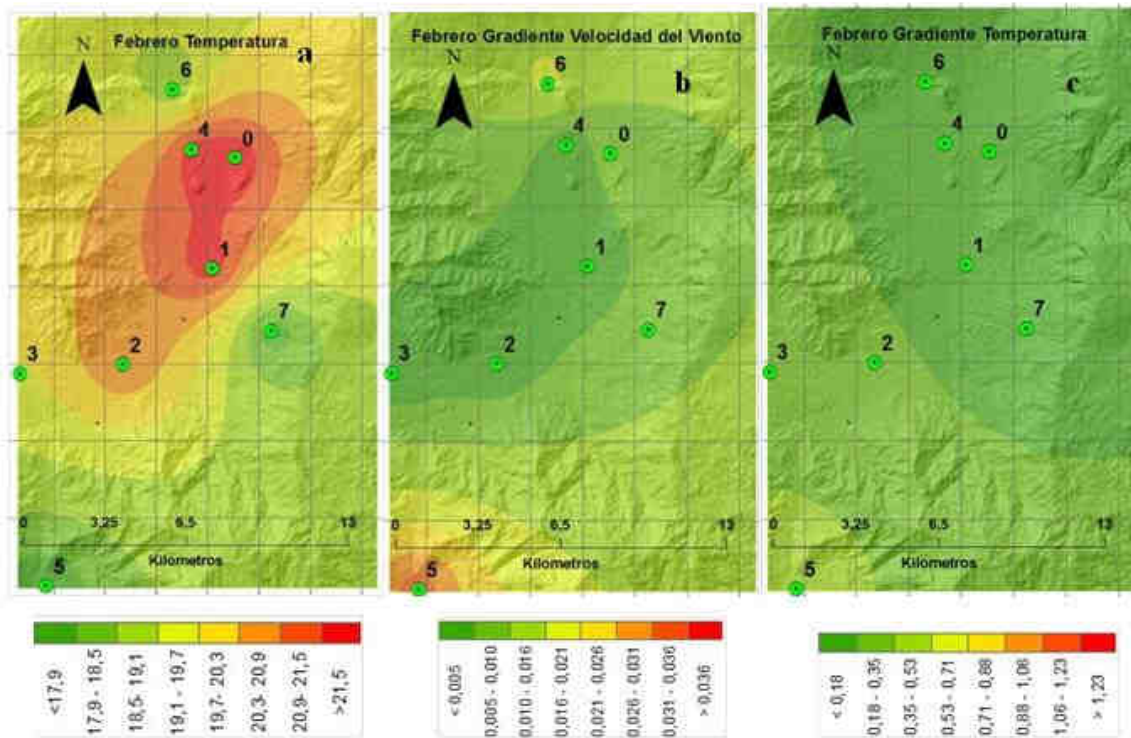


Figura 29. Mapas de caracterización espacial de las variables explicativas de PM2.5 período más contaminado.

Al igual que para el caso de marzo, en junio la caracterización espacio-temporal de los datos estimados con el algoritmo econométrico (Figura 30a) muestra un patrón similar al dato observado en el mismo mes (Figura 30b). Comparando las Figuras 30 a y b y la Figura 30 c, se concluye que la concentración de PM2.5 en junio es menor a la que se observa en mayo, pero con un patrón similar que es, junto con julio, el mes de menor contaminación en la ciudad. Según la Tabla 55, las zonas menos contaminadas de la ciudad están en la vecindad de los sitios de monitoreo con ID 1 y 7 pero como se muestra en las Figura 30 a-b-c en general se muestran valores bajos de la concentración en toda la ciudad. En el centro de la ciudad (ID = 0) también se da una disminución importante de la contaminación ya que este período intersemestral coincide con la época de vacaciones escolares y se da una disminución en la densidad de flujo vehicular. Los patrones de la temperatura (Figura 31a) y de ∇Vv de mayo (Figura 31b), muestran que la temperatura tiene una tendencia del sur-occidente hacia el nor-occidente con valores más altos que en el mes de marzo y el ∇Vv una tendencia centro-occidente hacia el nor-occidente. El ∇T (Figura 31c) muestra una tendencia contraria a la del ∇Vv desde centro-occidente hacia nor-occidente, lo que favorece la baja

contaminación. El mapa estimado de PM_{2.5} mediante GWR en mayo (Figura 30d) se calculó con el efecto de VT. En general en junio los valores de la concentración de PM_{2.5} están por debajo de 25 $\mu\text{g}/\text{m}^3$ en toda la ciudad.

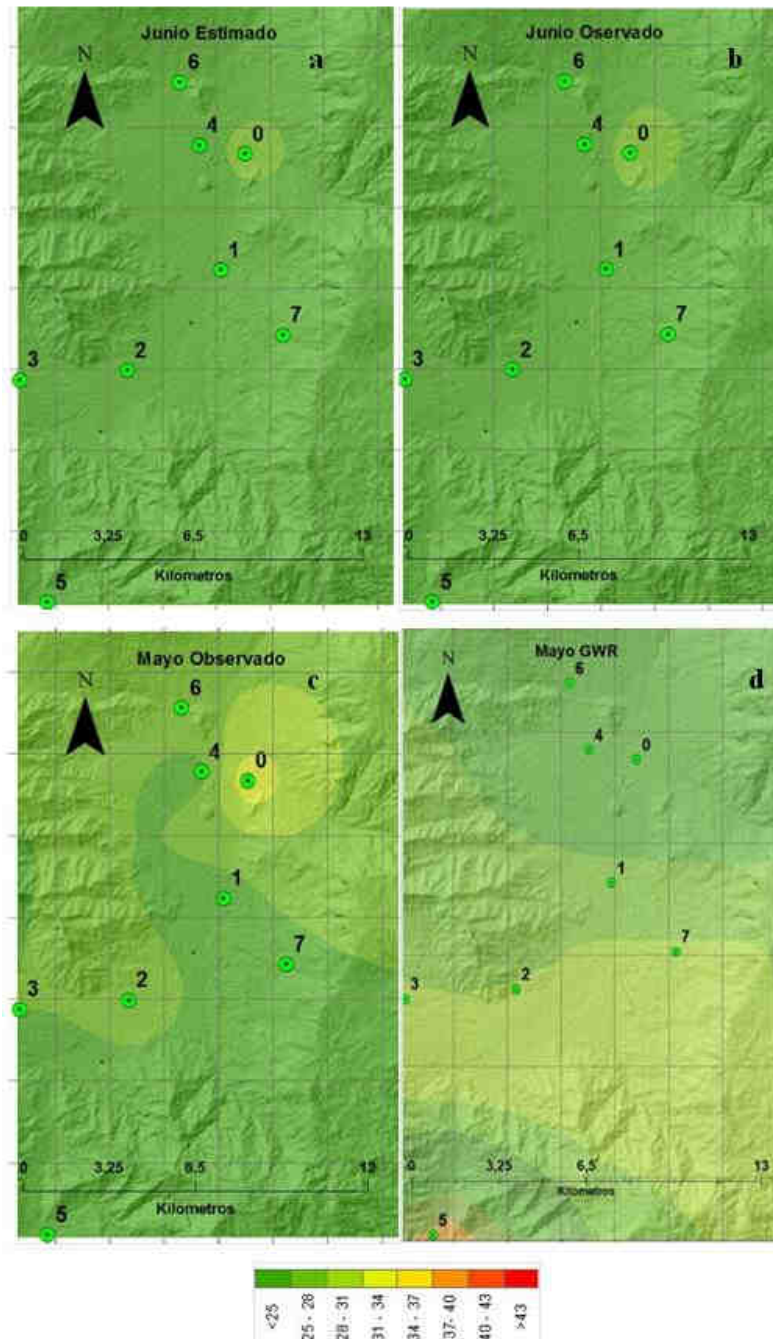


Figura 30. Mapas de caracterización espacial de PM_{2.5} período menos contaminado.

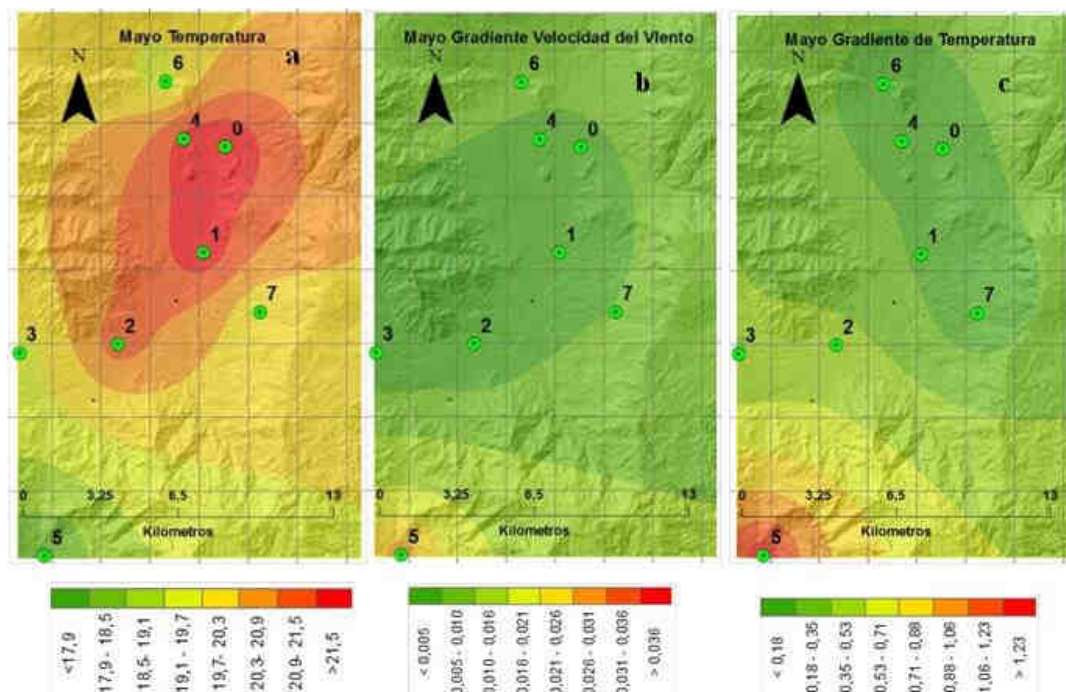


Figura 31. Mapas de caracterización espacial de las variables explicativas de PM2.5 período menos contaminado.

Al igual que en los casos anteriores, en octubre la caracterización espacio-temporal de los datos estimados con el algoritmo econométrico (Figura 32a) muestra un patrón similar al dato observado en el mismo mes (Figura 32b). Comparando las Figuras 32 a-b-c, se concluye que la concentración de PM2.5 en septiembre es menor a la que se observa en octubre (pero con un patrón similar) y en octubre el valor de la concentración tiende a aumentar al valor promedio anual. Como se muestra en la Figura 32a en la vecindad de los sitios con ID 0 y 4 se inicia el aumento en los valores de la concentración. Los patrones de temperatura (Figura 31a) y de ∇Vv de septiembre (Figura 31b), muestran que la temperatura vuelve a tomar una tendencia del nor-orienté hacia el sur-occidente con valores más altos generalizados en gran parte de la ciudad y el ∇Vv una tendencia centro-occidente hacia nor-orienté. El ∇T (Figura 33c) es casi el mismo lo cual puede favorecer el aumento en la contaminación. El mapa estimado de PM2.5 mediante GWR en septiembre (Figura 32d) se calculó con el ∇Vv . En general en octubre los valores de la concentración de PM2.5 están por encima de $28 \mu\text{g}/\text{m}^3$ en toda la ciudad.

Aplicando el MPE del capítulo 7, en la Figura 34, se muestran los datos estimados de PM2.5 mediante el algoritmo econométrico, corregidos por la proximidad a fuentes en el mes más contaminado (Figura 34a), el menos contaminado (Figura 34b) y el de contaminación promedio (Figura 34c).

En la Figura 35 (a – f) se muestran para el primer semestre los datos estimados de PM2.5 ajustados por la proximidad a las fuentes en primer semestre y en la Figura 36 (g a l) en el segundo semestre.

Al agregar las variables de proximidad a las fuentes de emisión y de mitigación, se obtienen detalles de la caracterización espacio-temporal de la concentración de PM2.5 asociados a la cercanía a las vías principales, a los usos industriales y a las zonas verdes; que se muestran en las Figura 35 y 36 y que no se percibían en las Figuras 28a para marzo, 30a para junio y 32a para octubre. De los resultados del MPE mostrados en las Figura 35 y 36 se pueden hacer los siguientes análisis:

- En las intersecciones de las vías principales y las rotondas se dan condiciones críticas para el aumento de la concentración del contaminante por el alto flujo vehicular sobre todo en horas pico.
- En las vías que atraviesan las industrias hay un aumento en la concentración del contaminante debido al doble efecto de la emisión de la vía y de la emisión de las industrias.
- En las vías que atraviesan zonas verdes, hay un efecto doble de aumento y atenuación simultánea de la concentración del contaminante.
- En las vías que atraviesan usos industriales y zonas verdes, hay un efecto triple que aumenta la concentración del contaminante por la proximidad a la vía y a las industrias, y simultáneamente la concentración se atenúa por proximidad a las zonas verdes.
- En el centro de los polígonos de usos industriales, hay un aumento en la concentración del contaminante por los deshechos de la producción industrial.

- En el centro de los polígonos de las zonas verdes hay una atenuación en la concentración del contaminante por absorción del contaminante en las plantas.
- En aproximadamente el centro de la intersección de los polígonos de usos industriales y de zonas verdes, hay un efecto doble de aumento y atenuación simultánea en la concentración del contaminante.

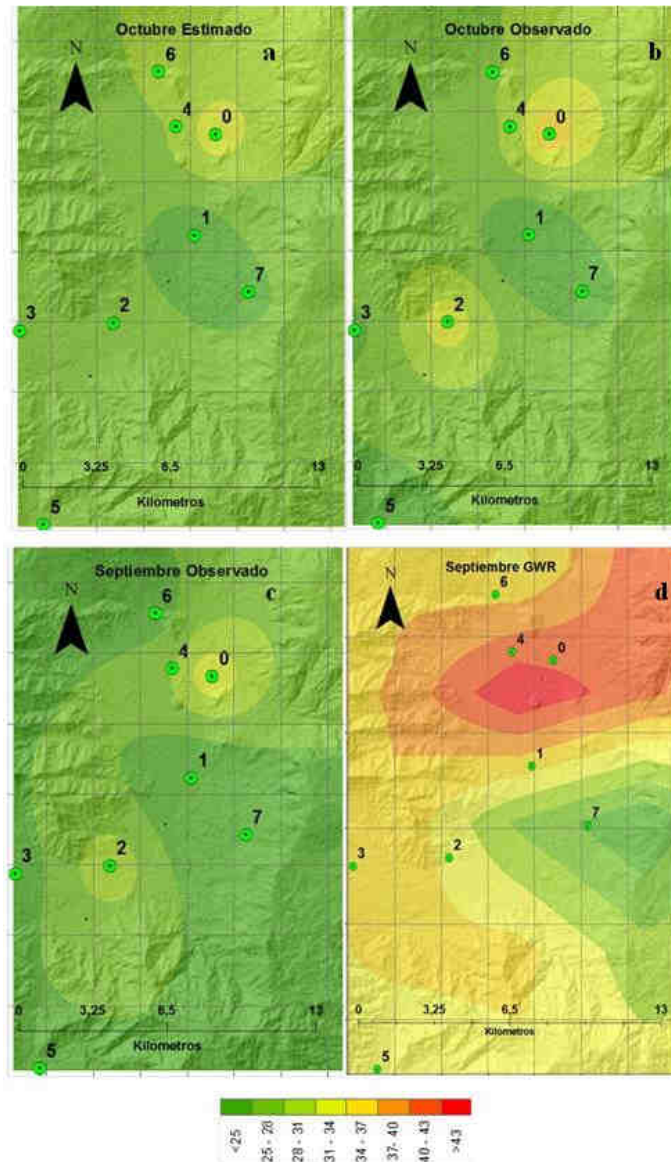


Figura 32. Mapas de caracterización espacial de PM_{2.5} período contaminación de promedio.

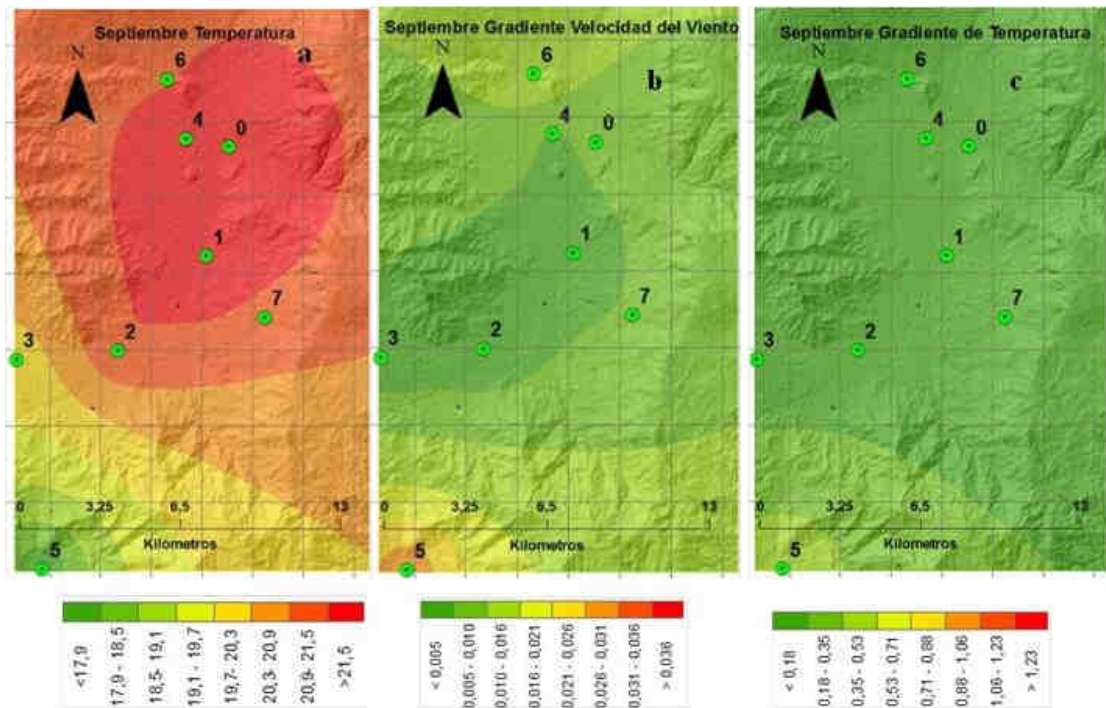


Figura 33. Mapas de caracterización espacial de las variables explicativas de PM_{2.5} período contaminación promedio.

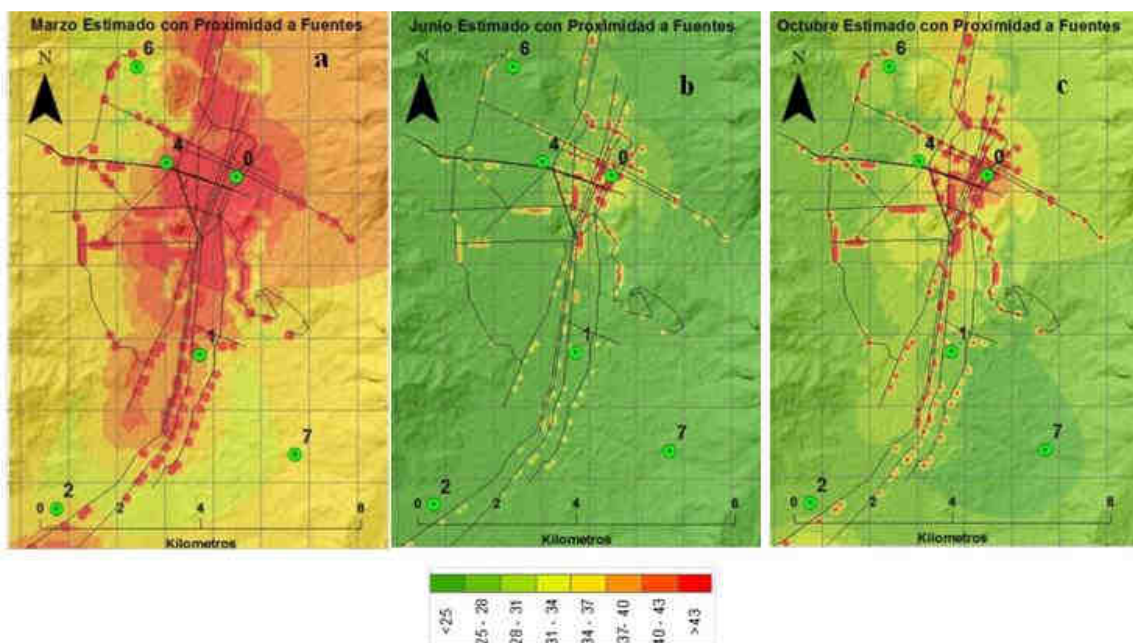


Figura 34. Mapas de caracterización espacial de PM_{2.5} con ajuste de proximidad a fuentes.

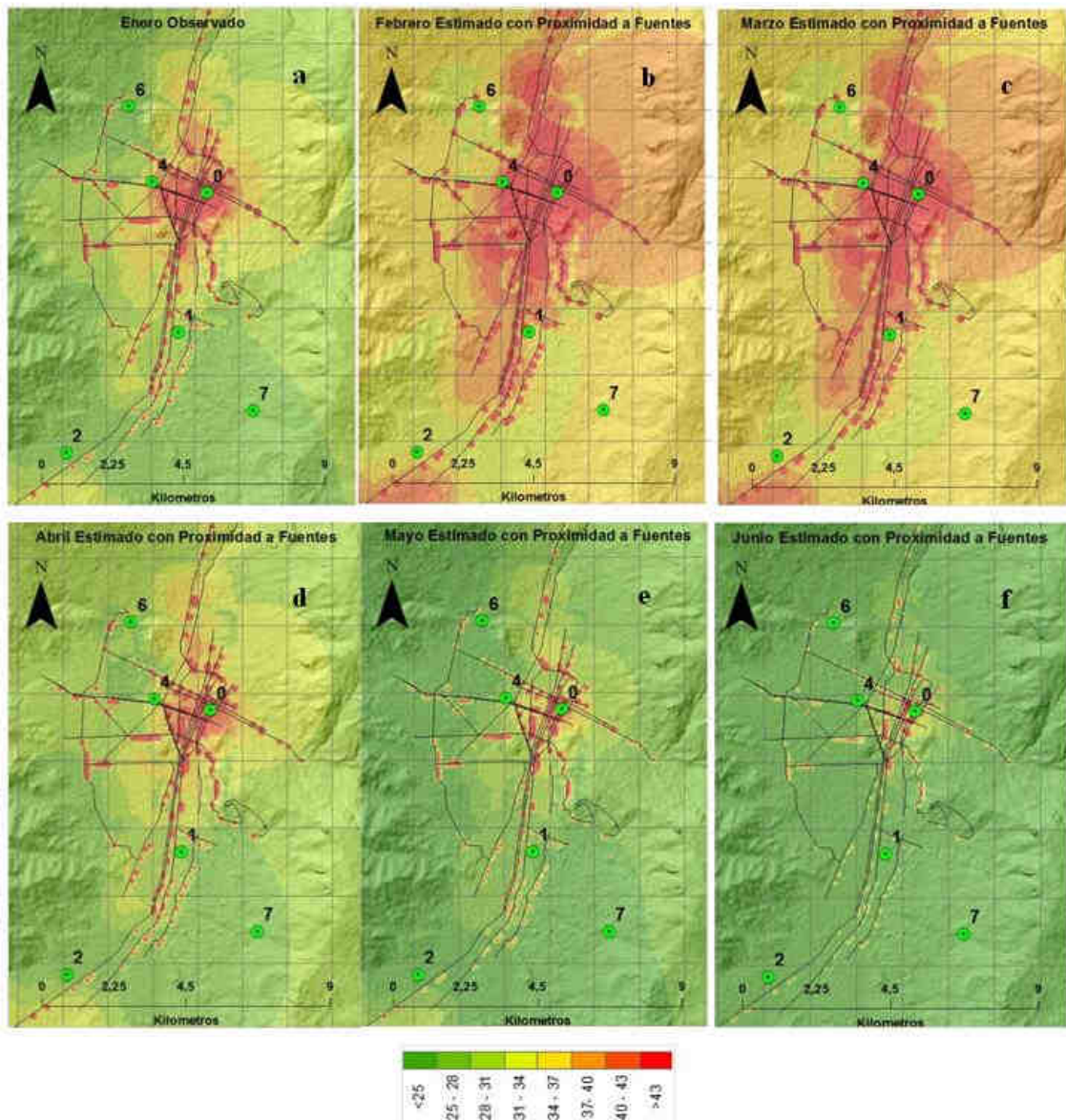


Figura 35. Mapas mensuales de caracterización espacial de PM2.5.

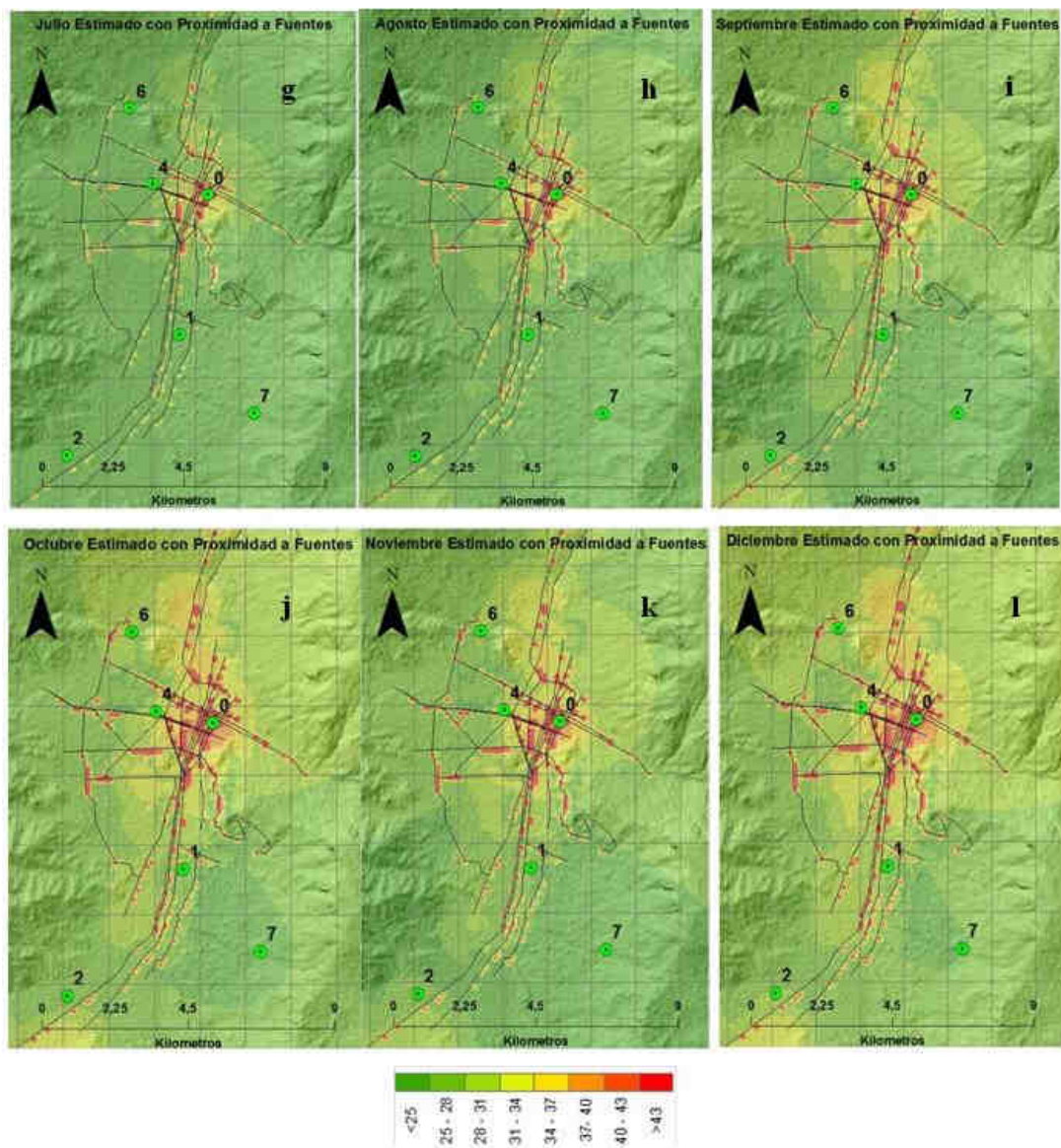


Figura 36. Mapas mensuales de caracterización espacial de PM2.5.

En las figuras 37, 38 y 39 se muestran, para cada cuatrimestre del año, los mapas de error estándar (SE) de los datos estimados con el algoritmo AR – RD con GWR. En general se muestran zonas de confianza de la estimación concéntricas a los sitios de medición de PM2.5 a un radio vector aproximado de 800 m, en donde el valor de SE es bajo y está entre 0,3 y 1,6 ($\mu\text{g}/\text{m}^3$)². En esta zona es donde el valor estimado de la concentración de PM2.5 tiene menos incertidumbre.

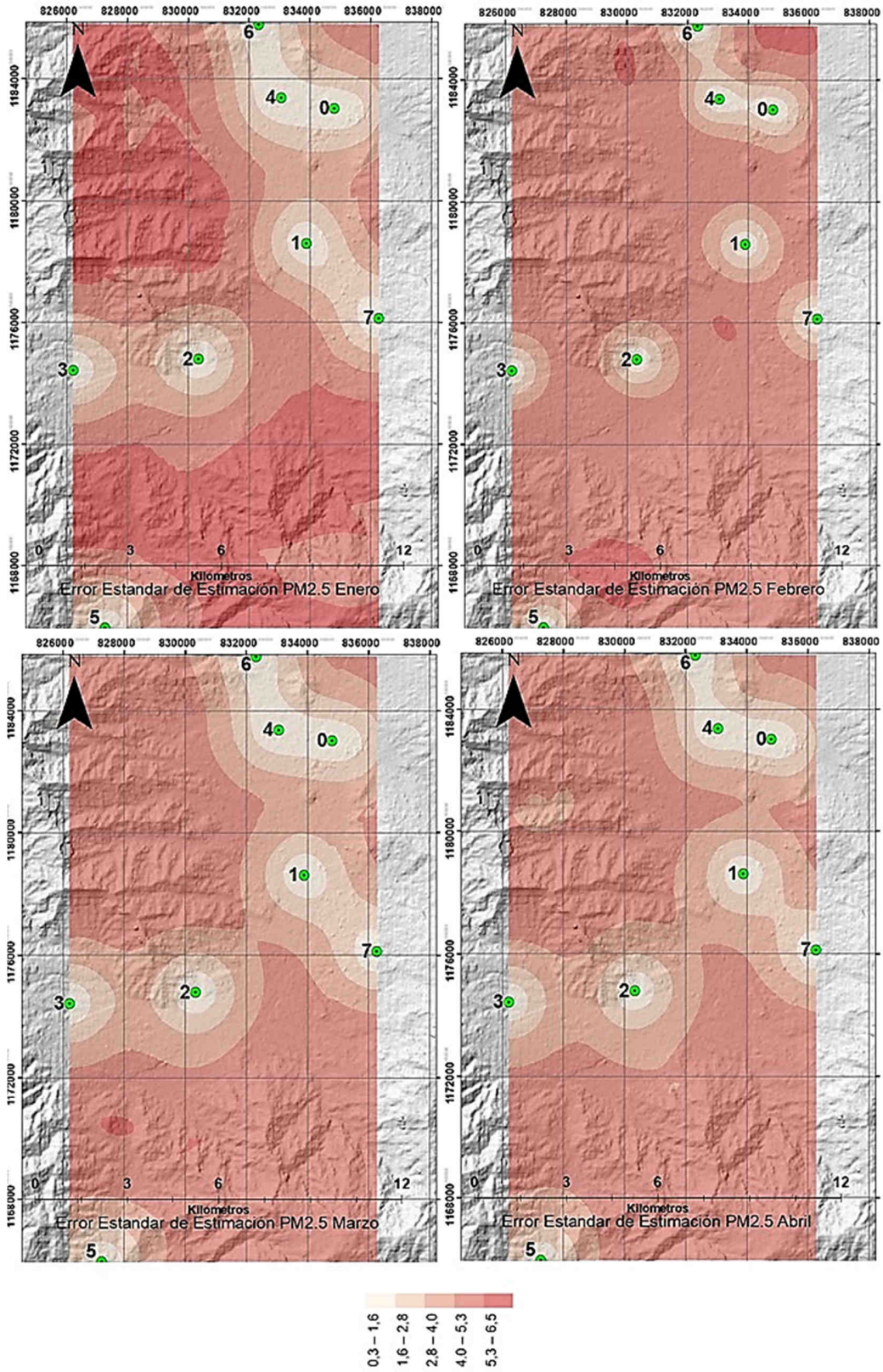


Figura 37. Error estándar de estimación de PM2.5 primer cuatrimestre del año.

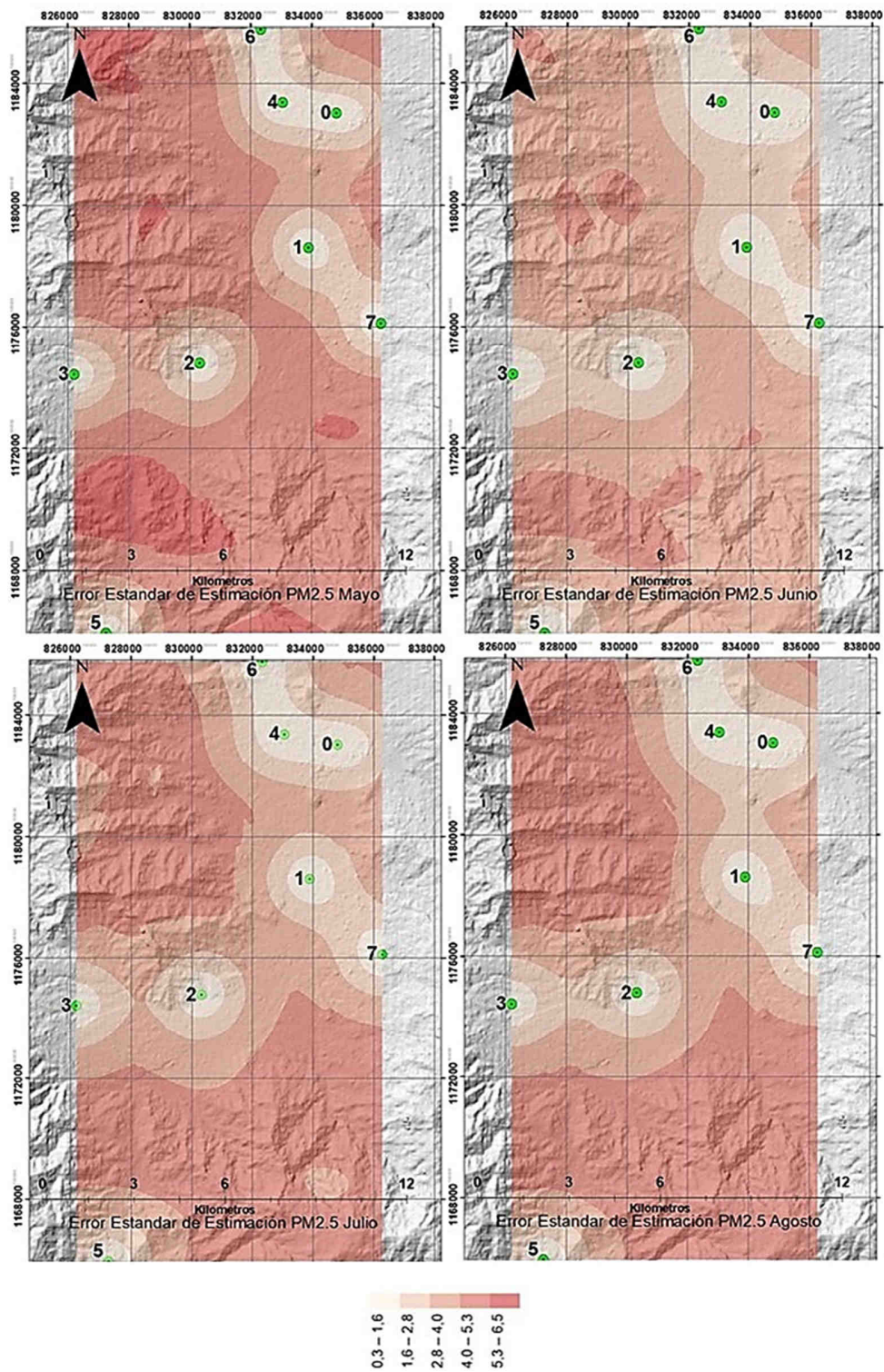


Figura 38. Error estándar de estimación de PM2.5 segundo cuatrimestre del año.

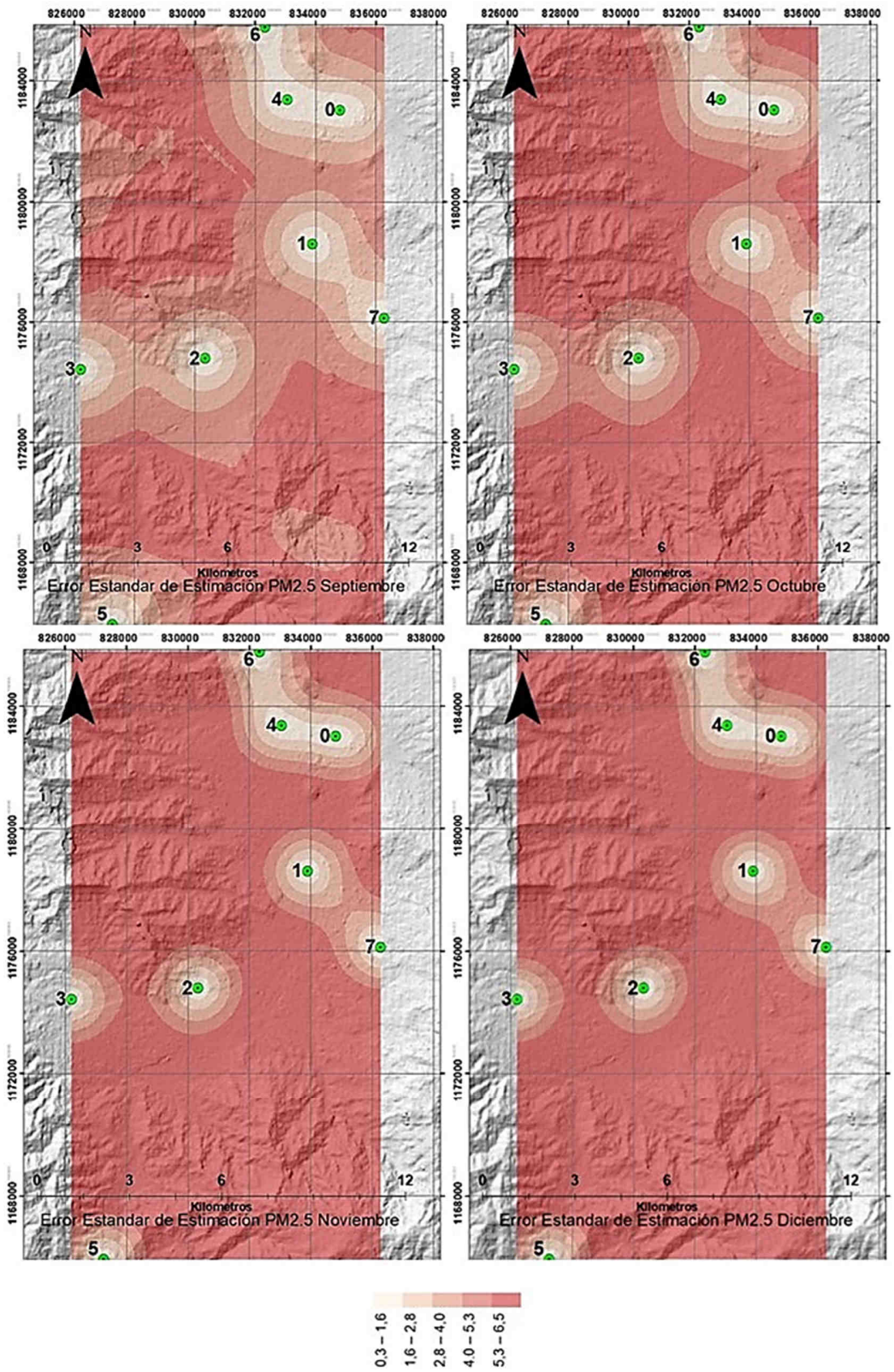


Figura 39. Error estándar de estimación de PM2.5 tercer cuatrimestre del año.

Con base en las zonas de confiabilidad de los mapas y los mapas de caracterización espacio-temporal de la concentración del contaminante en la ciudad se puede proponer la configuración de una red de sitios de monitoreo basada en los criterios de análisis espacial presentados durante el desarrollo de esta tesis.

En la Figura 40, se muestra y se propone la configuración de una red de monitoreo de concentración de material particulado para la ciudad de Medellín.

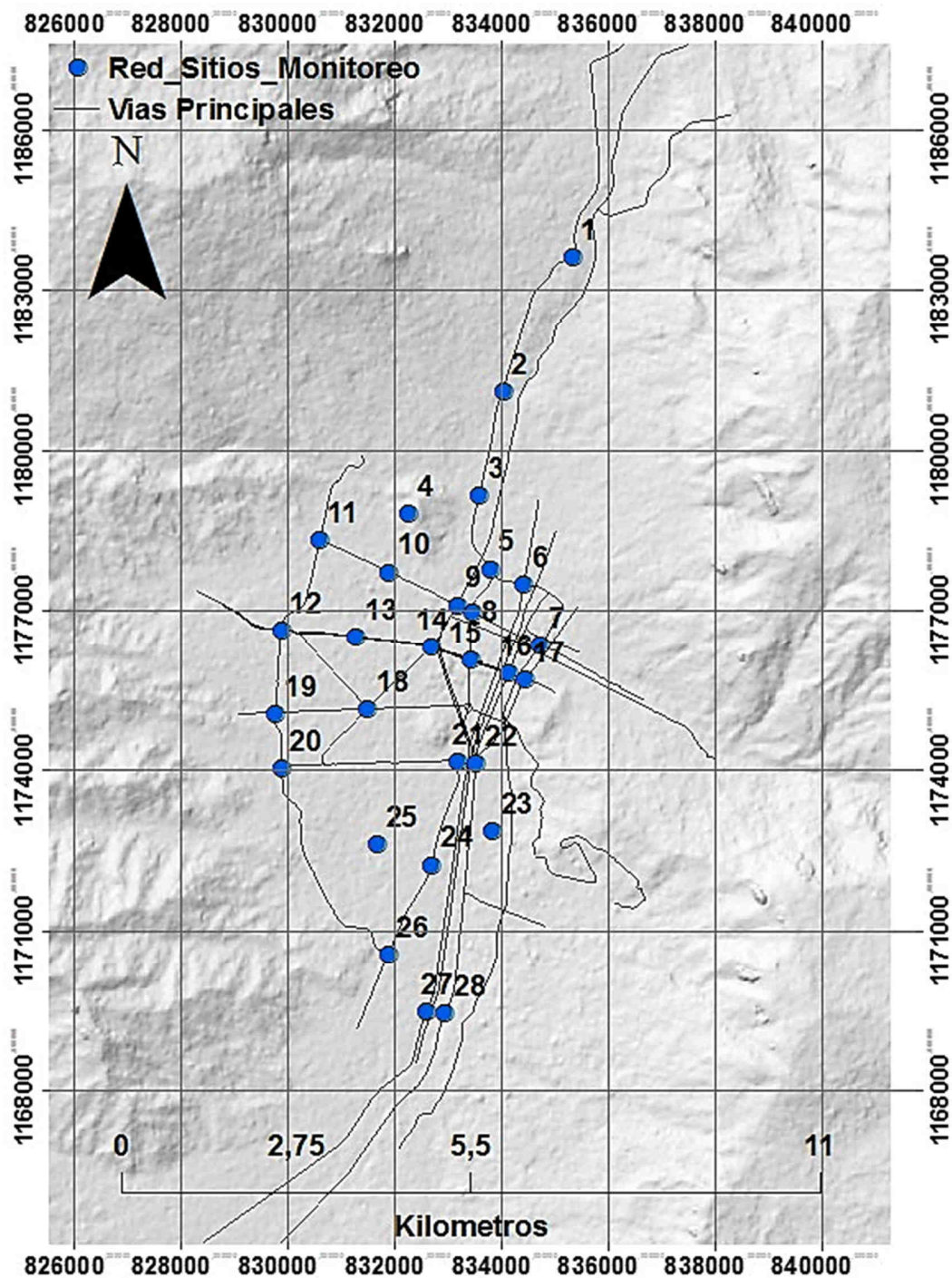


Figura 40. Propuesta de red de monitoreo de concentración de material particulado en Medellín (Londoño et al., 2017)

Como se muestra en la Figura 41, la red de sitios de monitoreo propuesta se localiza, casi en su totalidad, en la zona de confiabilidad para minimizar la incertidumbre de los valores que se puedan medir en esos sitios.

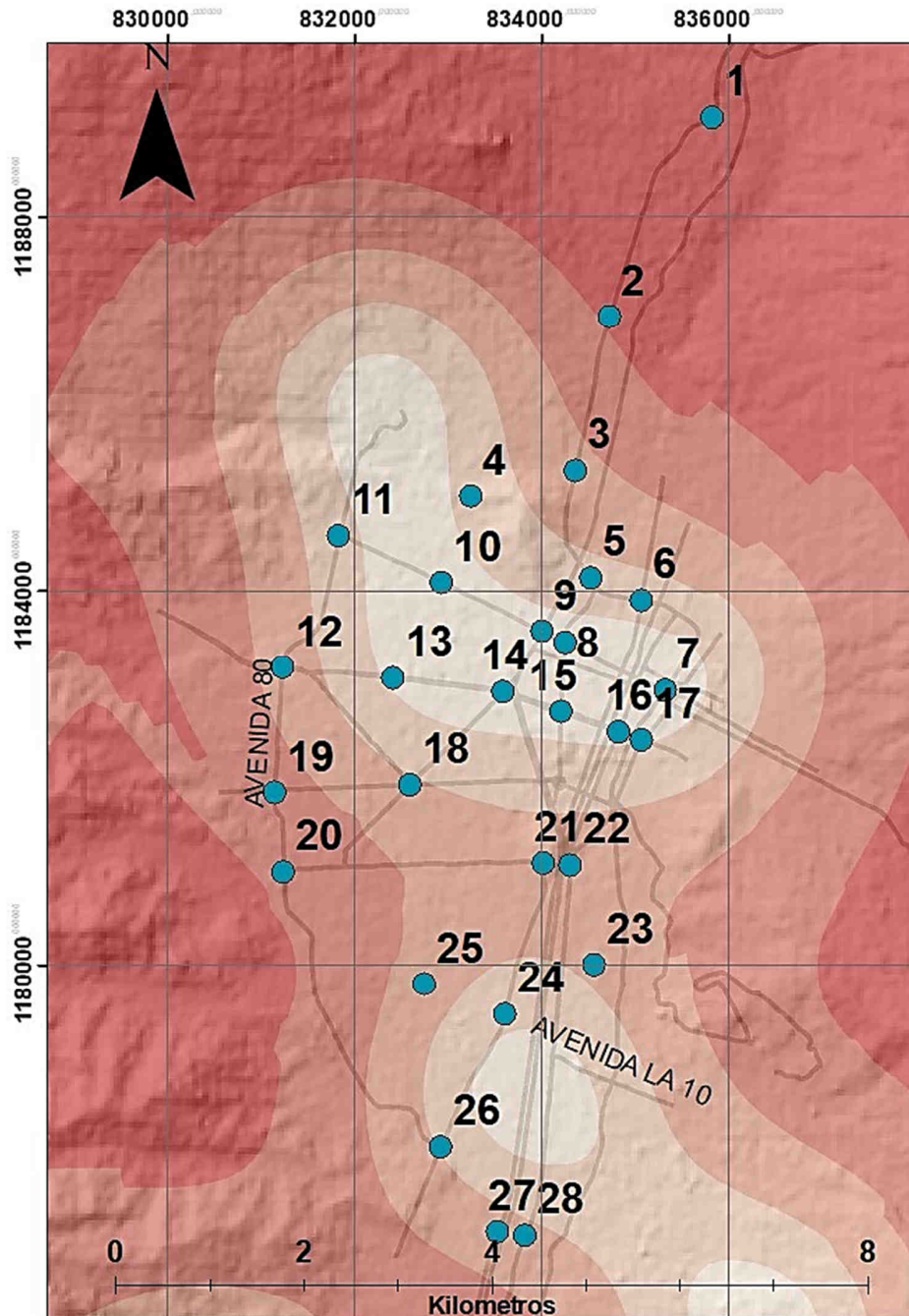


Figura 41. Confiabilidad de la Red de monitoreo de concentración de material particulado propuesta para Medellín.

En la Tabla 57 se hace una descripción de la tipología de los sitios de monitoreo propuestos en la red de la Figura 40.

Tabla 57. Tipología de los sitios de monitoreo de la Red de la Figura 40.

Sitio de monitoreo	X	Y	Nombre vía	Tipología del Sitio
1	835821,4600	1189056,5159	Autopista Norte con Calle 111	Punto Crítico - Efecto Industrial -Efecto zona verde
2	834727,8849	1186933,0582	Autopista Norte con puente Madre Laura	Punto Crítico
3	834363,3883	1185282,4638	Rompoy Punto Cero	Punto Crítico - Efecto Industrial
4	833245,6555	1185008,4703	Cerro Volador	Fondo - Efecto zona verde
5	834528,9797	1184135,1562	Rompoy Plaza Minorista	Punto Crítico - Fondo
6	835066,2914	1183894,6824	Av. Oriental con Carabobo	Punto Crítico - Fondo
7	835323,7063	1182929,3051	Av. Oriental con Colombia	Punto Crítico
8	834253,8655	1183444,3812	Av. Colombia con Av. Ferrocarril	Punto Crítico
9	834006,5425	1183564,7015	Av. Colombia con Autopista Norte	Punto Crítico
10	832932,4610	1184084,5628	Av. Colombia con Carrera 70	Punto Crítico
11	831832,5504	1184585,3842	Rompoy de la 80 con Colombia	Punto Crítico
12	831234,0130	1183174,7061	Rompoy de la 80 con San Juan	Punto Crítico
13	832410,9776	1183060,7730	Av. San Juan con Carrera 70	Punto Crítico – Efecto zona verde
14	833589,5533	1182912,8689	Av. San Juan con Av. Bolivariana	Punto Crítico
15	834209,6940	1182700,7333	Av. San Juan con Av. Ferrocarril	Punto Crítico - Efecto Industrial
16	834824,2007	1182481,7239	Av. San Juan con Av. Palacé	Punto Crítico
17	835067,7625	1182396,6045	Av. San Juan con El Palo	Punto Crítico - Efecto Industrial
18	832595,3343	1181918,8572	Rompoy Av. 33 con Av. Bolivariana	Punto Crítico
19	831148,9472	1181837,4766	Rompoy Av. 33 con Av. 80	Punto Crítico
20	831241,6636	1180988,9103	Rompoy Av. 30 con Av. 80	Punto Crítico
21	834021,0609	1181085,5504	Av. 30 con Av. Guayabal	Punto Crítico - Efecto Industrial -Efecto zona verde
22	834305,5055	1181062,6192	Av. 30 con Av. Industriales	Punto Crítico - Efecto Industrial
23	834562,4839	1179999,3490	Ciudad de rio	Efecto Industrial
24	833611,1277	1179468,0478	Rompoy Av. Guayabal con Calle 10	Punto Crítico - Efecto Industrial
25	832748,5981	1179793,4145	Aeropuerto Olaya Herrera	Fondo - Efecto Industrial
26	832921,9827	1178053,1046	Rompoy Av. Guayabal con Av. 80	Punto Crítico - Efecto Industrial
27	833531,5713	1177153,1965	Autopista sur con la Aguacatala	Punto Crítico - Efecto Industrial
28	833820,5123	1177114,6665	Av. Las Vegas con la Aguacatala	Punto Crítico

Con el modelo AR – RD GWR se estiman, para la red propuesta, los valores de la concertación de PM2.5 en cada uno de los sitios de monitoreo (ID) y en cada mes y se calcula el intervalo de confianza para cada dato estimado. En la Tabla 58 se muestran los resultados del primer semestre y en la Tabla 59 los del segundo semestre.

Tabla 58. Intervalos de confianza de los valores estimados en los sitios de monitoreo de la Red de la Figura 40 para el primer semestre.

ID	En	Linf	Lsup	Fb	Linf	Lsup	Mz	Linf	Lsup	Ab	Linf	Lsup	My	Linf	Lsup	Jn	Linf	Lsup
1	50,7	45,5	55,8	68,3	65,2	71,3	67,3	61,4	73,2	54,6	49,8	59,4	47,3	43,2	51,5	40,4	36,8	44,0
2	50,0	44,9	55,2	69,3	66,2	74,4	60,8	54,8	66,7	49,6	44,8	54,4	43,1	38,9	47,2	37,0	33,4	40,6
3	37,3	32,2	42,5	70,3	67,2	75,4	45,7	39,7	51,6	37,3	32,5	42,1	32,3	28,1	36,4	27,5	23,9	31,1
4	30,7	25,5	35,8	71,3	68,2	76,4	39,3	33,4	45,3	32,1	27,3	36,9	27,5	23,3	31,7	23,2	19,6	26,8
5	69,6	64,5	74,7	72,3	69,2	77,4	71,8	65,8	77,7	58,6	53,8	63,4	50,8	46,6	55,0	43,5	39,9	47,1
6	70,9	65,8	76,1	73,3	70,2	78,4	83,5	77,6	89,4	68,5	63,7	73,3	59,6	55,4	63,7	51,3	47,7	54,9
7	42,8	37,6	47,9	74,3	71,2	79,4	46,5	40,5	52,4	38,1	33,3	42,9	33,2	29,1	37,4	28,7	25,1	32,3
8	57,3	52,1	62,4	75,3	72,2	80,4	75,1	69,1	81,0	61,1	56,3	65,9	52,9	48,8	57,1	45,3	41,7	48,9
9	65,6	60,5	70,8	76,3	73,2	81,4	80,9	75,0	86,8	65,3	60,5	70,1	56,5	52,3	60,6	48,1	44,5	51,7
10	54,7	49,5	59,8	77,3	74,2	82,4	68,1	62,2	74,0	54,4	49,6	59,2	46,6	42,4	50,8	39,1	35,5	42,7
11	55,6	50,5	60,7	78,3	75,2	83,4	34,9	28,9	40,8	28,5	23,7	33,3	24,4	20,2	28,6	20,4	16,8	24,0
12	50,4	45,2	55,5	79,3	76,2	84,4	62,4	56,5	68,4	50,3	45,5	55,1	43,4	39,2	47,5	36,2	32,6	39,8
13	59,6	54,5	64,8	80,3	77,2	85,4	71,8	65,8	77,7	56,9	52,1	61,7	48,9	44,7	53,1	40,9	37,2	44,5
14	73,5	68,4	78,7	81,3	78,2	86,4	72,2	66,3	78,1	57,4	52,6	62,2	49,5	45,3	53,7	41,8	38,2	45,4
15	64,0	58,8	69,1	82,3	79,2	87,4	88,9	83,0	94,9	72,0	67,2	76,8	62,4	58,3	66,6	53,3	49,7	56,9
16	80,8	75,6	85,9	83,3	80,2	88,4	48,7	42,7	54,6	39,8	35,0	44,6	34,7	30,5	38,9	29,9	26,3	33,5
17	72,5	67,4	77,7	84,3	81,2	89,4	79,7	73,7	85,6	65,1	60,3	69,9	56,8	52,6	60,9	48,9	45,3	52,5
18	55,3	50,2	60,4	85,3	82,2	90,4	59,7	53,8	65,6	47,3	42,5	52,1	40,8	36,6	45,0	34,0	30,4	37,6
19	57,2	52,1	62,4	86,3	83,2	91,4	35,3	29,4	41,3	28,3	23,5	33,1	24,5	20,3	28,6	20,4	16,8	24,0
20	57,1	51,9	62,2	87,3	84,2	92,4	70,5	64,5	76,4	56,2	51,4	61,0	48,7	44,5	52,9	40,5	36,9	44,1
21	65,3	60,2	70,5	88,3	85,2	93,4	77,5	71,6	83,4	61,5	56,7	66,3	53,5	49,4	57,7	44,7	41,1	48,3
22	63,5	58,4	68,6	89,3	86,2	94,4	73,6	67,6	79,5	58,6	53,8	63,4	51,0	46,9	55,2	42,8	39,2	46,4
23	35,5	30,4	40,7	90,3	87,2	95,4	43,0	37,0	48,9	33,8	29,0	38,6	29,6	25,4	33,7	24,5	20,9	28,1
24	33,6	28,5	38,8	91,3	88,2	96,4	41,0	35,1	46,9	31,6	26,8	36,4	27,6	23,4	31,8	22,1	18,5	25,7
25	33,1	28,0	38,3	92,3	89,2	97,4	40,5	34,6	46,4	31,7	26,9	36,5	27,6	23,4	31,8	22,6	19,0	26,2
26	58,2	53,1	63,4	93,3	90,2	98,4	74,3	68,3	80,2	57,7	52,9	62,5	50,3	46,2	54,5	40,9	37,3	44,5
27	58,3	53,2	63,5	94,3	91,2	99,4	63,4	57,4	69,3	49,3	44,5	54,1	43,1	39,0	47,3	35,4	31,8	39,0
28	31,6	26,4	36,7	95,3	92,2	100,4	65,8	59,8	71,7	51,1	46,3	55,9	44,8	40,7	49,0	36,9	33,3	40,5

Tabla 59. Intervalos de confianza de los valores estimados en los sitios de monitoreo de la Red de la Figura 40 para el segundo semestre.

ID	Jl	Linf	Lsup	Ag	Linf	Lsup	Sp	Linf	Lsup	Oc	Linf	Lsup	Nv	Linf	Lsup	Dc	Linf	Lsup
1	41,8	38,0	45,5	44,9	40,9	48,9	48,1	43,9	52,3	51,3	46,8	55,7	49,2	45,0	53,5	52,4	47,9	57,0
2	38,2	34,5	41,9	40,9	36,9	44,9	43,3	39,1	47,5	46,1	41,6	50,5	44,3	40,0	48,6	47,1	42,5	51,6
3	28,5	24,7	32,2	30,7	26,7	34,7	33,3	29,1	37,5	35,5	31,1	39,9	34,0	29,7	38,3	36,3	31,7	40,8
4	24,2	20,5	27,9	26,4	22,4	30,3	29,2	25,0	33,4	31,2	26,8	35,7	29,9	25,6	34,2	32,0	27,4	36,5
5	45,0	41,3	48,7	48,3	44,3	52,3	51,6	47,4	55,8	54,9	50,5	59,4	52,8	48,5	57,1	56,1	51,6	60,7
6	52,9	49,2	56,6	56,5	52,6	60,5	59,9	55,7	64,1	63,6	59,2	68,1	61,2	56,9	65,4	65,0	60,4	69,5
7	29,5	25,8	33,2	31,5	27,5	35,4	33,1	28,9	37,3	35,1	30,7	39,5	33,8	29,5	38,0	35,9	31,3	40,4
8	46,9	43,2	50,6	50,3	46,3	54,3	53,6	49,4	57,7	57,0	52,6	61,5	54,8	50,6	59,1	58,3	53,8	62,9
9	49,9	46,1	53,6	53,7	49,7	57,7	57,7	53,5	61,9	61,6	57,1	66,0	59,1	54,8	63,4	63,0	58,4	67,5
10	40,9	37,1	44,6	44,6	40,6	48,5	49,3	45,1	53,4	52,8	48,4	57,3	50,6	46,3	54,9	54,1	49,6	58,7
11	21,3	17,6	25,1	23,3	19,4	27,3	26,2	22,0	30,4	28,1	23,6	32,5	26,8	22,5	31,1	28,7	24,1	33,2
12	37,6	33,9	41,3	41,0	37,1	45,0	45,5	41,3	49,6	48,7	44,2	53,1	46,4	42,1	50,7	49,7	45,1	54,2
13	42,6	38,8	46,3	46,4	42,5	50,4	51,2	47,0	55,4	54,9	50,5	59,4	52,5	48,3	56,8	56,2	51,7	60,8
14	43,4	39,7	47,1	46,9	43,0	50,9	50,8	46,6	54,9	54,3	49,9	58,8	52,1	47,9	56,4	55,7	51,1	60,2
15	55,1	51,3	58,8	59,1	55,1	63,1	62,9	58,7	67,0	67,0	62,6	71,5	64,4	60,2	68,7	68,6	64,0	73,1
16	30,8	27,1	34,5	32,8	28,9	36,8	34,5	30,3	38,7	36,6	32,2	41,1	35,3	31,0	39,5	37,5	32,9	42,0
17	50,3	46,6	54,0	53,7	49,7	57,6	56,4	52,2	60,6	59,9	55,5	64,3	57,6	53,4	61,9	61,3	56,7	65,8
18	35,2	31,5	39,0	38,4	34,4	42,4	42,0	37,8	46,2	45,1	40,6	49,5	43,1	38,8	47,4	46,2	41,6	50,7
19	21,1	17,4	24,8	23,0	19,0	27,0	25,4	21,2	29,5	27,1	22,7	31,6	25,9	21,6	30,1	27,7	23,1	32,2
20	41,8	38,1	45,5	45,6	41,7	49,6	50,2	46,0	54,4	53,7	49,3	58,2	51,2	46,9	55,4	54,8	50,3	59,4
21	46,0	42,3	49,8	49,9	45,9	53,8	53,4	49,3	57,6	57,3	52,9	61,8	55,0	50,8	59,3	58,9	54,4	63,5
22	44,0	40,3	47,7	47,6	43,6	51,5	50,8	46,6	55,0	54,4	50,0	58,9	52,3	48,0	56,5	55,9	51,3	60,4
23	25,0	21,3	28,7	27,2	23,2	31,2	29,0	24,9	33,2	31,2	26,8	35,7	30,0	25,7	34,3	32,2	27,7	36,8
24	22,7	18,9	26,4	25,0	21,0	29,0	27,0	22,8	31,2	29,4	25,0	33,8	28,3	24,0	32,6	30,5	25,9	35,0
25	23,2	19,5	27,0	25,5	21,5	29,4	27,7	23,5	31,9	29,9	25,4	34,3	28,6	24,3	32,9	30,7	26,2	35,3
26	41,8	38,1	45,5	46,0	42,0	49,9	50,0	45,8	54,1	54,0	49,6	58,5	51,7	47,4	56,0	55,7	51,1	60,2
27	36,0	32,2	39,7	39,4	35,4	43,3	42,5	38,3	46,7	45,8	41,3	50,2	43,8	39,5	48,0	47,2	42,7	51,7
28	37,4	33,7	41,1	40,9	36,9	44,8	43,9	39,8	48,1	47,3	42,9	51,8	45,3	41,0	49,5	48,8	44,3	53,4

Con la configuración de sitios de monitoreo propuesta en la Figura 40, la tipología de los mismos de la Tabla 57 y el valor estimado de concertación junto con el intervalo de confianza (Tablas 58 y 59) se podrían relocalizar y complementar los sitios de monitoreo existentes, hacer mediciones de control para compararlas con los valores estimados y correlacionar las mediciones con los efectos sobre la salud debidos a la exposición directa y de largo plazo de PM2.5 en la ciudad de Medellín.

9 CONCLUSIONES

En la presente tesis se propone una metodología que permite hacer una la caracterización espacial (univariada - multivariada) y espacio-temporal para la concentración del material particulado en centros urbanos intramontanos. La metodología integra el enfoque estocástico, de interpolación espacial y el econométrico. Para su desarrollo usa el análisis geoestadístico, el análisis espacial y los SIG. La caracterización espacial permite visualizar, a través de los mapas resultantes, cómo se comporta la concentración del material particulado en el área Metropolitana de la ciudad de Medellín. Este insumo es importante para tener una idea global de las zonas de mayor o menor contaminación en la ciudad con base en las mediciones hechas en algunos sitios de monitoreo.

Para el caso de PM10, con la caracterización espacial univariada se encuentra que el modelo *J-Bessel* es el que tiene mejor desempeño y para PM2.5 el *Hole Effect*. Sin embargo, el análisis de frecuencias de estos modelos no mostró diferencias significativas entre ambos, por lo que sería posible usar cualquiera de ellos para hacer este tipo de caracterización. En la caracterización espacial multivariada, el algoritmo GWR mostró buen desempeño en la estimación de la concentración. Se encontró que las variables explicativas más importantes fueron la velocidad del viento y la temperatura, pero también fueron muy importantes los gradientes espaciales de estas variables. Este hallazgo es un aporte de la tesis toda vez que en la literatura no se reporta el uso de ellas. En las Tablas 50 y 51, para cada mes se muestran las variables explicativas más relevantes.

La metodología propuesta tiene tres componentes fundamentales: el espacial univariado, el espacial multivariado y el espacio-temporal (ver Figura 42). Cada componente posee una serie de procedimientos que se esquematizan en los diagramas de flujo de las Figuras 13, 14, 15 y 16; complementados con los diagramas de flujo de las Figuras 2, 3 y 19.

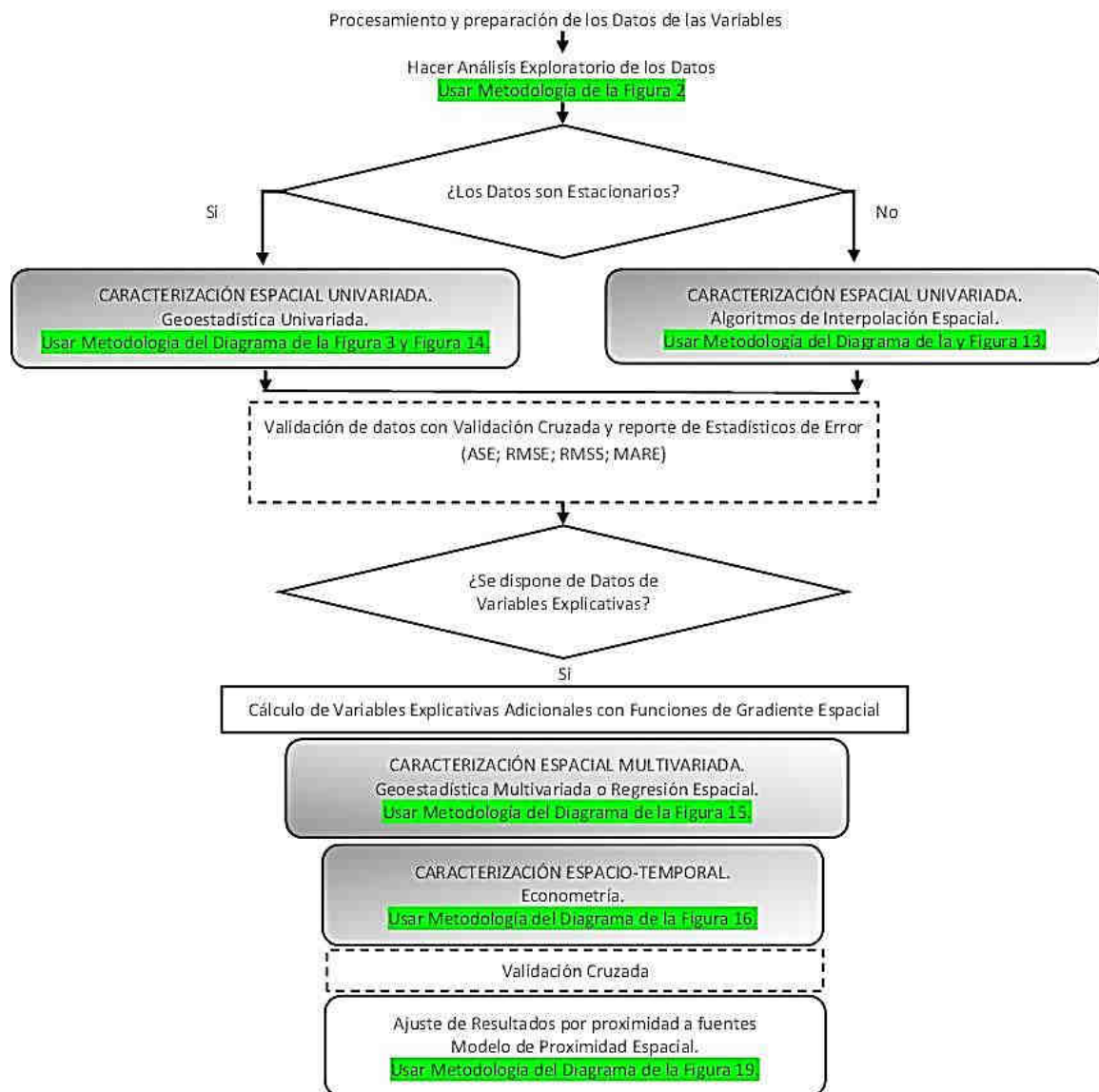


Figura 42. Metodología Final. Fuente propia.

El uso de un periodo de análisis mensual, permitió determinar la estacionalidad de la concentración del material particulado con tres episodios destacados: el de mayor contaminación que se da entre los meses de febrero y marzo, el de menor contaminación entre junio y julio y el de contaminación promedio que se da entre los meses de septiembre y octubre. Este patrón se encontró a partir de estimaciones hechas con datos de 2013-2014, pero recientemente en los últimos tres años (2015 a 2017) se ha venido repitiendo el episodio de contaminación máxima en el mes de marzo a valores de muy alta contaminación en la ciudad con los efectos inmediatos sobre la salud de sus habitantes. Igualmente, las

mediciones recientes han mostrado que junio es el mes menos contaminado. Lo anterior demuestra que los resultados obtenidos con la metodología planteada en esta tesis son válidos a pesar de que la densidad de muestreo de los datos no fue favorable.

En la caracterización espacio – temporal, se exploraron tres modelos econométricos básicos: AR, RD y AR – RD. Adicionalmente, se probaron otros tres modelos como el propuesto en la Figura 8, que fueron los modelos básicos más un componente espacial de rezago distribuido cuyo dominio espacial son los datos calculados con la metodología de caracterización espacial multivariada usando GWR con rezago 1 (RD, AR y AR-RD con GWR como variable explicativa). De estos modelos el de mejor desempeño fue un AR-RD con GWR como variable explicativa que logro explicar la variabilidad espacio-temporal de la concentración del material particulado en un 88% para PM10 (Tabla 42) y un 87% para PM2.5 (Tabla 47). Para ambos contaminantes las variables explicativas más significativas tuvieron un rezago de 1 y fueron: el dominio espacial calculado con GWR, el gradiente de la velocidad del viento, la temperatura y la concentración del material particulado. Sin embargo, al comparar los resultados de estos modelos (Tablas 42 y 47) con el modelo AR, la variable más importante fue el rezago 1 de la concentración del material particulado (en ambos casos las demás variables solo aportaron un 1% de más).

En las Figuras 17 (para PM10) y 18 (para PM2.5) se muestra la estacionalidad de la concentración del material particulado en cada sitio de monitoreo. Para PM10 en cada uno de los 13 sitios de monitoreo (Tabla 13) se presentan los tres episodios de contaminación, excepto en el sitio con ID = 0 (Universidad de San Buenaventura – Bello) que muestra dos episodios de máxima contaminación en marzo y julio y uno de mínima en junio. Lo mismo se muestra para PM2.5 en cada uno de los 8 sitios de monitoreo (Tabla 7), en donde se observa la estacionalidad del episodio de máxima y el de mínima (en los sitios con ID 0, 2, 3, 5, 6, y 7). Los tres episodios se ven en los sitios con ID 1 y 4.

Otro hallazgo de la implementación de la metodología fue que, durante el episodio de mayor contaminación, los gradientes espaciales cambian de manera inversa y en su mínima expresión, es decir, mientras que ∇V_v tiene su mínima disminución, ∇T tiene su mínimo

aumento. En otras palabras, en marzo no hay cambios en ∇Vv ni en ∇T , y como estas variables afectan fuertemente la dispersión del material particulado, entonces se favorece el aumento significativo de la contaminación. Otro aspecto importante que se encontró fue que, entre los dos meses anteriores al episodio más contaminado, se dio la máxima disminución en ∇Vv y no hubo cambios significativos en la temperatura. Esto no se encontró en los cambios intermensuales de estas variables en los demás meses del año.

Otra regularidad se da en el período menos contaminado, en donde los gradientes nuevamente cambian de manera inversa, pero en sus valores máximos, se da el máximo aumento en ∇Vv y la máxima disminución en ∇T . En otras palabras, en junio se dan los máximos cambios en ∇Vv y en ∇T lo que favorece la dispersión del contaminante. Adicionalmente en los dos meses anteriores al episodio de menor contaminación se dio el máximo aumento de T y un aumento de ∇Vv . Estas singularidades tampoco se encontraron en los cambios intermensuales de estas variables en los demás meses del año.

En el segundo semestre del año la ciudad empieza gradualmente a contaminarse hasta que en octubre se da otro episodio destacado y es que se alcanza un valor de concentración similar al del valor promedio anual. En este mes, a diferencia de los casos anteriores, los gradientes cambian directamente (∇Vv y ∇T disminuyen simultáneamente). En otras palabras, en octubre los gradientes empiezan nuevamente a bajar. En los dos meses anteriores se da el máximo aumento de T y un aumento de ∇Vv . Al igual que en los casos anteriores, tampoco se encontraron estos cambios intermensuales de las variables en los demás meses del año.

Por lo anterior sería muy importante a futuro hacer un seguimiento del comportamiento de las variables ∇Vv y ∇T , vigilando si el patrón detectado se sigue repitiendo antes de los episodios destacados. Si así fuese, se tendrían criterios importantes de decisión para la implementación oportuna de medidas preventivas y efectivas que apunten al mejoramiento de la calidad del aire de la ciudad y que además deberían aplicarse de manera distinta en cada mes del año debido a la estacionalidad de los episodios destacados detectados en esta tesis. Otro asunto importante tiene que ver con la medición. Se sugiere que simultáneamente a la medición de la concentración del material particulado en algún sitio de monitoreo, se

tomen también medidas de temperatura y velocidad del viento para mejorar las estimaciones ya que actualmente en general no se miden estas variables de manera simultánea en los mismos sitios de monitoreo.

Un asunto para reflexionar es que, en la ciudad de Medellín la mayor parte del año, el valor de la concentración de PM_{2.5} está por encima del valor máximo promedio anual permisible establecido en Colombia por la normativa de la Resolución 2254 de 2017 del Ministerio de Ambiente, Vivienda y Desarrollo Territorial, que es de 25 µg/m³. Es imperativo aplicar medidas urgentes para resolver esta situación, que ya se ha vuelto repetitiva en los últimos tres años, ya que, de no ser así, existe la tendencia a futuro de que en octubre ocurran episodios de alta contaminación como los de marzo que muy probablemente afectarían el episodio de menor contaminación de junio.

Otro aporte importante de la metodología es la propuesta del modelo de proximidad espacial (MPE), que permite hacer ajustes a las estimaciones de la caracterización espacial de la concentración del material particulado al incluir la variable de proximidad a fuentes de emisión y de mitigación. Esto hace que la caracterización espacial tenga menos incertidumbre y se aproxime más a la realidad del fenómeno. Adicionalmente con el MPE se pudieron definir unos criterios espaciales para localizar sitios de monitoreo y su tipificación. Estos criterios consideran la cercanía a las fuentes de emisión y de mitigación como se indica en la Figura 43.

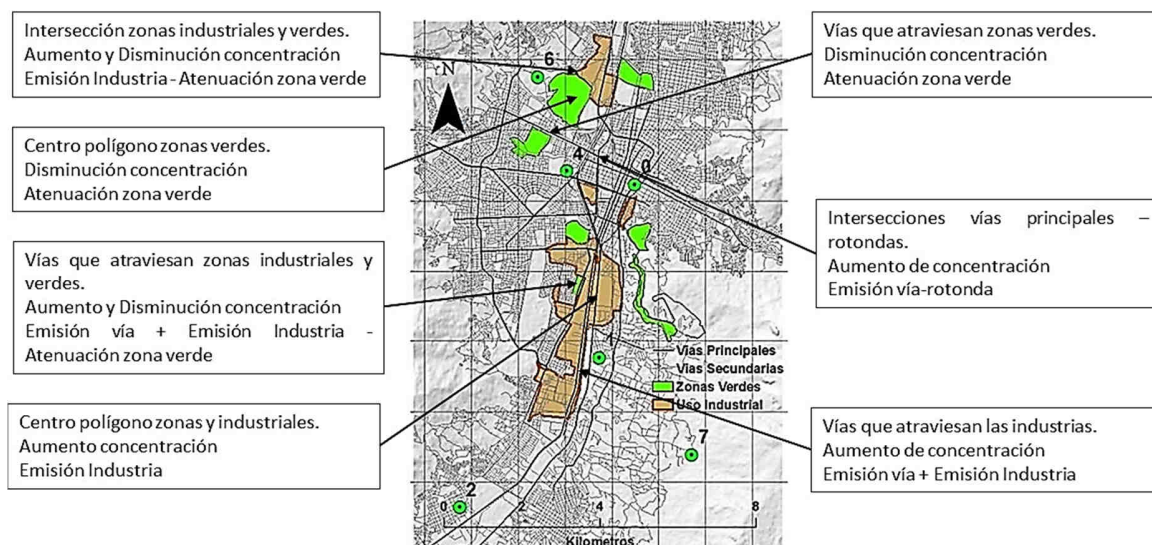


Figura 43. Criterios espaciales para la ubicación de la red de sitios de monitoreo propuesta. Fuente propia.

Adicionalmente a los criterios espaciales para la ubicación de los sitios de monitoreo de la red propuesta, se tuvo en cuenta que éstos estuvieran en la región de menor incertidumbre de estimación como se mostró en la Figura 41. Esta región de confianza es concéntrica a los sitios de medida de PM_{2.5} en la ciudad y tiene en general un radio vector aproximado de 800m, pues es allí en donde el valor del error estándar de estimación es bajo y está entre 0,3 y 1,6 ($\mu\text{g}/\text{m}^3$)².

La red que se propone en esta tesis, puede servir como insumo para la configuración de una red de vigilancia de la concentración de PM_{2.5} en la ciudad de Medellín que permita tomar mediciones y, con base en la metodología de esta tesis, hacer una gestión epidemiológica del territorio orientada a la minimización del riesgo por exposición directa y de largo plazo a este tipo de contaminantes.

Finalmente, se destaca que los resultados de esta metodología se pueden obtener de manera rápida y económica, en relación con algoritmos de enfoque determinístico que requieren de los datos de muchas variables para su implementación.

REFERENCIAS

- Alam, M., & McNabola, A. (2015). Exploring the modeling of spatiotemporal variations in ambient air pollution within the land use regression framework: Estimation of PM10 concentrations on a daily basis. *J. Air Waste Manag. Assoc.*, *65*, 628–640.
- Alt, F. (1942). Distributed Lags. *Econometrica*, *10*, 113-128.
- AMVA. (2007). *Informe final convenio 661 de 2004*. Área Metropolitana del Valle de Aburra. Dirección de comunicaciones., Medellín.
- AMVA. (2008). *Área Metropolitana del Valle de Aburra. Plan de Gestión 2008 - 2011*. Área Metropolitana del Valle de Aburra. Dirección de comunicaciones, Medellín.
- Baklanov, A., Hänninen, O., Slørdal, L., Kukkonen, J., Bjergene, N., Fay, B., y otros. (2007). Integrated systems for forecasting urban meteorology, air pollution and population exposure. *Atmos. Chem. Phys.*, *7*, 855-874.
- Bedoya, J., & Martínez, E. (2009). Calidad del aire en el Valle de Aburrá Antioquia – Colombia. *Revista Dyna*. Vol 76, No 158. pp. 7-15. *Revista Dyna*, *76*(158), 7-15.
- Beelen, R., Voogt, M. M., Duyzer, J., Zandveld, P., & Hoek, G. (2010). 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.
- Bennett, N., Croke, B., Guariso, G., Guillaume, J., Hamilton, S., Jakeman, A., y otros. (2013). Characterising performance of environmental models. *Environ. Model. Softw.*, *40*, 1-20.
- Bogaert, P., Mahau, P., & Beckers, F. (1995). *The Spatial Interpolation of Agroclimatic Data. Cokriging Software and Source Code*. Rome: FAO.
- Borges, R., Colmenares, G., & Paredes, D. (2009). Comparación de técnicas de imputación de datos faltantes con respecto al modelo de regresión logística multinomial. *XXXI Congreso Nacional de Estadística e Investigación Operativa*. Murcia.
- Borrego, C., Monteiro, A., Ferreira, J., Miranda, A., Costa, A., Carvalho, A., y otros. (2008). Procedures for estimation of modelling uncertainty in air quality assessment. *Environ. Int.*, *34*, 613–620.
- Brunelli, U., Piazza, V., Pignato, L., & Sorbello, F. (2007). Two-days ahead prediction of daily maximum concentrations of SO₂, O₃, PM₁₀, NO₂, CO in the urban area of Palermo, Italy. *Atmos. Environ.*, *41*, 2967–2995.
- Burrough, P., & Mcdonnell, R. (1998). *Principles of geographical information systems*. New York: Oxford University Press.
- Chen, L., Bai, Z., Kong, S., Han, B., Usted, Y., Ding, X., y otros. (2010). LUR Model to predict NO₂ and PM₁₀ concentrations at different seasons of the year in the Tianjin region, China. *J. Reinar. Sci.*, *22*, 1364 -1373.
- Chow, J., Watson, J., Mauderly, J., Costa, D., Wyzga, R., Vedal, S., y otros. (2006). Health effects of fine particulate air pollution: Lines that connect. *J. Air Waste Manag. Assoc.*, *56*, 1368–1380.
- Clougherty, J., Kheirbek, I., Eisl, H., Ross, Z., Pezeshki, G., Gorczyński, J., y otros. (2013). Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: The New York City Community Air Survey (NYCCAS). *Journal of Exposure Science and Environmental Epidemiology*, *23*, 232-240.

- Corantioquia. (2005). *Estudio Conceptual, Numérico y Experimental del Modelo RAMS en la Jurisdicción de Corantioquia. Informe de Investigación. Medellín, Colombia.* Corporación Autónoma del Centro de Antioquia. Universidad de Antioquia. Medellín: Universidad de Antioquia.
- Dadvand, P., Rushton, S., Diggle, P., Goffe, L., Rankin, J., & Pless-Mullooli, T. (2011). Using spatio-temporal modeling to predict long-term exposure to black smoke at fine spatial and temporal scale. *Atmospheric Environment*, 45, 659-664.
- Daly, A., & Zannetti, P. (2007). *Air Pollution Modeling – An Overview. Chapter 2 of Ambient Air Pollution.* Zannetti, P, Al-Ajmi, D and Al-Rashied, S, Editors. Published by The Arab School for Science and Technology (ASST).
- Deligiorgi, D., & Philippopoulos, K. (2011). *Spatial Interpolation Methodologies in Urban Air Pollution Modeling: Application for the Greater Area of Metropolitan Athens, Greece, Advanced Air Pollution.* Athens: Dr. Farhad Nejadkoorki.
- Dias de Freitas, E., Droprinchinski, L., Leite, P., & Andrade, M. (2005). A simple photochemical module implemented in RAMS for tropospheric ozone concentration forecast in metropolitan area of Sao Paulo, Brazil: Coupling and validation. *Atmospheric Environment*, 39, 6352-6361.
- Diaz-Robles, L., Ortega, J., Fu, J., Reed, G., Chow, J., Watson, J., y otros. (2008). A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmos. Environ*, 42, 8331-8340.
- Dickey, D., & Fuller, W. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, 74, 427-431.
- Diebold, F. (1971). *Elements of Forecasting, 2a. ed.* South Western.
- Dons, E., Van Poppe, M., Panis, L., De Prins, S., Berghmans, P., Koppen, G., y otros. (2014). 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.
- Dons, E., Van Poppel, M., Kochan, B., Wets, G., & Panis, L. (2013). Modeling temporal and spatial variability of traffic-related air pollution: hourly land use regression models for black carbon. *Atmospheric Environment*, 74, 237-246.
- Eeftens, M., Beekhuizen, J., Beelen, R., Wang, M., Vermeulen, R., Brunekreef, B., y otros. (2013). Quantifying urban street conFiguration for improvements in air pollution models. *Atmospheric Environment*, 72, 1-9.
- EPA. (1994). *Quality Assurance Handbook for Air Pollution Measurement Systems. Volume I: A Field Guide to Environment Quality Assurance. Office of Research and Development. IBSN EPA/600/R-94/038a.* Washington.: EPA.
- EPA. (1998). *Quality Assurance Handbook for Air Pollution Measurement Systems. Volume II: Part I Ambient Air Quality monitoring Program Quality System Development. Office of Air Quality Planning and Standards Research Triangle Park. IBSN EPA-454/R-98-004.* Washington: EPA.
- Fischer, M., & Getis, A. (2010). *Handbook of Applied Spatial Analysis.* Springer.
- Fotheringham, A., Brunson, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Chichester: John Wiley & Sons, Ltd.

- García, J., Morales, A., Gómez, J., Gómez, D., Tamayo, I., & Pardo, P. (2016). Association between residential proximity to environmental pollution sources and childhood renal tumors. *Environmental Research*, *147*, 405-414.
- Gass, S., & Fu, M. (2013). *Saaty, T. Analytic Hierarchy Process Encyclopedia of Operations Research and Management Science*. Springer.
- Giraldo, R. (2003). *Introducción a la Geoestadística: Teoría y aplicación*. Bogota: Departamento de Estadística. Universidad Nacional de Colombia.
- Gómez, M., Saldarriaga, J., Correa, M., Posada, E., & Castrillón, F. (2007). Modelación estadística de los aportes de las vías como fuentes de emisión a las partículas totales en suspensión según modelo MCF, zona centro de Medellín Antioquia Colombia, 2004. *Revista Dyna*, *74*(153), 35-40.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford Univ. Press.
- Goyal, P., Chan, A., & Jaiswal, N. (2006). Statistical models for the prediction of respirable suspended particulate matter in urban cities. *Atmos. Environ*, *40*, 2068–2077.
- Granger, C. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, 424-438.
- Gulliver, J., Morris, C., Lee, K., Vienneau, D., Briggs, D., & Hansell, A. (2011). Land use regression modeling to estimate historic (1962-1991) concentrations of black smoke and sulfur dioxide for Great Britain. *Environmental Science & Technology*, *45*, 3526-3532.
- Hengl, T. (2007). *A Practical Guide to Geostatistical Mapping of Environmental Variables*. Luxembourg: Office for Official Publication of the European Communities.
- Hooyberghs, J., Mensik, C., Dumont, G., Fierens, F., & Brasseur, O. (2005). A neural network forecast for daily average PM10 concentrations in Belgium. *Atmospheric Environment*, *39*, 3279 - 3289.
- Jerrett, M., Potoglou, D., Arain, A., Kanaroglou, P., Sahuvaroglu, T., Beckerman, B., y otros. (2005). A review and evaluation of intraurban air pollution exposure models. *Journal of Exposure Analysis and Environmental Epidemiology*, *15*, 185–204.
- Jiménez, J. (2016). *Altura de la Capa de Mezcla en un area urbana, montañosa y tropical Caso de estudio: Valle de Aburrá (Colombia)*. Revisado el 2017, de file:///C:/Users/POSGRADOS/Downloads/JimenezJose_2016_AlturaCapaMezcla.pdf
- Jiménez, J., & Mejía, O. (2002). Un modelo estacionario de circulación atmosférica diurna en el Valle de Aburrá para época de verano. Medellín, Colombia.
- Johnston, K., Ver Hoef, J., Krivoruchko, K., & Lucas, N. (2003). *Using ArcGis. Geostatistical Analyst*. ESRI.
- Kousa, A., Kukkonen, J., Karppinen, A., Aarnio, P., & Koskentalo, T. (2001). Statistical and diagnostic evaluation of a new generation urban dispersion modeling system against an extensive dataset in the Helsinki area. *Atmospheric Environment*, *35*, 4617–4628.
- Lam, N. (1983). Spatial interpolation methods: a review. *The American Cartographer*, *10*(2), 129-149.
- Laumbach, R., & Kipen, H. (2012). Respiratory health effects of air pollution: update on biomass smoke and traffic pollution. *J Allergy Clin Immunol*, *129*, 3–11.
- Lee, M., Brauer, M., Wong, P., Tang, R., Tsui, T., Choi, C., y otros. (2017). Land use regression modelling of air pollution in high density high rise cities: A case study in Hong Kong. *Science of the Total Environment*, *592*, 306–315.

- Lennon, J., & Turner, R. (1995). Predicting the spatial distribution of climate: temperature in Great Britain. *J. Anim. Ecol.*, 64, 370 – 392.
- Li, J., & Heap, A. (2011). A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecol. Inform.*, 6, 228-241.
- Li, J., & Heap, D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, 173 – 189.
- Liu, B., Wu, J., Zhang, J., Wang, L., Yang, J., Liang, D., y otros. (2017). Characterization and source apportionment of PM2.5 based on error estimation from EPA PMF 5.0 model at a medium city in China. *Environmental Pollution*, 222, 10-22.
- Londoño, L., & Cañón, J. (2015a). Imputation of spatial air quality data using GIS-Spline and the index of agreement in sparse urban monitoring networks. *Revista Facultad de Ingeniería Universidad de Antioquia*.(76), 73-81.
- Londoño, L., & Cañón, J. (2015b). 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*, 11(21), 29-40.
- Londoño, L., & Valdés, J. (2015). *Geoestadística Aplicada: Generación de mapas de interpolación para el estudio de fenómenos distribuidos espacialmente*. Madrid: Editorial Académica Española.
- Londoño, L., Cañón, J., & Giraldo, J. (2017). Modelo de proximidad espacial para definir sitios de muestreo en redes urbanas de calidad de aire. *Revista Facultad Nacional de Salud Pública*, 35(1), 111-122.
- Londoño, L., Cañón, J., Villada, R., & López, L. (2015). Caracterización espacial de PM10 en la ciudad de Medellín mediante modelos geoestadísticos. *Revista Ingenierías USBMED*, 6(2), 26-35.
- Lozano, A., Usero, J., Vanderlinden, E., Ruez, J., Contreras, J., Navarrete, B., y otros. (2009). Design of air quality monitoring networks and its application to NO2 and O3 in Cordova, Spain. *Microchemical Journal*, 93(2), 211-219.
- Marcon, A., Pesce G, G., Girardi, P., Marchetti, P., Blengio, G., & Sappadina, S. (2014). Association between PM10 concentrations and school absences in proximity of a cement plant in northern Italy. *International Journal of Hygiene and Environmental Health*, 217(2-3), 386-391.
- MinAmbiente (2007). *Ministerio de Ambiente Vivienda y Desarrollo Sostenible*. Revisado el 7 de abril de 2007, de Protocolo para el monitoreo y seguimiento de la calidad del aire. Bogotá:
https://www.cortolima.gov.co/sites/default/files/images/stories/calidadAire/prtocolo_monitoreo_calidad_aire.pdf
- MinAmbiente (2016). *Ministerio de Ambiente y Desarrollo Sostenible, República de Colombia*. Revisado el 1 de agosto de 2016, de Diagnostico Nacional de Salud Ambiental. .pp 93-137:
<https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/INEC/IGUB/Diagnostico%20de%20salud%20Ambiental%20compilado.pdf>
- MinAmbiente (2017). *Contaminación Atmosférica en Colombia*. Revisado el 23 de agosto de 2017, de Ministerio de Ambiente y Desarrollo Sostenible:
<http://www.minambiente.gov.co/index.php/component/content/article/1801-plantilla-#1-1-normativa>

- Mofarrah, A., & Husain, T. (2010). A holistic approach for optimal design of air quality monitoring network expansion in an urban area. *Atmospheric Environment*, 44(3), 432-440.
- Nejadkoorki, F., Nicholson, K., & Hadad, K. (2011). The design of long-term air quality monitoring networks in urban areas using a spatiotemporal approach. *Environmental Monitoring and Assessment*, 172(1-4), 215-223.
- Papanastasiou, D., Melas, D., & Kioutsioukis, I. (2007). Development and assessment of neural network and multiple regression models in order to predict PM10 levels in a medium-sized Mediterranean city. *Water, Air, & Soil Pollution*, 182(1-4), 325-340.
- Papanastasiou, D., Melas, D., & Kioutsioukis, I. (2007). Development and assessment of neural network and multiple regression models in order to predict PM10 levels in a medium-sized Mediterranean city. *Water Air Soil Pollut.*, 182, 325-334.
- Paschalidou, A., Karakitsios, S., Kleanthous, S., & Kassomenos, P. (2011). Forecasting hourly PM10 concentration in Cyprus through artificial neural networks and multiple regression models: Implications to local environmental management. *Environ. Sci. Pollut. Res*, 18, 316-327.
- Pérez, J., Palacio, C., & Jiménez, J. (2011). Aspectos descriptivos de la circulación en la capa límite atmosférica del Valle de Aburra. Revista Politécnica ISSN 1900-2351, Año 7, Número 13. *Revista Politécnica*, 7(13), 35-47.
- Pilsung, K. (2013). Locally linear reconstruction based missing value imputation for supervised learning. *Neurocomputing*, 118, 65-78.
- Pollice, A., & Jona, G. (2009). Two Approaches to Imputation and Adjustment of Air Quality Data from a Composite Monitoring Network. *Journal of Data Science*, 7, 43-59.
- Pope, C., & Dockery, D. (2006). Health effects of fine particulate air pollution: Lines that connect. *J. Air Waste Manag. Assoc*, 56, 709-742.
- Pope, R., & Wu, J. (2014). Characterizing air pollution patterns on multiple time scales in urban areas: A landscape ecological approach. *Urban Ecosyst*, 17, 855-874.
- Quiroz, M., Martínez, D., Massone, H., Londoño, L., & Dapeña, C. (2015). Spatial distribution of electrical conductivity and stable isotopes in groundwater in large catchments: a geostatistical approach in the Quequén Grande River catchment, Argentina.
- Ramsay, J., Ramsay, T., & Sangalli, L. (2011). Spatial functional data analysis. *Recent Advances in Functional Data Analysis and related topics*. (págs. 269-275). Springer.
- RedAire. (2015). Laboratorio de Calidad del Aire (CALAIRE). Universidad Nacional de Colombia Sede Medellín, Facultad de Minas. Medellin, Colombia.
- Rendón, A., Salazar, F., Palacio, C., Wirth, V., & B, B. (2014). Effects of Urbanization on the Temperature Inversion Breakup in a Mountain Valley with Implications for Air Quality. *J. Appl. Meteor. Climatol*, 53, 840-858.
- Rooney, M., Dionisio, K., Arku, R., Paciorek, C., Friedman, A., Carmichael, H., y otros. (2012). Spatial and temporal patterns of particulate matter sources and pollution in four communities in Accra, Ghana. *Science of The Total Environment*, 435-436(1), 107-114.
- Şahin, Ü., Bayat, C., & Uçan, O. (2011). Application of cellular neural network (CNN) to the prediction of missing air pollutant data. *Atmospheric Research*, 101(1-2), 314-326.

- Sangalli, L., Ramsay, J., & Ramsay, T. (2013). Spatial splines regressions models. *J. Roy. Statisc. Soc. Ser. B*, 75, 681-803.
- Sanhueza, P., Torreblanca, M., Diaz-Robles, L., Schiappacasse, L., Silva, M., & Astete, T. (2009). Particulate air pollution and health effects for cardiovascular and respiratory causes in Temuco, Chile: A wood-smoke-polluted urban area. *J. Air Wate Manag. Assoc.*, 59, 1481-1488.
- Sayegh, A., Munir, S., & Habeebullah, T. (2014). Comparing the performance of statistical models for predicting PM10 concentrations. *Aerosol and Air Quality Research*, 14(3), 653 - 665.
- Schmitz, R. (2004). Modelling of air pollution dispersion in Santiago de Chile. *Atmospheric Environment*, 39, 2035-2047.
- Scott, L., Rosenshein, L., & Janikas, M. (2011). Modeling Spatial Relationships using Regression Analysis. *ESRI International User Conference. Technical Workshops*. San Diego, CA.
- Shahraiyni, H., & Sodoudi, S. (2016). Statistical Modeling Approaches for PM10 Prediction in Urban Areas; A Review of 21st-Century Studies. *Atmosphere*, 7(2)(15).
- SIATA. (2016). Sistema de Alerta Temprana de Medellín y el Valle de Aburrá. *Sistema de Alerta Temprana de Medellín y el Valle de Aburrá*. Medellín, Colombia.
- Singh, N., Murari, V., Kumar, M., Barman, S., & Banerjee, T. (2017). Fine particulates over South Asia: Review and meta-analysis of PM2.5 source apportionment through receptor model. *Environmental Pollution*, 223, 121-136.
- Slini, T., Kaprara, A., Karatzas, K., & Moussiopoulos, N. (2006). PM10 forecasting for Thessaloniki, Greece. *Environmental Modelling & Software*, 21(4), 559 - 565.
- Song, C., He, J., Wu, L., Jin, T., Chen, X., Li, R., y otros. (2017). Health burden attributable to ambient PM2.5 in China. *Environmental Pollution*, 223, 575-586.
- Stadlober, E., Hörmann, S., & Pfeiler, B. (2008). Quality and performance of a PM10 daily forecasting model. *Atmospheric Environment*, 42, 1098–1109.
- Stern, R., Builtjes, P., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., y otros. (2008). A model inter-comparison study focusing on episodes with elevated PM10 concentrations. *Atmos. Environ*, 42, 4567-4588.
- Taheri Shahraiyni, H., Sodoudi, S., Cubasch, U., & Kerschbaumer, A. (2015). The influence of the plants on the decrease of air pollutants (Case study: Particulate matter in Berlin). *In Presented at the Euro-American Conference for Academic Disciplines*. Paris, France.
- Taheri, H., & Sodoudi, S. (2016). Statistical Modeling Approaches for PM10 Prediction in Urban Areas; A Review of 21st-Century Studies. *Atmosphere*, 7(15).
- Theil, H. (1971). *Principles of Econometrics*. New York: John Wiley & Sons.
- Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(2), 234-240.
- Toro, V., & Marín, A. (2006). Índice de Calidad del aire para el Valle de Aburrá. http://www.lasallista.edu.co/fxcul/media/pdf/revistalimpia/vol1n1/pl_v1_n1_102_ica.pdf. *Producción + Limpia*, 1(1), 102-107.
- Tuna, F., & Buluç, M. (2015). Analysis of PM10 pollutant in Istanbul by using Kriging and IDW methods: Between 2003 and 2012. *Int. J. Comput. Inf. Technol.*, 4, 170–175.
- Urrutia, J., Palomino, R., & Salazar, H. (2010). Metodología para la imputación de datos faltantes en Metereología. *Scientia Et Technica*, 3(46), 44 - 49.

- US EPA. (1998). *Quality Assurance Handbook for Air Pollution Measurement Systems. Volume 11: Part 1 Ambient Air Quality monitoring Program Quality System Development. Office of Air Quality Planning and Standards Research Triangle Park. IBSN EPA-454/R-98-004*. Washington: EPA.
- Wang, P., Zhang, H., Qin, Z., & Zhang, G. (2017). A novel hybrid-Garch model based on ARIMA and SVM for PM_{2.5} concentrations forecasting. *Atmospheric Pollution Research, In Press, Corrected Proof*.
- Webster, R., & Oliver, M. (2001). *Geostatistics for environmental scientists*. Great Britain: John Wiley & Sons Inc.
- Wiener, N. (1956). *The Theory of Prediction. Modern Mathematics for Engineers*. New York: McGraw-Hill.
- Willmott, C., Robeson, S., & Matsuura, K. (2012). A refined index of model performance. *International Journal of Climatology*, 32(13), 2088–2094.
- Woolridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Massachusetts: MIT Press.
- Zafra, C., Ángel, Y., & Torres, E. (2017). ARIMA analysis of the effect of land surface coverage on PM₁₀ concentrations in a high-altitude megacity. *Atmospheric Pollution Research, In Press, Corrected Proof*.
- Zhang, G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175.
- Žukovič, M., & Hristopulos, D. (2008). Environmental time series interpolation based on Spartan random processes. *Atmospheric Environment*, 42(33), 7669-7678.