

**APLICACIÓN DE MÉTODOS DE INTERPOLACIÓN Y MODELAMIENTO  
GEOESTADÍSTICO EN LA EVALUACIÓN DE LA CALIDAD DEL AIRE EN  
BOGOTÁ D.C.**

**APPLICATION OF INTERPOLATION METHODS AND GEOSTATISTICAL  
MODELING IN THE EVALUATION OF AIR QUALITY IN BOGOTÁ D.C.**

Heyner M. Fernando Rodríguez Rodríguez  
Ing. Catastral y Geodesta  
[comoelsuyo@gmail.com](mailto:comoelsuyo@gmail.com)

**UNIVERSIDAD MILITAR NUEVA GRANADA  
FACULTAD DE INGENIERIA  
ESPECIALIZACIÓN EN GEOMATICA  
2014**

# **APLICACIÓN DE MÉTODOS DE INTERPOLACIÓN Y MODELAMIENTO GEOESTADÍSTICO EN LA EVALUACIÓN DE LA CALIDAD DEL AIRE EN BOGOTÁ D.C.**

## **APPLICATION OF INTERPOLATION METHODS AND GEOSTATISTICAL MODELING IN THE EVALUATION OF AIR QUALITY IN BOGOTÁ D.C.**

Heyner M. Fernando Rodríguez Rodríguez  
Ing. Catastral y Geodesta  
[comoelsuyo@gmail.com](mailto:comoelsuyo@gmail.com)

### **RESUMEN**

La mayoría de las emisiones de contaminantes presentes en la atmosfera son provocadas por acción humana y que al ser difundidas por el territorio generan afectación para la vida de humanos, animales y plantas. En el Distrito Capital existe la Red de Monitoreo de Calidad del Aire de Bogotá RMCAB, que se encarga de la recolección de la información que se refiere al comportamiento de las concentraciones de contaminantes y de las variables meteorológicas.

Pero estas mediciones ambientales sólo proporcionan información acerca de las condiciones atmosféricas en un momento y ubicación en particular y las mediciones por sí solas no pueden ser utilizadas para establecer una estrategia eficaz para la solución de problemas de calidad del aire.

Es aquí donde se vale de la geoestadística como técnica de predicción que permite determinar una continuidad espacial de las diferentes variables y así poder predecir o estimar un valor muy cercano al real en un punto desconocido.

En el presente trabajo se presenta la aplicación de diferentes métodos de interpolación espacial de las mediciones de concentraciones de contaminantes atmosféricos obtenidos por la Red de Monitoreo de Calidad del Aire en las estaciones fijas ubicadas en diferentes puntos del distrito capital, para obtener superficies completas del territorio con el objeto de determinar las diferencias entre los métodos y realizar predicciones más precisas del nivel de contaminación en cualquier parte del área de estudio que puedan proporcionar información más realista de la calidad del aire en Bogotá DC.

Para lograr esto se inicia por un análisis exploratorio de los datos obtenidos por la red de calidad del aire y la georreferenciación de las estaciones de la misma, se aplican los diversos métodos de interpolación y realiza la comparación de estos,

para posteriormente analizar los resultados obtenidos por cada uno de ellos y posteriormente se representan con ayuda de mapas de predicción y error, para así poder determinar cuál de los métodos empleados presenta una mejor predicción.

**Palabras clave:** Geoestadística, Predicción, Interpolación, kriging, IDW, Material Particulado, Calidad del aire.

## ABSTRACT

Most of emissions from pollutant present in the atmosphere are caused by human action, and when they spread they affect the human life, animals and plants. In DC there is an institution which measure the quality of the air: Network Monitoring Air Quality (Red de Monitoreo de Calidad del Aire de Bogotá RMCAB) Which collects data from the behavior and concentration of contaminants and the weather variables.

But this measuring from the atmosphere just provides data of the weather conditions in a time and space in particular. Also these measures cannot be used to establish an efficient strategy to solve the problem with the quality of the air. That is why we use Geostatistics as a predicting technic, which allows us to define a spatial continuity on the different variables and predict or estimate a close value in any point.

In this paper we present how to use different methods of spatial interpolation of the measurements of concentrations of air pollutants given by the RMCAB in the fixed stations in Bogota and obtain wider surfaces of the territory. All of this to determine the differences between them and be able to make more accurate predictions of the contamination in any area, this will result in a more precise data about the quality of the air in Bogota, DC.

In order to achieve this we start an exploratory analysis of the data obtained by the RMCAB and the Georeference of the stations. The methods of interpolation are applied and the comparison made will result on predictions and error maps that will bring which method makes a better prediction.

**Key Words:** Geostatistics, prediction, interpolation, kriging, IDW, Particulate Matter, Air Quality.

## INTRODUCCIÓN

La geoestadística forma parte de la geografía matemática y puede ser definida como la parte de la estadística que se relaciona con los datos espaciales o espacio temporales, y que se vale de técnicas y herramientas que permiten analizar y realizar simulaciones, estimaciones y predicciones sobre una variable que se encuentra distribuida en el espacio o el tiempo de forma continua.

La geoestadística fue desarrollada inicialmente en operaciones mineras, pero se aplica en la actualidad en una variedad de disciplinas tales como la geología, hidrología, climatología, silvicultura, agricultura y epidemiología entre otras, ya que su importancia se basa en la interpolación, que permite crear continuidad espacial en las variables y de esta forma predecir o estimar su valor en lugares desconocidos, por consiguiente si no existe la continuidad espacial obtendríamos estimaciones sin sentido. [1]

El tema ambiental es uno de los de mayor relevancia actualmente, lo que ha promovido a que nuevas metodologías en el campo estadístico se hayan creado a través de los años y que profesionales en estadística enfoquen sus esfuerzos en el desarrollo de técnicas más apropiadas para el análisis de la información que se refiere a este tema.

El distrito capital de Bogotá DC (el cual es la zona de estudio para la elaboración de este trabajo) posee la RMCAB (Red de Monitoreo de Calidad del Aire) la cual es la encargada de recolectar toda la información referente a la concentración de contaminantes ya sean de origen humano o natural además de las variables meteorológicas en Bogotá. Dichos datos son capturados en diversos puntos de la ciudad en estaciones de monitoreo y agrupados en un nivel central donde se realizan diversos análisis para establecer y evaluar el cumplimiento de los estándares y normatividad vigente en el manejo de la calidad del aire. [2]

En el marco del presente trabajo se estableció como objetivo el aplicar métodos de interpolación y modelamiento geoestadístico de las mediciones de concentraciones de contaminantes atmosféricos obtenidos por la Red de Monitoreo de Calidad del Aire durante el año de 2013, que permitan la estimación de áreas de predicción de concentración de contaminantes partiendo de mediciones anuales en puntos específicos para obtener valores cercanos a la realidad en los lugares desconocidos.

Tomando que el ente encargado de la regulación y cumplimiento de los estándares de calidad para el año de estudio (2013) es la Secretaria de Ambiente (SDA) usando como herramienta la RMCAB, con la normatividad dada por la Resolución 601 de 2006 del Ministerio de Ambiente, Vivienda y Desarrollo, la cual fue modificada por la Resolución 610 de 2010. Durante el año 2013 dicha red estaba conformada por 13 estaciones de monitoreo fijas y una estación móvil, las cuales se encuentran ubicadas en diferentes lugares de la ciudad, provistas de equipos de alta tecnología que realizan el monitoreo en forma continua de concentraciones de contaminantes como el material particulado que se clasifica en PM10, PM2.5 y PST, también el monitoreo de gases contaminantes como el SO<sub>2</sub>,

NO<sub>2</sub>, CO y O<sub>3</sub>, además de algunas variables meteorológicas como lo son la velocidad y dirección del viento, temperatura y humedad relativa y presión barométrica . [2]

Se destacan los materiales particulados como el PM<sub>10</sub> y PM<sub>2.5</sub> que son todas la partículas que tienen diámetro inferior a 10 micrómetros y 2.5 micrómetros respectivamente, la partículas suspendidas totales PST y el Ozono O<sub>3</sub>, los cuales son los contaminantes que se ha encontrado que en algunas zonas están por encima de los límites permitidos por la norma, en cuanto a los demás contaminantes como el CO, NO<sub>2</sub> y SO<sub>2</sub>, aunque se encuentran presentes se presentan en niveles permisibles, por lo que no representarían un riesgo directo para la salud de la población. [2]

## **1. MATERIALES**

Aplicar a la información varios de los métodos existentes para realizar interpolaciones y tomando algunas consideraciones para el caso específico de Bogotá realizar estimaciones y comparar los resultados obtenidos por cada uno de los métodos utilizados.

### **1.1 INFORMACIÓN UTILIZADA**

La información utilizada para este proyecto fue extraída de la base de datos de la Red de Calidad del Aire de Bogotá que contiene las mediciones de las estaciones de la red en el periodo del 1 de enero de 2013 al 31 de diciembre de 2013, que como se mencionó anteriormente, contiene información de 13 estaciones fijas y una estación móvil, ubicadas en diferentes lugares de la ciudad que capturan datos de las variables material particulado (PM<sub>10</sub>, PST, PM<sub>2.5</sub>), de gases contaminantes (SO<sub>2</sub>, NO<sub>2</sub>, CO, O<sub>3</sub>) y de variables, velocidad y dirección del viento, temperatura, radiación solar, humedad relativa y presión barométrica.

La base de datos con las mediciones de las variables está disponible para descarga al público a través de la página web de la Red [3] y contiene registros por hora y día para cada mes, aunque existen algunos periodos de tiempo para los que no se encuentran mediciones, esto debido a diferentes factores tales como fallas en los medidores de las estaciones, o procesos de mantenimiento en las mismas. [2]

### **1.2 NORMATIVIDAD VIGENTE**

Para el año de estudio de los datos utilizados la normatividad vigente de calidad del aire que fue adoptada por el distrito como nivel de referencia en la evaluación

de los niveles máximos de concentración de los diferentes contaminantes en la ciudad fue la establecida por la Resolución 601 del 4 de abril de 2006 expedida por el entonces Ministerio de Ambiente, Vivienda y Desarrollo Territorial (MAVDT), la cual fue modificada mediante Resolución 610 de 2010. [4]

**Tabla 1** Niveles máximos permisibles para contaminantes.

CONTAMINANTE	NIVEL MÁXIMO PERMISIBLE ( $\mu\text{g}/\text{m}^3$ )	TIEMPO DE EXPOSICIÓN	CÁLCULO
PST	100	Anual	Promedio aritmético de las concentraciones diarias en 365 días.
	300	24 horas	Concentración promedio en 24 horas.
PM <sub>10</sub>	50	Anual	Promedio geométrico de las concentraciones diarias en 365 días.
	100	24 horas	Concentración promedio en 24 horas.
PM <sub>2.5</sub>	25	Anual	Promedio aritmético de las concentraciones diarias en 365 días.
	50	24 horas	Concentración promedio en 24 horas.
SO <sub>2</sub>	80	Anual	Promedio aritmético de las concentraciones diarias en 365 días.
	250	24 horas	Concentración promedio en 24 horas.
	750	3 horas	Concentración promedio horaria en 3 horas.
NO <sub>2</sub>	100	Anual	Promedio aritmético de las concentraciones diarias en 365 días.
	150	24 horas	Concentración promedio en 24 horas.
	200	1 hora	Concentración promedio horaria.
O <sub>3</sub>	80	8 horas	Concentración promedio en 8 horas
	120	1 hora	Concentración promedio horaria
CO	10000	8 horas	Concentración promedio en 8 horas
	40000	1 hora	Concentración promedio horaria

Fuente: Autor Heyner Rodríguez

## 2. METODOLOGÍA

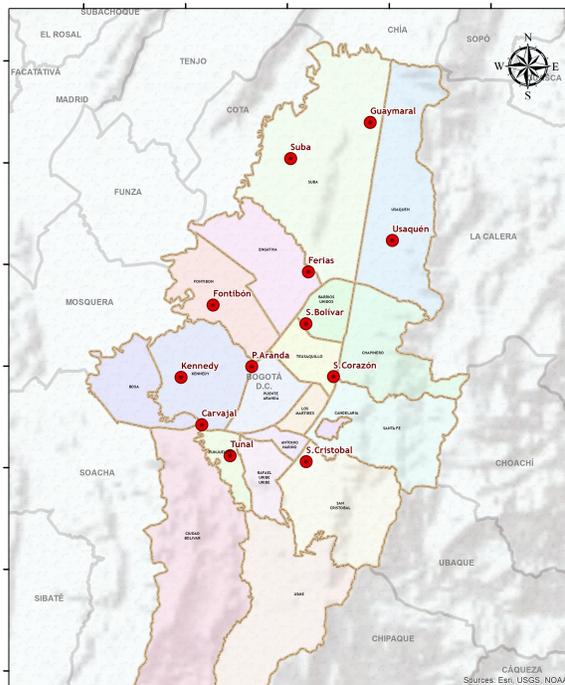
En la búsqueda de desarrollar un buen análisis geoestadístico es recomendable la creación de un ciclo metodológico que permita un proceso lógico de la información y así poder obtener mejores resultados. En la figura 1. Se muestra una representación de dicho ciclo. [5]



**Figura 1** Ciclo metodológico.  
Fuente: Autor Heyner Rodríguez

## 2.1 INFORMACIÓN BASE

En un estudio de este tipo donde existe correlación espacial es necesario realizar la espacialización del área de estudio, en este caso realizar la georreferenciación de cada una de las estaciones de monitoreo, y de igual manera se entiende que debe realizarse una muestra de la variable o variables a investigar que sea representativa es decir que cuente con la información suficiente para realizar una estimación de calidad.



ESTACION	LATITUD	LONGITUD	ALTITUD
GUAYMARAL	4,7837556	-74,0441833	2580
USAQUEN	4,71035	-74,0304167	2570
SUBA	4,7612472	-74,0934611	2571
FERIAS	4,6907	-74,0824833	2552
S.BOLIVAR	4,6584667	-74,0839667	2577
S.CORAZON	4,6254861	-74,0669806	2621
FONTIBON	4,6701	-74,14155	2591
P.ARANDA	4,6317667	-74,1174833	2590
KENNEDY	4,62505	-74,1613333	2580
CARVAJAL	4,5956167	-74,1485833	2563
TUNAL	4,576225	-74,1309556	2589
S.CRISTOBAL	4,5725528	-74,0838139	2688

**Figura 2** Ubicación de las estaciones de monitoreo.  
Fuente: Autor Heyner Rodríguez

## 2.2 SELECCIÓN DE VARIABLES

Cuando se realiza la selección de variables se debe tomar en cuenta aquellas que presenten un tamaño de muestra que sea representativo y que además tengan mayor influencia en el área de estudio. [5]

Tomando esta premisa como base y teniendo en cuenta que la normatividad vigente establece que el porcentaje mínimo de representatividad temporal es de 75%. [4] La tabla 2. indica el comportamiento porcentual de datos validos por estación y por parámetro, con esto podemos descartar los datos de los contaminantes CO y PM2.5 para el estudio ya que se encuentran en porcentajes de información inferiores a lo que indica la norma, el déficit de información de estas variables se debieron a fallas en los monitores de las estaciones y a la instalación de nuevos equipos de monitoreo [2].

Ahora si tomamos en cuenta además que los contaminantes que más destacan son PM10, el PM2.5 y el Ozono ya que sus concentraciones superan los límites máximos permitidos y que los demás contaminantes como CO, NO2, SO2, presentan concentraciones más bajas y en ningún valor ni estación monitoreada superó el máximo permisible por la norma podemos descartar el SO2 ya que no representa un peligro grave para la población. [2, 6]

Es así como definimos la muestra de los datos de los contaminantes: material particulado PM10 y Ozono O3, ya que son los que poseen un porcentaje de datos más alto, también son parte de los contaminantes que pueden producir más daño en la población.

**Tabla 2** Porcentaje de datos válidos por estación y contaminante.

ESTACION	PM10	OZONO	CO	SO2	PM2.5
CARVAJAL	99,36%	90,73%	95,45%	97,42%	26,38%
FERIAS	85,09%	84,89%	68,34%	76,47%	19,62%
FONTIBON	83,60%	81,56%	82,74%	73,88%	N.A
GUAYMARAL	92,72%	92,72%	N.A	N.A	N.A
KENNEDY	96,97%	N.A	93,97%	60,31%	96,86%
P.ARANDA	94,89%	83,15%	95,87%	98,55%	N.A
S.CORAZON	96,00%	52,63%	N.A	N.A	N.A
S.CRISTOBAL	79,01%	65,95%	22,60%	70,53%	N.A
S.BOLIVAR	86,29%	73,63%	85,05%	62,74%	27,23%
SUBA	91,76%	81,05%	N.A	91,72%	N.A
TUNAL	80,32%	91,07%	49,58%	73,41%	21,36%
USAQUEN	93,52%	88,54%	5,47%	N.A	23,09%
TOTAL	89,96%	80,54%	66,56%	78,34%	35,76%

Fuente: Autor Heyner Rodríguez

## 2.3 ANÁLISIS EXPLORATORIO DE LOS DATOS

Es de gran importancia este proceso de exploración y análisis gráfico, al igual que los próximos procedimientos geostadísticos, pues en este proceso se trata de observar y evaluar la forma de la distribución de la información, determinar la correlación, la variabilidad, buscar tendencias, encontrar valores atípicos y determinar su causa y si deban ser incluidos o no dentro de la muestra que ayudará a establecer supuestos y definir que método de predicción es el más conveniente en cada caso específico. [1] [5]

Uno de los métodos de exploración gráfico de datos es el diagrama de caja o diagrama de caja y bigote, el cual es un método de representación univariante que da una visión de la simetría de distribución de los datos y hace posible la identificación de datos atípicos y suministra información acerca de la dispersión con respecto a la mediana, la media, valores máximos y mínimos además de los cuartiles. [6, 1]

El diagrama de caja está conformado por varios elementos, un rectángulo o caja que representa a la varianza, los brazos o bigotes representan los valores máximo y mínimo de la muestra, los puntos son valores atípicos y la línea que indica la mediana, cuando la línea está en el centro de la caja significa que la media y la mediana son muy cercanas, indicando una distribución normal. [5]

El comportamiento de material particulado PM<sub>10</sub>, se realizó por el seguimiento en 12 estaciones de monitoreo durante todo el 2013 como se refleja en la Tabla 2. Las mediciones se realizan con una periodicidad de una hora, luego se promediaron por día, para realizar la exploración de los datos ya que los límites permitidos para este contaminante se definen para promedios diarios y anuales como se indica en la Tabla 1. de acuerdo con la normatividad [4] tenemos que el nivel máximo permisible en 50  $\mu\text{g}/\text{m}^3$  para un periodo de exposición anual y 100  $\mu\text{g}/\text{m}^3$  para un promedio diario.

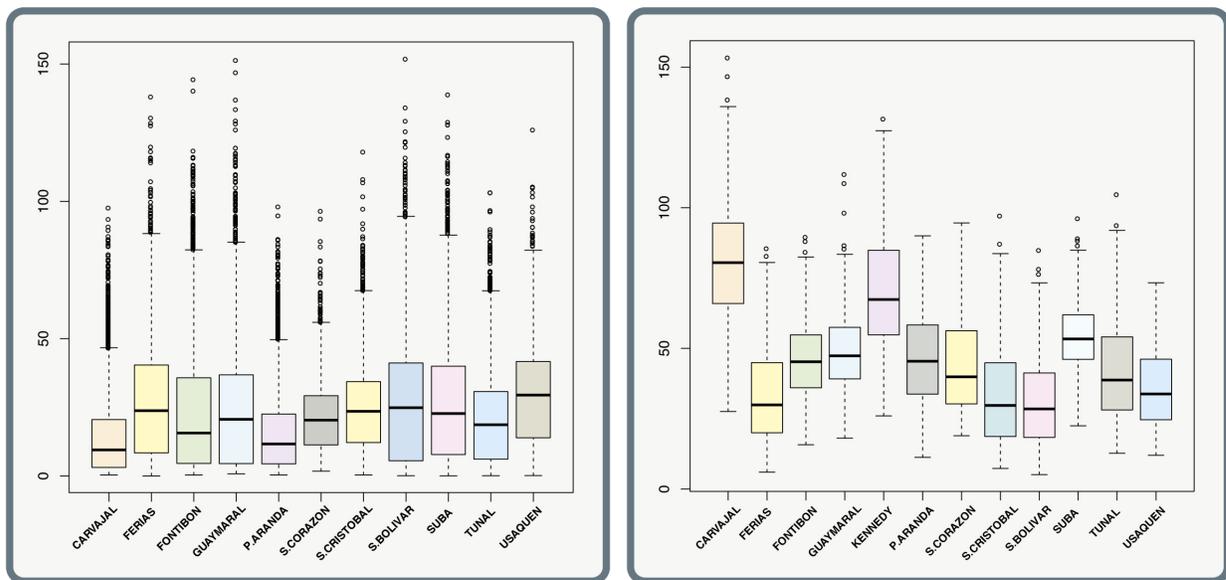
Ahora, nos valemos del diagrama de caja para determinar el comportamiento de las concentraciones de PM<sub>10</sub> como se muestra en la Figura 3. podemos observar que la distribución en todas las estaciones de monitoreo es bastante homogénea, vemos como en la mayoría de las estaciones la línea se encuentra en el centro de la caja, lo que indica que la media y la mediana tienen valores similares, es decir que la distribución de los datos es normal lo que significa que para estas zonas de la ciudad la distribución de la concentración de PM<sub>10</sub> es uniforme para todo el 2013. Tenemos como excepción las estaciones Kennedy y Sagrado corazón que presentan distribuciones no tan homogéneas.

También es posible identificar los límites máximos de concentración que se presentó en la estación Carvajal, así como la existencia de datos atípicos en casi todas las estaciones con excepción de Puente Aranda, Sagrado Corazón y Usaquén. Estos datos atípicos no son representativos ya que se presentan en un porcentaje muy bajo, por lo que no se requiere que sean excluidos de la muestra.

De la misma forma se procedió con los datos de concentraciones de Ozono, para tener mayor idea de su comportamiento, para este se realizó el seguimiento en 11 estaciones de monitoreo en el 2013 según la base de datos de la Red, para este contaminante la normatividad [4] establece que los límites se calculan por periodicidad horaria o por ciclos de 8 horas y establece límites permitidos máximos de  $120 \mu\text{g}/\text{m}^3$  y  $80 \mu\text{g}/\text{m}^3$  respectivamente. Para el ejercicio de exploración se analizaron datos horarios.

El comportamiento gráfico de la distribución de Ozono se ve en la Figura 3. Que muestra el diagrama de caja en el que vemos una distribución un poco menos homogénea que con el PM10, pero que sigue mostrando una tendencia normal que indica una distribución uniforme de la concentración en las estaciones monitoreadas, en cuanto a los valores atípicos vemos que parecen más representativos que en el PM10 debido a que la muestra es mucho más grande, como se mencionó anteriormente se tomó periodicidad horaria, también al calcular los valores atípicos se encontró que estos están distribuidos en todos los días de todos los meses del año, lo que indica que no se deben a algún tipo de falla del sensor y que si deben mantenerse dentro de la muestra, la causa real de los valores debe ser estudiada con mayor enfoque.

Cabe notar que aunque los promedios no superan en ningún caso los límites permisibles, si existen excedencias en 6 de las estaciones monitoreadas y representan menos del 1% de los datos de la muestra.



**Figura 3** Diagramas de caja para los contaminantes para Ozono y de material particulado PM10  
Fuente: Autor Heyner Rodríguez

Ahora se procede a calcular los promedios anuales de las variables seleccionadas Ozono y PM10 por cada estación de monitoreo y realizar el análisis de distribución y tendencia de los datos a interpolar.

**Tabla 3** Promedio anual de concentraciones de Ozono y material particulado PM10

ESTACION	PM10	OZONO
GUAYMARAL	49,4	23,68
USAQUEN	36,5	28,73
SUBA	54,1	26,06
FERIAS	33,3	26,40
S.BOLIVAR	31,4	26,66
S.CORAZON	43,8	21,38
FONTIBON	46,2	22,72
P.ARANDA	46,5	15,29
KENNEDY	70,2	N.A
CARVAJAL	80,1	13,86
TUNAL	42,2	20,31
S.CRISTOBAL	33,5	24,64

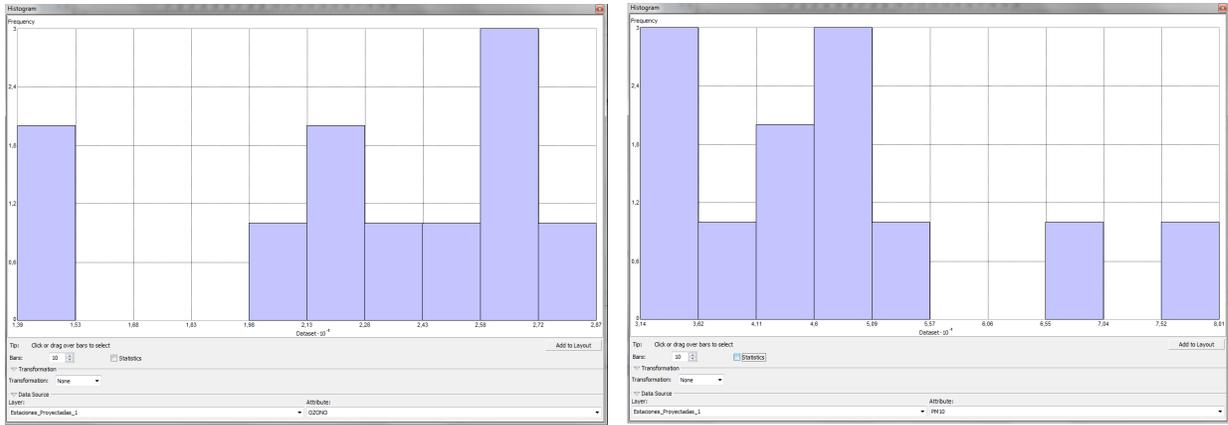
Fuente: Autor Heyner Rodríguez

Los métodos de interpolación se utilizan para generar una superficie de predicción para lugares en los que no se tiene información, ahora para tomar mejores decisiones al crear una superficie, primero debe explorar el conjunto de datos para obtener una mejor comprensión de la misma. Al explorar sus datos, lo que se busca es localizar errores en los valores que pueden afectar la superficie de predicción de salida; examinar la distribución de los datos; buscar tendencias globales e influencias direccionales.

Para esto nos valemos de las herramientas de la extensión análisis geoestadístico de la aplicación Arcgis 10.2 en el módulo de exploración de datos.

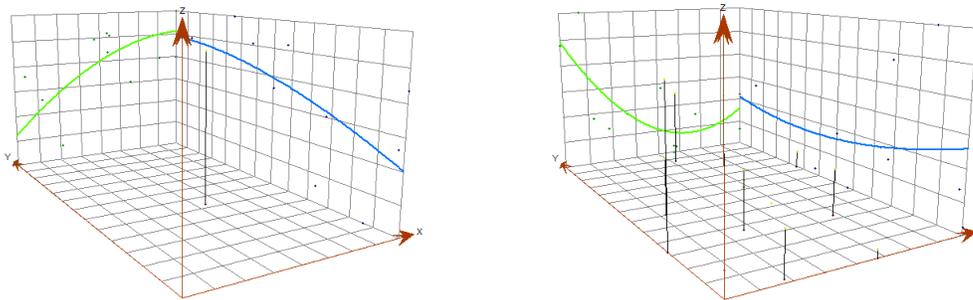
Los histogramas de frecuencia son una herramienta gráfica de los atributos del conjunto de datos, lo que le permite examinar la distribución para cada uno de estos atributos.

Se generaron histogramas para cada una de las variables estudiadas estas se dividieron en 10 rangos de valores, la altura de cada barra corresponde a la frecuencia. Lo que buscamos es percibir tendencias debidas a la propagación y simetría de los datos, como se ve en la Figura 4. Ninguna de las dos variables se encuentra cerca de distribución normal, lo que contribuiría a la creación de una mejor superficie esto también se pudo observar usando la herramienta *Normal QQ Plot* ya que la trama de puntos no forma una línea recta.



**Figura 3** Histogramas de frecuencia contaminantes para Ozono y de material particulado PM10  
Fuente: Autor Heyner Rodríguez

Continuando con el proceso de exploración se busca encontrar una tendencia global en los datos que pueda ser una representación no aleatoria del comportamiento de los datos, lo que significa que está representada por una ecuación matemática y puede ser eliminada para una mejor predicción. Para esto se usa la herramienta Trend Analysis y se puede percibir que en cuanto a las variables Ozono y PM10 (Figura 4.), estas presentan un comportamiento en U pero no muy bien definido, este comportamiento indicaría una función de crecimiento representada por un polinomio de segundo orden, sin embargo el que no esté muy bien definido puede significar que al eliminar la tendencia se produzcan superficies demasiado suaves que no representen de forma adecuada la superficie.



**Figura 4** Tendencia global en los datos contaminantes para Ozono y de material particulado PM10  
Fuente: Autor Heyner Rodríguez

Finalmente en la exploración se busca identificar las influencias direccionales y la correlación espacial de los datos, para eso se usó la herramienta *Semivariogram Cloud*, correlación supone que los elementos que están más cerca uno de otros más se parecen, el semivariograma experimental es construido gráficamente con una nube de puntos que reflejan las distancias de las muestras, y el modelo del

semivariograma a través del ajuste de funciones matemáticas que describen el crecimiento, existen varios modelos de semivariograma entre los más comunes y usados se encuentran el Modelo esférico, Modelo Gaussiano y el Modelo exponencial [5], en las variables estudiadas gráficamente se percibe que no existe un comportamiento claro que defina una dirección, pero se percibe una disminución progresiva en la autocorrelación espacial hasta acercarse a cero, característica de un modelo esférico. Cuando la correlación espacial depende de en qué dirección se calcula se dice que los datos tienen el comportamiento con anisotropía [1], la herramienta puede calcular la correlación en todas las direcciones y permite manipular la dirección para observar el comportamiento y en el caso de estudio nos muestra que debería considerarse la anisotropía en la interpolación.

## 2.4 SELECCIÓN DEL MÉTODO Y AJUSTE DEL MODELO

El enfoque de la investigación es predominantemente metodológico y busca la aplicación de métodos de interpolación espacial diferentes para generar superficies completas del territorio (área urbana de Bogotá DC) que representen de la forma más precisa posible el comportamiento de las variables de estudio (Concentración de Ozono y material particulado PM10), existen diversos métodos de interpolación, algunos son flexibles y otros no tanto y requieren que se cumplan ciertas condiciones en el comportamiento de los datos, el análisis se basa en escoger varios de estos y comparar los resultados obtenidos en su aplicación. Con tal fin se aplicaron los métodos de interpolación Inverso de la distancia IDW de tipo determinístico y los métodos Kriging ordinario y Kriging Simple de tipo estocástico.

Se utiliza el método Inverso de la Distancia o Distancia Inversa por su sencillez, no requiere parámetros complejos, que se basa en la estimación a partir de la semejanza del valor de las muestras más cercanas, la estimación se genera a partir de la ponderación de los valores de las muestras dependiendo de la distancia, asigna mayores pesos a las muestras cercana y menor a las más lejanas, el peso es inversamente proporcional a la distancia. [7]

En cuanto al kriging este es un método o conjunto de métodos de interpolación espacial, que se basa en predicciones realizadas a través de la minimización del error medio cuadrático. [1] Y que toma la correlación espacial para explicar las variaciones en una superficie de predicción.

**Kriging Ordinario:** Para este caso particular se desconoce la media de un proceso estacionario que se supone constante y se utilizan combinaciones de valores observados. [7]

$$Z^*(X_0) = \lambda_1 * Z(X_1) + \lambda_2 * Z(X_2) + \lambda_3 * Z(X_3) + \dots + \lambda_n * Z(X_n)$$

**Kriging Simple:** Usado si los fenómenos son estacionarios, con varianzas y esperanzas (m) conocidas y constantes. [7]

$$Z^*(X_0) = m + \sum \lambda_i * z_i(X_i)$$

**Ajuste del modelo al semivariograma empírico:** para ajustar los datos se selecciona una función que sirva como modelo, como se dijo anteriormente existen varios tipos de modelos a elegir tales como el esférico, circular, lineal, gaussiano y exponencial entre otros. Cabe notar que el modelo seleccionado tendrá fuerte influencia en la predicción de los valores desconocidos, cada modelo se ajusta a diferente tipo de comportamiento de los datos. Es acá donde el proceso de exploración de datos contribuye significativamente. De este proceso tenemos:

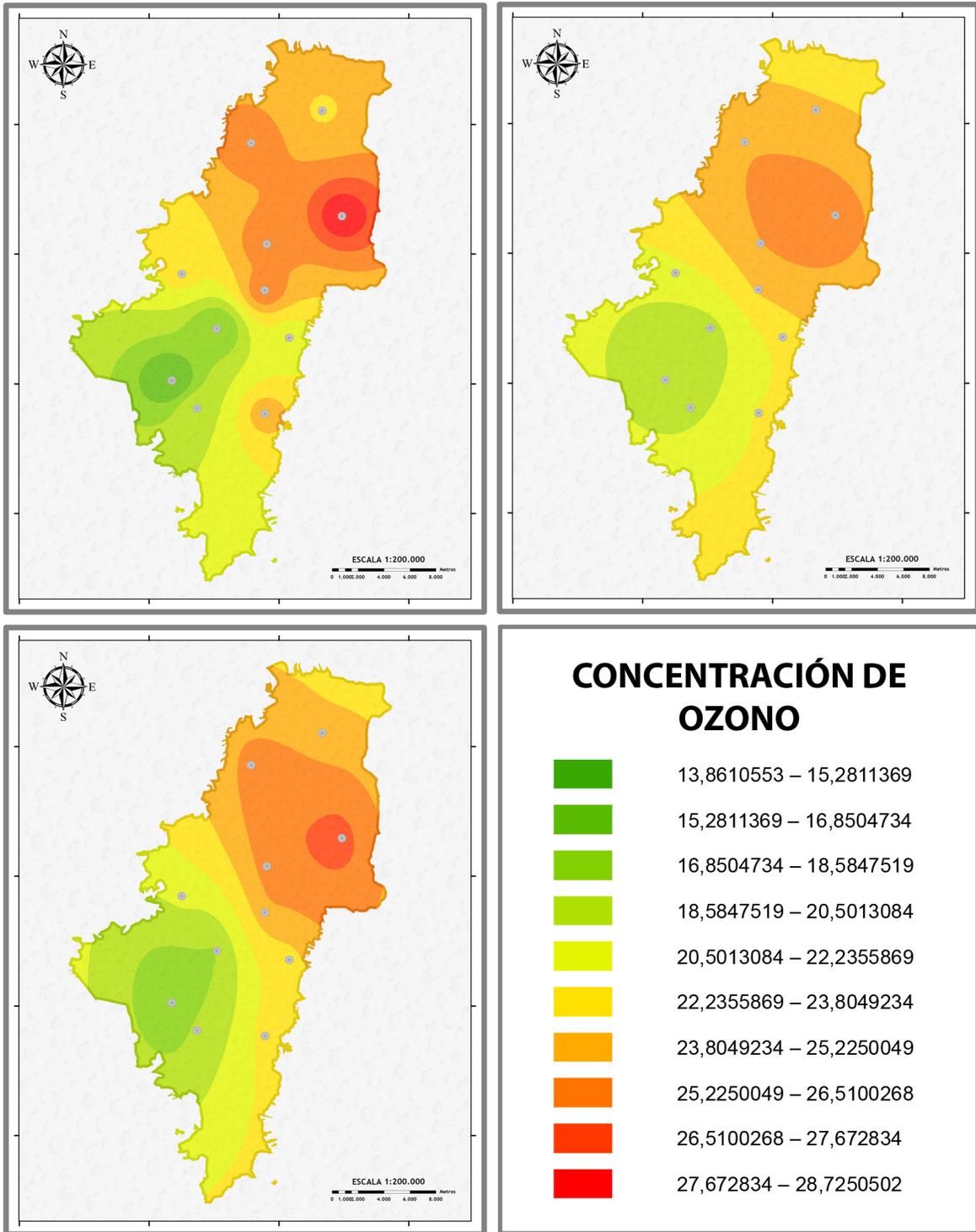
- Los datos de las variables analizadas no están cerca de una distribución normal.
- Una transformación de datos puede ser necesaria.
- Con la herramienta de Análisis de Tendencias, que vio que los datos mostraron una tendencia, se identificó que la tendencia de crecimiento en “U” podría contribuir un ajuste por un polinomio de segundo orden.
- El análisis indica que el modelo de interpolación debería tener en cuenta la anisotropía.
- La superficie del variograma se percibe una disminución progresiva en la autocorrelación espacial hasta acercarse a cero, característica de un modelo esférico.

Tomando en cuenta estos aspectos se puede proceder con confianza a la interpolación de superficie.

## 2.5 RESULTADOS DE LA ESTIMACIÓN

Los mapas, tablas y figuras que se muestran a continuación representan los resultados del proceso de interpolación de las variables de concentración de contaminantes de Ozono y Material Particulado PM10, creadas tomando en cuenta las tendencias de los datos, anisotropía y ajuste al modelo por función esférica, enunciados anteriormente.

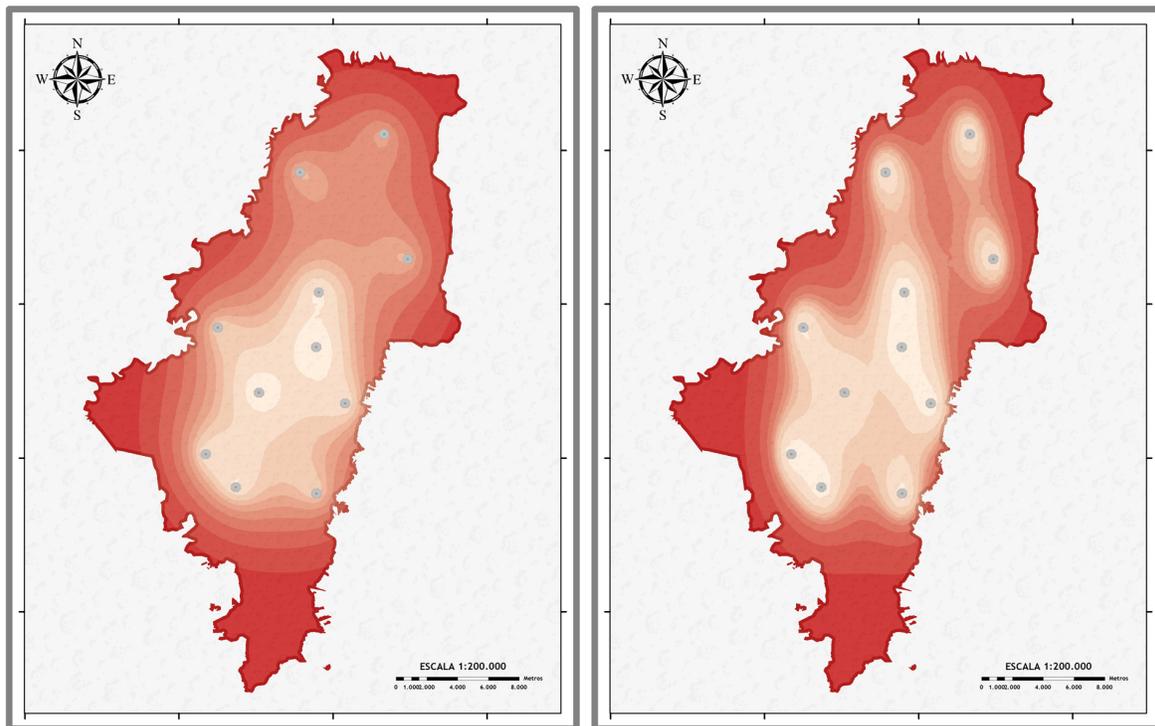
De estos resultados uno de los más relevantes son los mapas de predicción puestos muestran gráficamente como están distribuidos espacialmente las estimaciones y predicciones para el comportamiento de las variables estudiadas por cada uno de los diferentes métodos. Para estos se definen 10 niveles de concentración, tomando como unidad de medida  $\mu\text{g}/\text{m}^3$ , en la que el color verde representa los niveles más bajos y el rojo el rango con los niveles de concentración más altos.



**Figura 5** Mapa de predicción para concentraciones de Ozono Usando IDW, Kriging Ordinario y Kriging Simple  
Fuente: Autor Heyner Rodríguez

Si analizamos los mapas resultantes de concentración de Ozono en el área urbana de Bogotá podemos ver que en general tenemos la misma distribución espacial resultante, las zonas de color ocupan aproximadamente las mismas localizaciones, los niveles más bajos se encuentran en la zona sur occidental y valores más altos en las zonas más nororientales, pero si observamos con más detalle podemos percibir las diferencias, en el mapa IDW encontramos una mancha naranja en la zona suroriental que no aparece en ninguno de los mapas de predicción por kriging, ya que han sido eliminada por estos métodos debido a que está rodeada de estaciones con valores más bajos, igual sucede con un punto amarillo que aparece en la zona más al norte del mapa, que no puede percibirse en los otros dos métodos, entre los mapas del Kriging Ordinario y Simple las diferencias son menores, exhiben comportamientos muy parecidos, solamente se perciben variaciones en las áreas de cada uno de los rangos y en los valores máximos y mínimos.

A continuación se presentan los mapas de predicción del error de las concentraciones de Ozono de Bogotá interpoladas por los métodos Kriging Ordinario y Kriging simple, el método Distancia Inversa no permite la creación de estos mapas.



**Figura 6** Mapa de predicción del error para concentraciones de Ozono por Kriging Ordinario y Kriging Simple

Fuente: Autor Heyner Rodríguez

**Tabla 4** Resultados de estimación y error para concentraciones de Ozono por IDW, Kriging Ordinario y Kriging Simple

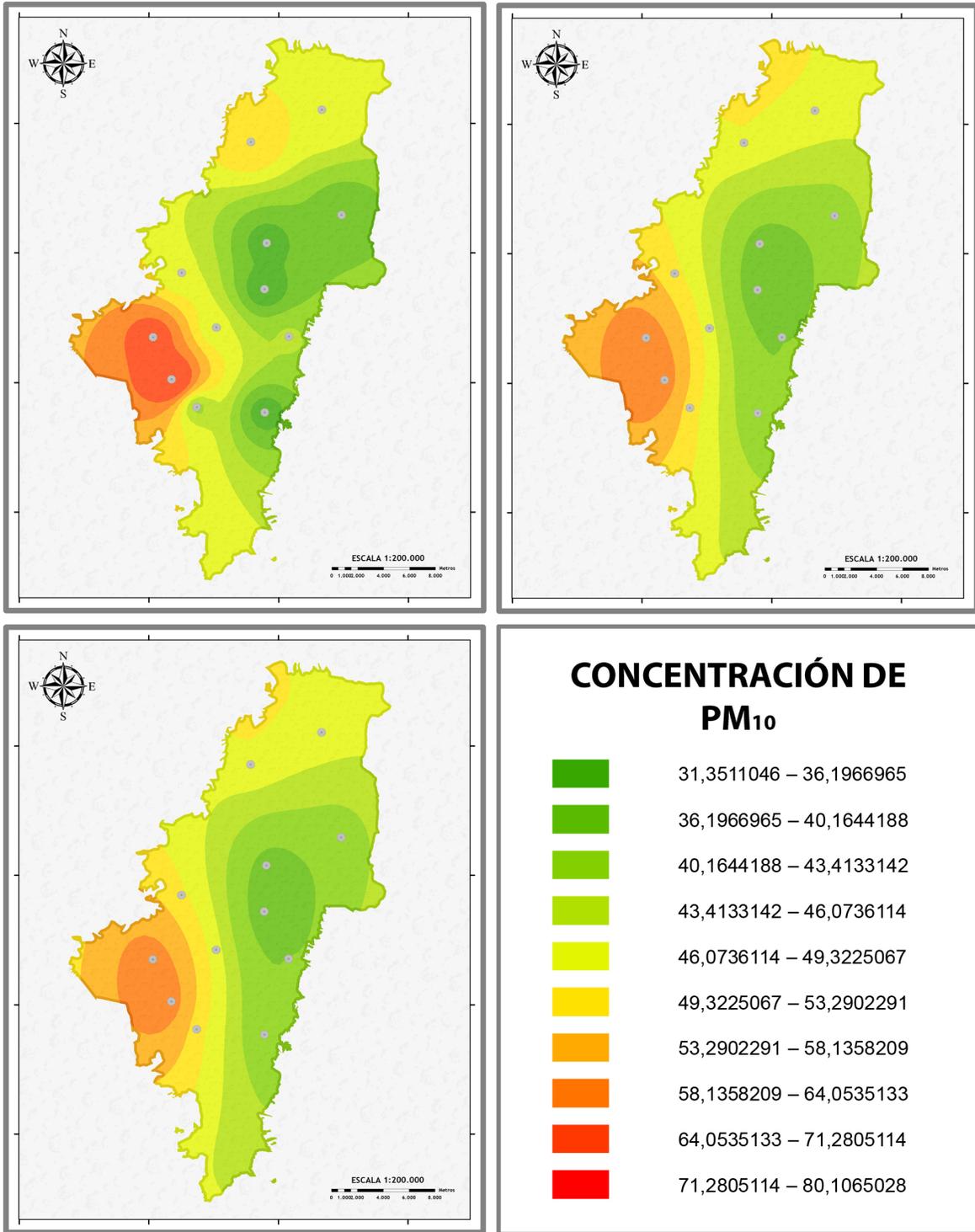
ESTACION	OZONO	IDW		ORDINARIO				SIMPLE			
		Predicted	Error	Predicted	Error	St_Error	Stdd_Error	Predicted	Error	St_Error	Stdd_Error
GUAYMARAL	23,67567	25,395479	1,719809	24,36496	0,689289	4,975248	0,138544	25,684742	2,009072	4,735327	0,424273
USAQUEN	28,72505	24,000814	-4,724237	23,829586	-4,895464	4,876707	-1,003846	23,954847	-4,770203	4,689102	-1,017296
SUBA	26,063888	24,140477	-1,923411	24,199931	-1,863956	4,878504	-0,382075	24,75084	-1,313047	4,664098	-0,281522
FERIAS	26,398625	24,301098	-2,097526	24,550886	-1,847739	4,528245	-0,408047	25,175965	-1,22266	4,012997	-0,304675
S.BOLIVAR	26,664476	22,374398	-4,290078	22,776367	-3,888109	4,415056	-0,880648	23,183667	-3,480809	3,863405	-0,900969
S.CORAZON	21,383629	23,086405	1,702776	23,677457	2,293827	4,58707	0,500064	25,347631	3,964002	4,175822	0,949275
FONTIBON	22,723106	21,200251	-1,522855	21,083573	-1,639533	4,739997	-0,345893	18,846477	-3,876629	4,475571	-0,866175
P.ARANDA	15,294742	22,306324	7,011582	22,053237	6,758495	4,442531	1,521316	21,992155	6,697413	4,03862	1,658342
CARVAJAL	13,861055	20,712668	6,851613	21,547598	7,686542	4,597965	1,671727	20,424766	6,563711	4,197017	1,563899
TUNAL	20,310127	17,774574	-2,535553	19,763264	-0,546863	4,565444	-0,119783	17,63311	-2,677016	4,128233	-0,648465
S.CRISTOBAL	24,639659	20,320404	-4,319255	20,812533	-3,827126	4,745624	-0,806454	20,470314	-4,169345	4,478087	-0,931055

Fuente: Autor Heyner Rodríguez

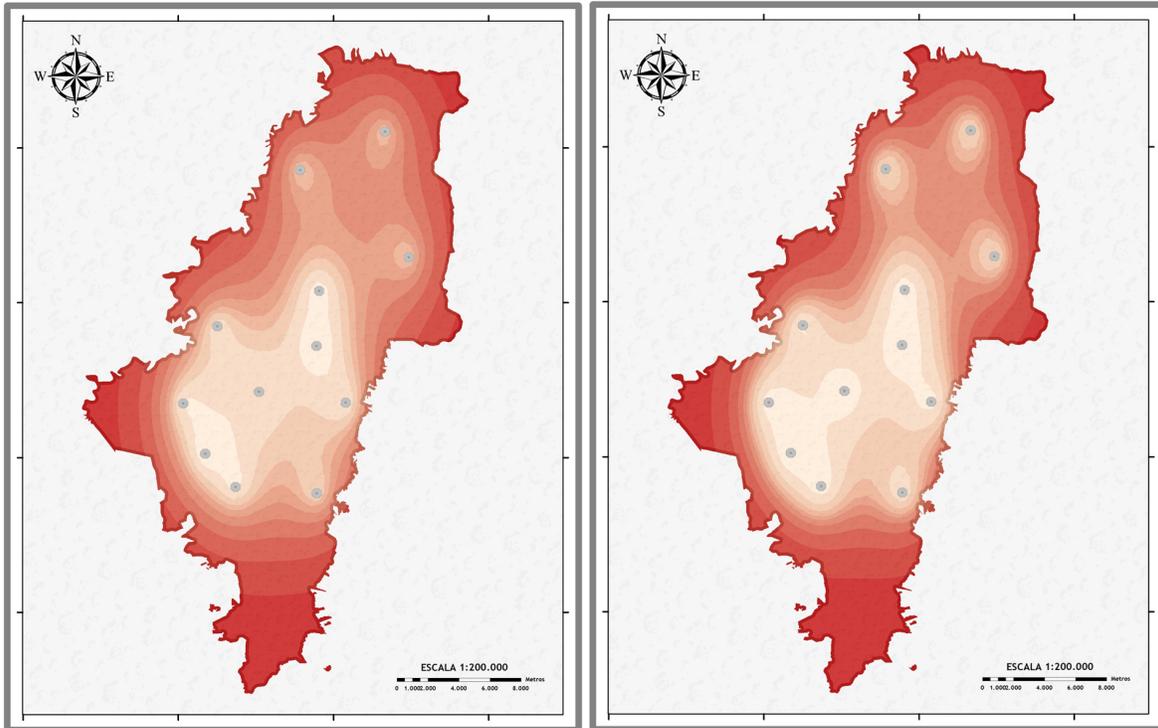
Realizando el mismo análisis para los mapas de predicción de la variable de Material particulado PM10. Encontramos al igual que en las concentraciones de Ozono un comportamiento similar de distribución entre los tres mapas, se observa mayor concentración en la zona Suroccidental y menores concentraciones en la zona Oriental, igual encontramos manchas que se visualizan en el mapa de predicción IDW, que no aparecen en los mapas de Kriging como un punto verde en la zona Suroriental de la ciudad. En este caso de concentración PM10 son aun menores la diferencias entre los dos mapas de predicción por Kriging, solamente existen pequeñas variaciones en las áreas de cada rango, mostrando áreas ligeramente mayores en el kriging ordinario.

Tomando que en las dos variables analizadas es comportamiento entre los diferentes métodos es similar, podemos concluir que los métodos de Kriging en mayor rango el Ordinario para lograr minimizar el error de predicción elimina los valores máximos y mínimos, esto se refleja en un suavizado generalizado en la estimación de los valores. A su vez el método de Distancia Inversa tiende a generar puntos focales o islas en los lugares de las muestras.

De igual forma se presentan también los mapas de predicción del error de las concentraciones de PM10 de Bogotá interpoladas por los métodos Kriging Ordinario y Kriging simple.



**Figura 7** Mapa de predicción para concentraciones de PM10 Usando IDW, Kriging Ordinario y Kriging Simple  
Fuente: Autor Heyner Rodríguez



**Figura 8** Mapa de predicción del error para concentraciones de PM10 por Kriging Ordinario y Kriging Simple  
Fuente: Autor Heyner Rodríguez

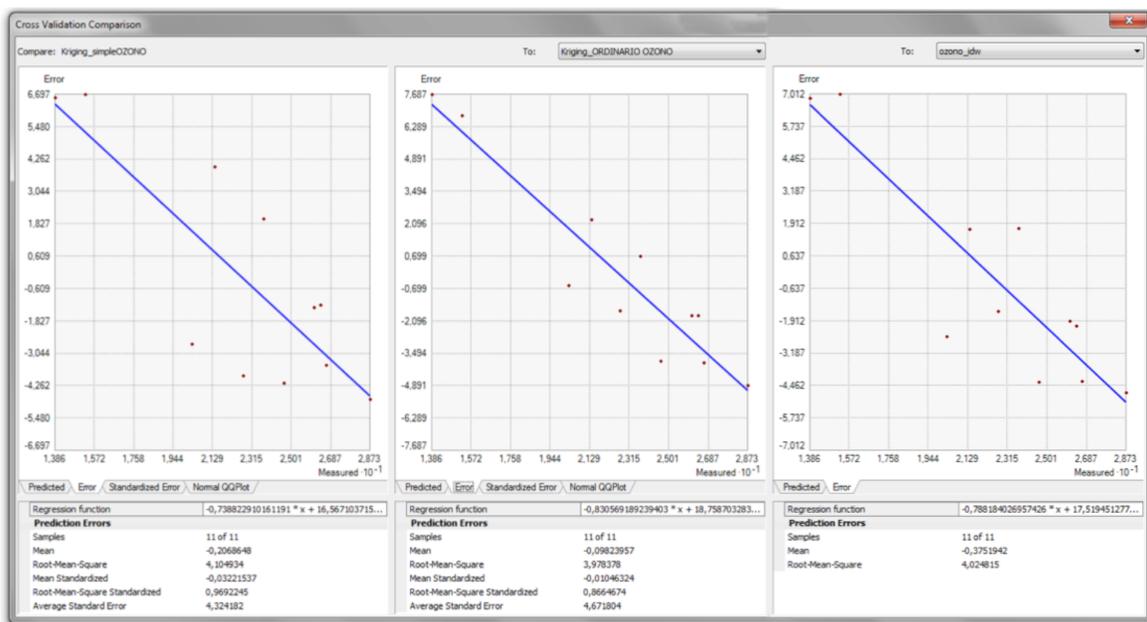
**Tabla 4** Resultados de estimación y error para concentraciones de PM10 por IDW, Kriging Ordinario y Kriging Simple

ESTACION	PM10	IDW		ORDINARIO				SIMPLE			
		Predicted	Error	Predicted	Error	St_Error	Stdd_Error	Predicted	Error	St_Error	Stdd_Error
GUAYMARAL	49,441271	45,879867	-3,561404	46,537657	-2,903614	16,196491	-0,179274	46,522232	-2,919039	15,435128	-0,189117
USAQUEN	36,475708	43,176295	6,700587	46,575435	10,099727	16,009498	0,630858	45,639982	9,164274	15,391313	0,595419
SUBA	54,115825	43,755202	-10,360623	43,307197	-10,80863	16,001458	-0,675478	44,218466	-9,897359	15,374595	-0,643748
FERIAS	33,294339	41,393005	8,098666	41,972311	8,677973	14,442945	0,600845	41,936776	8,642438	14,436872	0,598636
S.BOLIVAR	31,351105	43,385109	12,034005	41,408549	10,057445	14,082839	0,714163	41,447121	10,096016	14,130762	0,714471
S.CORAZON	43,787634	41,300992	-2,486642	37,422327	-6,365307	14,829999	-0,429218	38,069409	-5,718224	14,641041	-0,390561
FONTIBON	46,21016	49,454952	3,244792	53,463717	7,253557	15,086309	0,480804	52,45362	6,24346	14,846373	0,420538
P.ARANDA	46,467757	49,643122	3,175364	47,73858	1,270823	14,219651	0,089371	47,690592	1,222835	14,280487	0,08563
KENNEDY	70,244144	55,941353	-14,302791	55,46967	-14,77447	14,762416	-1,000817	55,412253	-14,83189	14,513181	-1,02196
CARVAJAL	80,106503	49,44939	-30,657113	50,956933	-29,14957	14,239483	-2,047095	51,246784	-28,85972	14,16401	-2,037539
TUNAL	42,246305	62,052249	19,805944	57,102877	14,856572	14,627486	1,015661	56,114517	13,868212	14,493024	0,956889
S.CRISTOBAL	33,482589	48,743058	15,260468	46,607829	13,12524	15,522538	0,84556	45,707036	12,224447	15,073289	0,811001

Fuente: Autor Heyner Rodríguez

## 2.6 VALIDAR Y COMPARAR LOS MODELOS

Una vez aplicados los diferentes métodos se procede a la realización de la comparación y la evaluación de las diferencias obtenidas, así como la pertinencia de los mismos en la estimación de las superficies, para esto podemos valernos de el método de Validación cruzada, que se basa en excluir uno de los puntos de la muestra y predecir el valor de la variable en este punto a partir de los demás puntos muestreados, buscando una variación mínima entre la predicción y el valor medido.



**Figura 9** Comparación de validación cruzada para concentraciones de Ozono entre Kriging Simple, Kriging Ordinario e IDW  
Fuente: Autor Heyner Rodríguez

La aplicación Arcgis posee la herramienta “Compare” que permite realizar la validación cruzada y comparar las predicciones que aparecen en las superficies generadas por los diferentes métodos y tomar una decisión de cual de ellos proporciona predicciones más exactas es decir que describe un comportamiento mas próximo a la realidad.

Algunos de los criterios determinísticos en la comparación y selección del método mas apropiado son las que se describen a continuación:

- Las predicciones deben realizarse de forma imparcial, indicado por un error medio cuadrático lo más cercano posible a cero (0).
- Error de predicción de la media cuadrática estandarizada cercana a uno (1).

- Las predicciones no deben alejarse mucho de los valores medidos, indicados por el error medio cuadrático y el error medio estándar, los cuales deben ser tan pequeños como sea posible.
- Se puede partir además de la comparación de los gráficos de predicción, error y QQPlot para comparar la calidad de la predicción.

En base a estos criterios evaluamos y tenemos que el método IDW solo proporciona información del error medio cuadrático y la media del error de la predicción, es así que obtenemos que el modelo Kriging Ordinario se comporta mejor, produce resultados más satisfactorios que los métodos Kriging Simple y Distancia Inversa IDW, generando una mejor superficie de predicción .

### **3. CONCLUSIONES**

Tomando en cuenta la muestra tomada para el análisis de las variables seleccionadas concentración de Ozono y Material Particulado PM10 en la zona urbana de Bogotá DC. Para el periodo de 1 de enero 2013 a 31 de diciembre del mismo año, tenemos que el método que mejor se comporta de los métodos utilizados es el Kriging Ordinario que proporciona resultados más satisfactorios, al tener errores de predicción más reducidos y un error medio cuadrático menor, como se pudo observar con la aplicación de la validación cruzada, además si aumenta el número de estaciones de monitoreo se producirían aún mejores resultados por este método.

Los métodos de Kriging Ordinario y simple, presentan comportamientos muy similares, como se ve en la tabla de errores y a través de los mapas de predicción, especialmente en la variable de concentración de PM10, presentando únicamente pequeñas variaciones en las áreas de cada rango, mostrando áreas ligeramente mayores en el kriging ordinario.

Los métodos de Kriging buscan minimizar el error de predicción para lo que eliminan los valores máximos y mínimos, representado suavizado generalizado en la estimación de los valores visible en los mapas de predicción. A su vez el método de Distancia Inversa tiende producir islas en los lugares de las muestras lo que puede ser útil para visualizar puntos críticos, en el caso de la calidad del aire,, los lugares puntuales en los que se sobrepasan los límites permisibles de concentración de un contaminante.

La exploración de datos es vital importancia dentro del proceso de interpolación, ya que permite una mejor comprensión del comportamiento de la información, al explorar los datos podemos identificar tendencias globales en los datos, examinar la distribución, influencias direccionales y localizar errores en datos o datos atípicos entre otros que puedan afectar la superficie de predicción.

En lo referente a la calidad del aire de un territorio pudo verse como la implementación de métodos geostadísticos es de gran validez en la representación de superficies espacialmente continuas del comportamiento de las diferentes concentraciones de contaminantes, lo que contribuye a establecer estrategias y tomar decisiones respecto a los problemas ambientales.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] R. Giraldo Henao, *Introducción a la Geoestadística Teoría y Aplicación*, Facultad de Ciencias Departamento de Estadística ed., Bogotá: Universidad Nacional de Colombia, 2002, p. 94.
- [2] Red de Monitoreo de Calidad del Aire de Bogotá, «Informe Anual de Calidad de Aire de Bogotá,» Secretaría Distrital de Ambiente, Bogotá, 2013.
- [3] Secretaría Distrital de Ambiente, «Red de Calidad del Aire,» 2014. [En línea]. Available: <http://ambientebogota.gov.co/red-de-calidad-del-aire>. [Último acceso: 2 Diciembre 2014].
- [4] Ministerio de Ambiente, Vivienda y Desarrollo Territorial, *Manual de Operación de Sistemas de Vigilancia de la Calidad del Aire Resolución 650 de 2010*, Bogotá: Ministerio de Ambiente, Vivienda y Desarrollo Territorial, 2010, p. 141.
- [5] J. W. Cely Pulido, W. L. Siabato Vaca , A. H. Sánchez Ipia y A. P. Rangel Sotter, «Geoestadística aplicada a estudios de contaminación ambiental,» *Facultad de Ingeniería de la Universidad Distrital Francisco José de Caldas. Bogotá*, vol. VII, nº 2, pp. 31-38, 2002.
- [6] C. C. Yrigoyen, *Métodos Gráficos del Análisis Exploratorio de Datos Espaciales*, Madrid: Universidad Autónoma de Madrid, 2002, p. 25.
- [7] J. A. Valero Fandiño, «Teorías Clásicas de Estimación Espacial y Temporal,» de *Tópicos de la geoestadística*, Bogotá, 2014, pp. 15 - 55.