

Comportamiento del material particulado 2.5 en el Valle de Aburrá y su relación con variables climáticas

Paulina Cardona Osorio ^a

^a Departamento de geociencias y medio ambiente, Especialización aprovechamiento de recursos hidráulicos, Universidad Nacional de Colombia, Medellín, Colombia. pcardona@unal.edu.co

Abstract

The Aburrá Valley has experienced various environmental contingency situations due to poor air quality, especially during the month of March. In this context, PM_{2.5} particulate matter is considered a key variable for its analysis, as it is influenced by multiple climatological factors. In this study, temperature and precipitation were considered as predictor variables to understand how they interact with the behavior of PM_{2.5}.

For this purpose, generalized linear models were applied, that adequately fit the data, allowing the exploration of the relationship between the variables. Additionally, models that consider heterogeneity and spatial dependence were employed in order to analyze the distribution of particulate matter in the Aburrá Valley. Finally, spatial statistics were used to visualize its behavior in the territory and facilitate the prediction of PM_{2.5} levels based on climatic variables and geographic location.

Keywords: Air quality; predictor variables; Aburrá Valley; spatial statistical models; particulate matter 2.5.

Resumen

El Valle de Aburrá ha experimentado diversas situaciones de contingencia ambiental debido a la mala calidad del aire, especialmente durante el mes de marzo. En este contexto, el material particulado PM_{2.5} se considera una variable clave para su análisis, ya que está influenciado por múltiples factores climatológicos. En este estudio, se consideraron la temperatura y la precipitación como variables predictoras para comprender cómo interactúan con el comportamiento del PM_{2.5}.

Para ello, se aplicaron modelos lineales generalizados que se ajustaron adecuadamente a los datos, permitiendo explorar la relación entre las variables. Adicionalmente, se emplearon modelos que consideran la heterogeneidad y la dependencia espacial con el fin de analizar la distribución del material particulado en el Valle de Aburrá. Finalmente, se utilizó estadística espacial para visualizar su comportamiento en el territorio y facilitar la predicción de los niveles de PM_{2.5} a partir de variables climáticas y de ubicación geográfica.

Palabras clave: Calidad del aire; variables predictoras; Valle de Aburrá; modelos estadística espacial; material particulado 2.5.

1 Introducción

El material particulado 2.5, son partículas pequeñas presentes en el aire que tienen un diámetro menor de 2.5 micrómetros, siendo una mezcla que puede incluir diversas sustancias, como lo son sustancias químicas orgánicas, polvo, hollín, entre otras cosas, que pueden provenir de la industria, el tráfico, quemas y demás actividades. Estas partículas, pueden llegar a traer problemas a la salud de las personas, puesto que por su tamaño pueden desplazarse dentro de los pulmones ocasionando enfermedades pulmonares o incluso cardíacas.[1] En Colombia se presenta una normativa de límites máximos permisibles dada por la Resolución 2254 del 2017, la cual fija máximos permisibles de manera anual y diaria de diversos tipos de contaminantes entre los cuales se encuentra el PM 2.5 con

un límite de 50 $\mu\text{g}/\text{m}^3$ por día, sin embargo, este valor es considerablemente más alto que los recomendables por la organización mundial de la salud (OMS) el cual es de 15 $\mu\text{g}/\text{m}^3$ en un lapso de unas 24 horas[2].

A medida que los municipios del Valle de Aburrá se han urbanizado, se ha presentado una preocupación por la mala calidad del aire, lo cual ha sido foco de estudios y monitoreos, por parte de investigadores y entes administrativos por medio de proyectos estratégicos como el Sistema de Alertas Tempranas de Medellín y el Valle de Aburrá (SIATA), permitiendo tener un seguimiento de diferentes variables climáticas como lo es el material particulado de 2.5 μm .

En particular, durante los meses de marzo, octubre y noviembre se ha observado un incremento en las concentraciones de este material particulado en las estaciones de monitoreo de calidad del aire del SIATA. Estas alzas están asociadas a las condiciones geográficas y meteorológicas propias del valle, lo cual resalta la importancia de comprender cómo influyen variables como la temperatura y la precipitación en la dinámica del material particulado [3]. Las concentraciones elevadas generan preocupación y motivan la implementación de acciones de mitigación, por lo que su estudio resulta fundamental para orientar decisiones informadas en materia de salud y planificación territorial.

En este trabajo se analiza el comportamiento del material particulado en el Valle de Aburrá, con énfasis en su relación con variables meteorológicas como la temperatura y la precipitación. El objetivo es comprender la variabilidad horaria del PM_{2.5} durante el mes de marzo de 2022, mediante la aplicación de modelos estadísticos y espaciales. Entre las metodologías empleadas se incluyen: modelos lineales generalizados, métodos de interpolación, y modelos espaciales como la regresión geográficamente ponderada multivariada (MGR) y modelos de dependencia espacial (SAR y SEM). También se incorpora un modelo gaussiano multivariado, con el fin de evaluar si las variables climáticas seleccionadas permiten explicar y predecir el comportamiento del PM_{2.5} en el territorio.

2 Objetivos

2.1 Objetivo general

Entender el comportamiento del material particulado en el Valle de Aburrá y su relación con temperatura y precipitación.

2.2 Objetivos específicos

Utilizar el modelo general linealizado y modelo jerárquico que más se acople para entender el relacionamiento del material particulado en el Valle de Aburrá y su relación con variables climáticas.

Utilizar modelos de dependencia espacial y heterogeneidad espacial para entender patrones en el material particulado.

Utilizar modelos que permitan visualizar e incluso predecir por medio de la precipitación, temperatura y ubicación el comportamiento del material particulado.

3 Materiales y métodos

Para el desarrollo de la presente investigación se utilizó la información disponible en la plataforma de descarga de datos del SIATA, correspondiente a estaciones de precipitación, temperatura y material particulado PM_{2.5}. Se logró recopilar información de 15 estaciones de calidad del aire ubicadas en municipios del Valle de Aburrá. Sin embargo, no fue posible incluir la estación localizada en

Girardota debido a la ausencia de datos para el periodo de análisis. Se procuró que las estaciones climatológicas seleccionadas (precipitación y temperatura) estuvieran ubicadas lo más cerca posible de las estaciones de monitoreo de material particulado. Finalmente, se obtuvieron datos horarios correspondientes al mes de marzo, los cuales fueron validados para asegurar su confiabilidad, siguiendo las recomendaciones y criterios establecidos por el SIATA en su documentación técnica para cada una de las variables meteorológicas.

Luego de obtener los datos horarios se realizó un promedio de dichos datos por estaciones para así empezar a analizar la variabilidad horaria entre estaciones e identificar cambios significativos en los datos. Además, se realizó un análisis de probabilidad de excedencia por municipio, el cual permitiera conocer siguiendo los límites máximos permisibles de los lineamientos de la OMS la probabilidad de excedencia por municipio. Todos estos cálculos fueron realizados mediante la herramienta de código abierto Jupyter Lab con el lenguaje de programación Python en diferentes versiones.

4 Metodología

A continuación, se explicarán los tipos de modelos utilizados para el presente estudio:

Modelos lineales generalizados: Son modelos que permiten modelar una variedad amplia de tipo de datos, incluyendo los que no tienen una distribución normal. Para el caso de estudio se utilizó el modelo lineal mixto y el modelo logístico binomial. Los cuales permitieron conocer la probabilidad de excedencia del límite máximo permisible recomendado por la OMS. A continuación, se explicará más a detalle el modelo utilizado.

Modelo de regresión logística con familia binomial:

Es útil para predecir la probabilidad de que ocurra un evento binario. En el caso de estudio que se exceda el límite máximo permisible por la OMS. Este usa un modelo Logit para mapear la probabilidad (0 o 1).

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = B_0 + B_1 \text{temp} + B_2 \text{pp} + v \text{cat} \quad (1) \quad (5)$$

De la familia binomial, lo cual asume que la probabilidad de exceso siga una distribución de Bernoulli. El modelo se basó en variables predictoras como lo es la temperatura(temp), la precipitación(pp) y un efecto fijo categórico que indica el Código de ubicación de las estaciones de análisis (v cat). Luego se obtuvieron coeficientes de las variables predictoras que indican la relación que tiene estas con el material particulado. Además de la significancia estadística y cuánto explica dicho modelo el comportamiento del material particulado.

Modelo jerárquico:

Este tipo de modelo también se denominan multinivel o mixtos y permiten modelar datos que no son independientes. El uso de estos modelos está recomendada para datos que suelen tener una organización jerárquica, en este caso, es por esto que resulta lógico para el análisis planteado puesto que están organizados por estaciones y de manera horaria.

El modelo utilizado tiene la siguiente forma:

$$\log(\text{PM}_{2.5ij}) = B_0 + B_1 \text{temp}_{ij}^2 + B_2 \text{pp}_{ij} + B_3 \cos(\text{hora}_{ij}) + B_4 \sin^2(\text{hora}_{ij}) + u_{0j} + \varepsilon_{ij} \quad (2)$$

Donde, i, j son el índice de observación y el índice de estación, B_0 es el intercepto global y $B_{1 a 4}$ son los coeficientes fijos de las variables, los senos y cosenos de las horas necesarios para capturar los ciclos diarios completos.

Se utilizó logaritmo para que el material particulado tuviera una distribución normal, ya que el material particulado tenía una distribución sesgada hacia un lado. De igual manera, la temperatura y el seno se modificaron para que fuera mejor al capturar relaciones no lineales, lo cual mejoró notablemente el modelo sin modificaciones.

Modelos de heterogeneidad espacial: Está relacionado con la diferencia espacial o regional de las unidades geográficas, esto implica que, en los modelos espaciales, las formas funcionales y los parámetros variarán con la con la localización geográfica, no siendo homogénea para toda la matriz de datos. El método utilizado fue **el modelo de regresión geográficamente ponderada multiescala (MGWR)**, el cual permite capturar la heterogeneidad espacial. En dicho modelo, cada variable tiene su propio ancho de banda, permitiendo que algunas relaciones sean más locales, puesto que este modelo optimiza un ancho de banda para cada variable. [4]

$$y_i = B_0(u_i, v_i) + \sum_{k=1}^p B_k(u_i, v_i, h_k) * x_{ik} + \varepsilon_i \quad (3)$$

Dónde y_i es la variable dependiente en la ubicación i , x_{ik} De la variable independiente k en la ubicación i , $B_k(u_i, v_i, h_k)$ coeficiente de la variable k en la ubicación (u_i, v_i) con su propio ancho de banda h_k y ε_i es el término de error. [5]

Modelos de dependencia espacial: Estos modelos se consideraron debido a la naturaleza de la dispersión de contaminantes, lo cual puede generar dependencia espacial incluso en estaciones de monitoreo relativamente alejadas.

Para esto se utilizó **el índice de moran** para observar efectos espaciales, ayudando a obtener el gráfico de dispersión que ilustra la relación entre cada observación y el promedio de sus vecinos, ayudando a identificar conglomerados locales y valores atípicos espaciales. El estadístico I de Moran es la pendiente del ajuste lineal al gráfico de dispersión, su fórmula es la siguiente:

$$MI = n \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

Donde n es el número de observaciones, y_i es el valor de la variable respuesta en la ubicación i . Este índice varía entre -1 y 1, donde cero indica ausencia de autocorrelación espacial, +1 agrupación perfecta y -1 dispersión. [6]

Antes de aplicar los demás modelos, se aplicó un modelo de mínimos cuadrados ordinarios para verificar en los residuos qué conjunto de datos presentaban posible dependencia, es así que se utilizaron los siguientes modelos:

Modelo de rezago espacial (SAR): Asume que la variable relacionada al material particulado depende de los valores en puntos vecinos.

Modelo de error espacial (SEM): Se modela el término de error con autocorrelación espacial, lo cual permite conocer si pueden haber otras variables que permitan modelar en ese conjunto de datos que permitan explicar el comportamiento del PM 2.5.

Modelos de interpolación: Se utilizaron métodos de interpolación como **el inverso de la distancia ponderadas**, el cual se describe como un método determinístico que asigna pesos a puntos conocidos en función de su distancia al punto a interpolar, no requiere distribución estadística de los datos y asume que la influencia decae con la distancia. Su fórmula matemática es la siguiente:

$$z(x) = \sum_{i=1}^n w_i z(x_i) \quad (6)$$

En donde $z(x)$ es el valor interpolado en el punto x , $z(x_i)$ es el valor observado en el punto x_i , n es el número de observaciones para interpolar y w_i es el factor de ponderación [7].

Para realizar el método de interpolación Kriging, se empleó una regresión polinómica con el fin de captar relaciones no lineales entre PM 2.5 y los predictores climáticos, especialmente la temperatura, cuya relación no lineal con el material particulado se analizará más adelante. Además, se aplicó Kriging ordinario por hora para interpolar las variables de temperatura y precipitación, generando superficies continuas en forma de grilla para cada una.

Posteriormente, se calcularon los residuos del modelo polinómico, es decir, las diferencias entre los valores observados y los estimados por la regresión. Estos residuos se interpolaron mediante Kriging, lo cual permitió incorporar las variaciones locales no explicadas por la regresión.

Finalmente, la predicción de PM2.5 se obtuvo aplicando el modelo polinómico a la grilla interpolada de temperatura y precipitación, y luego sumando los residuos interpolados. Este enfoque mejora la precisión de la estimación, ya que el kriging de residuos permite corregir errores locales,

añadiendo componentes espaciales que la regresión no logra capturar por sí sola.

Por último, se utilizó un modelo **gaussiano multivariado**, la cual es una técnica de regresión no paramétrica utilizada en aprendizaje automático y estadística, es útil cuando se involucran datos continuos, y la relación entre las variables de entrada y salida no se conoce explícitamente. En el caso de estudio, las variables están correlacionadas entre sí y múltiples variables sirven para explicar el comportamiento del pm 2.5

$$f\left(\begin{matrix} \rightarrow \\ x \\ \mu \end{matrix} \rightarrow \Sigma\right) = \frac{1}{\sqrt{(2\pi^d)(\det(\Sigma))}} \exp\left(-\frac{1}{2}\left(\begin{matrix} \rightarrow \\ x \\ \mu \end{matrix} - \begin{matrix} \rightarrow \\ x \\ \mu \end{matrix}\right)^T \Sigma^{-1} \left(\begin{matrix} \rightarrow \\ x \\ \mu \end{matrix} - \begin{matrix} \rightarrow \\ x \\ \mu \end{matrix}\right)\right)$$

(6)

Donde Σ^{-1} es la matriz de covarianza y d es el número de vectores o dimensiones.[6]

5 Resultados y discusión

En el presente literal se discutirán los resultados obtenidos producto de utilizar las metodologías explicadas.

Inicialmente, se analizó la probabilidad de excedencia al límite permitido en un 24 por la OMS el cual es $15 \mu\text{g}/\text{m}^3$ para las 15 estaciones, posteriormente se realizó un mapa, el cual mostrara por municipio la probabilidad de excedencia a superar este límite recomendado.

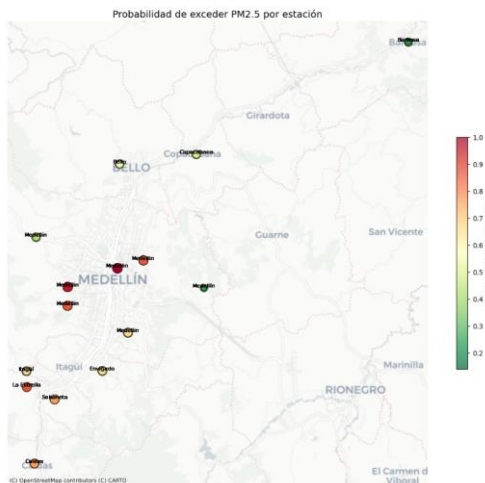


Figura 1: Probabilidad de excedencia PM 2.5 por estación según límites OMS

Fuente: Elaboración propia

En la figura, se observa que el municipio de Medellín concentra la mayor cantidad de estaciones de monitoreo, las cuales presentan una alta variabilidad en la probabilidad de excedencia. Esta variabilidad se refleja en un promedio ponderado no tan elevado para el municipio, principalmente porque algunas de sus estaciones, ubicadas en zonas rurales o periféricas como Santa Elena y San Cristóbal, registran valores bajos de excedencia. Por su parte, en municipios como Copacabana y Barbosa también se observan valores

bajos, mientras que en el sur del valle de Aburrá, se presentan probabilidades medias a altas, especialmente en el municipio de La Estrella el cual presenta el mayor valor de las estaciones en el sur.

Los resultados muestran que en el mes de marzo del año de análisis hubo estaciones que excedieron el 100% de las veces el límite permisible por la OMS, siendo esto un panorama preocupante.

Ahora, teniendo esta información en mente y en miras de responder adecuadamente el objetivo del presente trabajo, se realizó el modelo lineal mixto, el cual busca explicar la relación entre el material particulado y variables climáticas de manera horaria, por lo cual se pudo observar que en especial la temperatura y el sin(hora) presentan una relación cuadrática con el material particulado y del cual se obtuvieron resultados que representan la naturaleza de este tipo de fenómenos.

Dicho modelo encontró la relación entre el log(pm2.5) y las variables climáticas que se presentan en la siguiente tabla.

Tabla
Resultados de modelo lineal mixto.

1.

Variable	Coefficiente	Error estándar	p- valor	Intervalo de confianza
<i>Temperatura</i> ²	-0.002	0	0	-0.002 a -0.001
Precipitación	-0.063	0.019	0.001	-0.099 a -0.026
Cos(hora)	-0.215	0.014	0	-0.242 a -0.188
<i>sen</i> ² (hora)	-0.366	0.019	0	-0.403 a -0.329
Varianza entre grupos	0.093	0.308	-	
intercepto	3.629	0.089	0	3.454 a 3.803

Fuente: Elaboración propia.

En donde las variables de seno y coseno de la hora representan la variación cíclica diaria, y la precipitación tiene un efecto de lavado sobre el material particulado, mientras que la temperatura presenta una relación negativa cuadrática, esto quiere decir que presenta un punto de inflexión, en donde al llegar a dicho punto, la relación con el material particulado cambia. Para verificar esto se exponen las gráficas a continuación:

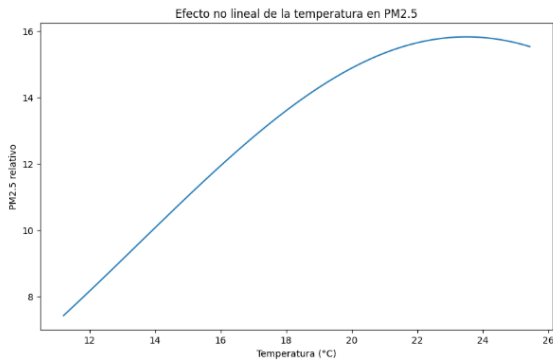


Figura 2: Relación cuadrática negativa de la temperatura con el material particulado.

Fuente: Elaboración propia.

En la Figura 2 se puede observar que se presenta un punto de inflexión a 23.5 °C, en donde antes de este punto, al incrementar la temperatura incrementa el material particulado, pero al pasar este límite debido a una mayor estabilidad atmosférica o mezcla vertical por fenómenos convectivos hay un escape de dichos contaminantes a las capas más altas de la atmósfera.

Asimismo, el seno y el coseno muestran cómo es el comportamiento del material particulado según la hora del día, este presenta mayores concentraciones al medio día, y luego vuelven a incrementar en las horas de la noche, estas concentraciones altas se deben a que hay una estabilidad atmosférica, este estado favorece la acumulación de partículas en zonas cercanas a la superficie porque reduce la dispersión vertical de contaminantes, esto si no hay precipitación significativa que favorezca su lavado.[8]

Teniendo en cuenta todo lo explicado anteriormente, a continuación, se observa cómo se acopla dicho modelo a lo observado en las estaciones del Valle de Aburrá.

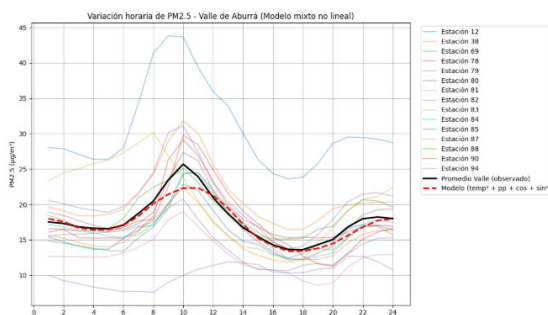


Figura 3 : Variación horaria de PM2.5 según modelo mixto no lineal.

Fuente: Elaboración propia

Además, dicho modelo presentó un R^2 de 0.85, lo cual sugiere que explica de una manera adecuada la variabilidad de los valores de PM 2.5, además de esto el RMSE y MAE presentaron valores pequeños esto quiere decir que no hay muchos errores extremos.

Para visualizar el comportamiento del PM 2.5 según dicho modelo, se realizaron mapas de 1 a 24 horas en los cuales

por municipio se muestran los valores predichos de PM 2.5, donde los máximos se alcanzan a las 10 y 11 de la mañana. Con los municipios más contaminados siendo La Estrella, Medellín y Caldas.

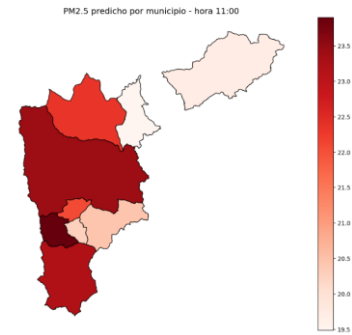
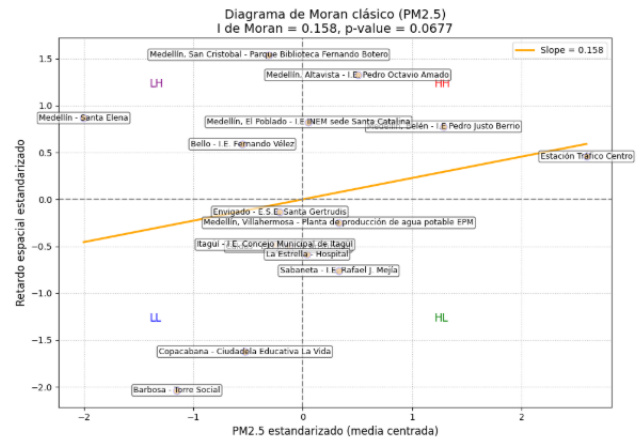


Figura 4: Ejemplo de hora 11 con el modelo lineal mixto aplicado.

Fuente: Elaboración propia.

Estos gráficos se pueden observar en el portal de GitHub, donde está presente el script con el que se realizó el presente trabajo.

Luego de obtener dichos modelos, se procedió a analizar la dependencia espacial por medio del índice de moran inicialmente, para esto se calculó la matriz de pesos espaciales basada en vecinos más cercanos, con un número de vecinos de 3, luego se obtuvo el índice de Moran con un resultado de 0.158 y nivel de significancia de 0.068 para el promedio por estación para dicho mes de marzo, por lo cual aunque exhibe correlación espacial positiva, o sea dependencia espacial, presenta una significancia estadística.



Sin embargo, en este gráfico se puede observar que las zonas con mayor cantidad de material particulado se encuentran en Medellín y están rodeadas de otras con altos niveles de contaminación, y como se mencionó anteriormente las de menor contaminación se encuentran más alejadas y en zonas rurales.

A pesar de obtener dichos resultados, se realizó el mismo análisis de manera horaria y se aplicó el moran a los residuos del modelo OLS, en donde se observó que si bien no se presenta dependencia espacial en todas las horas para las estaciones de análisis, sí se presenta en algunas, por lo

cual se decidió utilizar modelos SAR Y SEM para determinar la dependencia que presentan las estaciones de medición en algunas horas del día.

En las horas e las cuales se presento un índice de moran estadísticamente significativo fueron para las 1, 2, 5, 6, 15, 21, 22 y 24.

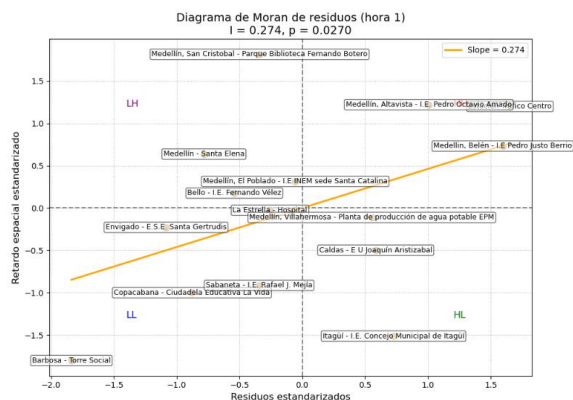


Figura 5: Diagrama de Moran para residuos de OLS (hora 1)

Fuente: Elaboración propia

Es así, que al obtener los resultados expuestos anteriormente, se procede a realizar modelo SAR y SEM que permitan conocer mejor la dependencia espacial del modelo.

En el modelo SAR se incorpora la autocorrelación espacial de la variable dependiente, es decir, se tiene en cuenta la influencia que pueden ejercer unas estaciones de calidad del aire sobre otras. A través del coeficiente de correlación espacial, es posible identificar si, en determinadas horas, algunas estaciones presentan comportamientos similares o si sus niveles de PM2.5 dependen en parte de las estaciones vecinas. En la gráfica se muestran los valores estimados de este coeficiente, lo que permite visualizar la intensidad y variación de dicha dependencia a lo largo del tiempo.

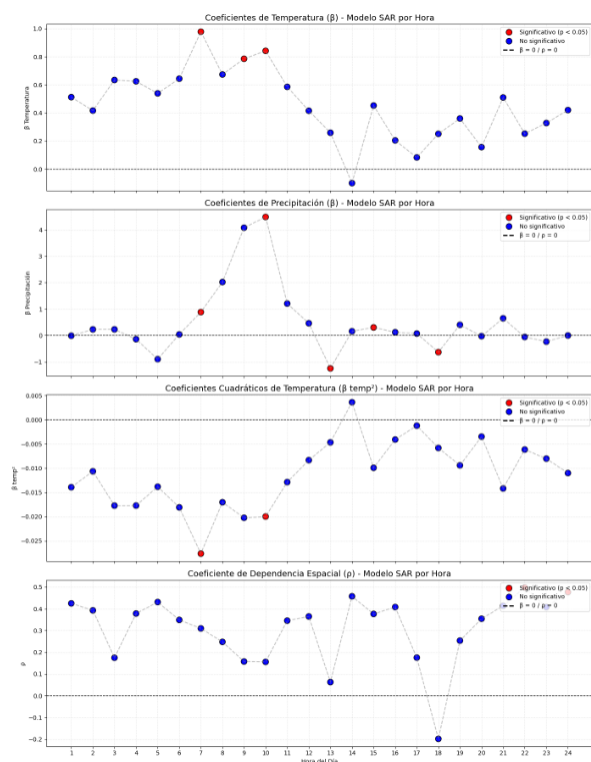


Figura 6: Resultado modelos SAR.

Fuente: Elaboración propia.

Esto muestra una dependencia espacial baja, en especial en las horas de la noche a las 22 y 24 horas.

Asimismo, se estimaron los coeficientes para las variables de precipitación y temperatura. En el caso de la temperatura, se encontraron valores estadísticamente significativos tanto para el término lineal como para el cuadrático durante las horas de la mañana. En el modelo sin el término cuadrático, la temperatura mostró una correlación positiva a las 7, 9 y 10 a.m.; sin embargo, al incluir el término cuadrático, se observaron correlaciones negativas a las 7 y 10 a.m. Esto sugiere una relación no lineal, en la que el aumento de la temperatura inicialmente eleva la concentración de material particulado, pero al superar cierto umbral, esta concentración comienza a disminuir.

En cuanto a la precipitación, se evidencian variaciones en su relación con el material particulado. Esto podría explicarse porque, en algunos casos, la precipitación estabiliza tempranamente la atmósfera y favorece la formación de una inversión térmica, lo cual impide la dispersión de contaminantes y genera una correlación positiva. Por el contrario, cuando ocurre el lavado atmosférico, la relación se torna negativa.

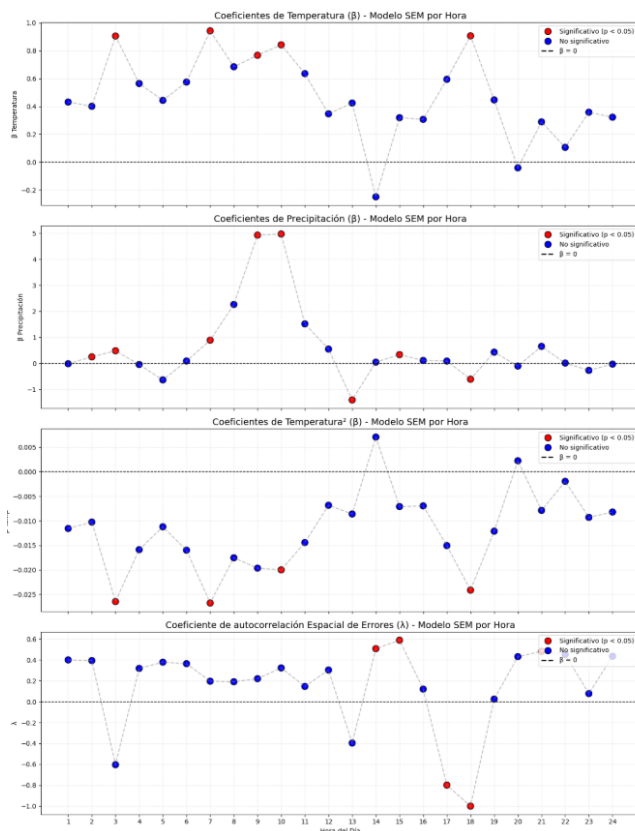


Figura 7: Resultados modelos SEM.

Fuente : Elaboración propia.

Para el modelo SEM, se observa que en varias horas el coeficiente λ , el cual mide la autocorrelación espacial de los errores, toma valores tanto positivos como negativos. Un valor positivo de λ indica que los errores están espacialmente correlacionados de forma positiva, es decir, tienden a agruparse; mientras que un valor negativo sugiere que los errores están correlacionados de forma negativa, es decir, se distribuyen de manera alternante en el espacio. En varias horas, λ presenta correlaciones espaciales positivas estadísticamente significativas, lo cual sugiere que existen variables relevantes no incluidas en el modelo que están espacialmente correlacionadas.

Sin embargo, también se identificaron casos en los que el coeficiente λ toma valores negativos, lo cual indica una autocorrelación espacial negativa en los errores del modelo. En estos casos, los errores tienden a alternarse espacialmente: cuando el modelo sobreestima en una zona, tiende a subestimar en las zonas vecinas, y viceversa. Este comportamiento puede reflejar una dinámica espacial más compleja del fenómeno, asociada a efectos locales no capturados. Por ejemplo, en algunas zonas la precipitación puede favorecer el lavado atmosférico y disminuir la concentración de material particulado, mientras que en zonas cercanas sin lluvia, las condiciones pueden favorecer su acumulación. Estas diferencias pueden generar patrones espaciales opuestos en los errores, especialmente si la matriz de pesos espaciales utilizada no representa

adecuadamente la estructura geográfica y climática del área de estudio.

Respecto a los coeficientes, se identificaron tanto correlaciones positivas como negativas en relación con la temperatura y la precipitación, lo que indica relaciones no lineales y efectos heterogéneos. Estos resultados evidencian que, para modelar de manera más precisa la concentración de material particulado, sería necesario incorporar más variables tanto climáticas como antrópicas. No obstante, en el presente trabajo se decidió enfocar el análisis únicamente en la relación entre precipitación y temperatura.

Es así, que para entender el efecto diferenciado que pueden tener las dos variables climatológicas de estudio sobre el PM 2.5, se realizó un modelo de heterogeneidad espacial, el cual se denomina el modelo de regresión geográficamente ponderada multiescala, este tiene como ventaja permitir que cada variable explicativa tenga su propio ancho de banda espacial, o sea su propio grado de variación espacial, de esta manera, se prepararon los datos en donde se especificaron las variables independientes como las variables climáticas y la dependiente el pm 2.5, este proceso se realizó de manera horaria, ósea un modelo por las 24 horas del día de las 15 estaciones de análisis con lo cual se obtuvieron los siguientes resultados por hora en cuanto al R^2 .

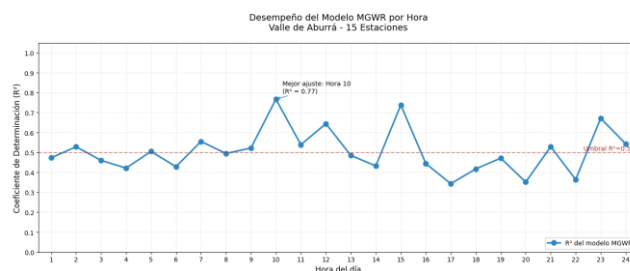


Figura 8 : Desempeño modelo MGWR.

Fuente: Elaboración propia

De igual manera, se realizaron 24 figuras para observar el comportamiento y los valores predichos por horas según el modelo MGWR.

En la Figura 9 se observa los valores predichos para las estaciones y su promedio por municipios en las 24 horas del día.

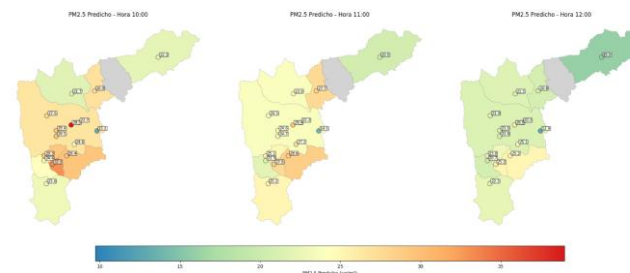


Figura 9 : Mapas de MGWR.

Fuente: Elaboración propia.

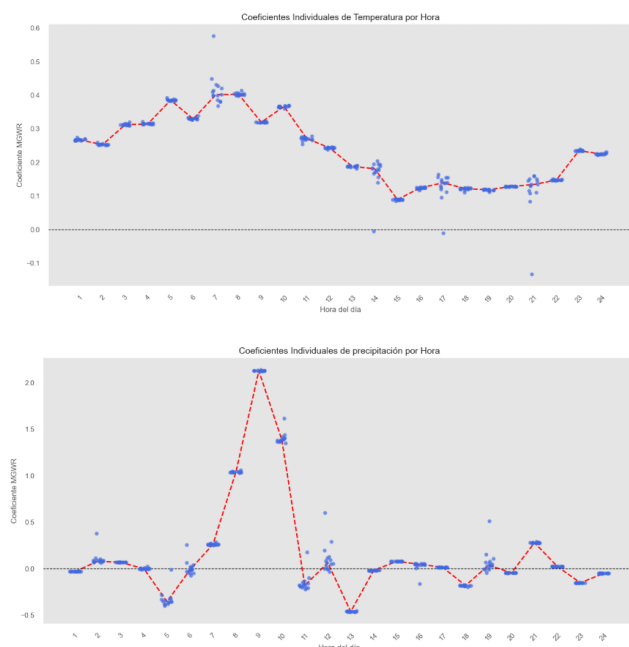


Figura 10: Coeficientes de precipitación y temperatura MGWR.
Fuente: Elaboración propia.

Los gráficos muestran, puntos azules que son los coeficientes individuales del modelo por hora para cada una de las variables climáticas y una línea roja la cual “promedia” dichos valores el cual se asigna al resultado del modelo MGWR por hora. Los coeficientes para este modelo presentan valores negativos y positivos para las dos variables climáticas, sin embargo, si bien para la temperatura se muestran valores negativos, no son tan significativos en su peso. Por otro lado, para la precipitación se puede observar un posible lavado del material particulado en algunas horas del día por sus coeficientes negativos, además para la hora 9 se observa valor positivo significativo, lo cual indica que a mayor precipitación en esa hora en específico mayor material particulado. Este resultado es similar al encontrado en el modelo SAR propuesto, en donde el comportamiento de la precipitación se asemeja mucho en ambos modelos.

En cuanto a los anchos de banda, se observa que tanto el intercepto como la precipitación presentan una mayor variabilidad espacial en comparación con la temperatura, la cual muestra un comportamiento más estable. La precipitación exhibe un efecto regional en algunas zonas, mientras que en otras actúa de manera más localizada, lo cual es coherente con su naturaleza estocástica. Por su parte, el intercepto refleja fuertes efectos locales, posiblemente asociados a la ubicación de estaciones con mayores niveles de contaminación, aunque en ciertas horas también manifiesta un patrón más regional, posiblemente influenciado por condiciones meteorológicas.

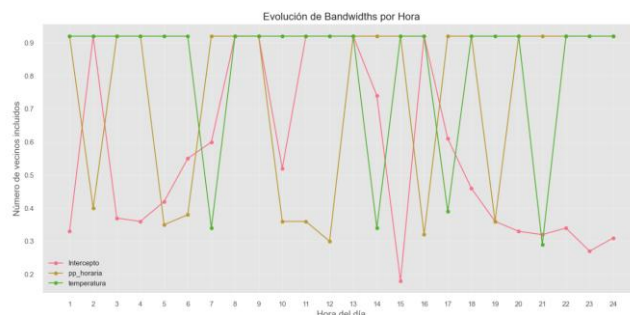


Figura 11: Anchos de banda de variables modelo MGWR.
Fuente: Elaboración propia.

Finalmente, para observar el comportamiento del material particulado de una manera más continua y real se utilizó métodos de interpolación y modelos gaussianos multivariados.

Primeramente, se analizó el IDW en donde se utilizaron los valores observados de estaciones de material particulado para cada hora, obteniendo así 24 mapas que muestran continuamente el comportamiento en el Valle de Aburrá del material particulado, precipitación y temperatura usando únicamente las estaciones de análisis o sea los datos observados. Esto se utilizó para comparar con los resultados obtenidos con los demás métodos a utilizar las cuales usaron las variables climáticas predictoras.

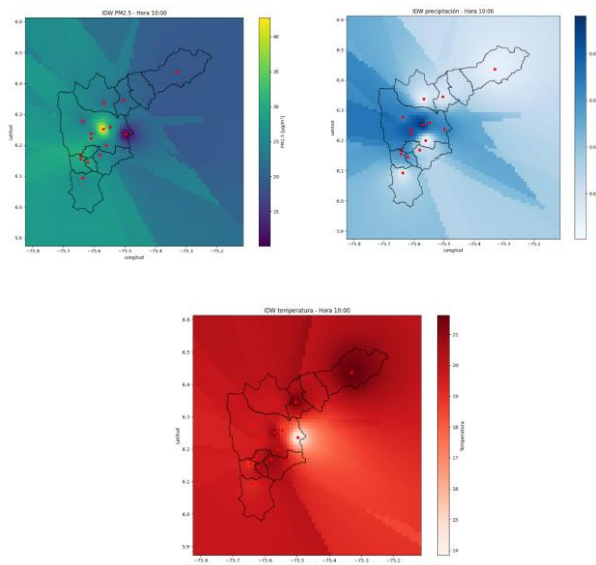


Figura 12: IDW de precipitación, temperatura y material particulado.
Fuente: Elaboración propia.

De igual manera, se realizó un diagrama de Voronoi para analizar el área de influencia de las estaciones de análisis con respecto al Valle de Aburrá obteniendo lo que se observa en la siguiente figura.

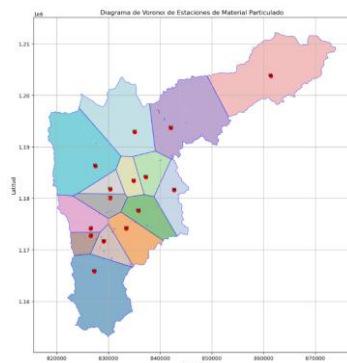


Figura 13: Diagramas de Voronoi estaciones de medición PM 2.5.

De esta manera, se procedió con los demás análisis para intentar obtener un mapa continuo tipo ráster que refleje el comportamiento del material particulado, teniendo en cuenta su relacionamiento también con las variables climáticas del estudio.

Se llevó a cabo la interpolación espacial horaria de PM2.5 mediante una combinación de regresión polinómica y Kriging de residuos. En primer lugar, se construyó una malla regular de interpolación sobre la región de estudio. Posteriormente, se ajustó un modelo de regresión polinómica, seleccionado por ofrecer un mejor desempeño en comparación con otros enfoques y por su capacidad para capturar relaciones no lineales observadas previamente en los datos. En particular, variables como la temperatura mostraron comportamientos no lineales respecto a PM2.5, evidenciando puntos de inflexión y cambios en la concavidad de la relación.

Una vez ajustado el modelo polinómico, se calcularon los residuos, los cuales representan la componente espacial no explicada por la tendencia general del modelo. Estos residuos contienen información local relevante que la regresión por sí sola no logra capturar. Paralelamente, se interpolaron espacialmente las variables predictoras (temperatura y precipitación) sobre la malla utilizando Kriging ordinario.

Finalmente, se aplicó Kriging a los residuos para interpolarlos sobre la misma malla y estos se sumaron a la predicción del modelo polinómico. Este proceso de corrección permite afinar la estimación espacial de PM2.5, al integrar tanto la tendencia general como las variaciones locales, mejorando así la precisión y realismo de la predicción final.

Si bien si se observaron valores creíbles, se obtuvo en las horas 13 y 23 valores alejados de la espacialidad del comportamiento del PM 2.5, si bien en la hora 13 se observa un aumento en la concentración del material particulado por el escape a zonas más altas de la atmósfera de este contaminante, los valores pueden ser altos para dicha hora, además de estar estas zonas rodeadas de valores bajos. Es así, que para las demás horas se observaron valores un poco

más constantes alrededor del Valle de Aburrá, como se observa en la figura.

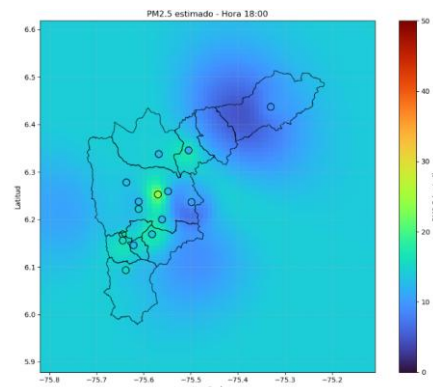


Figura 14: Kriging PM 2.5.

Fuente: Elaboración propia.

Por último, se utilizó un modelo Gaussiano multivariado, el cual es un modelo probabilístico no paramétrico, el cual asume que los datos siguen una distribución normal. Para esto se incluyeron cuatro variables predictoras las cuales fueron la precipitación, temperatura y la latitud y longitud para que el modelo pudiera aprender que estaciones cercanas a estas tendrían PM 2.5 similares, luego se definió un kernel compuesto, el cual tuviera en cuenta un factor constante, RBF el cual es un núcleo gaussiano radial y White kernel para modelar el ruido o la variabilidad que no puede ser explicada; es así que este kernel le dice al modelo que los valores de PM 2.5 deben cambiar suavemente según la temperatura, precipitación y ubicación.

Luego se escalaron las variables predictoras y el valor objetivo, y define las matrices de variables predictoras y objetivo, para luego entrenar el modelo aprendiendo cómo se comporta el PM 2.5 en función del clima y la ubicación para cada una de las horas.

Es así, que se generaron gráficos de valores predichos por el modelo de PM 2.5 y valores observados para las 24 horas., estos valores se asemejan mucho entre sí en comportamiento y valor de concentración de material particulado como se observa en la siguiente Figura 15. También se realizaron las matrices de covarianza para cada hora (ver Figura 16), acá se puede observar que para las horas 10, 12, 15 y 21 hay patrones más estructurados, esto quiere decir que muchas de las estaciones están relacionadas entre sí, o sea que hay un patrón regional de contaminación, este hallazgo conversa un poco con lo encontrado previamente en los demás análisis realizados, mientras que para otras matrices de covarianza casi vacías fuera de la diagonal, exhiben fenómenos tal vez más locales.

De igual manera, hay algunas estaciones que según el modelo y las matrices de covarianza muestra una gran relación entre los valores de PM 2.5, como por ejemplo la estación 83 y 79, esto tiene sentido debido a que están cercanas.

Por último, se realizaron mapas en formato ráster (ver Figura 17) para mostrar el comportamiento del material particulado según este modelo para las 24 horas.

Sin embargo, para la hora 11 los resultados no fueron satisfactorios, lo cual también puede observarse en la matriz de covarianza correspondiente a dicha hora.

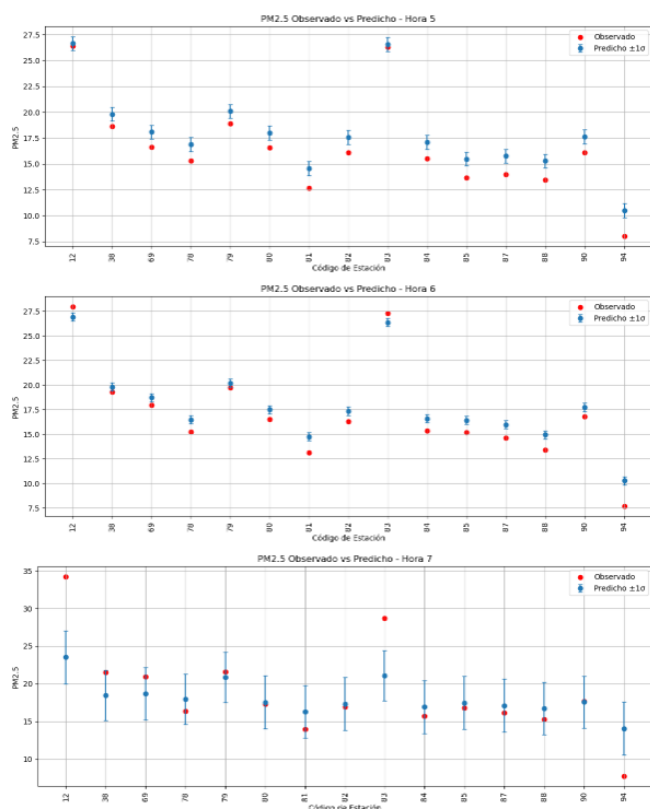


Figura 15: Valores predichos y observados modelo Gaussiano multivariado.

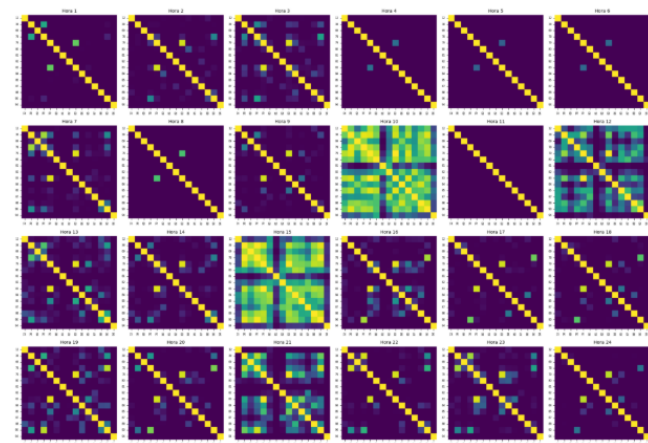


Figura 16: Matrices de covarianza modelo gaussiano multivariado.
Fuente: Elaboración propia.

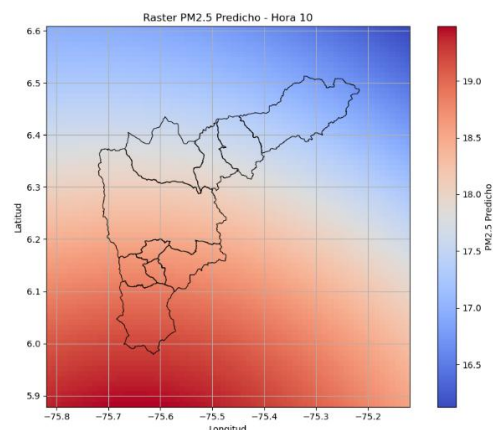


Figura 17: Mapa ráster valores predichos de modelo gaussiano multivariado (hora 10).

Fuente: Elaboración propia.

Finalmente, para observar si el modelo puede predecir bien, se descargaron datos aleatoriamente dentro del polígono del Valle de Aburrá por medio de la herramienta de Google earth engine, en donde se utilizó información satelital a escala temporal horaria de ERA5- land en formato de punto (para cada punto había un valor de ubicación y de variables climatológicas), de esta manera se obtuvo datos de entrada para observar como el modelo gaussiano permitía conocer el PM 2.5 dependiendo de la temperatura, precipitación y ubicación.

Si bien se presentan en su mayoría valores cercanos a la realidad, en otros casos para algunas horas como lo es la 1, 9, 13 y en pequeñas proporciones la 16 y 14 se presenta una subestimación de los valores.

De igual manera, es importante observar que los valores por horas predichos, presentaron en su mayoría valores sin variabilidad a diferencia de los datos reales, sin embargo algunos como por ejemplo la Figura 19 presento gran variabilidad.

Es fundamental, para llevar a cabo un análisis riguroso con nuevos puntos, contar con datos puntuales provenientes de estaciones de monitoreo que no hayan sido incluidas en el estudio original. Esto se debe a que la baja resolución espacial de los datos satelitales descargados puede afectar la precisión de los resultados e introducir incertidumbre en las estimaciones

Es importante observar que los picos de contaminación cambian de hora al utilizar los datos satelitales de precipitación y temperatura y usando el modelo gaussiano multivariado, con nuevos datos de registro, como se observa en la Figura 20, en donde el pico de contaminación pasa de las 9 a 11 horas a las 15 horas.

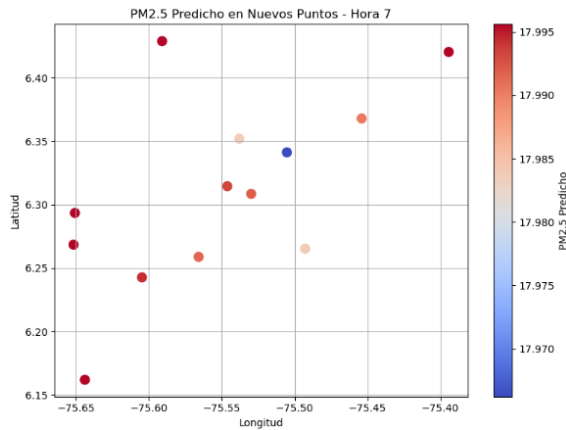


Figura 18: Valores predichos con de material particulado con datos satelitales de modelo Gaussiano multivariado.
Fuente: Elaboración propia.

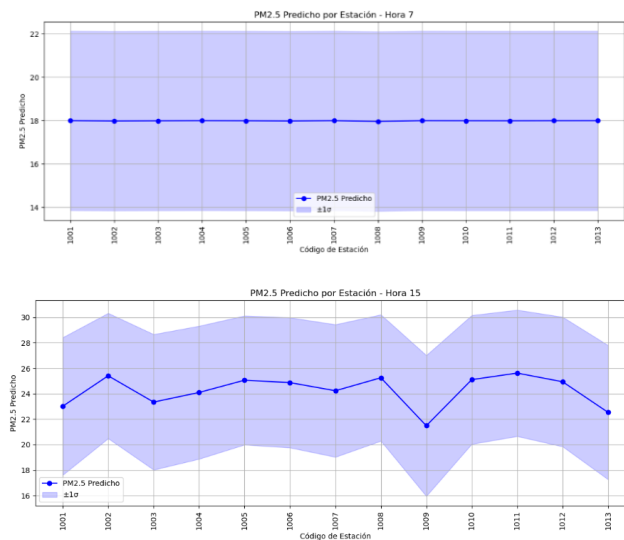


Figura 19: Predicho horario por estación descargadas con datos satelitales.
Fuente: Elaboración propia.

Por último, aunque los datos para probar el modelo son puntuales, estos provienen de productos satelitales, diferentes a los datos in situ, como lo son las estaciones de calidad del aire del SIATA, por lo cual esto puede afectar la precisión de los datos y se recomienda usar datos puntuales de estaciones en un futuro, debido a que en el presente trabajo no fue posible realizarlo por falta de estaciones de medición.

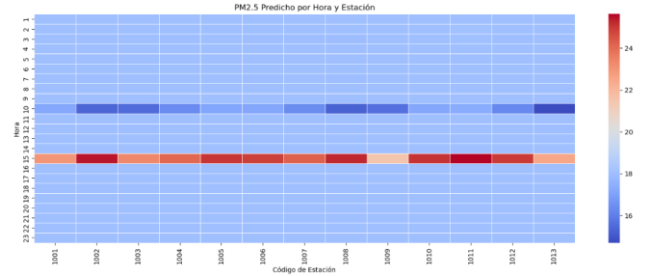


Figura 20: Valores predichos PM 2.5 para estaciones satelitales con modelo gaussiano multivariado.
Fuente: Elaboración propia.

6 Conclusiones

De acuerdo con los resultados obtenidos en el análisis de las variables climáticas, precipitación y temperatura, se concluye que el modelo lineal mixto se ajusta adecuadamente a las variables explicativas. Los resultados evidencian que la relación entre estas variables y el material particulado no es estrictamente lineal. En particular, la temperatura muestra una relación cuadrática inversa: a partir de los 23.5 °C, se observa una disminución en la concentración de material particulado a medida que la temperatura continúa aumentando.

En los modelos y estadísticos utilizados se observan los picos de contaminación entre las 9 y 11 horas, coincidiendo con los picos observados.

Por su parte, en los modelos de dependencia espacial SAR y SEM se evidencia que las estaciones presentan relaciones entre sí, especialmente en horas de la noche, lo que refuerza la idea de que existen dinámicas espaciales en el comportamiento del material particulado. La autocorrelación en los errores sugiere la necesidad de incorporar otras variables climáticas y antrópicas, como presión atmosférica, viento, humedad relativa, radiación solar e incluso usos del suelo, para lograr una comprensión más completa del fenómeno en el Valle de Aburrá.

En línea con esto, se observa que la temperatura y la precipitación comparten comportamientos coherentes con lo reportado en diversos estudios [8]. La precipitación puede actuar de manera dual: facilitando el lavado atmosférico o, por el contrario, estabilizando la atmósfera y atrapando el material particulado dentro del valle. La temperatura, por su parte, puede promover procesos convectivos que permiten la dispersión del material particulado, hasta llegar a un punto de inflexión a partir del cual esta relación cambia. Así, mientras la temperatura muestra un patrón más homogéneo a escala regional, la precipitación, por su naturaleza estocástica, puede tener efectos tanto locales como globales sobre la concentración de PM2.5.

Es así, que en futuros estudios se recomienda utilizar las demás variables climáticas para tener un mejor acercamiento a la naturaleza del comportamiento en el Valle de Aburrá.

Referencias

- [1] Agencia de Protección Ambiental de Estados Unidos (EPA)(2025). Efectos del material particulado PM) sobre la salud y el medio ambiente. Recuperado de <https://espanol.epa.gov/espanol/efectos-del-material-particulado-pm-sobre-la-salud-y-el-medioambiente>
- [2] Organización mundial de la Salud (OMS)(2024),Contaminación del aire ambiente(exterior) y salud. Recuperado de [https://www.who.int/es/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/es/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
- [3] SIATA (2024). ¿Por qué se generan dos períodos de altas concentraciones de PM 2.5 en el Valle de Aburrá?. Recuperado de https://siata.gov.co/sitio_web/index.php/noticia22
- [4] Farzane Omrani, Rouzbeh Shad, Seyed Ali Ziaee. (2025). A multiscale geographically weighted regression approach to emphasize the effects of traffic characteristics on vehicular emissions, Atmospheric Environment: X, Volume 25, <https://doi.org/10.1016/j.aeaoa.2025.100315>.
- [5] Chasco, Coro. (2004). Modelos de heterogeneidad espacial. EconWPA, Econometrics.
- [6] Aristizábal, Edier. 2024. Libro Guía del curso Análisis Geoespacial. Universidad Nacional de Colombia, Departamento de Geociencias y Medio Ambiente, Facultad de Minas. Recuperado de https://edieraristizabal.github.io/Libro_AnalisisGeoespacial/intro.html
- [7] International Research institute for climate and society (IRI). (S.F). Spatial interpolation methods. Recuperado de https://iri.columbia.edu/~rijaf/CDTUserGuide/html/interpolation_methods.html
- [8] Roldan-Henao N, Hoyos C.D, Herrera-Mejía L, Isaza A. (2020). An investigation of the precipitation net effect on the particulate matter concentration in a narrow valley: Role of lower - trophosphere stability. <https://doi.org/10.1175/JAMC-D-18-0313.1>
- [9] Christensen, M. W. and Wu, P. and Varble, A. C. and Xiao, H. and Fast, J. D. (2024). Aerosol-induced closure of marine cloud cells: enhanced effects in the presence of precipitation. Atmospheric Chemistry and Physics, Volume 24. Pp. 6455-6476. <https://doi.org/10.5194/acp-24-6455-2024>

Cardona-Osorio Paulina: Ingeniera ambiental graduada en 2022, estudiante especialización posgrado aprovechamiento de recursos hidráulicos.