LA MEJOR TESIS DEL MUNDO MUNDIAL

Andres Arevalo

**Resumen**

Dwd

Palabras clave:

Índice

[1. Introducción y objetivos 1](#_Toc170117839)

[1.1 Introducción 1](#_Toc170117840)

[1.2 Objetivos 1](#_Toc170117841)

[1.2.1 Objetivo general 1](#_Toc170117842)

[1.2.2 Objetivos específicos 1](#_Toc170117843)

[2. Antecedentes 1](#_Toc170117844)

[3. Revisión de literatura 1](#_Toc170117845)

[4. Metodología 1](#_Toc170117846)

[Diagrama de flujo 1](#_Toc170117847)

[Estructuración de datos 1](#_Toc170117848)

[Variables y Unidades de Medida 1](#_Toc170117849)

[Naturaleza de los Datos 2](#_Toc170117850)

[Desafíos en la Recolección de Datos 2](#_Toc170117851)

[Cobertura de Variables por Estación 2](#_Toc170117852)

[Selección de variables para el Análisis de la Calidad del Aire en Bogotá 3](#_Toc170117853)

[Exclusión de Variables 3](#_Toc170117854)

[Inclusión de Variables 3](#_Toc170117855)

[Datos y procesamiento 6](#_Toc170117856)

[Estadísticas descriptivas 6](#_Toc170117857)

[Proceso de Estequiometria 18](#_Toc170117858)

[Análisis exploratorio de la correlación entre las variables 21](#_Toc170117859)

[Análisis de Contaminantes Atmosféricos 23](#_Toc170117860)

[Modelos propuestos 29](#_Toc170117861)

[Modelo Probit 29](#_Toc170117862)

[Modelo de Vectoriales autorregresivos (VAR) 40](#_Toc170117863)

[Modelo de Vector Error Correction Model (VECM) 43](#_Toc170117864)

[Modelo de Autoregresivo de Rezagos Distribuidos (ARDL) 48](#_Toc170117865)

[Random forest 52](#_Toc170117866)

[Resultados 55](#_Toc170117867)

[Conclusiones y recomendaciones 55](#_Toc170117868)

[Referencias 55](#_Toc170117869)

[Anexos 57](#_Toc170117870)

# 1. Introducción y objetivos

## 1.1 Introducción

## 1.2 Objetivos

### 1.2.1 Objetivo general

### 1.2.2 Objetivos específicos

# 2. Antecedentes

# 3. Revisión de literatura

# 4. Metodología

## Diagrama de flujo

## Estructuración de datos

El presente estudio se enfoca en la evaluación de la calidad del aire en Bogotá, Colombia, utilizando datos provenientes de la RMCAB (Red de Monitoreo de Calidad del Aire de Bogotá, 2024). Para ello, se analizaron registros continuos desde el 1 de enero de 2021 hasta el 31 de mayo de 2024, obtenidos de trece estaciones de monitoreo distribuidas estratégicamente a lo largo de la ciudad. Las estaciones se localizan en las siguientes áreas:

* Zona norte: Guaymaral, Min Ambiente, Suba, Usaquén y Las Ferias.
* Zona sur: San Cristóbal y Tunal.
* Zona oeste: Bolivia, Carvajal-Sevillana, Fontibón, Kennedy y Puente Aranda.
* Zona centro: Centro de Alto Rendimiento.

### Variables y Unidades de Medida

Se obtuvieron datos correspondientes a las siguientes variables ambientales, cada una con sus respectivas unidades de medida:

* PM10, PM2.5, Black Carbon (µg/m³)
* CO, CO2 (ppm)
* SO2, NO, NO2, NOX, O3 (ppb)
* Velocidad del Viento (m/s)
* Dirección del Viento (grados)
* Temperatura (°C)
* Presión Barométrica (mmHg)
* Radiación Solar (W/m²)
* Precipitación (mm)
* Humedad Relativa (HR) (%)
* Precipitación Fronteriza en Masa (BBP) (%)

### Naturaleza de los Datos

Los datos se estructuran en series de tiempo, con observaciones registradas cronológicamente e intervalos regulares de una hora. Cada estación proporcionó aproximadamente 30,000 observaciones por variable, sumando un total aproximado de 390,000 observaciones para cada variable a lo largo de todas las estaciones.

### Desafíos en la Recolección de Datos

Los equipos de monitoreo, al estar expuestos a la intemperie, son susceptibles a daños físicos que pueden requerir la importación de piezas de repuesto o, en casos extremos, el envío del equipo al exterior para su reparación. Este proceso no solo implica un considerable tiempo de inactividad sino también la potencial pérdida de datos.

Además, los equipos requieren recalibraciones periódicas para asegurar la precisión de los datos, un proceso que puede variar desde algunas horas hasta varios días, dependiendo de la complejidad del ajuste y la capacidad técnica disponible, introduciendo otro factor de pérdida de datos.

Cabe resaltar que antes de la publicación en el sitio web de la RMCAB, los datos son rigurosamente verificados por ingenieros especializados para confirmar su fiabilidad.

### Cobertura de Variables por Estación

No todas las estaciones cuentan con el equipo necesario para monitorear las 18 variables mencionadas, principalmente debido al alto costo de los equipos especializados. Sin embargo, cada zona cuenta con al menos una estación capaz de medir un mínimo de 17 de estas variables. Variables como PM10, PM2.5, velocidad del viento, precipitación, O3, CO, NO y NO2 son monitoreadas por todas las estaciones.

## Selección de variables para el Análisis de la Calidad del Aire en Bogotá

En el proceso de determinar cuáles variables incorporar en el análisis de los contaminantes en Bogotá, se efectuó una cuidadosa selección basada tanto en la relevancia ambiental como en la consistencia de los datos disponibles. A continuación, se detallan las variables seleccionadas y las razones para excluir variables:

### Exclusión de Variables

* **Dirección del Viento:** Esta variable fue excluida debido a que las mediciones se realizan a 10 metros del suelo, una altura que puede no reflejar adecuadamente las condiciones de dispersión de contaminantes a nivel de respiración humana. Estudios indican que la dirección del viento a esta altura puede ser significativamente diferente de la experimentada a nivel del suelo debido a efectos de edificación y otros obstáculos urbanos, lo que podría llevar a interpretaciones erróneas en el contexto de la calidad del aire urbana.
* **Precipitación Fronteriza en Masa (BBP):** Esta variable se excluyó debido a su escasa disponibilidad a través de las estaciones de monitoreo, lo que limita la capacidad para realizar análisis comparativos y de tendencias a largo plazo.
* **NOX:** Aunque NOX es un precursor clave en la formación de ozono y otras partículas finas, su inclusión fue considerada redundante dado que se optó por analizar NO y NO2 individualmente, ambos con límites estándar definidos en las nuevas Directrices Mundiales de la OMS sobre la calidad del aire (World Health Organization, 2021). Además, la desagregación en NO y NO2 permite un análisis más detallado del comportamiento y origen de estos óxidos en el ambiente urbano.

### Inclusión de Variables

Los contaminantes seleccionados para el estudio incluyen CO, CO2, SO2, NO, NO2 y O3. Estas variables son esenciales por sus conocidos efectos adversos tanto para la salud humana como para el ambiente:

* **Monóxido de Carbono (CO):** es un gas incoloro e inodoro que se considera altamente tóxico. Su peligrosidad radica en su capacidad para unirse a la hemoglobina en la sangre, formando carboxihemoglobina, lo que reduce drásticamente la capacidad de transporte de oxígeno por parte de los glóbulos rojos. Este fenómeno puede llevar a la hipoxia tisular, provocando asfixia y daño cerebral irreversible, e incluso la muerte en exposiciones prolongadas a altas concentraciones. A nivel ambiental, el CO contribuye a la contaminación del aire, participando en reacciones fotoquímicas que generan ozono troposférico, un componente nocivo del smog urbano. (Benavides Ballesteros & León Aristizabal, 2007)
* **Dióxido de Azufre (SO2):** es un gas irritante que puede tener serios impactos en la salud y el medio ambiente. Desde el punto de vista de la salud, el SO2 puede causar problemas respiratorios, especialmente en personas con enfermedades pulmonares preexistentes como el asma. La exposición a altas concentraciones de SO2 puede provocar inflamación de las vías respiratorias y reducir la función pulmonar. Además, el SO2 contribuye a la formación de lluvia ácida, la cual tiene efectos perjudiciales sobre los ecosistemas acuáticos y terrestres, y puede deteriorar edificios y monumentos históricos al reaccionar con materiales de construcción como la piedra caliza. (Benavides Ballesteros & León Aristizabal, 2007)
* **Óxidos de Nitrógeno (NO y NO2):** son contaminantes del aire que presentan serios riesgos para la salud humana y el medio ambiente. En términos de salud, la exposición a estos gases puede irritar las vías respiratorias y empeorar condiciones pulmonares preexistentes como el asma y la bronquitis. (Benavides Ballesteros & León Aristizabal, 2007)
* **Ozono Troposférico (O3):** a nivel del suelo, es un contaminante secundario formado por reacciones fotoquímicas entre los óxidos de nitrógeno y compuestos orgánicos volátiles en presencia de luz solar. Este gas es especialmente dañino para la salud humana, causando problemas respiratorios en poblaciones vulnerables como niños, ancianos y personas con enfermedades respiratorias crónicas. La exposición al ozono troposférico puede resultar en inflamación y daño a los tejidos pulmonares, reduciendo la función pulmonar y exacerbando enfermedades como el asma. Además, el ozono troposférico contribuye a la degradación de materiales, afectando la durabilidad de edificios y otros bienes materiales. (Benavides Ballesteros & León Aristizabal, 2007)
* **PM10, PM2.5 y Black Carbón (BC):** Además, para el análisis de la calidad del aire en Bogotá, se utilizan variables de material particulado como. Según la EPA, el tamaño de las partículas está directamente relacionado con su potencial para causar problemas de salud. Las partículas pequeñas, de menos de 10 micrómetros de diámetro, son las más preocupantes, ya que pueden penetrar profundamente en los pulmones e incluso alcanzar el torrente sanguíneo. La exposición a estas partículas puede afectar tanto a los pulmones como al corazón. (Agencia de Protección Ambiental de Estados Unidos, 2024)

Numerosos estudios científicos han vinculado la exposición a la contaminación por partículas con una variedad de problemas de salud, incluyendo muerte prematura en personas con enfermedades cardíacas o pulmonares, infartos de miocardio no mortales, latidos cardíacos irregulares, agravamiento del asma, reducción de la función pulmonar y aumento de síntomas respiratorios, como irritación en las vías respiratorias, tos o dificultad para respirar. (Agencia de Protección Ambiental de Estados Unidos, 2024)

La exposición a la contaminación por partículas tiende a afectar principalmente a personas con enfermedades cardíacas o pulmonares, niños y adultos mayores. Por lo tanto, es un factor crucial a considerar en el análisis de la calidad del aire en Bogotá.

* **Velocidad del Viento (ws):** La velocidad del viento es una de las variables más críticas en estudios de calidad del aire, ya que una mayor velocidad puede dispersar rápidamente los contaminantes, reduciendo su concentración en un área específica pero potencialmente aumentando la exposición en áreas más amplias. Inversamente, velocidades de viento bajas pueden llevar a una acumulación de contaminantes, elevando la concentración local y exacerbando los efectos nocivos sobre la salud y el ambiente. La dirección y velocidad del viento también son esenciales para la modelización de la dispersión de contaminantes y para entender las fuentes y sumideros de contaminación en áreas urbanas.
* **Presión Barométrica (pressure):** La presión barométrica influye en el comportamiento de los contaminantes en la atmósfera. Cambios en la presión pueden alterar la dinámica de la capa límite atmosférica, que a su vez afecta la dispersión vertical de los contaminantes. Una presión alta generalmente se asocia con condiciones meteorológicas más estables, que pueden llevar a una acumulación de contaminantes cerca del suelo, mientras que una presión baja puede indicar condiciones más turbulentas, favoreciendo la dispersión de estos.
* **Temperatura (tmp):** La temperatura afecta las reacciones químicas en la atmósfera, incluyendo la formación de ozono y otros contaminantes secundarios. Temperaturas más altas pueden acelerar estas reacciones, aumentando la concentración de ciertos contaminantes durante episodios de calor. Además, la temperatura también influye en la estabilidad de la atmósfera; un gradiente de temperatura decreciente con la altura puede llevar a una mayor estabilidad atmosférica y menor mezcla vertical, atrapando contaminantes cerca del suelo.
* **Humedad Relativa (rh):** La humedad afecta la capacidad de la atmósfera para dispersar o retener contaminantes. Altos niveles de humedad pueden facilitar la formación de aerosoles y smog fotoquímico, especialmente en combinación con otros contaminantes como NOx y compuestos orgánicos volátiles. La humedad también juega un papel crucial en la formación de lluvia ácida, al interactuar con contaminantes como el SO2 y NOx.
* **Radiación Solar (radsolar):** La radiación solar es un factor clave en la química atmosférica, particularmente en la producción de ozono a nivel del suelo a través de la fotólisis de contaminantes como NO y NO2. Días soleados pueden llevar a mayores concentraciones de ozono, mientras que condiciones nubladas pueden reducir este efecto.
* **Precipitación (rain):** La precipitación juega un papel crucial en el lavado de los contaminantes atmosféricos, reduciendo su concentración en el aire.

## Datos y procesamiento

Dado que las observaciones se registraron en intervalos horarios, el primer paso en el procesamiento consistió en calcular el promedio diario de cada una de las variables seleccionadas para cada una de las 13 bases de datos. Esto permitió transformar los datos de series de tiempo de alta frecuencia en una forma más manejable y representativa de las condiciones promedio diarias.

Posteriormente, las 13 bases de datos fueron consolidadas en una única base de datos. Este paso fue crucial para permitir un análisis integrado y comparativo de las dinámicas de contaminación a lo largo de toda la ciudad. La unificación de los datos facilitó el cálculo de promedios diarios agregados y proporcionó una base robusta para análisis estadísticos y modelado posterior.

### Estadísticas descriptivas

Para cada una de las 13 bases de datos, se calcularon estadísticas descriptivas básicas, incluyendo la media, mediana, desviación estándar, valores mínimos y máximos, y rangos intercuartílicos. Estas estadísticas proporcionan una visión inicial del comportamiento de los contaminantes y las condiciones meteorológicas en cada zona de la ciudad, permitiendo identificar patrones, tendencias y posibles anomalías en los datos.

#### Estadísticas descriptivas por zona

En la zona de Guaymaral, las estadísticas descriptivas muestran que los contaminantes principales presentan niveles variables pero significativos. El PM2.5 tiene una media de 14.373 µg/m³, con un pico máximo registrado de 44 µg/m³. Por otro lado, el PM10 exhibe un promedio de 28.02 µg/m³, con un máximo de 69.67 µg/m³, revelando una variabilidad mayor en sus niveles. En cuanto al ozono (O3), la media es de 11.015 ppb, alcanzando un máximo de 35.133 ppb. El dióxido de nitrógeno (NO2) presenta un promedio de 9.96 ppb, con un máximo registrado de 27.857 ppb. Respecto al monóxido de carbono (CO), la media es de 0.3265 ppm, con un máximo de 0.8111 ppm. [Ver Salida de la consola 1. Estadísticas descriptivas Guaymaral]

En términos de variables meteorológicas, Guaymaral experimenta temperaturas que oscilan entre 11.64°C y 17.93°C, con una media que indica un clima moderado. La humedad relativa promedio es del 68.66%, con variaciones entre 45.27% y 84.28%, indicando condiciones de humedad moderada a alta. La precipitación media registrada es de 0.115343 mm, con un máximo de 2.183333 mm. La radiación solar promedio es de 160.024 W/m², alcanzando hasta 330.5 W/m² en ciertos momentos, lo que sugiere una exposición solar significativa durante el día. La presión barométrica media es de 563.9 mmHg, con variaciones mínimas entre 561.7 y 566.2 mmHg, indicando condiciones atmosféricas estables. Finalmente, la velocidad del viento presenta un promedio de 1.2367 m/s, con un máximo registrado de 3.3792 m/s, mostrando condiciones de viento moderadas en la zona.

En la zona de Min Ambiente, los niveles de contaminantes principales muestran una variabilidad considerable. El PM2.5 presenta una media de 14.267 µg/m³, alcanzando un máximo de 48.875 µg/m³, indicando fluctuaciones significativas en la concentración de partículas finas en el aire. Por otro lado, el PM10 registra un promedio de 22.239 µg/m³, con un pico máximo de 77.833 µg/m³, lo que sugiere una presencia notable de partículas más gruesas. En cuanto al ozono (O3), se observa una media de 13.071 ppb, con un máximo de 30.921 ppb, indicando niveles moderados con episodios esporádicos de mayor concentración. El dióxido de nitrógeno (NO2) presenta un promedio de 15.127 ppb, con un máximo registrado de 40.450 ppb, señalando una presencia significativa de este contaminante, posiblemente asociado al tráfico vehicular. En términos de monóxido de carbono (CO), se registra una media de 0.7371 ppm, con un máximo de 1.7044 ppm. [Ver Salida de la consola 2. Estadísticas descriptivas Min Ambiente]

En cuanto a las variables meteorológicas, la zona de Min Ambiente experimenta una media de precipitación de 0.134503 mm, con un máximo de 2.166667 mm, señalando condiciones de lluvia moderadas, pero con posibles eventos intensos. La velocidad del viento muestra un promedio de 1.2350 m/s, con un máximo registrado de 2.6417 m/s, indicando condiciones de viento relativamente bajas, lo que podría contribuir a la acumulación de contaminantes en la zona.

Es importante destacar que se observa una cantidad significativa de valores faltantes (NA's) en varias variables, como SO2, black carbon, temperatura, humedad relativa, radiación solar, CO2 y presión atmosférica.

En la zona de Suba, los niveles de contaminantes principales muestran una variabilidad notable. El PM2.5 presenta una media de 15.154 µg/m³, con un máximo registrado de 44.792 µg/m³. Por otro lado, el PM10 registra un promedio de 33.39 µg/m³, alcanzando hasta 80.61 µg/m³ en momentos pico, lo que sugiere presencia de partículas más gruesas. En cuanto al ozono (O3), se observa una media de 11.183 ppb, con un máximo de 31.488 ppb, indicando niveles moderados con episodios esporádicos de concentraciones elevadas. El dióxido de nitrógeno (NO2) presenta un promedio de 12.085 ppb, con un máximo registrado de 24.106 ppb, mostrando concentraciones variables. En términos de monóxido de carbono (CO), se registra una media de 0.4515 ppm, con un máximo de 0.8833 ppm, indicando niveles moderados. [Ver Salida de la consola 3. Estadísticas descriptivas Suba]

En cuanto a las variables meteorológicas, la temperatura oscila entre 10.67°C y 18.20°C, con una media de 14.76°C, reflejando las típicas variaciones climáticas de Bogotá. La precipitación media es de 0.154056 mm, con un máximo de 3.033333 mm, indicando un patrón variable con algunos eventos de lluvia intensa. La velocidad del viento muestra un promedio de 1.4205 m/s, con un máximo registrado de 3.2792 m/s, lo que sugiere condiciones moderadas que podrían favorecer la dispersión de contaminantes en la zona.

Es importante mencionar que se observan valores faltantes (NA's) en varias variables, como black carbon, humedad relativa, radiación solar y CO2, lo cual limita el análisis detallado sobre cómo estas condiciones meteorológicas específicas influyen en los niveles de contaminación en Suba.

En la zona de Usaquén en Bogotá, se observan niveles variables de contaminantes atmosféricos y condiciones meteorológicas significativas. El PM2.5 presenta una media de 11.6842 µg/m³, con un máximo registrado de 43.6667 µg/m³. Por otro lado, el PM10 muestra un promedio de 22.097 µg/m³, con un pico máximo de 62.809 µg/m³, sugiriendo una presencia notable de partículas más gruesas. En cuanto al ozono (O3), se registra una media de 17.9344 ppb, alcanzando un máximo de 58.2923 ppb, indicando niveles más altos en comparación con otras zonas y episodios esporádicos de concentraciones elevadas. El dióxido de nitrógeno (NO2) presenta un promedio de 10.375 ppb, con un máximo registrado de 31.846 ppb, mostrando concentraciones variables, pero generalmente moderadas. En términos de monóxido de carbono (CO), se observa una media de 0.4655 ppm, con un máximo de 1.5791 ppm, indicando niveles moderados. [Ver Salida de la consola 4. Estadísticas descriptivas Usaquén]

En cuanto a las variables meteorológicas, la temperatura varía entre 11.53°C y 20.5°C, con una media de 14.52°C, reflejando las típicas fluctuaciones climáticas de Bogotá. La precipitación media es de 0.13708 mm, con un máximo registrado de 15.67083 mm, señalando un patrón variable con algunos eventos de lluvia intensa. La velocidad del viento muestra un promedio de 1.2907 m/s, alcanzando un máximo notable de 7.3417 m/s, lo que sugiere condiciones variables que podrían influir en la dispersión de contaminantes en la zona.

Es importante mencionar que se observan valores faltantes (NA's) en varias variables, como black carbon, humedad relativa, radiación solar, CO2 y presión atmosférica, lo cual limita el análisis detallado sobre cómo estas condiciones meteorológicas específicas afectan los niveles de contaminación en Usaquén.

En la zona de las Ferias en Bogotá, se observan niveles variables de contaminantes atmosféricos y condiciones meteorológicas significativas. El monóxido de carbono (CO) muestra niveles relativamente bajos, con un máximo registrado de 1.59500 ppm. Las partículas gruesas (PM10) exhiben niveles moderados, alcanzando un máximo de 137.144 μg/m³, mientras que las partículas finas (PM2.5) varían de moderadas a altas, con un máximo de 50.833 μg/m³. [Ver Salida de la consola 5. Estadísticas descriptivas Las Ferias]

En cuanto a las variables meteorológicas, la temperatura oscila entre 11.77°C y 18.79°C, con una media de 14.75°C, reflejando un clima templado característico de la altitud de Bogotá. La humedad relativa varía del 23.50% al 81.25%, con una media de 63.40%, indicando condiciones de humedad moderada. La presión barométrica se mantiene estable alrededor de 564 mmHg, lo cual sugiere una ubicación elevada. La velocidad del viento muestra un promedio de 1.7971 m/s, con máximos registrados de hasta 13.3917 m/s, indicando condiciones variables que pueden influir en la dispersión de contaminantes.

En la zona de San Cristóbal al sur de Bogotá, se observa una variabilidad en los niveles de contaminantes atmosféricos y condiciones meteorológicas significativas. El monóxido de carbono (CO) muestra niveles bajos, con un máximo registrado de 1.2208 ppm. El dióxido de azufre (SO2) se presenta con datos lógicos, posiblemente indicando presencia o ausencia sin especificar niveles numéricos. Las partículas gruesas (PM10) alcanzaron un máximo de 167.458 μg/m³, mientras que las partículas finas (PM2.5) varían de moderadas a altas, con un máximo de 46.7692 μg/m³. [Ver Salida de la consola 6. Estadísticas descriptivas San Cristóbal]

En cuanto a las variables meteorológicas, la temperatura varía entre 10.35°C y 16.58°C, con una media de 13.88°C, reflejando un clima templado característico de la altitud de Bogotá. La humedad relativa oscila entre el 37.33% y el 91.00%, con una media de 69.70%, indicando condiciones de humedad moderada a alta. La presión barométrica se presenta como datos lógicos, lo cual es inusual y podría indicar un error en los datos o en su presentación. La velocidad del viento muestra un promedio de 1.3930 m/s, con máximos registrados de hasta 2.7583 m/s.

Otros datos como el ozono (O3) muestran niveles moderados, con un máximo de 27.637 ppb, y el dióxido de nitrógeno (NO2) alcanzó un máximo de 32.871 ppb. La radiación solar varía de 0 a 333.0 W/M², indicando condiciones variables de exposición solar.

Es importante destacar la presencia de datos faltantes (NA's) en varias variables, lo que puede sugerir interrupciones en la recolección de datos o problemas con los sensores utilizados para monitorear la calidad del aire y las condiciones meteorológicas en San Cristóbal.

Comparado con la zona de Las Ferias, San Cristóbal parece mostrar niveles ligeramente más bajos de algunos contaminantes como PM2.5 y NO2, pero niveles similares o incluso más altos de otros como PM10. Las temperaturas son comparables, aunque San Cristóbal parece tener una humedad relativa ligeramente mayor y vientos menos intensos, lo que puede influir en la dispersión de contaminantes en la zona.

En la zona de Tunal, ubicada en el sur de Bogotá, el monóxido de carbono (CO) tiene un promedio de 0.8826 ppm, con un máximo registrado de 1.7797 ppm. El dióxido de azufre (SO2) presenta un promedio de 2.058 ppb, alcanzando un máximo de 5.707 ppb. Las partículas gruesas (PM10) muestran un promedio de 37.107 μg/m³, con un máximo de 109.625 μg/m³, mientras que las partículas finas (PM2.5) tienen un promedio de 17.625 μg/m³ y un máximo de 58.000 μg/m³. El ozono (O3) tiene un promedio de 13.820 ppb, con un máximo de 35.837 ppb, y el dióxido de nitrógeno (NO2) presenta un promedio de 13.597 ppb, con un máximo de 35.829 ppb. El carbono negro (BC) muestra un promedio de 4.573 μg/m³, con un máximo de 13.287 μg/m³. [Ver Salida de la consola 7. Estadísticas descriptivas Tunal]

En cuanto a las variables meteorológicas, la temperatura promedio es de 15.10°C, con un rango que va desde 8.85°C hasta 18.87°C. La humedad relativa tiene un promedio de 62.81%, oscilando entre el 31.61% y el 81.66%. La presión barométrica promedio es de 563.3 mmHg, con variaciones entre 561.2 y 566.0 mmHg, reflejando la altitud de Bogotá. La velocidad del viento presenta un promedio de 1.2106 m/s, con un máximo registrado de 2.1375 m/s. La radiación solar promedio es de 153.1 W/m², con un pico máximo de 296.2 W/m², mostrando variaciones significativas en la exposición solar.

Los datos indican que la zona de Tunal experimenta niveles moderados de contaminación atmosférica, con episodios esporádicos de alta contaminación, especialmente para PM10 y PM2.5. Las condiciones meteorológicas muestran un clima templado con variaciones notables en la humedad y la radiación solar, mientras que la presión barométrica confirma la altitud de la ciudad de Bogotá.

En la zona de Bolivia, ubicada al oeste de Bogotá, se observan niveles variados de contaminantes atmosféricos y limitaciones en la presentación de variables meteorológicas. El PM2.5 presenta un promedio de 15.467 μg/m³, con un rango que va desde 2.454 hasta 42.767 μg/m³. El monóxido de carbono (CO) tiene un promedio de 0.4330 ppm, con un máximo registrado de 1.6250 ppm. El dióxido de azufre (SO2) muestra un promedio de 1.9332 ppb, alcanzando un máximo considerable de 44.8792 ppb. Las partículas gruesas (PM10) tienen un promedio de 34.830 μg/m³ y un máximo de 90.392 μg/m³. El ozono (O3) presenta un promedio de 6.37736 ppb, con un máximo de 17.15000 ppb, mientras que el dióxido de nitrógeno (NO2) tiene un promedio de 18.579 ppb, con un máximo de 30.350 ppb. [Ver Salida de la consola 8. Estadísticas descriptivas Bolivia]

En cuanto a las variables meteorológicas, como temperatura (tmp), humedad relativa (rh), presión barométrica (pressure), velocidad del viento (ws) y radiación solar (radsolar), se presentan como (NAs), lo cual indica una falta de instrumentos meteorológicos para la recolección y el procesamiento de datos en la estación de monitoreo de Bolivia. Esta limitación dificulta la interpretación precisa de las condiciones meteorológicas en la zona.

En comparación con otras zonas analizadas, los niveles de PM2.5 son similares a los de Tunal y ligeramente inferiores a los de Las Ferias. Sin embargo, los niveles de NO2 son más altos que en las otras zonas estudiadas, y el ozono (O3) muestra niveles más bajos en Bolivia.

En la zona de Carvajal, ubicada al oeste de Bogotá, se registran niveles significativos de contaminantes atmosféricos y limitaciones en la presentación de variables meteorológicas. El PM2.5 presenta un promedio de 34.65 μg/m³, con valores que oscilan entre 10.50 y 81.12 μg/m³. Las partículas gruesas (PM10) tienen un promedio de 70.96 μg/m³, alcanzando un máximo de 140.00 μg/m³. El monóxido de carbono (CO) muestra un promedio de 1.9090 ppm, con un máximo registrado de 4.3833 ppm. El dióxido de azufre (SO2) presenta un promedio de 5.3324 ppb, alcanzando un máximo de 20.5250 ppb. El ozono (O3) tiene un promedio de 6.440 ppb, con un máximo de 17.908 ppb, mientras que el dióxido de nitrógeno (NO2) registra un promedio de 24.50 ppb, con un máximo de 42.20 ppb. [Ver Salida de la consola 9. Estadísticas descriptivas Carvajal]

En cuanto a las variables meteorológicas, la temperatura promedio es de 16.09°C, con un rango entre 12.67°C y 19.35°C. La velocidad del viento promedio es de 1.9635 m/s, con un máximo de 3.5208 m/s. Sin embargo, la humedad relativa (rh), presión barométrica (pressure) y radiación solar (radsolar) se presentan como (NAs), dado que la estación de monitoreo no mide, lo cual limita su interpretación y análisis.

Observando los patrones de contaminación, se destaca que los niveles de PM2.5 y PM10 son altos, indicando una significativa contaminación por partículas en la zona de Carvajal.

Comparando con otras zonas analizadas, Carvajal presenta niveles significativamente más altos de PM2.5 y PM10 en comparación con Las Ferias, Tunal y Bolivia. Asimismo, los niveles de CO son más elevados que en las otras estaciones evaluadas. Los niveles de NO2 son similares a los de Bolivia, pero más altos que en Las Ferias y Tunal.

La calidad del aire en la zona de Carvajal parece estar más deteriorada, posiblemente debido a una mayor actividad industrial o tráfico vehicular en el área. La falta de datos numéricos para algunas variables meteorológicas limita un análisis más detallado de las condiciones ambientales, aunque se observa una velocidad del viento ligeramente mayor que podría influir en la dispersión de contaminantes.

En la zona de Fontibón, ubicada al oeste de Bogotá, se observan variaciones significativas en los niveles de contaminantes atmosféricos y condiciones meteorológicas. El PM2.5 tiene un promedio de 19.00 µg/m³. Sin embargo, se registró un máximo de 55.75 µg/m³, lo cual indica episodios de alta concentración que podrían afectar la calidad del aire durante esos períodos. El mínimo de 0.96 µg/m³ refleja condiciones relativamente limpias en momentos de baja contaminación. El monóxido de carbono (CO) muestra un promedio de 0.54212 ppm, con un máximo de 1.24583 ppm y un mínimo de 0.05556 ppm. El dióxido de azufre (SO2) tiene un promedio de 1.6310 ppb, con un máximo de 7.5792 ppb y un mínimo de 0.1286 ppb. Los valores máximos sugieren variabilidad en la concentración de SO2, posiblemente influenciada por factores como la actividad industrial o la combustión de combustibles fósiles. El ozono (O3) presenta un promedio de 16.68 ppb, con un máximo de 48.16 ppb. Aunque el mínimo registrado es 0.00 ppb, lo cual es inusual y podría indicar períodos sin presencia detectable de ozono. [Ver Salida de la consola 10. Estadísticas descriptivas Fontibón]

En términos de condiciones meteorológicas, la temperatura promedio es de 15.45°C, con un rango entre -12.80°C (un valor inusualmente bajo, posiblemente un error de medición) y 18.38°C. La humedad relativa varía entre un mínimo del 30.29% y un máximo del 86.62%, reflejando cambios significativos en la humedad ambiental. La precipitación promedio es baja, con 0.1378 mm, y los datos de radiación solar no fueron proporcionados, mientras que la presión barométrica promedio es de 564.6 mmHg, indicando condiciones estables acordes con la altitud de Bogotá.

En la zona de Kennedy, ubicada al oeste de Bogotá, los datos revelan condiciones ambientales que presentan tanto desafíos como características típicas de la ciudad. El PM2.5, que indica la concentración de material particulado fino, tiene un promedio de 21.300 µg/m³. Este valor promedio es significativamente alto y sugiere una exposición considerable a partículas finas en la zona. El máximo registrado alcanza los 56.842 µg/m³, indicando episodios de alta contaminación que pueden afectar la calidad del aire durante períodos específicos. Por otro lado, el mínimo de 1.875 µg/m³ refleja condiciones más limpias en momentos de menor contaminación. El monóxido de carbono (CO) muestra un promedio de 0.8361 ppm, con un máximo de 2.05750 ppm y un mínimo de 0.06667 ppm. Estos niveles indican presencia moderada de CO, con algunos picos que podrían estar asociados con el tráfico vehicular u otras fuentes locales de emisión. El dióxido de azufre (SO2) tiene un promedio de 2.3020 ppb, con un máximo de 9.0435 ppb y un mínimo de 0.1565 ppb. Los valores máximos muestran variabilidad en la concentración de SO2, posiblemente influenciada por actividades industriales o la combustión de combustibles. El ozono (O3) presenta un promedio de 13.651 ppb, con un máximo de 39.608 ppb y un mínimo de 1.562 ppb. Aunque el ozono es un contaminante secundario que se forma por reacciones químicas en presencia de luz solar y es menos preocupante que las partículas finas, los picos altos pueden impactar la salud respiratoria durante períodos prolongados de exposición. [Ver Salida de la consola 11. Estadísticas descriptivas Kennedy]

En cuanto a las condiciones meteorológicas, la temperatura promedio es de 15.69°C, con un rango entre 9.05°C y 20.16°C. Estas variaciones son típicas de Bogotá y reflejan cambios estacionales y diurnos normales. La humedad relativa varía considerablemente, con un promedio de 63.79% y valores que oscilan entre 30.88% y 82.50%, lo cual es común en áreas urbanas con influencias climáticas variadas. La precipitación promedio es baja, con 0.0614 mm, y la radiación solar presenta un promedio de 143.04 W/m², con picos que alcanzan los 354.50 W/m².

En la zona de Puente Aranda, ubicada al oeste de Bogotá, el PM2.5, que indica la concentración de material particulado fino, tiene un promedio de 16.730 µg/m³. Este valor promedio sugiere una exposición moderada a partículas finas en la zona, con un máximo registrado de 48.429 µg/m³ y un mínimo de 1.279 µg/m³. Estos datos indican variaciones significativas en la calidad del aire, con episodios de contaminación que pueden afectar la salud pública durante los picos más altos de concentración. El monóxido de carbono (CO) muestra un promedio de 0.7373 ppm, con un máximo de 3.7542 ppm y un mínimo de 0.1042 ppm. Estos niveles sugieren presencia moderada de CO en el área, con algunos. El dióxido de azufre (SO2) tiene un promedio de 1.2878 ppb, con un máximo de 4.7571 ppb y un mínimo de 0.2583 ppb. Aunque los niveles son relativamente bajos en promedio. El ozono (O3) presenta un promedio de 8.5172 ppb, con un máximo de 27.1375 ppb y un mínimo de 0.4143 ppb. [Ver Salida de la consola 12. Estadísticas descriptivas Puente Aranda]

En términos de condiciones meteorológicas, la temperatura promedio es de 15.46°C, con máximos y mínimos que reflejan las variaciones diurnas y estacionales típicas de Bogotá. La humedad relativa es alta en promedio, alcanzando un máximo del 96.79%, lo cual puede ser común en áreas urbanas que pueden experimentar variaciones climáticas rápidas. La radiación solar muestra un promedio considerable de 165.926 W/m², con picos que alcanzan los 313.708 W/m².

En el Centro de Alto Rendimiento, ubicado en el centro de Bogotá, el PM2.5 tiene un promedio de 14.191 µg/m³. Este valor promedio sugiere una exposición moderada a material particulado fino, con un máximo registrado de 47.611 µg/m³ y un mínimo de 0.700 µg/m³. El monóxido de carbono (CO) muestra un promedio de 0.5384 ppm, con un máximo de 1.5833 ppm y un mínimo de 0.1167 ppm. Estos niveles indican una presencia moderada de CO en el área, posiblemente asociada con el tráfico vehicular y otras fuentes de combustión urbana. El dióxido de azufre (SO2) tiene un promedio de 0.9686 ppb, con un máximo de 6.5667 ppb y un mínimo de 0.04286 ppb. Aunque los niveles son relativamente bajos en promedio, los picos altos pueden estar relacionados con actividades industriales o locales específicas. El ozono (O3) presenta un promedio de 12.847 ppb, con un máximo de 34.083 ppb y un mínimo de 2.567 ppb. El PM10, que mide partículas más gruesas, tiene un promedio de 24.133 µg/m³, con un máximo de 66.063 µg/m³ y un mínimo de 1.956 µg/m³. [Ver Salida de la consola 13. Estadísticas descriptivas Centro de Alto Rendimiento]

En términos de condiciones meteorológicas, la temperatura promedio es de 14.920°C, con máximos y mínimos típicos de Bogotá que reflejan variaciones diurnas y estacionales. La humedad relativa es alta en promedio (68.45%), alcanzando un máximo del 86.83%, lo cual es común en áreas urbanas densamente pobladas. La radiación solar muestra un promedio de 150.03 W/m², con picos que alcanzan los 357.29 W/m². Esto indica una exposición adecuada al sol, con variaciones normales a lo largo del día que pueden influir en la formación de ozono y otros procesos atmosféricos.

Finalmente, los niveles de CO2 son notables, con un promedio de 509.5 ppm y un máximo de 627.2 ppm, lo cual sugiere una influencia significativa de actividades urbanas en la zona.

#### Estadísticas descriptivas de Bogotá

Se examinan estadísticas descriptivas de contaminantes clave como PM2.5, PM10, O3, NO2, CO, SO2 y Black Carbon (BC), así como variables meteorológicas como temperatura, humedad relativa, precipitación, velocidad del viento, presión barométrica y radiación solar. Se realizo un análisis de estos resultados. [Ver Salida de la consola 14. Estadísticas descriptivas Bogotá]

* **PM2.5 (µg/m³):** El promedio de PM2.5 es de 16.57 µg/m³, lo cual está por encima de las recomendaciones de la Organización Mundial de la Salud (OMS), que sugieren un límite de 15 µg/m³ para la media diaria (World Health Organization, 2021). Esto indica una preocupación constante por la calidad del aire en Bogotá, con niveles que podrían tener efectos adversos en la salud de la población. El valor máximo registrado es de 48.88 µg/m³, un nivel alarmante que podría exacerbar problemas respiratorios y cardiovasculares, mientras que el mínimo es de 3.23 µg/m³, lo que muestra una variabilidad considerable en los niveles de contaminación por partículas finas.
* **PM10 (µg/m³):** El PM10 presenta un promedio de 31.90 µg/m³, superando el límite anual recomendado por la OMS de 20 µg/m³. Esto subraya la presencia de una contaminación significativa por partículas más grandes que también pueden tener efectos perjudiciales en la salud. El valor máximo de 77.775 µg/m³ y el mínimo de 6.597 µg/m³ reflejan picos de contaminación que podrían estar relacionados con episodios específicos, como alta actividad vehicular o condiciones meteorológicas desfavorables.
* **O3 (ppb):** El ozono (O3) troposférico muestra un promedio de 12.136 ppb, con un rango que va desde 3.420 ppb hasta un máximo de 27.848 ppb. Aunque los niveles promedio de ozono son moderados, los picos máximos podrían ser preocupantes, especialmente durante episodios de alta radiación solar que facilitan la formación de ozono troposférico, un contaminante secundario conocido por sus efectos negativos en la salud respiratoria.
* **NO2 (ppb):** El dióxido de nitrógeno (NO2) tiene un promedio de 15.10 ppb, con un valor máximo de 29.46 ppb y un mínimo de 4.36 ppb. Estos niveles indican una exposición moderada a este contaminante, principalmente generado probablemente por las emisiones de vehículos y actividades industriales. Los picos altos sugieren la necesidad de estrategias para mitigar la contaminación vehicular, especialmente en áreas de alta congestión.
* **CO (ppm):** El monóxido de carbono (CO) presenta un promedio de 0.6380 ppm, con un valor máximo de 1.9417 ppm y un mínimo de 0.2417 ppm. Estos niveles son relativamente bajos y dentro de los límites de seguridad establecidos, lo cual es positivo en términos de exposición a este gas tóxico. Sin embargo, los picos podrían indicar áreas locales de alta emisión que requieren monitoreo.
* **SO2 (ppb):** El dióxido de azufre (SO2) tiene un promedio de 1.7835 ppb, con un rango de 0.5799 ppb a 8.0799 ppb. Los niveles bajos de SO2 son favorables, ya que este gas puede contribuir a problemas respiratorios y la formación de lluvia ácida. La presencia de valores máximos relativamente bajos sugiere una fuente de emisiones controlada o limitada.
* **Black Carbon (µg/m³):** El black carbon tiene un promedio de 3.769 µg/m³, con un máximo de 11.160 µg/m³ y un mínimo de 1.113 µg/m³. Esta variable muestra niveles moderados de contaminación por material carbonoso, aunque presenta 1002 datos faltantes (NAs), puesto que se comenzó a tomar en 2023. La falta de datos puede afectar la precisión del análisis.
* **Temperatura (°C):** La temperatura promedio es de 15.04°C, con un rango de -12.16°C a 34.79°C. La gran variabilidad, incluyendo extremos inusuales. Las temperaturas extremas podrían influir en la formación y dispersión de contaminantes.
* **Humedad Relativa (%):** La humedad relativa promedio es de 66.88%, con un rango de 33.70% a 85.49%. Estos niveles moderados a altos son típicos de Bogotá y pueden influir en la dispersión y concentración de contaminantes, especialmente partículas.
* **Precipitación (mm):** La precipitación promedio es de 0.13469 mm, con un rango de 0.000000 mm a 1.823333 mm. La variabilidad en la precipitación refleja períodos secos y lluviosos, lo que puede afectar la eliminación de contaminantes atmosféricos a través de la deposición húmeda.
* **Velocidad del Viento (m/s):** La velocidad del viento promedio es de 1.4522 m/s, con un máximo de 2.4644 m/s y un mínimo de 0.6051 m/s. Una velocidad del viento moderada puede influir en la dispersión de contaminantes, mitigando la acumulación en áreas específicas.
* **Presión Barométrica (mmHg):** La presión barométrica promedio es de 563.8 mmHg, con un rango de 561.6 mmHg a 566.0 mmHg. Estos valores son consistentes con la altitud de Bogotá y pueden tener implicaciones en la formación y dispersión de contaminantes.
* **Radiación Solar (W/m²):** La radiación solar promedio es de 155.79 W/m², con un rango de 36.43 W/m² a 283.30 W/m². La variabilidad en la radiación solar refleja los patrones diurnos y estacionales, que pueden influir en la formación de ozono troposférico y otros procesos fotoquímicos.

En un primer análisis, los datos indican problemas de contaminación del aire en Bogotá, especialmente con partículas finas (PM2.5 y PM10) y otros contaminantes como NO2 y O3. Las condiciones meteorológicas, como la velocidad del viento moderada y la variabilidad en la precipitación, pueden influir en la acumulación y dispersión de contaminantes. Por lo cual, se realizó un análisis más detallado de las tendencias temporales y la relación entre variables meteorológicas y contaminantes para desarrollar estrategias efectivas de mitigación y control de la contaminación atmosférica en Bogotá.

### Proceso de Estequiometria

Los datos referentes a la concentración de diversos contaminantes atmosféricos como ozono (O3), dióxido de nitrógeno (NO2) y dióxido de azufre (SO2) fueron obtenidos de la RMCAB (Red de Monitoreo de Calidad del Aire de Bogotá, 2024), sin embargo, estos datos se encuentran inicialmente expresados en partes por billón (ppb). Sin embargo, para realizar una comparación adecuada con los estándares de calidad del aire recomendados por la Organización Mundial de la Salud (World Health Organization, 2021), que están expresados en microgramos por metro cúbico (µg/m³), como se puede observar en Tabla 1. Niveles de AQG recomendados y objetivos provisionales, fue necesario convertir las unidades de medida de los contaminantes de ppb a µg/m³ mediante procedimientos de estequiometría.

|  |  |  |
| --- | --- | --- |
| Contaminante | Tiempo promedio | Nivel AQG |
|  | 24 horas | 15 |
|  | 24 horas | 45 |
|  | 8 horas | 100 |
|  | 24 horas | 25 |
|  | 24 horas | 40 |
|  | 24 horas | 4 |

Tabla 1. Niveles de AQG recomendados y objetivos provisionales (World Health Organization, 2021)

La conversión de las unidades se llevó a cabo utilizando el siguiente proceso para los contaminantes ozono (O3), dióxido de nitrógeno (NO2) y dióxido de azufre (SO2):

1. **Fundamento Teórico de la Conversión de Unidades**

La conversión de unidades de los contaminantes gaseosos de partes por billón (ppb) a microgramos por metro cúbico (µg/m³) se fundamenta en la ley de los gases ideales, que relaciona la presión, el volumen y la temperatura. Dicha ley establece que , donde:

* + P es la presión atmosférica en atmósferas (atm),
  + 𝑉 es el volumen en litros (L),
  + 𝑛 es la cantidad de sustancia en moles,
  + 𝑅 es la constante universal de los gases ideales, ,
  + 𝑇 es la temperatura en Kelvin (K)

A partir de esta relación, el volumen se expresa como:

1. **Conversión de la Presión Barométrica a Atmósferas**

Aunque los datos obtenidos están en presión barométrica (mmHg), la fórmula de los gases ideales requiere la presión en atmósferas. Por lo tanto, es esencial convertir estos valores utilizando la equivalencia entre mmHg y atm:

donde 760 mmHg equivale a 1 atm.

1. **Ajuste de la Temperatura a Escala Kelvin**

La fórmula de los gases ideales también requiere que la temperatura esté en grados Kelvin. Para convertir grados Celsius (C) a Kelvin (K), se utiliza la siguiente relación:

1. **Cálculo de Concentraciones en ppb**

Para calcular la concentración de contaminantes en partes por billón, se utiliza la siguiente relación, considerando la densidad y el peso molecular de cada gas:

1. **Cálculo de la masa molar**

El cálculo de la masa molar es fundamental para la conversión de las concentraciones de contaminantes de partes por billón (ppb) a microgramos por metro cúbico (µg/m³). Se inicia con el ozono (O3), que consiste en tres átomos de oxígeno. Dado que la masa atómica del oxígeno es de 16 gramos por mol, la masa molar del O3 se calcula como sigue:

De manera similar, calculamos la masa molar para los siguientes gases:

* + **Dióxido de nitrógeno (NO2):** Compuesto por un átomo de nitrógeno (14 g/mol) y dos átomos de oxígeno (16 g/mol cada uno), resultando en:
  + **Dióxido de azufre (SO2):** Compuesto por un átomo de azufre (32 g/mol) y dos átomos de oxígeno (16 g/mol cada uno), resultando en:

1. **Conversión de unidades**

ara realizar la conversión de las concentraciones de contaminantes de ppb a µg/m³, se utiliza la siguiente relación que incorpora la ley de los gases ideales y los ajustes de las unidades de medida:

Luego se realiza la conversión de unidades:

Adicionalmente, los datos nos proporcionan los niveles de los contaminantes CO y CO2 están en partes por millón (ppm). Por lo tanto, es necesario llevar a cabo un proceso similar al descrito anteriormente, donde para calcular la concentración de contaminantes en partes por millón, se utiliza la siguiente relación, considerando la densidad y el peso molecular de cada gas:

Además, de realizar el cálculo de la masa molar:

* **Monóxido de carbono (CO):** Formado por un átomo de carbono (12 g/mol) y un átomo de oxígeno (16 g/mol), resultando en:
* **El dióxido de carbono (CO2):** Formado por un átomo de carbono (12 g/mol) y dos átomos de oxígeno (16 g/mol), resultando en:

Una vez convertidas las concentraciones de los contaminantes a µg/m³ o mg/m³, se puede proceder a comparar estos valores con los límites establecidos por la OMS para determinar si los niveles de contaminación superan los estándares de calidad del aire, lo cual indicaría un problema ambiental significativo.

### Análisis exploratorio de la correlación entre las variables

El análisis de correlación entre las variables es un paso fundamental en la comprensión de las interrelaciones y dinámicas dentro del conjunto de datos ambientales y de calidad del aire. La matriz de correlación de Pearson se utiliza para cuantificar la intensidad y la dirección de las relaciones lineales entre las variables. Esta herramienta estadística permite identificar asociaciones potenciales que pueden guiar posteriores análisis y modelos econométricos. El coeficiente de correlación de Pearson oscila entre -1 y 1, donde valores cercanos a 1 indican una fuerte correlación positiva, valores cercanos a -1 una fuerte correlación negativa, y valores cercanos a 0 una correlación débil o inexistente. En este estudio, la matriz de correlación ayudará a entender cómo se relacionan entre sí diversos contaminantes y factores meteorológicos, lo cual es crucial para el desarrollo del modelo Probit.

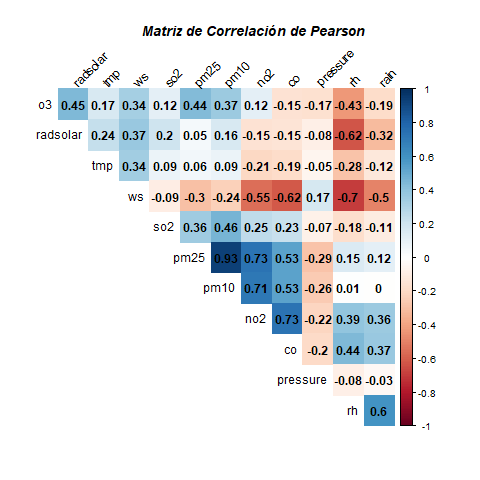


Ilustración 1. Matriz de correlación de Pearson

Al observar la matriz de correlación, se identificaron varias relaciones significativas entre las variables de estudio.

1. Correlaciones fuertes positivas

* PM2.5 y PM10 (0.93): Se encontró una fuerte relación positiva entre las partículas finas (PM2.5) y las partículas más gruesas (PM10). Esta alta correlación es esperada, ya que ambas variables son medidas de contaminación del aire por partículas, y es común que coexistan en la atmósfera debido a fuentes de emisión similares.
* NO2 y PM10 (0.71), NO2 y PM2.5 (0.73): Estas correlaciones indican una fuerte asociación entre el dióxido de nitrógeno y las partículas en suspensión. Esta relación probablemente se deba a fuentes comunes de emisión, como el tráfico vehicular, que es una fuente significativa tanto de NO2 como de partículas.

1. Correlaciones moderadas positivas

* SO2 y PM2.5 (0.46), SO2 y PM10 (0.46): Se observa una relación moderada entre el dióxido de azufre y las partículas en suspensión. Esta correlación sugiere que ambas variables pueden tener fuentes industriales comunes, como la quema de combustibles fósiles en plantas industriales.
* O3 y radiación solar (0.45): Esta correlación es esperada, ya que la radiación solar contribuye a la formación de ozono troposférico a través de reacciones fotoquímicas.

1. Correlaciones negativas notables

* Velocidad del viento (WS) y PM10 (-0.55), WS y PM2.5 (-0.55): La velocidad del viento muestra una correlación negativa significativa con las partículas en suspensión, lo que sugiere que mayores velocidades de viento ayudan a dispersar estas partículas, reduciendo su concentración en el aire.
* WS y NO2 (-0.62): De manera similar, la velocidad del viento también dispersa el dióxido de nitrógeno, resultando en una correlación negativa con esta variable.

1. Correlaciones débiles o no significativas:

* Muchas variables muestran correlaciones débiles entre sí (valores cercanos a 0), indicando la ausencia de relaciones lineales fuertes. Por ejemplo, la temperatura (tmp) tiene correlaciones débiles con la mayoría de los contaminantes, exceptuando una correlación negativa débil con SO2 (-0.21).
* La presión atmosférica muestra correlaciones débiles con la mayoría de las variables, sugiriendo que no es un factor determinante en los niveles de contaminación en este conjunto de datos.
* La humedad relativa (rh) presenta una correlación moderada positiva con el monóxido de carbono (CO) (0.6), lo cual podría indicar que condiciones más húmedas están asociadas con mayores niveles de CO, posiblemente debido a cambios en la dispersión y reactividad química en el aire.

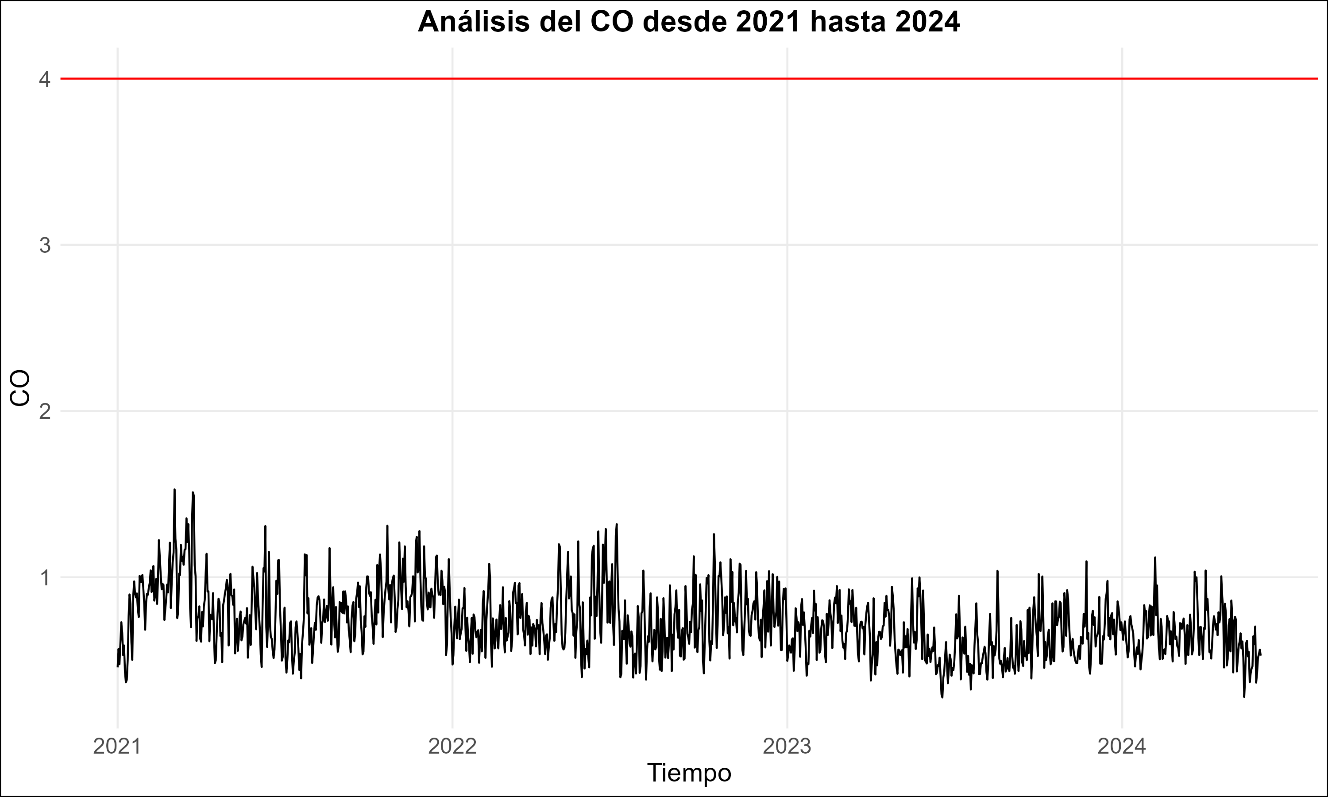
La matriz de correlación revela patrones importantes en la interrelación de contaminantes atmosféricos y factores meteorológicos. Las correlaciones más fuertes se observan entre diferentes tipos de partículas y entre NO2 y partículas, mientras que la velocidad del viento muestra un efecto dispersor significativo. Estos resultados son valiosos para comprender la dinámica de la contaminación del aire.

## Análisis de Contaminantes Atmosféricos

Un objetivo del presente estudio es evaluar la calidad del aire en Bogotá, con especial atención en cómo ciertos contaminantes atmosféricos exceden los niveles recomendados por la Organización Mundial de la Salud (OMS). Para este propósito, se han analizado datos de la Red de Monitoreo de la Calidad del Aire de Bogotá (RMCAB) en el periodo comprendido entre el 1 de enero de 2021 y el año 2024. Este análisis incluye la medición de varios contaminantes como el monóxido de carbono (CO), dióxido de nitrógeno (NO2), partículas finas PM2.5 y PM10, dióxido de azufre (SO2), ozono (O3), y carbono negro (BC), aunque este último no cuenta con un límite establecido por la OMS.

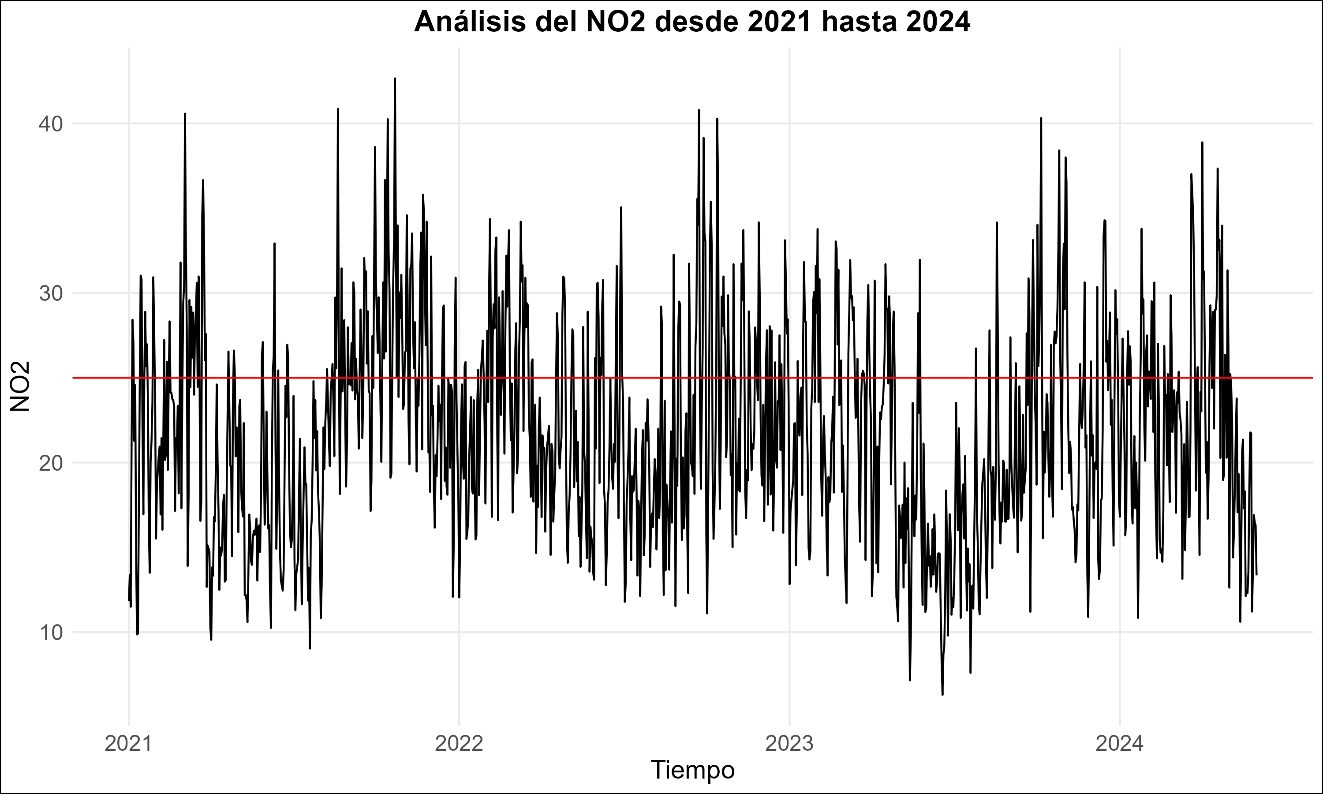
Para visualizar y analizar los niveles de contaminantes, se realizaron gráficos de series de tiempo que muestran las concentraciones diarias de cada contaminante. En cada gráfico, se ha trazado una línea roja representativa del límite máximo recomendado por la OMS para cada contaminante, excepto para el BC y el dióxido de carbono (CO2), donde el CO2 no cuenta con un límite OMS establecido y para el BC, el interés reside en observar su comportamiento a lo largo del tiempo. Cabe destacar que los datos de CO2 se dejaron de recopilar a comienzos de 2022 y los datos de BC están disponibles solo desde mediados de 2023.

* **Análisis del Monóxido de Carbono (CO) desde 2021 hasta 2024**

****

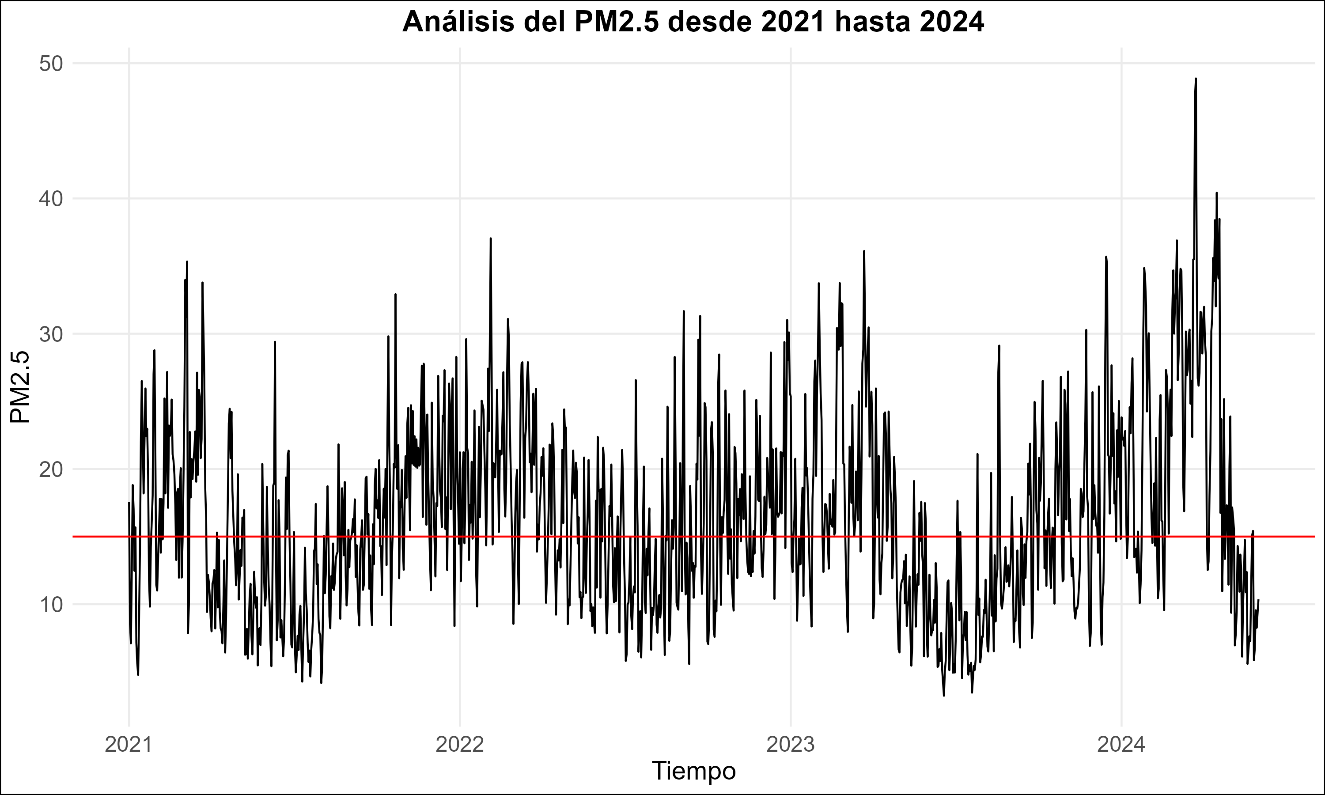
En la gráfica de CO, observamos que los niveles se mantienen por debajo del umbral de la OMS, fijado en 4 (World Health Organization, 2021). Esto indica un control general efectivo de este contaminante en el periodo analizado, aunque existen algunos picos esporádicos que sugieren incidentes de corta duración con niveles elevados.

* **Análisis del Dióxido de Nitrógeno (NO2) desde 2021 hasta 2024**

****

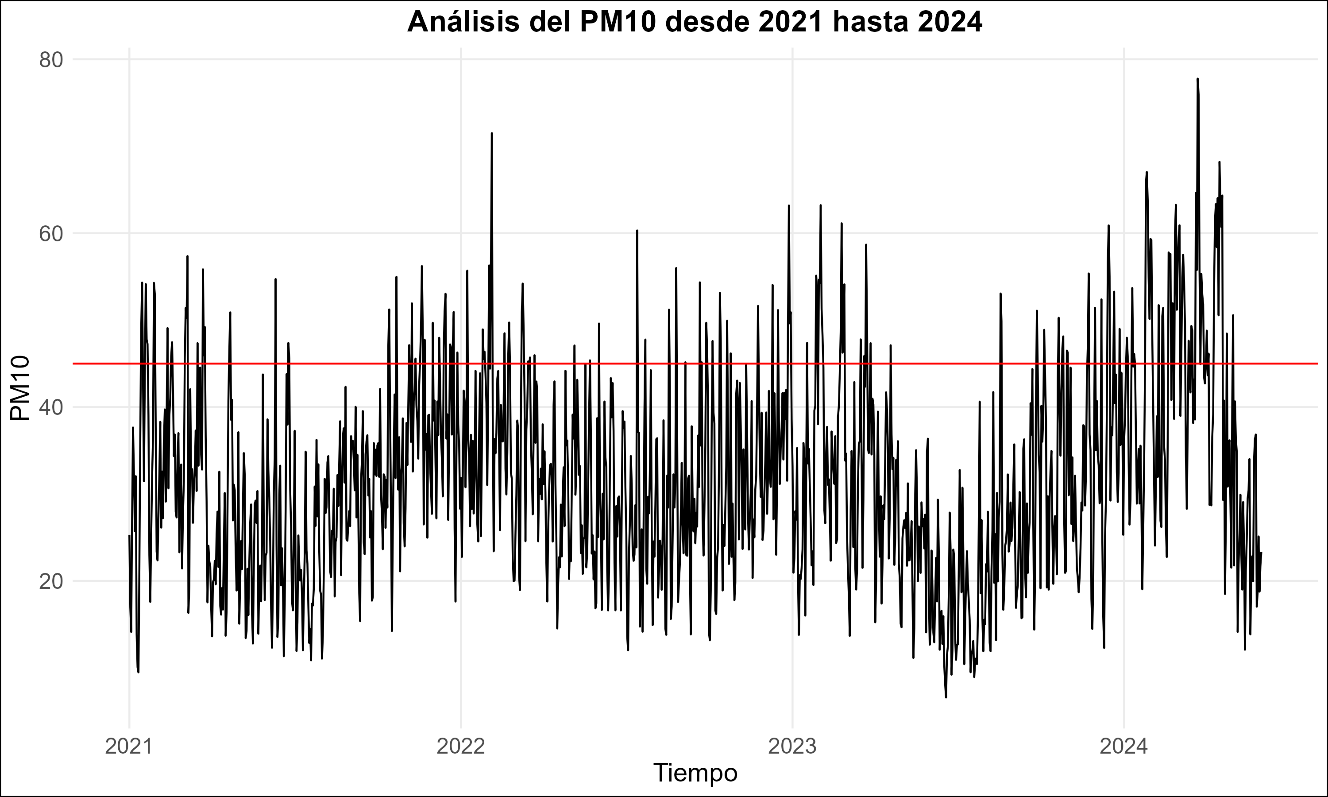
La gráfica del NO2 muestra numerosos picos que superan el límite establecido por la OMS de 25 (World Health Organization, 2021). Este patrón sugiere episodios recurrentes de alta contaminación por NO2 en Bogotá, lo que podría conllevar a graves problemas de salud pública relacionados con este contaminante.

* **Análisis del PM2.5 desde 2021 hasta 2024**

****

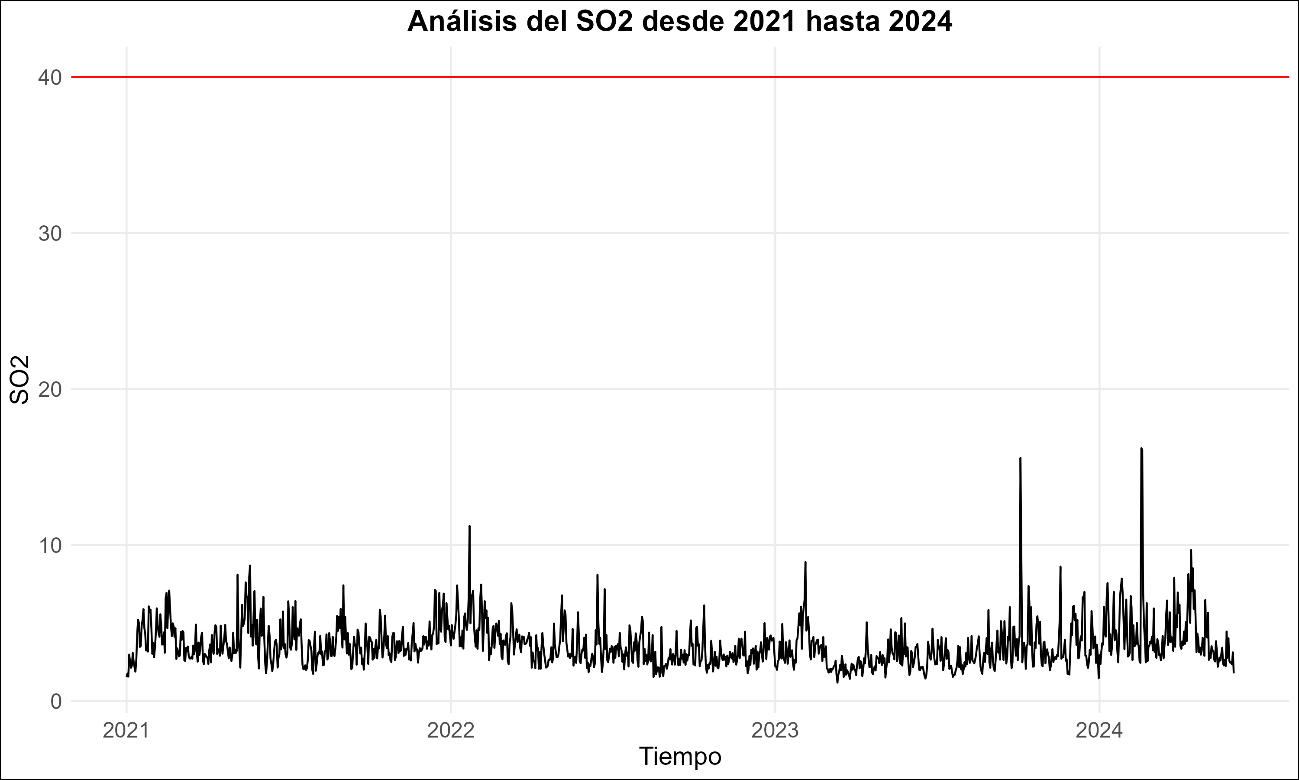
En la serie temporal de PM2.5, se aprecia una variabilidad significativa, con múltiples instancias donde las concentraciones exceden el límite de la OMS de 15 (World Health Organization, 2021). Esto refleja la presencia continua de partículas finas en el aire de Bogotá, posiblemente atribuible a fuentes como la combustión de vehículos y actividades industriales.

* **Análisis del PM10 desde 2021 hasta 2024**



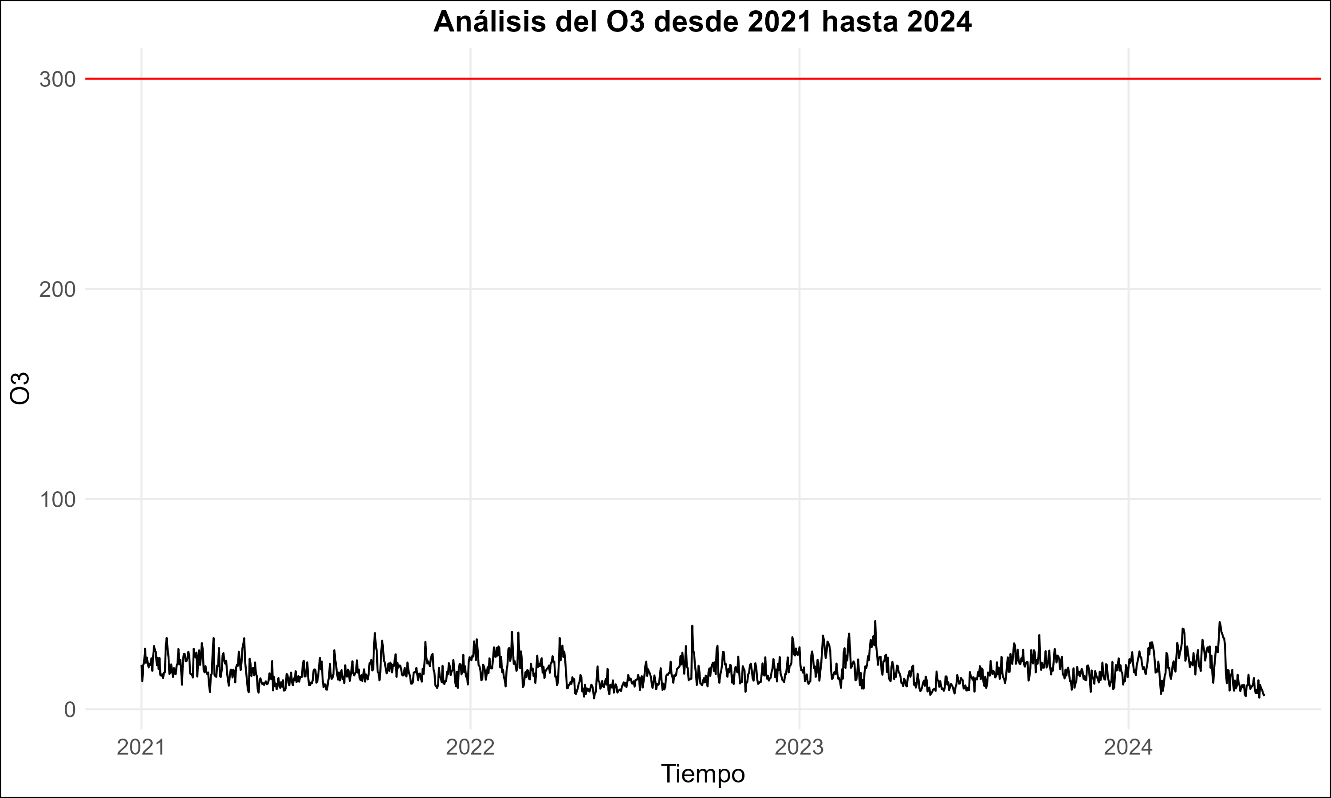
Similar al PM2.5, el PM10 también exhibe fluctuaciones significativas y frecuentemente sobrepasa el límite de la OMS de 45 (World Health Organization, 2021). Esto resalta la problemática persistente de partículas más gruesas en el ambiente, que pueden provenir de la construcción, el tráfico, y otras fuentes de polvo.

* **Análisis del Dióxido de Azufre (SO2) desde 2021 hasta 2024**

****

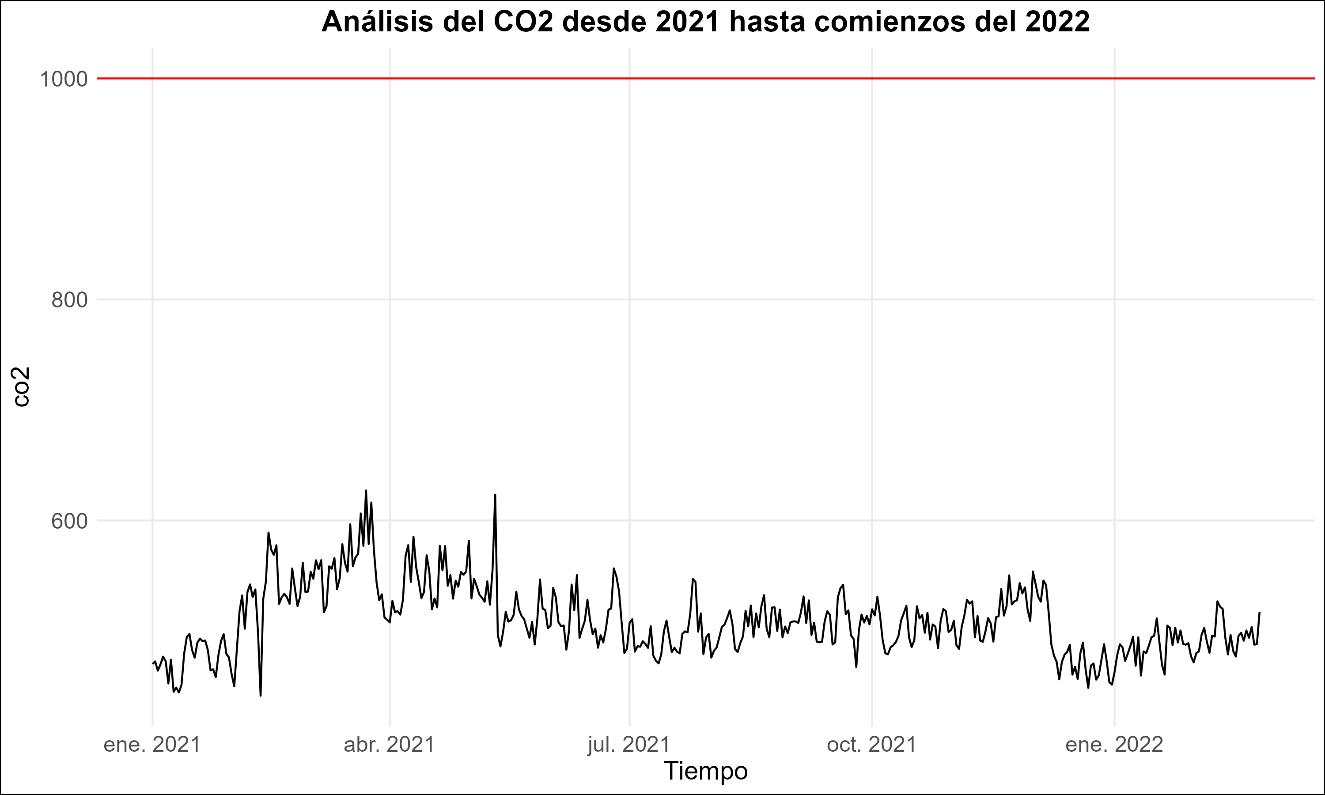
El SO2 muestra niveles bajos y controlados, permaneciendo en por debajo del límite de la OMS de 40 (World Health Organization, 2021). Los picos observados pueden estar vinculados a actividades industriales específicas y son eventos relativamente aislados.

* **Análisis del Ozono (O3) desde 2021 hasta 2024**



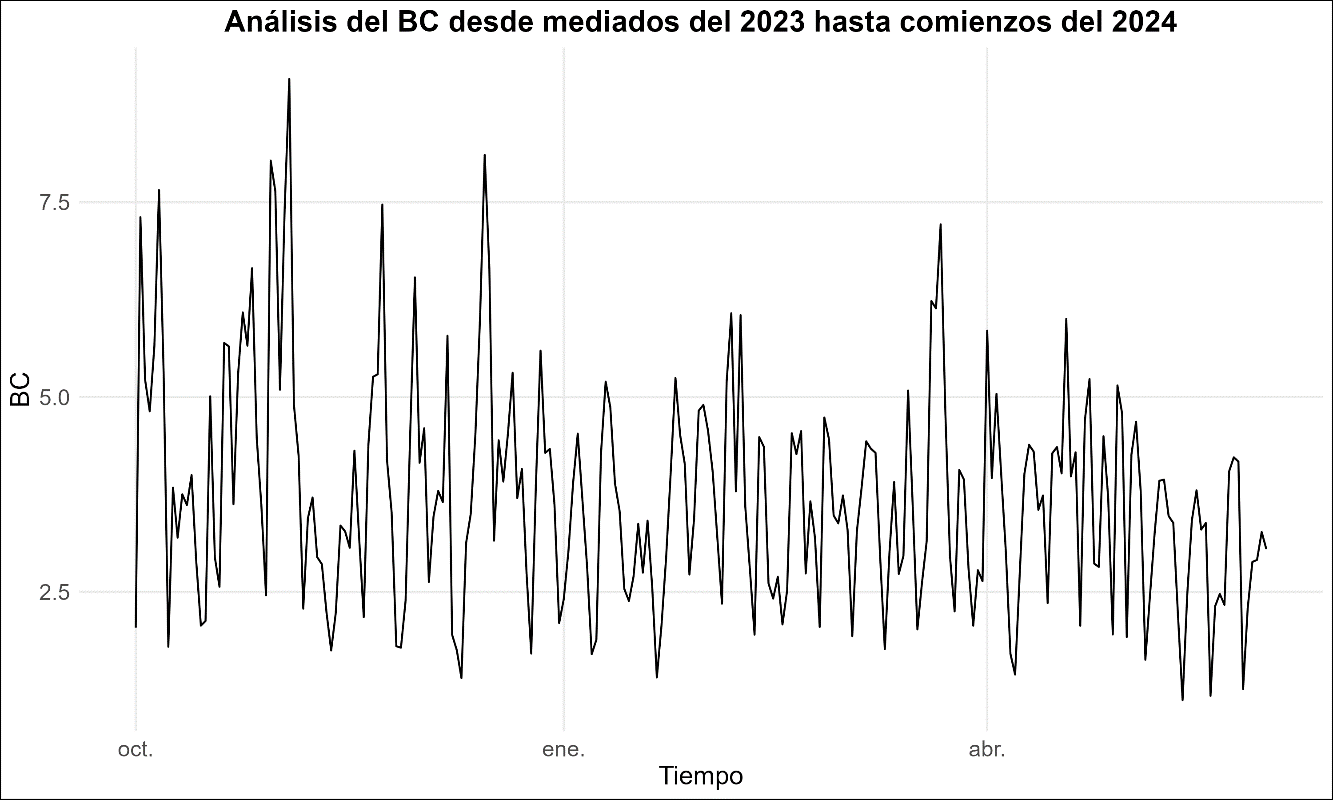
La gráfica del O3 destaca por mantenerse siempre por debajo del límite de la OMS de 300 (World Health Organization, 2021). Esto sugiere que, aunque el ozono puede ser un contaminante preocupante en muchos entornos urbanos, no parece ser un problema mayor en términos de exceder las directrices de la OMS en Bogotá durante el periodo estudiado.

* **Análisis del Dióxido de Carbono (CO2) desde 2021 hasta comienzos de 2022**

****

Aunque el CO2 no cuenta con un límite establecido por la OMS, es importante monitorear sus concentraciones por su relación con el cambio climático y la calidad del aire interior. La serie termina a comienzos de 2022, y los niveles mostrados no representan un riesgo directo para la salud bajo directrices de la OMS, por otro lado, empresas especializadas en ventilación establecen un límite de 1000 mg/m (Siber, 2024); sin embargo, este límite está muy lejos de alcanzarse en Bogotá.

* **Análisis del Carbono Negro (BC) desde mediados de 2023 hasta comienzos de 2024**



El BC no tiene un umbral establecido por la OMS, pero su monitoreo es crucial dado su impacto potencial en la salud respiratoria y cardiovascular. Los datos muestran una variabilidad considerable, lo que puede indicar fuentes de emisión no constante como el tráfico vehicular pesado y la quema de biomasa.

A través de las gráficas de series de tiempo, se observa que los niveles de varios contaminantes frecuentemente exceden los límites recomendados por la OMS, indicando periodos de alta contaminación que podrían representar un riesgo significativo para la salud pública en Bogotá. La variabilidad en las concentraciones sugiere la influencia de factores estacionales, así como de eventos específicos que pueden aumentar temporalmente la contaminación. El PM2.5, el PM10 y el NO2 muestran picos significativos que superan los límites establecidos, lo que refleja episodios de alta contaminación.

El análisis de las series temporales de los contaminantes en Bogotá ha mostrado que los niveles de PM2.5, PM10 y NO2 a menudo superan los límites recomendados por la Organización Mundial de la Salud (OMS). Esta observación ha motivado centrar el estudio en estos tres contaminantes, dado su impacto significativo en la salud pública. La elección se justifica por la necesidad de comprender a fondo los factores ambientales y meteorológicos que contribuyen a la superación de estos umbrales y el comportamiento de estos contaminantes. Por lo tanto, enfocarse en PM2.5, PM10 y NO2 permitirá un análisis más exhaustivo sobre las dinámicas específicas que los caracterizan.

## Modelos propuestos

### Modelo Probit

En el marco de esta investigación sobre la calidad del aire en Bogotá, se ha procedido a evaluar la probabilidad de que las concentraciones de ciertos contaminantes atmosféricos excedan los umbrales establecidos por la Organización Mundial de la Salud (OMS). El enfoque se centra en los contaminantes PM2.5, PM10 y NO2, dado que las series temporales indicaron que estos son los que más frecuentemente superan los niveles recomendados.

Para la modelación de estos eventos de excedencia, se ha empleado un modelo Probit, una técnica estadística frecuentemente utilizada en la econometría para modelar variables dependientes binarias. En nuestro contexto, estas variables binarias representan la ocurrencia o no de una excedencia del límite de contaminante. La variable dependiente se define como 1 si la concentración diaria del contaminante supera el umbral de la OMS, y 0 en caso contrario. Para el PM2.5, el umbral se estableció en 15 µg/m³, mientras que para el NO2 fue de 25 µg/m³ y para el PM10 de 45 µg/m³.

El modelo Probit se ajusta utilizando como variables independientes una serie de factores ambientales y temporales que se presume influyen en los niveles de contaminación. Estos incluyen la precipitación (rain), la temperatura (tmp) y la velocidad del viento (ws), junto con variables categóricas que representan el día de la semana (dow), el mes (month) y el año (year). Estas variables permiten controlar por efectos estacionales y otros patrones temporales que podrían afectar las mediciones de los contaminantes.

El ajuste del modelo se lleva a cabo a través de la función generalizada lineal (GLM) con enlace probit, diseñada específicamente para tratar con distribuciones de error no estándar y relaciones no lineales, adecuada para nuestro análisis de datos binarios.

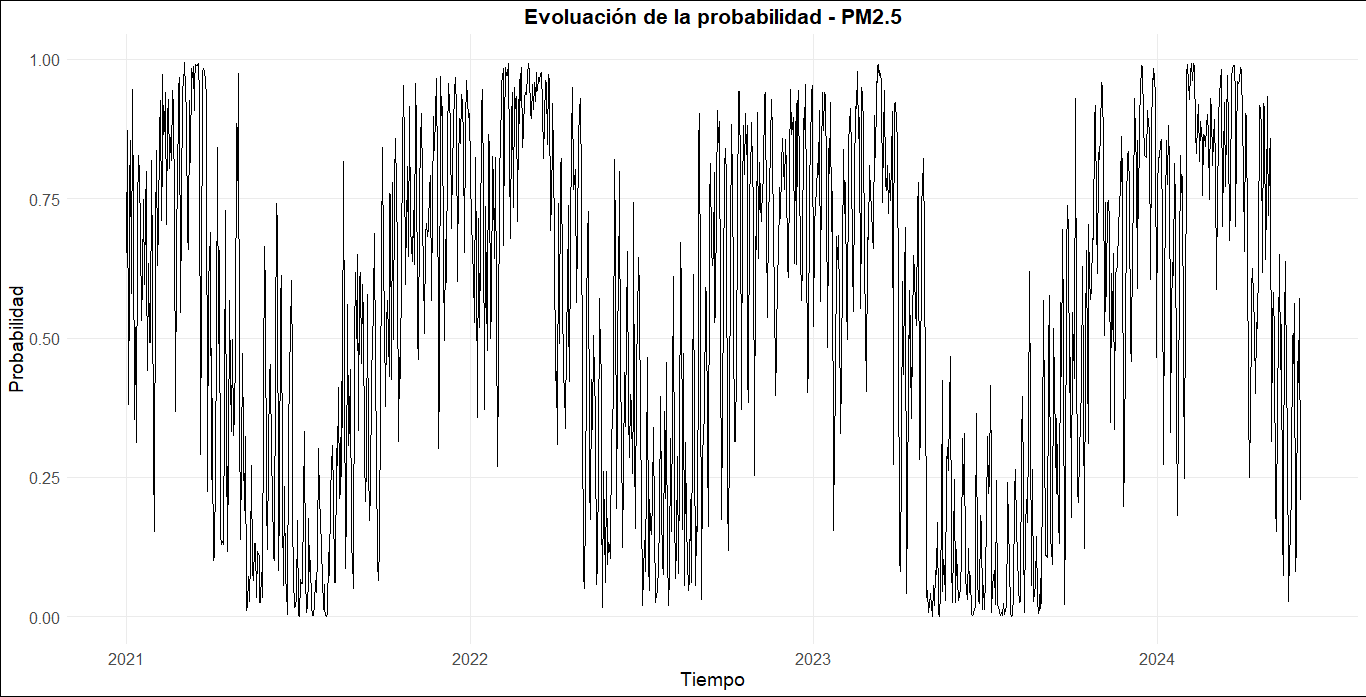
Una vez estimado el modelo, se procede a la evaluación de los ajustes y la significancia de los predictores, lo que permite obtener una comprensión más detallada de los factores que contribuyen a las excedencias. Además, se calculan las probabilidades ajustadas de excedencia para cada día del periodo de estudio, lo que facilita la visualización y la interpretación de las tendencias a lo largo del tiempo mediante gráficos de línea.

* **Modelo Probit para el PM2.5** [Ver Salida de la consola 15. Modelo Probit PM2.5]

A partir del modelo Probit para el PM2.5, al interpretar los coeficientes se encontró que:

* + **Rain (Lluvia):** El coeficiente para la lluvia es negativo (-0.199), aunque no es estadísticamente significativo (p>0.1). Esto sugiere que hay una ligera tendencia a que las excedencias disminuyan con la lluvia, pero no es una relación fuerte o consistente.
  + **Tmp (Temperatura):** La temperatura tiene un coeficiente positivo (0.039) con significancia estadística (p<0.05), indicando que aumentos en la temperatura están levemente asociados con un aumento en la probabilidad de excedencias de PM2.5.
  + **Ws (Velocidad del Viento):** Muestra un coeficiente significativamente negativo (-2.314, p<0.01), lo que indica que mayores velocidades de viento están fuertemente asociadas con menores probabilidades de excedencia de PM2.5. Esto es consistente con la idea de que el viento dispersa los contaminantes.
  + **Dow (Día de la Semana):** Varios días de la semana muestran diferencias significativas en la probabilidad de excedencias comparados con el lunes. Los viernes (dow5) muestran el mayor aumento (0.723, p<0.01), lo cual podría estar relacionado con patrones de tráfico y actividad industrial semanal.
  + **Month (Mes):** Algunos meses muestran aumentos significativos en las probabilidades de excedencia, como marzo (0.711, p<0.01) y diciembre (0.591, p<0.01), lo que podría reflejar condiciones estacionales específicas que afectan los niveles de PM2.5.
  + **Year (Año):** Cada año posterior al año base (2021) muestra un aumento en la probabilidad de excedencias, siendo más pronunciado en 2024 (0.641, p<0.01), lo que sugiere que los factores a lo largo de los años podrían estar incrementando los riesgos de contaminación o que las medidas de control existentes no están siendo efectivas en la reducción de eventos de excedencia.

Sumado a esto, la gráfica muestra la probabilidad predicha de excedencia de los niveles de PM2.5 sobre el umbral establecido por la OMS en Bogotá, según estimaciones del modelo Probit, a lo largo del periodo de 2021 a 2024.



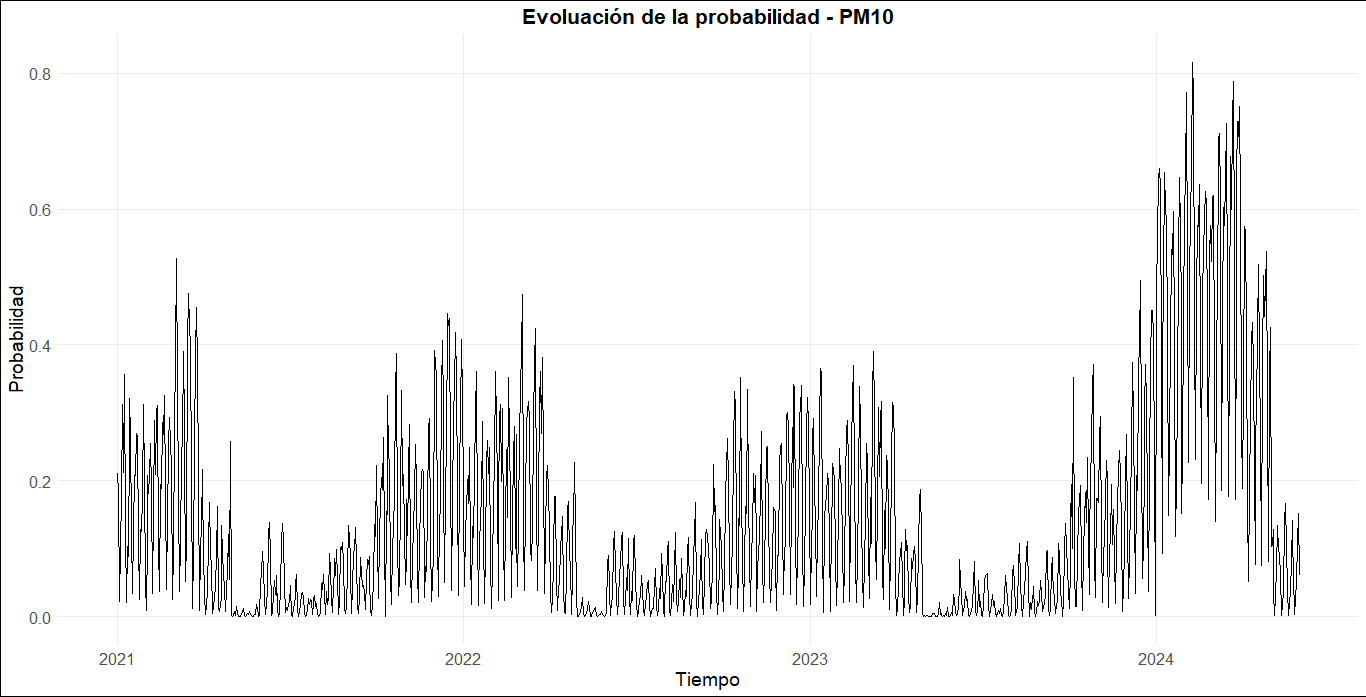
Se puede observar que:

* + **Variabilidad Temporal:** La línea en la gráfica oscila significativamente a lo largo del tiempo, lo que sugiere que la probabilidad de exceder los niveles recomendados de PM2.5 varía considerablemente de un día a otro. Esta variabilidad puede estar influenciada por cambios en las condiciones meteorológicas, actividades humanas y variaciones estacionales, como se refleja en los coeficientes del modelo Probit para variables como la temperatura, la velocidad del viento y las categorías temporales (días y meses).
  + **Rango de Probabilidades:** Las probabilidades varían entre casi 0 (cero) y 1 (uno), indicando días con muy baja probabilidad de excedencia hasta días con probabilidad casi cierta de exceder los niveles de PM2.5.
  + **Tendencias Anuales:** Aunque es difícil discernir una tendencia clara a largo plazo debido a la variabilidad diaria, se puede observar que, en algunos períodos específicos del año, como se ha visto reflejado en la configuración del modelo, la probabilidad tiende a ser más alta, posiblemente relacionada con condiciones estacionales específicas.
  + **Implicaciones para Política Ambiental y Salud Pública:** Esta gráfica puede servir como una herramienta valiosa para los responsables de la formulación de políticas y las autoridades de salud pública. La capacidad de prever días con alta probabilidad de excedencia puede facilitar la implementación de medidas preventivas, como advertencias a la población, restricciones temporales en actividades industriales o de construcción, y regulaciones de tráfico para mitigar los niveles de PM2.5.
* **Modelo Probit para el PM10** [Ver Salida de la consola 16. Modelo Probit PM10]

A partir del modelo Probit para el PM10, al interpretar los coeficientes se encontró que:

* + **Rain (Lluvia):** Coeficiente de -0.808 (p<0.01), indicando una relación negativa significativa. A medida que aumenta la lluvia, la probabilidad de exceder el umbral de PM10 disminuye, probablemente debido a que la lluvia ayuda a asentar las partículas en suspensión.
  + **Tmp (Temperatura):** Coeficiente positivo de 0.040, aunque no es estadísticamente significativo (p>0.05). Sugiere una leve tendencia a que la probabilidad de excedencias aumente con la temperatura, posiblemente relacionado con patrones de estabilidad atmosférica que favorecen la acumulación de partículas.
  + **Ws (Velocidad del Viento):** Coeficiente de -1.089 (p<0.01), similar al efecto en PM2.5, mostrando que velocidades más altas del viento reducen la probabilidad de excedencias de PM10. El viento dispersa las partículas, reduciendo las concentraciones.
  + **Dow (Día de la Semana):** Varias categorías muestran diferencias significativas. Los días entre semana (especialmente el miércoles, jueves y viernes) presentan un aumento en la probabilidad de excedencia en comparación con el lunes, lo que podría reflejar patrones de actividad semanal como tráfico o producción industrial.
  + **Month (Mes):** La probabilidad de excedencia tiende a ser más baja en mayo, junio, julio y agosto (todos con p<0.01), lo que podría estar asociado a patrones climáticos específicos como mayor precipitación o diferentes patrones de viento.
  + **Year (Año):** Los años posteriores al año base muestran un aumento gradual en la probabilidad, con un salto significativo en 2024 (coeficiente de 1.117, p<0.01), sugiriendo una tendencia creciente a lo largo del tiempo que podría estar influenciada por factores a largo plazo como el crecimiento urbano o cambios en normativas ambientales.

Adicionalmente, la gráfica muestra la probabilidad predicha de excedencia de los niveles de PM10 sobre el umbral establecido por la OMS en Bogotá, según estimaciones del modelo Probit, a lo largo del periodo de 2021 a 2024.



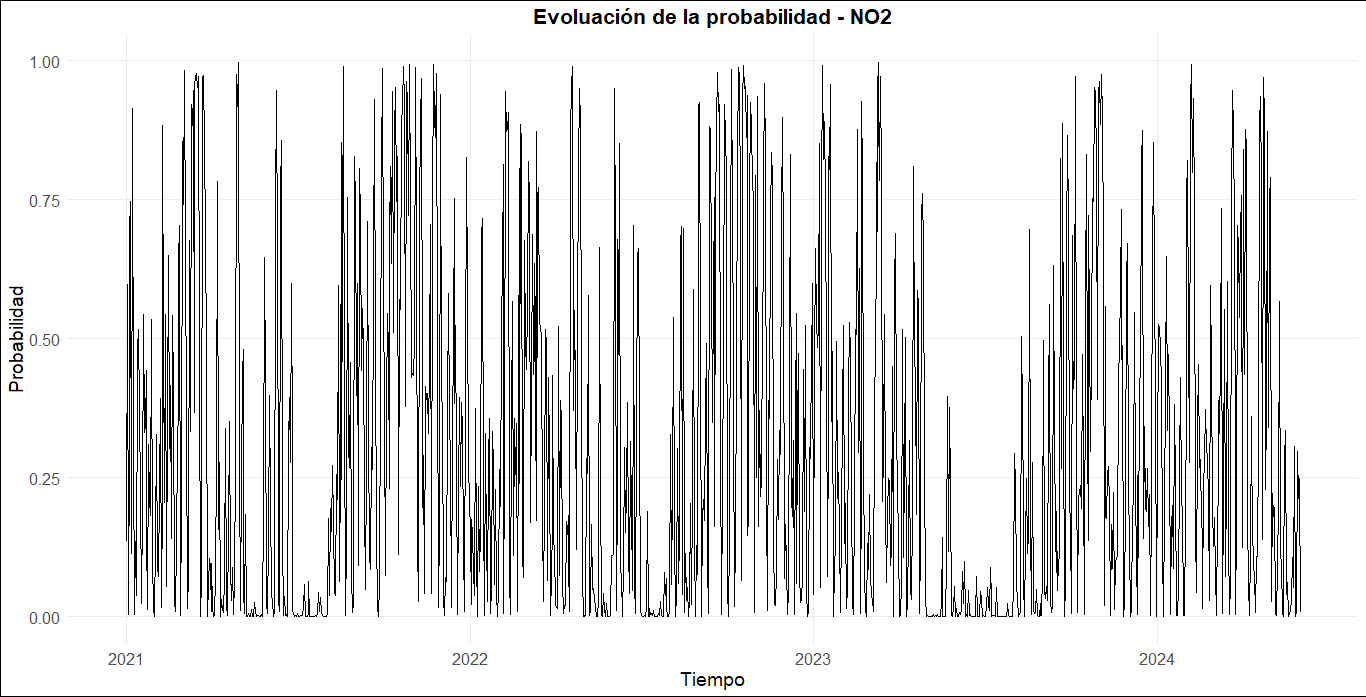
Se puede observar que:

* **Variabilidad y Tendencia:** La gráfica muestra una alta variabilidad diaria en la probabilidad de excedencia. Hacia el final del período, en 2024, hay un notable aumento en las probabilidades predichas, lo cual está en línea con el aumento significativo en el coeficiente del año 2024 observado en el modelo. Esto sugiere que los factores que contribuyen a las excedencias de PM10 podrían estar intensificándose con el tiempo.
* **Implicaciones:** El modelo y su visualización en la gráfica sugieren la necesidad de monitorizar continuamente las concentraciones de PM10, especialmente en anticipación a condiciones que se han identificado como contribuyentes a altas probabilidades de excedencia. La capacidad de prever estos eventos permite la implementación de medidas preventivas para proteger la salud pública y cumplir con las regulaciones ambientales.
* **Modelo Probit para el NO2** [Ver Salida de la consola 17. Modelo Probit NO2]

Estos coeficientes detallan cómo diferentes variables influyen en la probabilidad de que los niveles de NO2 excedan el umbral establecido.

* + **Rain (Lluvia):** Coeficiente positivo de 0.620 (p<0.05), lo que indica que, a diferencia de PM10 y PM2.5, la lluvia aumenta la probabilidad de excedencias de NO2. Esto podría deberse a reacciones químicas en la atmósfera que aumentan la concentración de NO2 durante episodios de lluvia.
  + **Tmp (Temperatura):** Coeficiente negativo significativo de -0.248 (p<0.01). Un descenso en la temperatura está asociado con un aumento en la probabilidad de excedencia de NO2, posiblemente debido a la mayor estabilidad atmosférica que atrapa contaminantes cerca del suelo durante condiciones más frías.
  + **Ws (Velocidad del Viento):** Coeficiente negativo de -3.299, indicando que un aumento en la velocidad del viento reduce significativamente la probabilidad de excedencias de NO2, similar a su efecto en las partículas PM10 y PM2.5.
  + **Dow (Día de la Semana):** Los días laborables tienden a tener una mayor probabilidad de excedencia en comparación con el lunes. El efecto es particularmente fuerte los miércoles y jueves (0.712 y 0.625 respectivamente, p<0.01), reflejando posiblemente el tráfico vehicular incrementado y la actividad industrial.
  + **Month (Mes):** Mayo muestra un coeficiente negativo significativo de -1.705 (p<0.01), lo que sugiere una menor probabilidad de excedencia. En contraste, septiembre y octubre muestran incrementos significativos en la probabilidad de excedencia.
  + **Year (Año):** Hay un incremento notable en la probabilidad de excedencia en 2023 y 2024 (0.675 y 1.022 respectivamente, p<0.01), indicando una tendencia creciente a lo largo del tiempo.

Adicionalmente, la gráfica muestra la probabilidad predicha de excedencia de los niveles de NO2 sobre el umbral establecido por la OMS en Bogotá, según estimaciones del modelo Probit, a lo largo del periodo de 2021 a 2024.



Se puede observar que:

* + **Variabilidad Temporal:** La línea muestra variaciones diarias significativas, con algunos periodos y días específicos donde la probabilidad de excedencia es particularmente alta, incluso en algunos días la probabilidad de exceder los niveles impuestos por la OMS es del 100%.
  + **Tendencias Anuales:** La gráfica muestra un aumento general en la probabilidad de excedencias hacia el final del período en 2024, lo que coincide con el análisis de los coeficientes que muestran un aumento en la probabilidad en los años posteriores.

Dado que se ha demostrado que la variable de velocidad del viento (ws) tiene una influencia significativa en la probabilidad de que los niveles de PM2.5, PM10 y NO2 excedan los umbrales establecidos, se decidió graficar esta probabilidad en función de la velocidad del viento para los tres contaminantes. Además, para observar mejor la tendencia, se optó por segmentar los datos y mostrar solo una línea de tendencia.

|  |  |
| --- | --- |
| E | e |

Tanto en la nube de puntos como en la línea de tendencia se observa una consistencia con los coeficientes analizados, mostrando que mayores velocidades de viento están fuertemente asociadas con menores probabilidades de excedencia de PM2.5. Este patrón es congruente con la hipótesis de que el viento actúa dispersando los contaminantes.

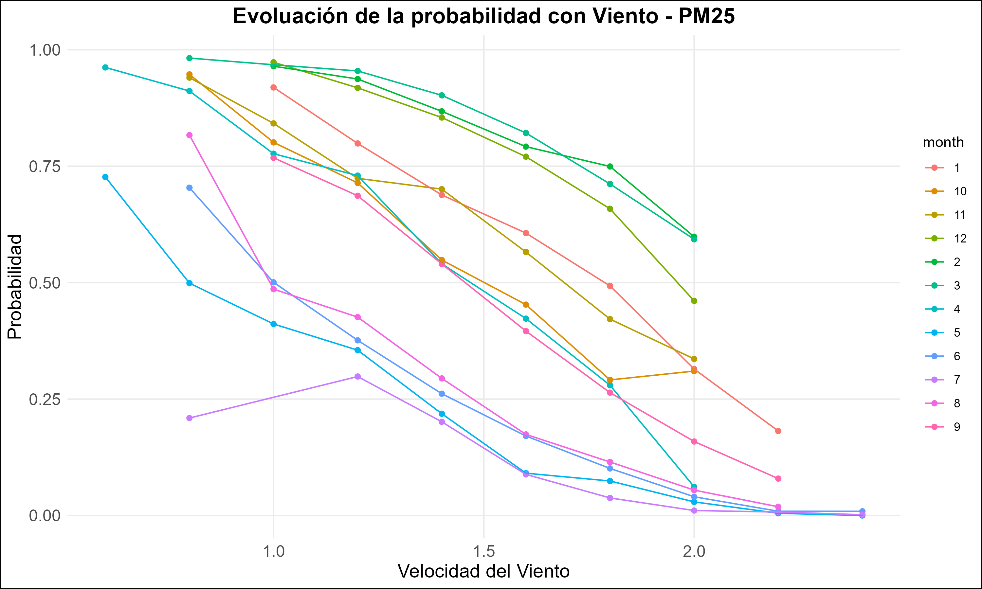
|  |  |
| --- | --- |
| e | E |

De manera similar, para el NO2 también se observa que a mayores velocidades del viento, la probabilidad de exceder los niveles establecidos por la OMS disminuye, siguiendo el mismo patrón que se observa en PM2.5.

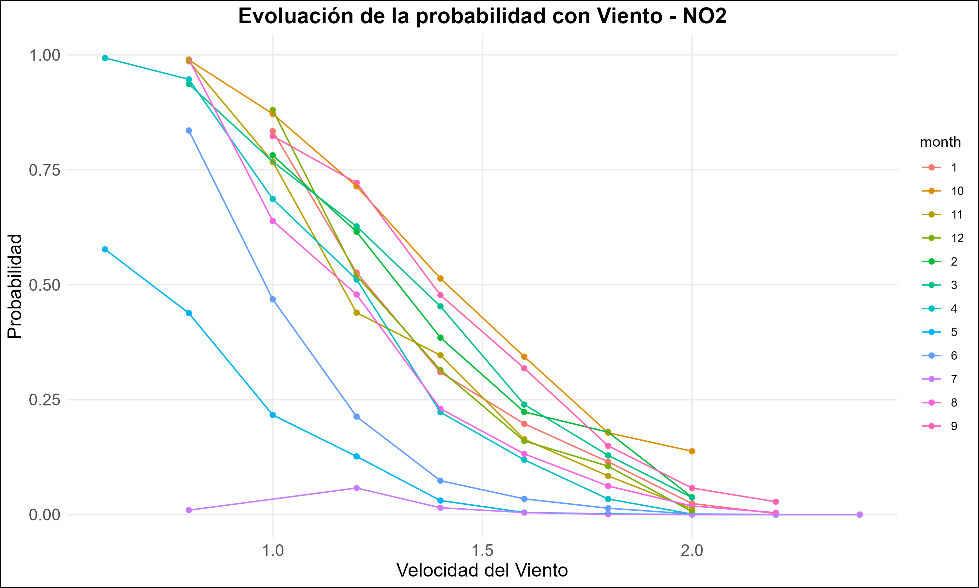
|  |  |
| --- | --- |
| e | e |

Para el PM10, aunque no es tan evidente en la gráfica de la nube de puntos, sí se observa en la gráfica de la línea de tendencia que, a mayores velocidades del viento, la probabilidad de exceder los niveles establecidos disminuye. Es posible que debido a que el PM10 es más pesado, se requiera una velocidad de viento significativamente mayor para reducir esta probabilidad.

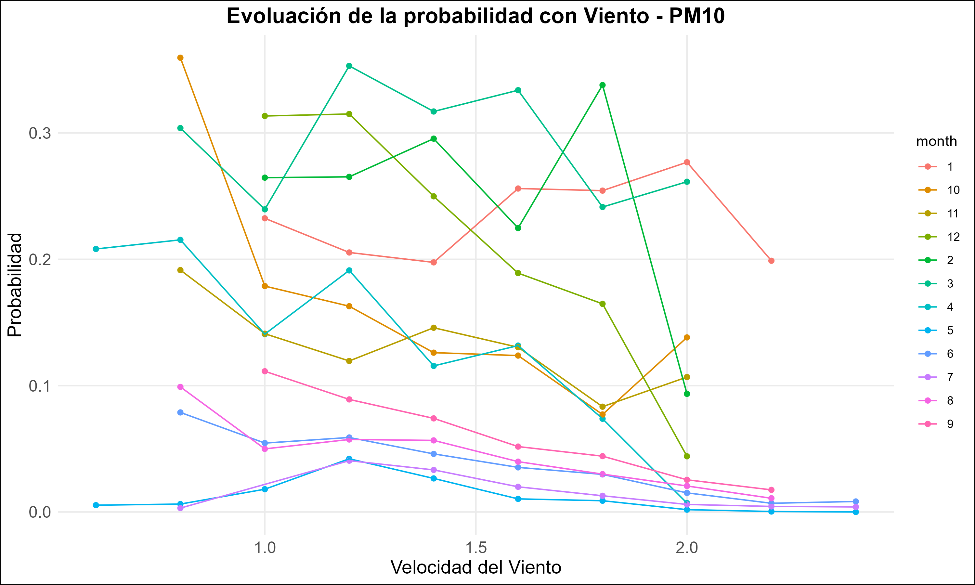
No obstante, dado que en los modelos Probit se pudo observar que los meses también ejercían una influencia significativa en los niveles de contaminación, se procedió a graficar la probabilidad de exceder los límites establecidos por la OMS en función de la velocidad del viento para cada mes. A continuación, se presentan las gráficas obtenidas:



La gráfica de PM2.5 muestra una clara relación negativa entre la velocidad del viento y la probabilidad de exceder los niveles de PM2.5 recomendados por la OMS. Cada línea representa un mes específico del año, revelando cómo la estacionalidad influye en esta relación. Para casi todos los meses, la probabilidad de excedencia disminuye notablemente a medida que aumenta la velocidad del viento. Este comportamiento sugiere que el viento juega un papel crucial en la dispersión de las partículas de PM2.5, reduciendo su concentración en el aire. Sin embargo, los meses de invierno, representados por líneas más altas en velocidades de viento bajas, indican que condiciones estacionales como la inversión térmica pueden inhibir este efecto dispersante, manteniendo altas concentraciones de PM2.5 a pesar de la presencia de viento.



En la gráfica correspondiente al NO2, la tendencia es similarmente decreciente, lo que demuestra que la velocidad del viento también reduce la probabilidad de excedencia de NO2, aunque las líneas de los diferentes meses están más agrupadas comparadas con las de PM2.5. Esto indica una menor variación estacional en el efecto del viento sobre las concentraciones de NO2. La reducción es particularmente notable en todos los meses a medida que la velocidad del viento se aproxima a 2.0 metros por segundo, resaltando el efecto significativo del viento en la dispersión de NO2. Este patrón reitera la importancia de la ventilación atmosférica en la gestión de la calidad del aire urbano.



La relación entre la velocidad del viento y la probabilidad de excedencia de PM10 por mes muestra una dinámica más variable. Aunque generalmente se observa una tendencia decreciente, algunos meses exhiben fluctuaciones o incluso un ligero aumento en la probabilidad a ciertas velocidades del viento. Esto podría reflejar la complejidad de las fuentes de PM10 y su comportamiento en el aire, que puede ser afectado por factores locales como la topografía y actividades urbanas específicas. Los meses de verano tienden a mostrar una disminución más constante y pronunciada, lo que podría asociarse con patrones climáticos que favorecen una mayor dispersión de partículas más grandes como las de PM10.

Estas gráficas son fundamentales para comprender cómo la velocidad del viento afecta la probabilidad de excedencia de diferentes contaminantes en Bogotá. Evidencian la importancia del viento como un factor natural de control de la calidad del aire, y cómo su efectividad puede variar considerablemente dependiendo del tipo de contaminante y la época del año.

#### Justificación de la metodología

En el contexto de la investigación sobre contaminación, el uso del modelo Probit para analizar la probabilidad de que las concentraciones de PM2.5 superen los umbrales recomendados por la Organización Mundial de la Salud se justifica por varias razones clave. El modelo Probit, perteneciente a la familia de modelos lineales generalizados, es especialmente adecuado para tratar con variables de respuesta binarias donde los eventos de interés tienen dos posibles resultados: ocurren o no ocurren. En este estudio, el evento de interés es la excedencia del umbral del contamínate, codificado como 1 si hay excedencia y 0 si no la hay.

El modelo Probit utiliza una función de enlace probit que asume que la variable latente subyacente, influida por las variables predictoras, sigue una distribución normal. Esta suposición es crucial porque permite una interpretación probabilística del efecto de las variables independientes sobre la probabilidad de excedencia, proporcionando así una base teórica sólida y coherente con los principios de normalidad distributiva en errores, comúnmente asumidos en econometría.

#### Ventajas del modelo Probit en Estudios de Contaminación

1. **Interpretación probabilística:** El modelo Probit facilita la interpretación de los coeficientes en términos de la probabilidad de que un evento ocurra. Esto es particularmente útil en estudios ambientales donde se necesita entender el impacto de factores como las condiciones meteorológicas y temporales en la probabilidad de eventos extremos de contaminación.
2. **Flexibilidad en la inclusión de variables:** El modelo Probit puede incorporar fácilmente variables categóricas y continuas. En el estudio de la contaminación, esto permite modelar la influencia de múltiples factores, incluidos patrones estacionales, condiciones diarias y variaciones anuales, proporcionando un análisis integral.
3. **Robustez frente a variaciones extremas:** Los modelos Probit son menos sensibles a las variaciones extremas de las variables predictoras en comparación con otros modelos binarios como el logit, lo que puede ser particularmente ventajoso cuando se manejan datos ambientales que a menudo presentan altos niveles de variabilidad y ruido.

#### Desventajas del modelo probit

1. **Suposición de normalidad:** Una de las principales limitaciones del modelo Probit es su suposición de que la variable latente subyacente sigue una distribución normal. Esta suposición puede no ser válida en todos los contextos, especialmente en datos ambientales que pueden no distribuirse normalmente debido a la presencia de picos o sesgos significativos en la respuesta a eventos específicos.
2. **Dificultad en la estimación:** La estimación de los parámetros en un modelo Probit puede ser más compleja y computacionalmente más intensiva en comparación con otros modelos lineales generalizados, como el modelo logístico. Esto puede resultar en un mayor tiempo de computación y la necesidad de mayor potencia computacional, especialmente con grandes conjuntos de datos.
3. **Sensibilidad a las especificaciones del modelo:** El desempeño del modelo Probit puede ser altamente sensible a la especificación correcta del modelo, incluyendo la elección y transformación de las variables predictoras. Una especificación inadecuada puede llevar a interpretaciones erróneas o a estimaciones sesgadas de los efectos de las variables.

### Modelo de Vectoriales autorregresivos (VAR)

En el marco del desarrollo metodológico para el análisis de la calidad del aire en Bogotá, se implementó un Modelo de Vectores Autorregresivos (VAR) para explorar las dinámicas temporales y las interdependencias entre diversas variables ambientales y meteorológicas, y su relación con los niveles de contaminación de PM2.5. La selección de variables para el modelo se basó en un análisis preliminar de sus distribuciones y correlaciones. Las variables PM2.5, PM10 y NO2 presentaron distribuciones asimétricas, con PM2.5 y PM10 siguiendo una distribución gamma y NO2 una distribución chi-cuadrado. Las variables meteorológicas (temperatura, radiación solar, presión atmosférica y velocidad del viento) se ajustaron a distribuciones normales.

Dada la alta correlación observada entre PM10 y NO2 con PM2.5, se excluyeron estas del modelo para evitar problemas de multicolinealidad, que podrían llevar a estimaciones ineficientes y sesgadas de los coeficientes del modelo. Se seleccionaron las variables de radiación solar (radsolar), velocidad del viento (ws), presión atmosférica (pressure) y temperatura (tmp) por su baja correlación mutua, lo que ayuda a evitar la sobre especificación del modelo y permite un análisis más claro del impacto independiente de cada variable sobre las concentraciones de PM2.5.

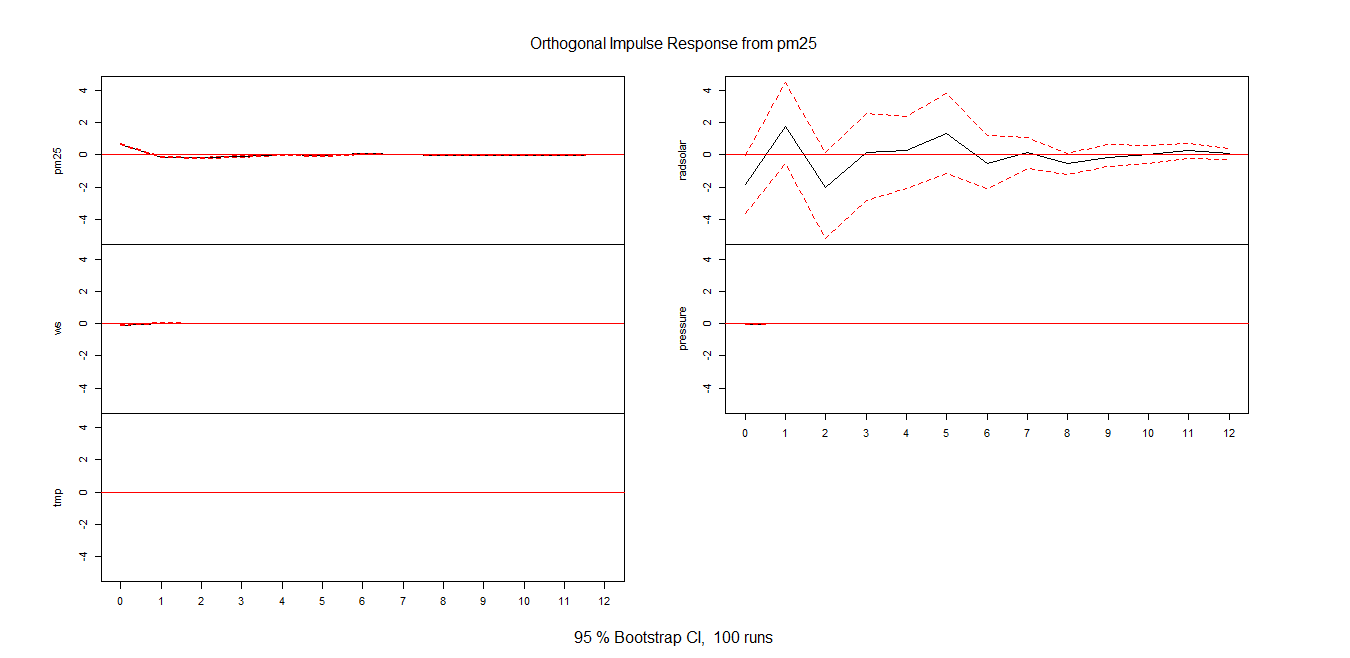
Para abordar la no estacionariedad observada en las series temporales, se procedió a diferenciar las series de tiempo. Este proceso ayuda a estabilizar la media de las series, eliminando tendencias y ciclos que podrían afectar la validez de los resultados del modelo VAR. La diferenciación transforma las series en estacionarias, condición necesaria para la aplicación efectiva de modelos VAR, asegurando que las relaciones estimadas no sean espurias, sino que reflejen verdaderas conexiones dinámicas entre las variables.

Se utilizó la función `VARselect` para determinar el número óptimo de retardos para el modelo, basándose en criterios como el Criterio de Información de Akaike (AIC), el Criterio de Hannan-Quinn (HQ), el Criterio de Schwarz (SC) y el Error de Predicción Final (FPE), resultando en recomendaciones variadas que oscilaron entre 3 y 6 retardos.

El análisis de Dickey-Fuller, aplicado posteriormente, confirmó la estacionariedad de todas las series después de la diferenciación, permitiendo proceder con la estimación del modelo VAR. Dado que el p-value es menor a una significancia del 5%, sugiere que tenemos evidencia significativa para rechazar la hipótesis nula de no estacionariedad (presencia de raíz unitaria) en favor de la hipótesis alternativa de estacionariedad.

La implementación del modelo VAR reveló que las concentraciones de PM2.5 son significativamente influenciadas por sus propios valores rezagados, así como por los rezagos de las variables meteorológicas seleccionadas, mostrando una dependencia particularmente fuerte del primer rezago de la radiación solar y de la velocidad del viento. Además, se observó que las concentraciones de PM2.5 son significativamente influenciadas por el segundo rezago de sí mismas, así como por el segundo rezago de la radiación solar y de la presión.

Este resultado sugiere un impacto directo y significativo de estos factores sobre la dispersión y concentración de partículas de PM2.5 en la atmósfera urbana.



Finalmente, la función de respuesta al impulso, generada a través de la librería `vars` en R, ofrece una visualización de cómo un choque en PM2.5 afecta sus propias dinámicas futuras y las de otras variables en el sistema a lo largo del tiempo. Los resultados indican que un aumento en PM2.5 tiende a ser moderado por las condiciones meteorológicas en sucesivos periodos, resaltando la interacción compleja y dinámica entre la contaminación del aire y los factores ambientales en un contexto urbano.

#### Justificación de la metodología

El modelo de Vectoriales Autorregresivos (VAR) se justifica como una metodología adecuada para este análisis por varias razones fundamentales:

* Modelo Multivariado: El VAR permite modelar simultáneamente múltiples series temporales, como los niveles de contaminación por PM2.5 y variables meteorológicas, que están interrelacionadas y pueden influenciarse mutuamente a lo largo del tiempo. Esto es crucial dado que estos contaminantes pueden compartir factores comunes, como la velocidad del viento y la estacionalidad, que afectan sus niveles.
* Flexibilidad en la Modelización: Permite capturar la dinámica compleja entre las variables al incorporar términos autorregresivos que reflejan la influencia pasada de cada variable sobre sí misma y sobre las otras variables en el modelo. Esto es esencial para entender cómo cambios en una variable pueden afectar las otras en el contexto del análisis de contaminación y calidad del aire.

#### Ventajas del modelo

* Captura de Interacciones Dinámicas: El VAR es efectivo para capturar las interacciones dinámicas entre las variables, como la velocidad del viento y los niveles de contaminación, permitiendo una evaluación precisa de cómo estos factores afectan los niveles de contaminación en diferentes períodos.
* Modelización Multivariada: Al considerar múltiples variables simultáneamente, el VAR evita problemas de sesgo y errores que podrían surgir al analizar cada serie de manera independiente, proporcionando una visión holística de la relación entre los contaminantes y sus factores determinantes.
* Flexibilidad en la Interpretación: Permite identificar y cuantificar rápidamente el impacto relativo de diferentes variables sobre los niveles de contaminación.

#### Desventajas del modelo

* Requiere Series Estacionarias: El VAR asume que las series temporales involucradas son estacionarias o pueden transformarse en estacionarias. Esto puede requerir preprocesamiento adicional de los datos, como la diferenciación, para cumplir con este requisito.
* Sensibilidad a Especificaciones Erróneas: La selección inadecuada de variables, rezagos o la especificación errónea del modelo pueden conducir a resultados sesgados o poco precisos. Se requiere un análisis cuidadoso y pruebas de diagnóstico para garantizar la adecuación del modelo.

### Modelo de Vector Error Correction Model (VECM)

Se empleó un Modelo de VECM para examinar la dinámica de ajuste a largo plazo entre las variables ambientales consideradas. Previo a la aplicación del VECM, se realizó un proceso de normalización y diferenciación de las series temporales para asegurar estacionariedad, un requisito previo fundamental para la aplicación de modelos de series temporales.

Para determinar la existencia de relaciones de equilibrio a largo plazo entre las series, se implementó la prueba de cointegración de Johansen. Esta técnica es esencial para identificar si un conjunto de variables no estacionarias están cointegradas, es decir, si comparten una dinámica común de largo plazo que las mantiene en equilibrio. Los resultados de la prueba indicaron la presencia de cointegración de rango completo, lo que sugiere que existe al menos un vector de cointegración que vincula a las series en el largo plazo.

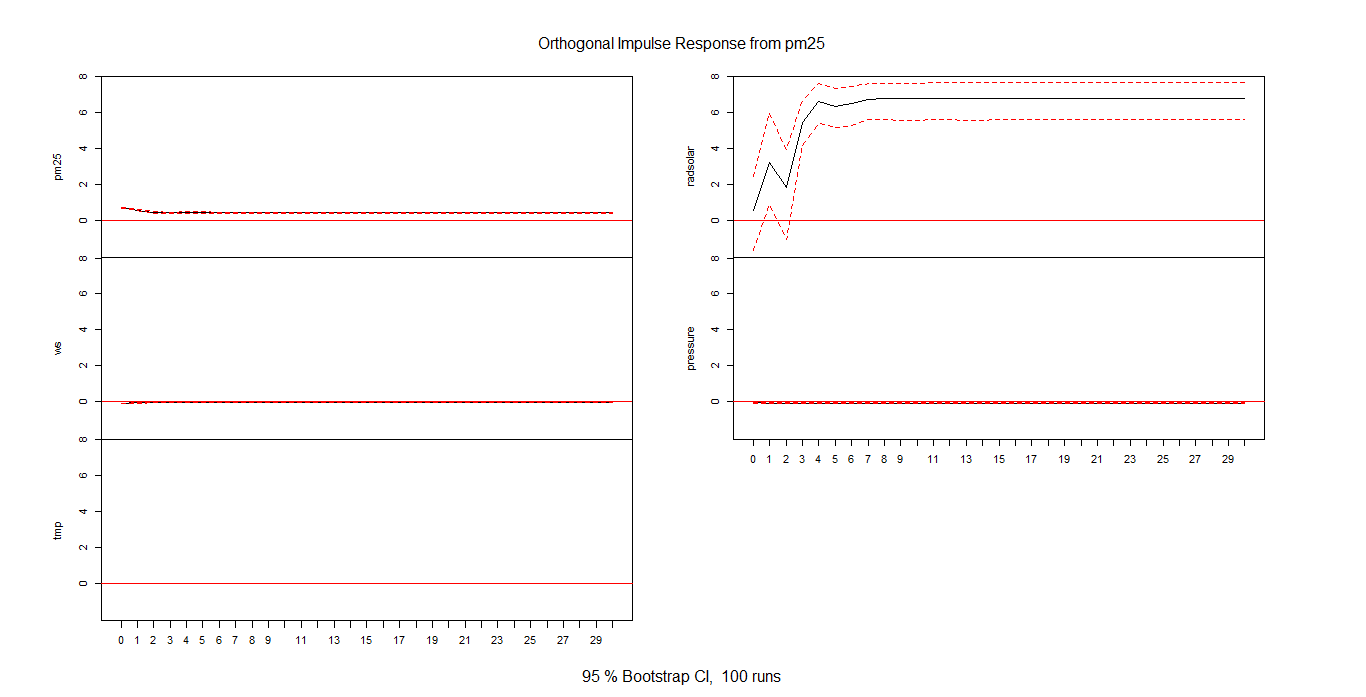
La implementación del modelo VECM reveló varias relaciones importantes a largo plazo entre las variables:

* **PM2.5 y sus propios rezagos:** Se observó un ajuste hacia el equilibrio en los rezagos de consigo mismo en el largo plazo.
* **PM2.5 y la temperatura (tmp):** Las interacciones a largo plazo entre PM2.5 y temperatura mostraron ajustes significativos en los rezagos, lo que implica que cambios en la temperatura pueden influir en los niveles de PM2.5.
* **PM2.5 y radiación solar (radsolar):** Similarmente, se detectó un equilibrio a largo plazo entre PM2.5 y radiación solar en sus rezagos, destacando la influencia de la radiación solar en los niveles de PM2.5.
* **PM2.5 y velocidad del viento (ws):** Se encontró que la velocidad del viento está cointegrada con PM2.5, sugiriendo que variaciones en la velocidad del viento pueden afectar la dispersión o acumulación de partículas finas.
* **PM2.5 y presión atmosférica (pressure):** La relación entre PM2.5 y la presión atmosférica también mostró ajustes a largo plazo en sus rezagos, indicando que los cambios en la presión atmosférica podrían influir en los niveles de PM2.5.

En cuanto al análisis del término de intercepción en el modelo VECM, no se encontraron contribuciones significativas de las variables al equilibrio de largo plazo, lo que sugiere que las dinámicas de corto plazo podrían estar más influenciadas por factores externos no incluidos en el modelo.

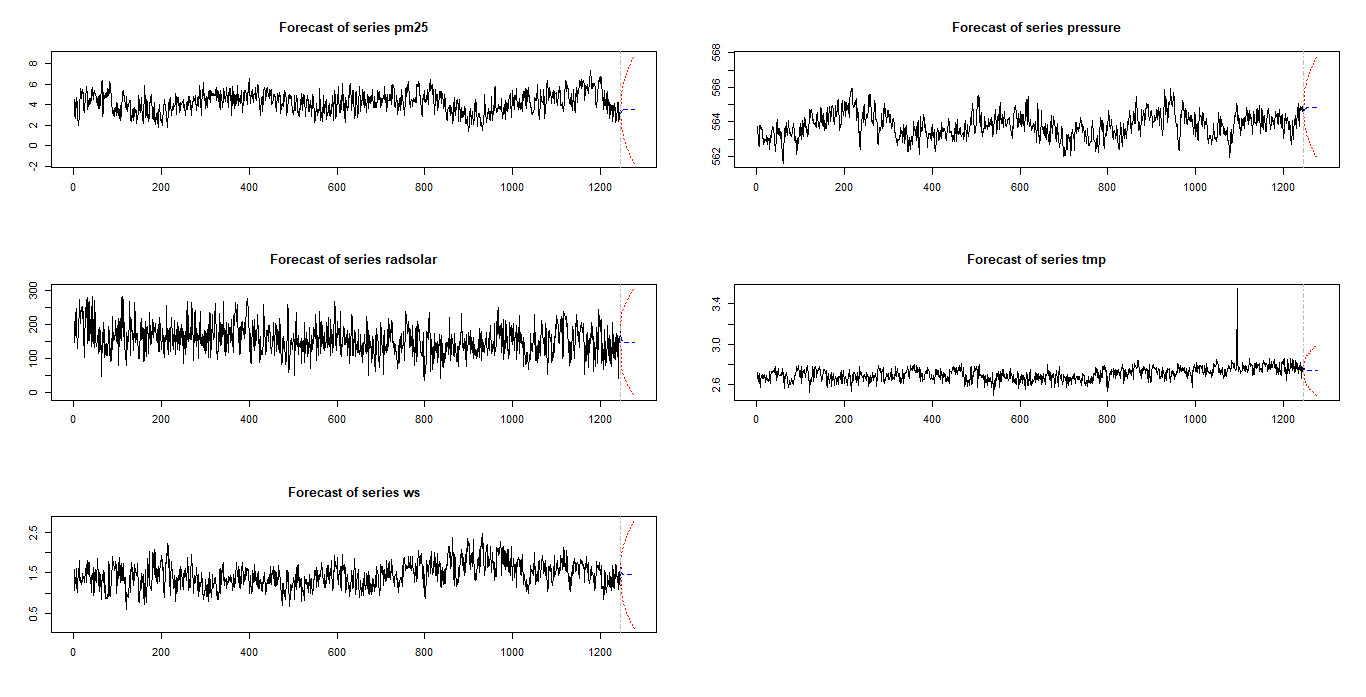
Además, el análisis de la primera diferencia de PM2.5 mostró correlaciones significativas con su propio primer rezago, así como con los primeros rezagos de radiación solar y velocidad del viento, lo que indica una respuesta dinámica de PM2.5 a cambios inmediatos en estas variables.

Para profundizar en la dinámica temporal y las interacciones entre las variables, se realizó un análisis de respuesta al impulso, que permite observar cómo una variable responde a lo largo del tiempo ante choques en otra variable.



En las gráficas muestran la respuesta de las variables a un impulso se puede observar que:

1. pm25: La gráfica en la parte superior izquierda muestra cómo pm25 responde a un impulso en sí mismo. La línea negra indica la respuesta estimada y las líneas rojas punteadas representan los intervalos de confianza del 95%. Vemos que la respuesta inicial de pm25 es positiva y se mantiene cerca del valor inicial, indicando que un shock en pm25 tiene un efecto persistente en la misma variable.
2. ws (velocidad del viento): La gráfica en el centro izquierdo muestra la respuesta de ws a un impulso en pm25. Observamos que la respuesta es prácticamente nula, lo que indica que un shock en pm25 no tiene un efecto significativo en la velocidad del viento.
3. tmp (temperatura): La gráfica en la parte inferior izquierda muestra la respuesta de tmp a un impulso en pm25. Al igual que con ws, la respuesta es nula, sugiriendo que un shock en pm25 no afecta significativamente la temperatura.
4. radsolar (radiación solar): La gráfica en la parte superior derecha muestra la respuesta de radsolar a un impulso en pm25. Observamos que inicialmente hay una respuesta positiva que se estabiliza con el tiempo. Esto podría indicar que un aumento en pm25 está asociado con un incremento temporal en la radiación solar.
5. pressure (presión): La gráfica en la parte inferior derecha muestra la respuesta de pressure a un impulso en pm25. La respuesta es esencialmente nula, indicando que un shock en pm25 no afecta significativamente la presión atmosférica.

Además, se efectuaron predicciones para las series consideradas utilizando el modelo ajustado, lo cual permitió proyectar el comportamiento futuro de las variables bajo estudio, con base en la información histórica y las dinámicas identificadas.

Las gráficas de pronóstico ilustran estas proyecciones, destacando las bandas de confianza que encapsulan la incertidumbre asociada a las predicciones.

#### Justificación de la metodología

La elección del Modelo de Corrección de Error Vectorial (VECM) para este estudio se fundamenta en su capacidad para modelar y comprender las relaciones de largo plazo entre series temporales múltiples que son integradas de orden uno, I(1). En el contexto de la calidad del aire y la contaminación, las variables como PM2.5, temperatura, radiación solar, velocidad del viento y presión atmosférica, a menudo exhiben propiedades de no estacionariedad y están influenciadas por interacciones complejas y dinámicas de largo plazo. El VECM es especialmente útil para discernir estas relaciones y para identificar ajustes dinámicos que las variables experimentan para retornar a un equilibrio de largo plazo después de perturbaciones.

#### Ventajas del modelo

* Identificación de Relaciones a Largo Plazo: El VECM permite identificar y cuantificar las relaciones de cointegración entre variables, lo cual es crucial en estudios ambientales donde los factores como los contaminantes y las condiciones meteorológicas interactúan a lo largo del tiempo.
* Ajuste Dinámico al Equilibrio: A diferencia de los modelos VAR estándar, el VECM incorpora mecanismos de ajuste que ayudan a entender cómo las variables se ajustan dinámicamente hacia el equilibrio a largo plazo, lo que es fundamental para analizar cómo las medidas de control de la contaminación y los cambios ambientales afectan la calidad del aire.
* Uso de Información de Largo y Corto Plazo: El modelo explota tanto las dinámicas de corto como de largo plazo entre las series, proporcionando una comprensión más completa sobre cómo los factores ambientales y los contaminantes interactúan entre sí en diferentes escalas temporales.
* Predicción de Respuestas a Choques: El análisis de respuesta al impulso en un VECM permite examinar el efecto de un choque en una variable sobre las demás a lo largo del tiempo, facilitando la evaluación de potenciales políticas ambientales y su impacto.

#### Desventajas del modelo

* Complejidad y Requisitos de Datos: La implementación del VECM requiere que las series temporales sean estacionarias en sus primeras diferencias y que exista al menos una relación de cointegración entre ellas, lo que puede no ser siempre el caso y puede requerir transformaciones previas de los datos.
* Interpretación de Resultados: Los resultados del VECM pueden ser difíciles de interpretar, especialmente cuando se incluyen múltiples variables y rezagos. Esto puede complicar la comunicación de los hallazgos a partes interesadas que no estén familiarizadas con técnicas econométricas avanzadas.
* Sensibilidad a la Especificación del Modelo: El VECM es sensible a la especificación correcta de los rezagos y la inclusión de términos de cointegración, lo que puede afectar la validez de las conclusiones si el modelo no está correctamente especificado.
* Riesgo de Sobreajuste: Dada la complejidad del modelo y la inclusión de múltiples variables y dinámicas, existe un riesgo de sobreajuste, donde el modelo se ajusta demasiado a los datos de muestra y no generaliza bien a nuevos datos.

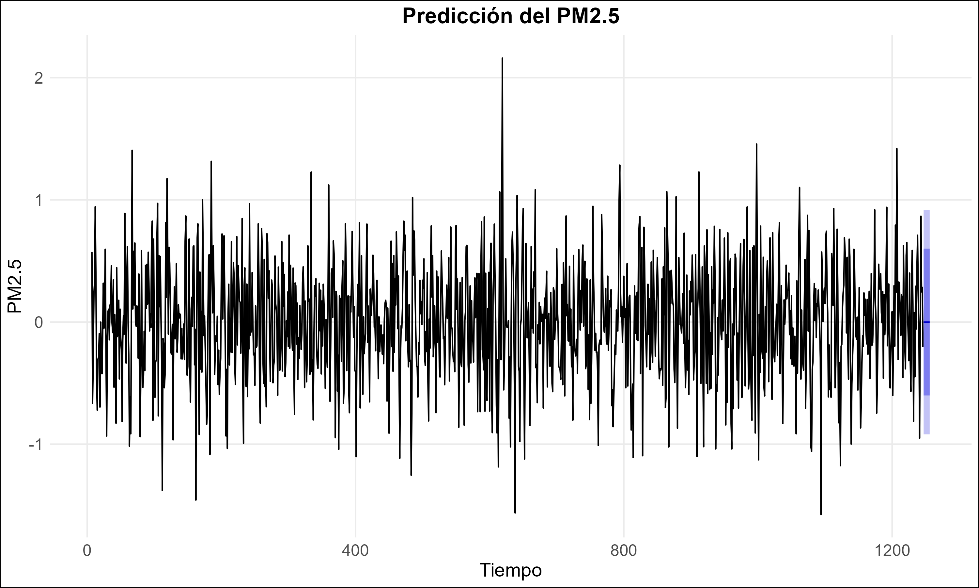
### Modelo de Autoregresivo de Rezagos Distribuidos (ARDL)

En el análisis de la relación entre la calidad del aire y variables ambientales clave, se implementó el Modelo Autoregresivo de Rezagos Distribuidos (ARDL). Este modelo se seleccionó por su robustez en el manejo de series temporales que podrían ser no estacionarias en sus niveles pero estacionarias en sus primeras diferencias. La variable dependiente designada fue la concentración de partículas PM2.5, mientras que las variables independientes incluidas fueron la radiación solar (radsolar), la velocidad del viento (ws), la presión atmosférica (pressure) y la temperatura (tmp).

Inicialmente, se normalizaron y diferenciaron todas las series temporales para garantizar la estacionariedad requerida para el análisis ARDL. Posteriormente, se empleó un enfoque de búsqueda tipo "grid search" para determinar la cantidad óptima de diferenciaciones necesarias para alcanzar un equilibrio a largo plazo entre las variables. Este método identificó que seis diferenciaciones eran necesarias para cada variable independiente para lograr la estabilidad deseada.

Con base en estos resultados, se construyó el modelo ARDL utilizando la mejor combinación de diferenciaciones. La estimación del modelo permitió evaluar tanto las dinámicas de corto plazo como las interacciones a largo plazo entre las variables.

Utilizando la función stats::predict, se realizaron predicciones a corto plazo para PM2.5 basadas en los valores ajustados por el modelo, específicamente proyectando los próximos 10 periodos. Este enfoque de predicción es crucial para evaluar la capacidad del modelo de replicar y anticipar el comportamiento de la variable dependiente basándose en los patrones históricos observados y las interacciones entre variables.



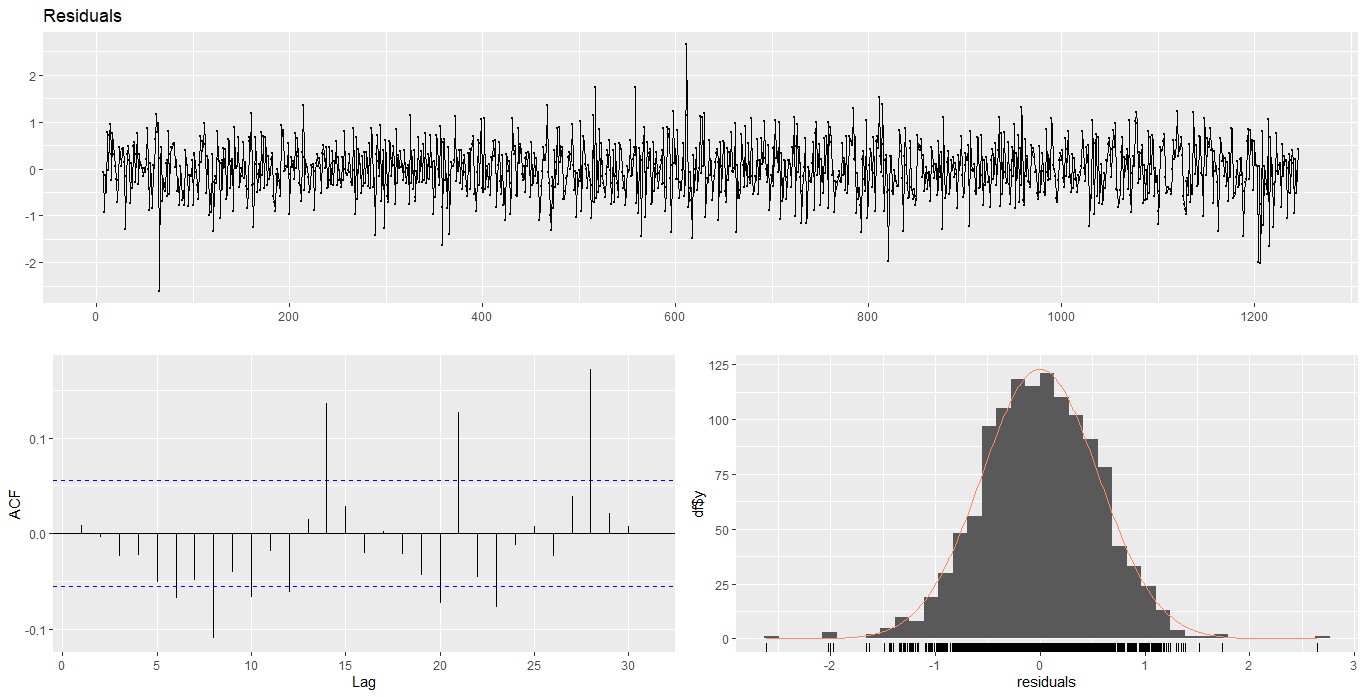
Además, se aplicó la prueba "bounds testing approach to cointegration" o prueba de límites de cointegración, utilizando la función bounds\_f\_test, para verificar la existencia de equilibrios a corto y largo plazo.

En ambos casos, los resultados mostraron p-values significativamente menores que el umbral del 5% (p-value = 1e-06), lo que indica la presencia de relaciones de equilibrio tanto en el corto como en el largo plazo entre las variables estudiadas.

La interpretación de los coeficientes obtenidos revela insights significativos sobre cómo las variables ambientales afectan la concentración de PM2.5:

* Corto Plazo:
  + Cada aumento de una unidad en la presión atmosférica (mmHg) se asocia con una disminución de 0.1544445652 µg/m3 en PM2.5.
  + Cada incremento de un grado Celsius en temperatura reduce PM2.5 en 0.8314739740 µg/m3.
  + Un aumento de una unidad en radiación solar (W/m²) incrementa PM2.5 en 0.0048481106 µg/m3.
  + Cada unidad adicional en velocidad del viento (m/s) reduce PM2.5 en 1.6913903054 µg/m3.
* Largo Plazo:
  + Por cada unidad de aumento en la presión atmosférica, PM2.5 disminuye en 0.0360737761 µg/m3.
  + Por cada grado de incremento en temperatura, PM2.5 disminuye en 1.9135865542 µg/m3.
  + Un aumento en radiación solar resulta en un incremento de 0.0108755033 µg/m3 en PM2.5.
  + Un aumento en la velocidad del viento se asocia con una reducción de 2.0467589077 µg/m3 en PM2.5.

Adicionalmente, se graficaron los residuales del modelo ARDL. El análisis incluye tres paneles esenciales para evaluar la calidad de los residuos:



1. Residuales en el Tiempo: Muestra la variación de los residuos a lo largo del tiempo. En esta gráfica, los residuos parecen centrarse alrededor de cero y no muestran tendencias evidentes ni patrones periódicos, lo cual es indicativo de que el modelo no sufre de problemas evidentes como heteroscedasticidad o cambios estructurales no capturados por el modelo.
2. Función de Autocorrelación (ACF): Este panel es crucial para determinar si existe autocorrelación en los residuos. Observamos que hay picos significativos fuera de los límites de confianza (líneas azules) en algunos retrasos. Esto sugiere la presencia de una leve autocorrelación residual, pues la mayoría de los retrasos están dentro de estos límites, indicando que no hay autocorrelación significativa en los residuos.
3. Histograma y Curva de Densidad Normal: La distribución de los residuos se compara con una distribución normal teórica (línea roja). El histograma parece tener una forma bastante simétrica y alineada con la curva normal, lo que indica que los residuos no están significativamente sesgados y son aproximadamente normales.

#### Justificación de la metodología

El modelo Autorregresivo de Retrasos Distribuidos (ARDL) es especialmente pertinente para analizar series temporales en estudios de contaminación y calidad del aire debido a su capacidad para manejar la naturaleza dinámica de los datos ambientales. Este modelo permite incorporar tanto variables independientes retardadas como la variable dependiente retardada, lo cual es crucial en el contexto ambiental donde los efectos pasados de los contaminantes pueden persistir y afectar las mediciones futuras.

En estudios de contaminación, es común que las concentraciones de contaminantes del aire no solo sean influenciadas por las condiciones actuales sino también por las concentraciones previas debido a la persistencia atmosférica y reacciones químicas lentas. El modelo ARDL puede captar estas dinámicas al incluir retrasos adecuados, proporcionando un entendimiento más completo y preciso de los factores que influyen en la calidad del aire.

#### Ventajas del modelo

* Flexibilidad en la Integración de Órdenes: El ARDL es adecuado para series temporales donde las variables pueden ser de diferentes órdenes de integración, es decir, I(0) o I(1). Esto es particularmente útil en estudios ambientales donde diferentes contaminantes pueden tener diferentes propiedades estacionarias, como es este caso.
* Corto y Largo Plazo: El modelo ARDL permite descomponer los efectos a corto y largo plazo de las variables independientes sobre la variable dependiente. Esto es vital para entender cómo intervenciones o políticas ambientales afectan la calidad del aire de manera inmediata y persistente.
* Eficiencia en Muestras Pequeñas: A diferencia de otros métodos cointegración, el ARDL puede ser implementado eficientemente incluso en muestras relativamente pequeñas, lo cual es una ventaja en estudios con datos limitados o cuando se estudian eventos de contaminación específicos.
* Inclusión de Variables Exógenas: Permite la inclusión fácil de variables exógenas (como condiciones meteorológicas) que pueden afectar directamente los niveles de contaminantes, mejorando así la especificidad del modelo.

#### Desventajas del modelo

* Selección de Retrasos: La necesidad de seleccionar el número adecuado de retrasos para cada variable puede ser complicada y está sujeta a error humano o sesgo. Una mala elección en los retrasos puede llevar a conclusiones erróneas sobre las relaciones entre variables.
* Riesgo de Sobreajuste: Al incluir múltiples retrasos y variables, existe el riesgo de sobreajustar el modelo a los datos específicos del estudio, lo cual puede reducir la generalización de los resultados a otros contextos o períodos.

### Random forest

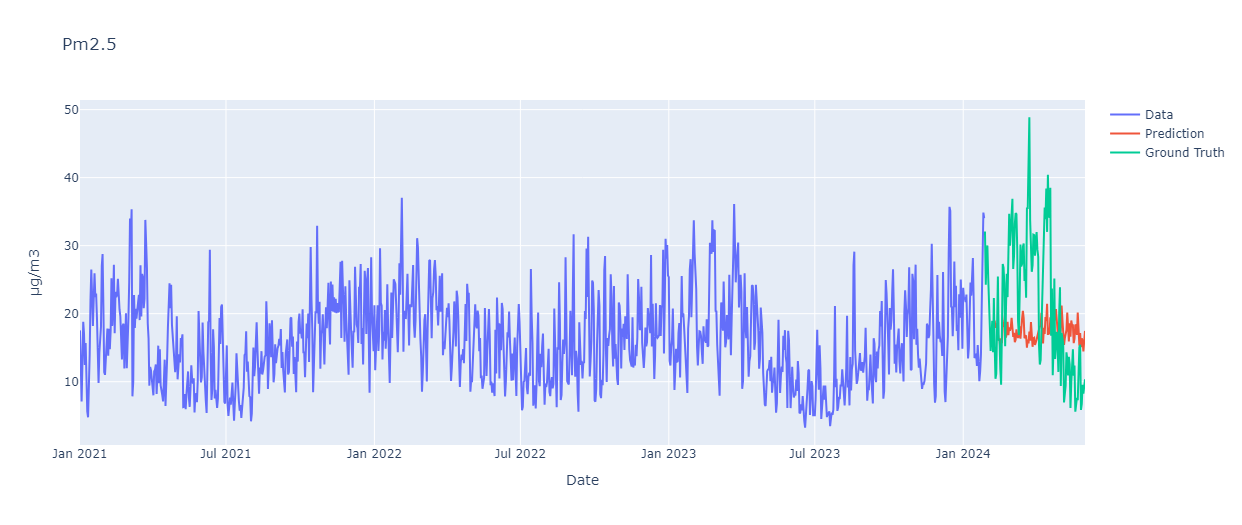
Se emplea el modelo de Random Forest para evaluar las relaciones entre diversas variables atmosféricas y los niveles de PM2.5, un contaminante crítico debido a sus impactos en la salud pública y el ambiente.

Inicialmente, se procede a la carga y estructuración de los datos, optando por la utilización de Python como lenguaje de programación dada su flexibilidad y las potentes librerías disponibles para el manejo y análisis de datos. Los datos son importados desde un archivo de Excel, asegurándose que la columna correspondiente a las fechas sea reconocida como tal, para facilitar su manipulación en series temporales. Con estos datos se construyen dos estructuras principales: una serie temporal exclusivamente de los valores de PM2.5, y un DataFrame con las variables explicativas, que incluyen temperatura, radiación solar, presión atmosférica y velocidad del viento.

Para adecuar los datos al análisis de series temporales, se transforman las fechas en índices temporales y las variables explicativas en un arreglo NumPy, lo cual permite su manipulación eficiente en los pasos subsecuentes. Se realiza entonces la división del conjunto de datos en subconjuntos de entrenamiento y prueba, manteniendo el 90% de los datos para entrenamiento y el 10% restante para la validación del modelo. La estructuración de estos conjuntos se facilita mediante la función `TimeseriesGenerator`, que permite generar lotes de datos temporales con un número específico de observaciones pasadas para prever los futuros valores de PM2.5.

El modelo de Random Forest se configura y entrena utilizando un enfoque de redes neuronales LSTM debido a su capacidad para manejar dependencias a largo plazo en datos secuenciales. La configuración incluye capas densas y parámetros específicos como la función de activación y el optimizador, esenciales para ajustar la red a las particularidades de los datos de contaminación atmosférica. Posteriormente, se realiza el entrenamiento del modelo a lo largo de 50 épocas, ajustando los parámetros del modelo para minimizar el error cuadrático medio, un indicador crítico del desempeño predictivo del modelo en datos temporales.

Finalmente, se evalúa la capacidad predictiva del modelo sobre el conjunto de datos de prueba. Las predicciones, junto con los datos reales, se visualizan en un gráfico que muestra las series temporales de entrenamiento, las predicciones y los valores reales de PM2.5.



La gráfica ilustra los resultados de la predicción del modelo de Random Forest sobre los niveles de PM2.5 en Bogotá, junto con los datos reales para comparación.

1. Serie de Datos (Azul): Representa los valores históricos de PM2.5, mostrando la variabilidad típica a lo largo del tiempo con fluctuaciones que parecen seguir un patrón estacional y/o dependiente de factores no modelados explícitamente.
2. Predicción del Modelo (Rojo): Los valores predichos por el modelo de Random Forest para el segmento de prueba se muestran en rojo. Estas predicciones intentan seguir las tendencias y variaciones de los datos reales.
3. Valores Reales (Verde): La serie en verde representa los valores reales de PM2.5 durante el período de prueba. Esta serie es crucial para evaluar la precisión de las predicciones del modelo.

Al observar la gráfica, se puede notar que las predicciones del modelo (rojo) logran seguir en cierta medida la tendencia general de los valores reales (verde), especialmente en la captura de los picos y las caídas. Sin embargo, hay áreas, especialmente hacia el final del período de prueba, donde las predicciones no coinciden exactamente con los valores reales, indicando posibles limitaciones en el modelo para adaptarse a cambios abruptos o a la volatilidad alta en los niveles de PM2.5.

En cuanto a variabilidad y precisión el modelo parece capturar adecuadamente la variabilidad media de los datos, pero falla en predecir los picos extremos con precisión. Esto podría deberse a la naturaleza de los modelos de Random Forest que, aunque efectivos en capturar relaciones no lineales y complejas entre variables, pueden no ser tan precisos en situaciones de alta volatilidad.

#### Justificación de la metodología

El modelo Random Forest es especialmente adecuado para estudios en el ámbito de la contaminación y la calidad del aire debido a su robustez y eficacia en el manejo de grandes conjuntos de datos con múltiples variables predictoras. En el contexto de la calidad del aire, los datos suelen ser complejos y presentan interacciones no lineales entre múltiples factores ambientales y meteorológicos. Random Forest, siendo un método de aprendizaje ensemble que combina múltiples árboles de decisión, es capaz de capturar estas complejidades a través de su estructura inherente, proporcionando así predicciones y análisis de alta precisión.

#### Ventajas del modelo

* Capacidad para Manejar Alta Dimensionalidad: Random Forest maneja eficientemente grandes volúmenes de datos con muchas variables, lo cual es común en estudios ambientales que integran diversas fuentes de datos como meteorología, emisiones industriales y tráfico vehicular.
* Robustez frente a Datos Atípicos y Ruido: Este modelo es menos susceptible a la influencia de datos atípicos y ruido, debido a que promedia múltiples árboles de decisión, lo cual resulta en un modelo más estable y confiable.
* No requiere Linealidad: A diferencia de los modelos estadísticos tradicionales, Random Forest no asume una relación lineal entre las variables, lo cual es una ventaja al modelar procesos ambientales que son típicamente no lineales y complejos.

#### Desventajas del modelo

* Complejidad Computacional: La construcción de múltiples árboles de decisión hace que Random Forest pueda ser computacionalmente intensivo, especialmente con conjuntos de datos muy grandes, lo que podría requerir recursos de hardware significativos y tiempos de procesamiento largos.
* Riesgo de Sobreajuste: Aunque generalmente Random Forest maneja bien el sobreajuste, en situaciones con conjuntos de datos muy ruidosos o con un número excesivo de árboles, puede sobreajustarse a los datos de entrenamiento, lo que afecta la capacidad de generalización del modelo.
* Dependencia de Parámetros: La eficacia del modelo puede depender en gran medida de la configuración de sus parámetros, como el número de árboles en el bosque o la profundidad de los árboles. Elegir estos parámetros de manera inadecuada puede llevar a un rendimiento subóptimo.

# Resultados

s

# Conclusiones y recomendaciones

.

# Referencias

Agencia de Protección Ambiental de Estados Unidos. (23 de 03 de 2024). *Efectos del material particulado (PM) sobre la salud y el medioambiente*. Obtenido de https://espanol.epa.gov/espanol/efectos-del-material-particulado-pm-sobre-la-salud-y-el-medioambiente

Benavides Ballesteros, H. O., & León Aristizabal, G. E. (2007). *Información tecnica sobre gases de efecto invernadero y el cambio climático.* Instituto de Hidrología, Meteorología y Estudios Ambientales.

Red de Monitoreo de Calidad del Aire de Bogotá (RMCAB). (01 de 06 de 2024). *Reporte de estaciones*. Recuperado el 01 de 06 de 2024, de http://rmcab.ambientebogota.gov.co/Report/stationreport

World Health Organization. (2021). *WHO global air quality guidelines: particulate matter (‎PM2.5 and PM10)‎, ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide.* World Health Organization.

# Anexos



Salida de la consola 1. Estadísticas descriptivas Guaymaral



Salida de la consola 2. Estadísticas descriptivas Min Ambiente



Salida de la consola 3. Estadísticas descriptivas Suba



Salida de la consola 4. Estadísticas descriptivas Usaquén



Salida de la consola 5. Estadísticas descriptivas Las Ferias



Salida de la consola 6. Estadísticas descriptivas San Cristóbal



Salida de la consola 7. Estadísticas descriptivas Tunal



Salida de la consola 8. Estadísticas descriptivas Bolivia



Salida de la consola 9. Estadísticas descriptivas Carvajal



Salida de la consola 10. Estadísticas descriptivas Fontibón



Salida de la consola 11. Estadísticas descriptivas Kennedy



Salida de la consola 12. Estadísticas descriptivas Puente Aranda



Salida de la consola 13. Estadísticas descriptivas Centro de Alto Rendimiento



Salida de la consola 14. Estadísticas descriptivas Bogotá

===============================================

Dependent variable:

---------------------------

ex\_pm

-----------------------------------------------

rain -0.199

(0.227)

tmp 0.039

(0.047)

ws -2.314\*\*\*

(0.229)

as.factor(dow)2 0.380\*\*

(0.158)

as.factor(dow)3 0.400\*\*

(0.159)

as.factor(dow)4 0.583\*\*\*

(0.161)

as.factor(dow)5 0.723\*\*\*

(0.161)

as.factor(dow)6 0.345\*\*

(0.158)

as.factor(dow)7 -0.395\*\*

(0.158)

as.factor(month)2 0.665\*\*\*

(0.196)

as.factor(month)3 0.711\*\*\*

(0.201)

as.factor(month)4 -0.496\*\*\*

(0.179)

as.factor(month)5 -1.585\*\*\*

(0.193)

as.factor(month)6 -1.204\*\*\*

(0.208)

as.factor(month)7 -1.550\*\*\*

(0.227)

as.factor(month)8 -1.084\*\*\*

(0.206)

as.factor(month)9 -0.417\*\*

(0.193)

as.factor(month)10 -0.340\*

(0.195)

as.factor(month)11 -0.161

(0.199)

as.factor(month)12 0.591\*\*\*

(0.210)

as.factor(year)2022 0.217\*\*

(0.110)

as.factor(year)2023 0.428\*\*\*

(0.125)

as.factor(year)2024 0.641\*\*\*

(0.166)

Constant 2.715\*\*\*

(0.671)

-----------------------------------------------

Observations 1,247

Log Likelihood -572.301

Akaike Inf. Crit. 1,192.601

===============================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Salida de la consola 15. Modelo Probit PM2.5

===============================================

Dependent variable:

---------------------------

ex\_pm

-----------------------------------------------

rain -0.808\*\*

(0.324)

tmp 0.040

(0.075)

ws -1.089\*\*\*

(0.274)

as.factor(dow)2 0.325\*

(0.196)

as.factor(dow)3 0.546\*\*\*

(0.190)

as.factor(dow)4 0.638\*\*\*

(0.188)

as.factor(dow)5 0.604\*\*\*

(0.188)

as.factor(dow)6 0.049

(0.205)

as.factor(dow)7 -0.783\*\*\*

(0.272)

as.factor(month)2 0.144

(0.195)

as.factor(month)3 0.206

(0.189)

as.factor(month)4 -0.512\*\*

(0.217)

as.factor(month)5 -1.749\*\*\*

(0.345)

as.factor(month)6 -0.777\*\*\*

(0.291)

as.factor(month)7 -0.966\*\*\*

(0.341)

as.factor(month)8 -0.667\*\*

(0.288)

as.factor(month)9 -0.523\*\*

(0.261)

as.factor(month)10 -0.085

(0.231)

as.factor(month)11 -0.207

(0.233)

as.factor(month)12 0.254

(0.214)

as.factor(year)2022 0.007

(0.136)

as.factor(year)2023 0.153

(0.152)

as.factor(year)2024 1.117\*\*\*

(0.191)

Constant -0.308

(1.044)

-----------------------------------------------

Observations 1,247

Log Likelihood -394.507

Akaike Inf. Crit. 837.014

===============================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Salida de la consola 16. Modelo Probit PM10

===============================================

Dependent variable:

---------------------------

ex\_no2

-----------------------------------------------

rain 0.620\*\*

(0.262)

tmp -0.248\*\*\*

(0.074)

ws -3.299\*\*\*

(0.288)

as.factor(dow)2 0.733\*\*\*

(0.179)

as.factor(dow)3 0.712\*\*\*

(0.179)

as.factor(dow)4 0.625\*\*\*

(0.180)

as.factor(dow)5 0.727\*\*\*

(0.178)

as.factor(dow)6 0.142

(0.181)

as.factor(dow)7 -1.522\*\*\*

(0.257)

as.factor(month)2 0.308

(0.202)

as.factor(month)3 0.277

(0.200)

as.factor(month)4 -0.249

(0.215)

as.factor(month)5 -1.705\*\*\*

(0.268)

as.factor(month)6 -0.860\*\*\*

(0.260)

as.factor(month)7 -1.952\*\*\*

(0.466)

as.factor(month)8 -0.080

(0.235)

as.factor(month)9 0.616\*\*\*

(0.219)

as.factor(month)10 0.749\*\*\*

(0.225)

as.factor(month)11 0.071

(0.223)

as.factor(month)12 0.187

(0.221)

as.factor(year)2022 -0.101

(0.127)

as.factor(year)2023 0.675\*\*\*

(0.145)

as.factor(year)2024 1.022\*\*\*

(0.194)

Constant 7.151\*\*\*

(1.069)

-----------------------------------------------

Observations 1,247

Log Likelihood -444.072

Akaike Inf. Crit. 936.145

===============================================

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Salida de la consola 17. Modelo Probit NO2

T