

Análisis de los cambios en la concentración de PM_{2.5} en Pekín en el periodo 2010-2014 y relación con otras variables meteorológicas.

Estefanía Hurtado Reina. estefania.hurtado@udea.edu.co

Abstract

Este estudio analiza la variabilidad espacial y temporal del material particulado fino (PM_{2.5}) en Pekín, con el objetivo de identificar los factores meteorológicos y estacionales que influyen en la ocurrencia de episodios de baja, moderada y alta polución. La metodología comprende limpieza y clasificación de datos, análisis exploratorio, correlaciones, visualización multivariada, técnicas de detección de anomalías (Isolation Forest), y la creación de categorías estacionales para evaluar patrones diferenciados. Los resultados evidenciaron que los eventos de alta polución se asocian principalmente con vientos del sureste y con un aumento de condiciones de viento en calma, especialmente en verano. Por el contrario, los eventos de baja polución se relacionan con vientos del noroeste, que presentan mayores velocidades y favorecen la dispersión del contaminante. Adicionalmente, se observó que los meses de febrero y octubre concentran los valores más altos de PM_{2.5}, mientras que mayo y agosto registran los más bajos. Las regresiones muestran que el PM_{2.5} aumenta con el punto de rocío y disminuye con la presión atmosférica. Se concluye que la dinámica del viento, en interacción con la topografía local, es el factor más determinante en la acumulación de PM_{2.5} en Pekín, por encima de otras variables meteorológicas analizadas.

1. Introducción

El PM_{2.5} se refiere a la materia particulada fina, y es un contaminante del aire que consiste en una mezcla de partículas sólidas y líquidas suspendidas en el aire con un diámetro aerodinámico de 2.5 micras (µm) o menos, es decir, son 30 veces más pequeñas que el diámetro promedio de un cabello humano. Esta es una mezcla que puede incluir sustancias químicas orgánicas, polvo, hollín y metales, entre otros. La fuente de origen puede ser de origen primario (es decir, que son emitidas directamente a la atmósfera) proveniente principalmente de quema de combustibles, procesos industriales, construcción o incendios forestales, o también puede ser de origen secundario (estas se forman en la atmósfera a partir de reacciones químicas entre gases). En Pekín, la contaminación atmosférica por material particulado fino (PM_{2.5}) es un problema de gran magnitud con implicaciones directas para la salud y el medio ambiente (EPA, 2023)

La pregunta central que guiará este análisis es: “¿Cuál es la influencia relativa de las condiciones meteorológicas (temperatura, presión, punto de rocío, dirección y velocidad acumulada del viento y precipitación) en la severidad y clasificación de los estados de PM_{2.5}?”. La identificación de eventos de concentraciones altas de PM_{2.5} es importante debido a varios factores; como menciona Marshall, J (2003), el PM_{2.5} penetra profundamente en el sistema respiratorio, provocando o agravando enfermedades respiratorias y cardiovasculares, por otro lado, la ciudad combina factores de alta emisión (tráfico, industria, uso de carbón en calefacción) con condiciones meteorológicas que pueden favorecer la acumulación de contaminantes. Analizar la influencia de variables climáticas como temperatura, punto de rocío, presión atmosférica, dirección del viento, y ocurrencia de precipitación sólida o líquida permite distinguir variaciones debidas a cambios meteorológicos de aquellas provocadas por modificaciones en las emisiones. Esta distinción es esencial para diseñar políticas

ambientales efectivas y evaluar el impacto real de medidas temporales (ej. APEC 2014) o estacionales (ej. calefacción invernal).

El proyecto tiene como objetivo explorar variables que permitan describir la evolución temporal del PM_{2.5}, indagar en cómo distintos factores meteorológicos pueden estar asociados a cambios en la concentración de partículas y comparar periodos controlados por ciertas condiciones climáticas. Las variables de temperatura, punto de rocío, presión atmosférica, dirección combinada del viento, velocidad acumulada del viento y ocurrencia de precipitación serán relacionadas con la concentración de PM_{2.5} tanto en su forma de variable continua como de variable categórica.

2. Metodología

2.1 Fuente de Datos

En este estudio se empleó el conjunto de datos *Beijing PM2.5 Data*, disponible en el repositorio de la University of California, Irvine (UCI Machine Learning Repository). Este conjunto abarca datos horarios de la concentración de partículas PM_{2.5} registradas por la Embajada de los Estados Unidos en Pekín, combinados con variables meteorológicas del Aeropuerto Capital de Pekín en el periodo comprendido entre el 1 de enero de 2010 y el 31 de diciembre de 2014. El archivo contiene 43.824 registros y 11 variables, entre ellas datos de año, mes, día y hora, valores de PM_{2.5}, punto de rocío, temperatura, presión atmosférica, dirección del viento, velocidad acumulada del viento, horas acumuladas de nieve y horas acumuladas de lluvia. Los valores ausentes están marcados como “NA”. El uso de esta base de datos permite integrar una variable de calidad del aire con agentes meteorológicos en un análisis de series de tiempo y modelización ambiental. Asimismo, su licencia (Creative Commons Atribución 4.0) permite su reutilización con el debido reconocimiento de la fuente. La base de datos del análisis tiene lecturas horarias de PM_{2.5} tomadas en la embajada de Estados Unidos en Pekín, ubicada en las coordenadas 116° 28' 12" E y 39° 57' 00" N, en conjunto con mediciones meteorológicas horarias tomadas en el Aeropuerto Internacional de Pekín Capital (BCIA) obtenidas de weather.nocrew.org. Aunque la embajada y el aeropuerto están a 17 km de distancia, se experimentan condiciones meteorológicas muy similares (Liang, X., et al., 2015)

2.2 EDA

Variable	Nombre y Unidades	Rango Típico	Información de la variable	Tipo de variable
pm2.5	Concentración de PM _{2.5} (µg/m³)	0 a 500µg/m³	Material Particulado Fino. Mide la concentración de partículas en el aire con un diámetro aerodinámico de 2.5 micrómetros o menos.	Continua, cuantitativa
DEWP	Punto de rocío (°C)	-30 a 30°C	La temperatura a la que el aire debe enfriarse, a presión constante, para que el vapor de agua se sature y comience la condensación (formación de rocío o niebla). Es una medida directa del contenido de humedad.	Continua, cuantitativa
TEMP	Temperatura (°C)	-15 a +40 °C	Medida de la energía térmica del aire. Indica qué tan caliente o frío está el .	Continua, cuantitativa
PRES	Presión atmosférica (hPa)	990 a 1040 hPa	El peso o la fuerza que ejerce el aire sobre una superficie debido a la gravedad. Los cambios indican la aproximación de sistemas meteorológicos (altas o bajas presiones).	Continua, cuantitativa

cbwd	Dirección combinada del viento.	NW, SW, NE, SE, N, S, E, W, cv	El origen geográfico (de dónde viene) del viento predominante.	Categoría, discreta
Iws	Potencia acumulada del viento (W/m ²)	-	Mide la energía total contenida en el viento, no solo su velocidad.	Continua, cuantitativa
Is	Horas acumuladas de nieve (h)	-	El tiempo total (en horas) durante el cual se ha observado precipitación en forma sólida (nieve) en un evento de precipitación.	Continua, cuantitativa
Ir	Horas acumuladas de lluvia (h)	-	El tiempo total (en horas) durante el cual se ha observado precipitación en forma líquida (lluvia) en un evento de precipitación.	Continua, cuantitativa

Tabla 1. Resumen de variables originales del Dataset

Las variables 'Iws', 'Is' e 'Ir' se presentan de manera acumulada. Iws representa la potencia acumulada del viento, y este valor se acumula hasta que este cambia de dirección. La duración de Iws está ligada a la estabilidad de la categoría en cbwd. Si cbwd no cambia, Iws aumenta. Is e Ir representan las horas acumuladas de precipitación de nieve y lluvia respectivamente, estas también reinician el conteo cuando el evento cesa. A partir de estas variables continuas (Is e Ir), se realizará un proceso de transformación de variables, donde se presenten como variables categóricas (ocurrencia/no ocurrencia, con valores de 1 y 0 respectivamente) para cada tipo de precipitación.

Como primer análisis, se presenta la Tabla 2. con la descripción estadística exploratoria de las variables numéricas:

Estadístico	pm2.5	DEWP	TEMP	PRES	Iws
Media	98.61	1.82	12.45	1016.45	23.89
Desviación Estándar	92.05	14.43	12.20	10.27	50.01
Mínimo	0	-40	-19	991	0.45
25%	29	-10	2	1008	1.79
50%	72	2	14	1016	5.37
75%	137	15	23	1025	21.91
Máximo	994	28	42	1046	585.6

Tabla 2. Resumen estadístico descriptivo de variables numéricas

Para clasificar los niveles de contaminación por material particulado fino, se tomaron como referencia los valores guía establecidos por entidades internacionales. La Agencia de Protección Ambiental de Estados Unidos (EPA) define un límite de 35 µg/m³ como el valor máximo permitido para la media móvil de 24 horas en aire ambiente, umbral por encima del cual se considera que la calidad del aire deja de ser aceptable para la población general (EPA, 2012). Por su parte, la Unión Europea ha establecido históricamente un valor límite anual de 25 µg/m³ para PM_{2.5}, utilizado como parámetro de referencia en evaluaciones de calidad del aire (European Commission, 2023). Asimismo, la Organización Mundial de la Salud (OMS) ha propuesto guías más estrictas, recomendando una concentración anual de 5 µg/m³ y un valor de 15 µg/m³ para el promedio de 24 horas, destacando la evidencia creciente sobre los efectos adversos incluso a concentraciones bajas (WHO, 2021).

Con base en estos valores de referencia, para este estudio se construyó una variable categórica que clasifica la concentración horaria de PM_{2.5} en tres niveles (clasificados como 1,2,3) relativos: baja contaminación (≤ 35 µg/m³), contaminación moderada o evento de contaminación

(> 35 y $\leq 150 \mu\text{g}/\text{m}^3$) y alta contaminación (> 150 $\mu\text{g}/\text{m}^3$). Aunque los valores superiores a 150 $\mu\text{g}/\text{m}^3$ no corresponden a un estándar normativo formal para PM_{2.5}, este umbral se utiliza en este trabajo como criterio operativo para identificar episodios de contaminación severa, dada la evidencia de impactos sanitarios significativos en concentraciones muy elevadas.

Como parte del análisis exploratorio de datos, se genera la Figura 1. donde se grafica la serie temporal de valores de PM_{2.5}, así como su distribución y serie multianual mensual para conocer el comportamiento temporal y cíclico de los datos.

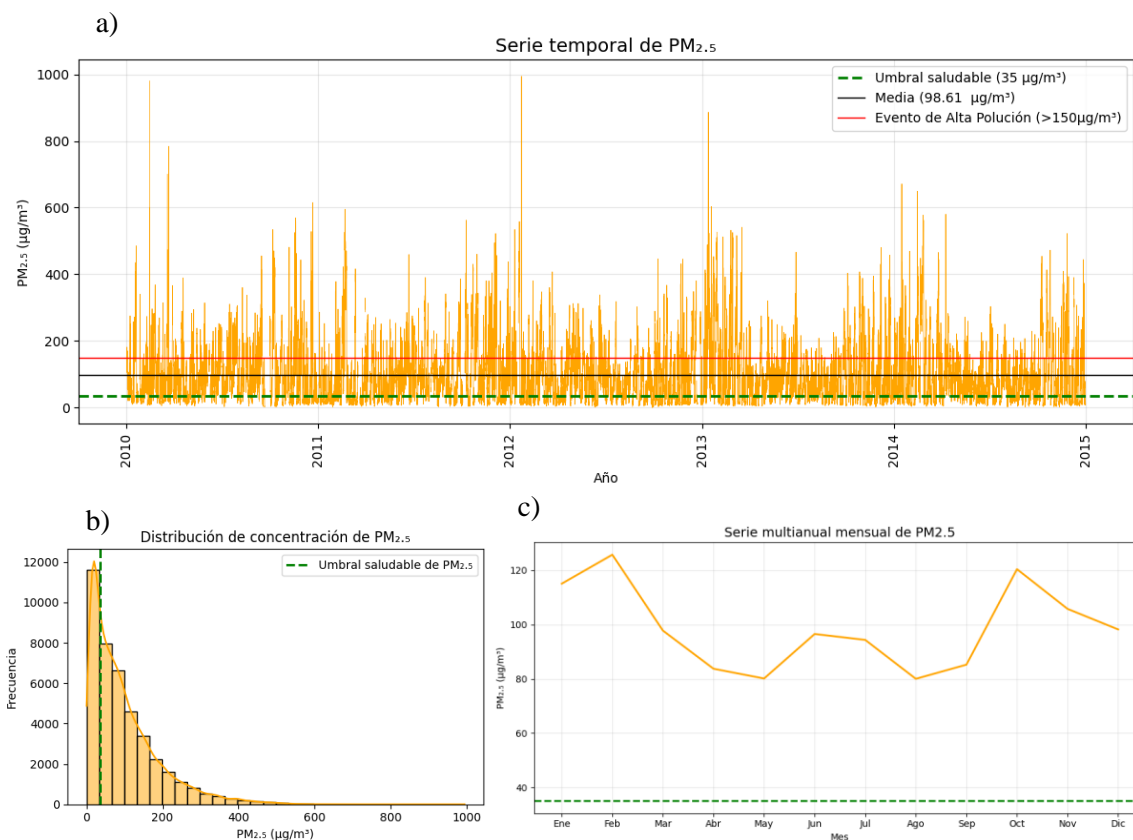


Figura 1. a) Serie temporal, b) Histograma de distribución, c) Serie mensual multianual de PM_{2.5}

Los datos no siguen una distribución normal, sino que presentan asimetría positiva (hacia la derecha), lo que indica que los valores extremadamente altos de concentración (por encima de 200 $\mu\text{g}/\text{m}^3$) son menos frecuentes, aunque sí se registran de forma ocasional. Por otro lado, se visualizan en la Figura 2. las tendencias horarias promedio tanto de PM_{2.5} como de las variables meteorológicas.

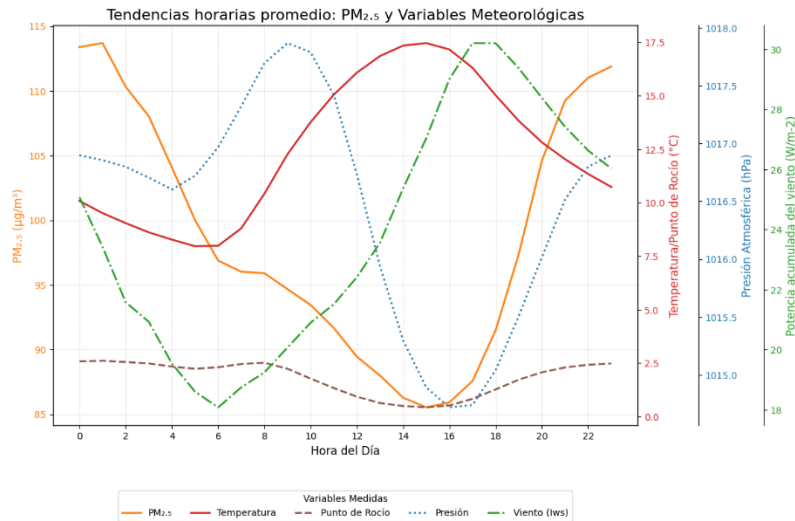


Figura 2. Tendencias horarias promedio de PM_{2.5} y variables meteorológicas

2.3 Valores Atípicos

Después del análisis exploratorio, se hace un análisis de valores atípicos, donde se usa el método de IQR para la detección de atípicos, debido a que es un método robusto frente a sesgos y asimetrías, característica que se identifica en la distribución de PM_{2.5}, la única variable que será sujeto a detección de valores atípicos con este método. En este caso, el tipo de datos atípicos que estamos buscando estarán relacionados a errores de medición (valores por encima de los máximos registrados y publicados en bibliografía). El límite inferior será programado en 0 µg/m³ debido a que no es posible que haya valores negativos de concentración. Se incluyó un umbral de $4.6 \times \text{IQR}$ debido a la presencia de una proporción considerable de datos altamente dispersos. Además, por encima de los 600 µg/m³ se evidencia una separación clara entre dos grupos de valores, los cuales se distancian de los outliers detectados mediante el umbral tradicional de $1.5 \times \text{IQR}$. Estos últimos, a su vez, muestran una mayor cohesión interna. Para añadir mayor robustez a la decisión a tomar para los valores atípicos, se utilizará en método de Isolation Forest en diagramas de Pairplot para las variables continuas. Teniendo en cuenta que la frecuencia de datos es horaria durante 5 años, el volumen de datos es grande. Además, ninguna de las variables numéricas se distribuye de manera normal, por lo que el método de Isolation Forest es conveniente para reforzar la detección de datos atípicos, esta vez incluyendo relación entre las variables. La tasa de contaminación utilizada será de 0.01 ya que los datos son algo dispersos.

2.4 Valores faltantes

Se identificaron 2.067 valores faltantes, equivalentes al 4,72 % del total de observaciones. Dado que el objetivo central del proyecto es determinar qué variables contribuyen con mayor importancia a la ocurrencia de distintos estados de polución, la imputación no resulta adecuada en esta fase, ya que introducir valores estimados podría generar relaciones falsas entre las variables y afectar las interpretaciones posteriores. Desde una perspectiva

estadística, los métodos de imputación deben aplicarse únicamente cuando los datos faltantes son lo suficientemente numerosos como para comprometer la representatividad del conjunto, o cuando se ha verificado formalmente el mecanismo de ausencia para evitar sesgos (Little & Rubin, 2019). En este caso, el porcentaje de datos faltantes es reducido y su distribución no altera la forma general de las variables, por lo que la alternativa metodológicamente más consistente es eliminar estas observaciones, dado que su exclusión no modifica de manera significativa la dispersión o estructura del conjunto de datos.

2.5 Aislamiento de casos especiales

2.5.1 Estacionalidad

Se creó una nueva columna categórica que agrupa los datos según la estación del año: junio, julio y agosto se clasificaron como verano; septiembre, octubre y noviembre como otoño; diciembre, enero y febrero como invierno; y marzo, abril y mayo como primavera. Esto permitió identificar tendencias diferenciadas según los grupos estacionales.

2.5.2 “Airpocalipsis”

En primer lugar, se busca el año con más casos de estado de polución alto y se encuentra que este fue el 2013 con 4.398 eventos de este tipo, seguido del año 2014 con 4.389 eventos. Como se menciona en la bibliografía asociada al “Airpocalipsis”, este fue un evento que se caracterizó por una rápida conversión de aerosoles primarios en secundarios, derivada de la combustión de combustibles fósiles y las emisiones vehiculares, que generaron niveles particularmente altos de óxidos de nitrógeno y facilitaron la transformación del dióxido de azufre en aerosoles secundarios de sulfato (Ji et al., 2014; Wang et al., 2014), por lo que se decide aislar los datos del año 2013 para comparar las condiciones meteorológicas con los otros años, mediante la diferencia entre los valores de estadística descriptiva para las variables numéricas.

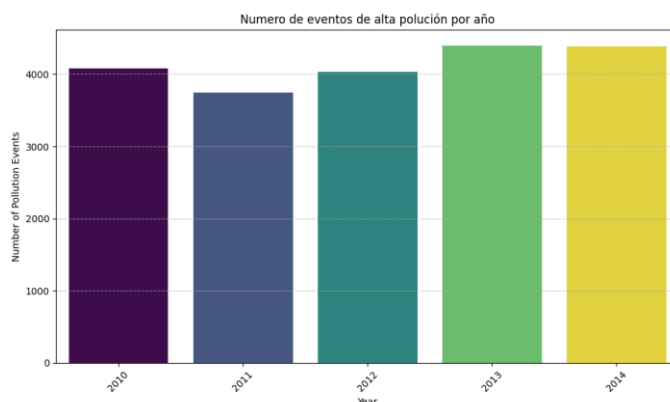


Figura 3. Número de eventos clasificados como “Alta polución” por año.

3. Resultados

3.1 Análisis exploratorio

A lo largo de los cinco años analizados, se observa que gran parte de los valores de concentración de $PM_{2.5}$ superan el umbral considerado saludable. Es común que se alcancen niveles clasificados como de alta polución, lo cual evidencia una persistente afectación en la calidad del aire. Asimismo, se identifican posibles ciclos anuales, donde las concentraciones máximas tienden a presentarse al final de cada año o al inicio del siguiente, mientras que los valores mínimos suelen registrarse a mitad de año. En particular, durante los meses de febrero y octubre se evidencian las mayores concentraciones de $PM_{2.5}$, mientras que en mayo y agosto se presentan los valores más bajos. Cabe destacar que la media mensual de todos los meses se mantiene por encima del umbral saludable, lo que sugiere una exposición continua a niveles de contaminación perjudiciales.

En cuanto a la frecuencia de los estados de polución, cerca del 50 % de las horas analizadas durante el periodo de estudio corresponden a un evento de polución (con concentraciones superiores a $35 \mu g/m^3$) como se logra evidenciar en la Figura 4. Los episodios de alta polución representan un porcentaje menor en comparación con los de baja polución, con una diferencia de aproximadamente 7.5 % entre ambos estados.

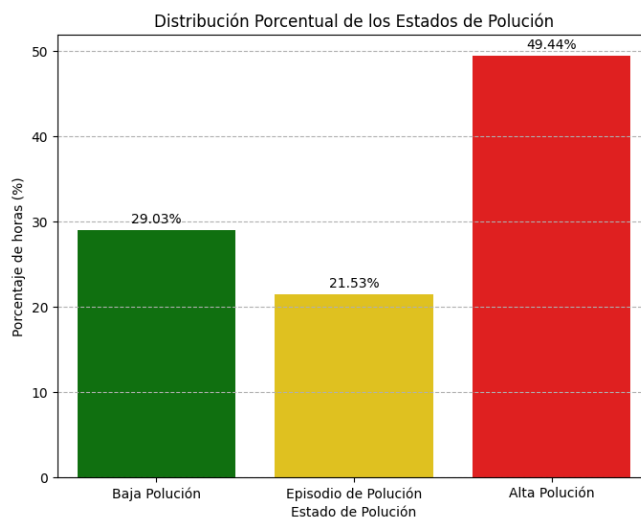


Figura 4. Distribución porcentual de estados de polución en la totalidad de registros.

$PM_{2.5}$

Se logran evidenciar valores altos en la noche entre las 0–6 h; los valores de $PM_{2.5}$ son elevados al inicio del día, es un comportamiento común asociado a la estabilidad atmosférica durante la noche, donde hay poca mezcla del aire, baja turbulencia y menor dispersión de contaminantes. Hay un descenso entre las 6–14; conforme aumenta la temperatura y el viento, la atmósfera se vuelve más inestable y se dispersan las partículas. Por último, se nota un aumento nocturno después de las 18 h; al caer la temperatura y disminuir el viento, los contaminantes vuelven a acumularse cerca del suelo.

Temperatura y punto de rocío

La temperatura muestra el patrón típico diario mínimo en la madrugada y máximo hacia las 14–16 h. El punto de rocío se mantiene bajo y bastante estable, con un ligero aumento nocturno, lo que indica mayor humedad relativa en la noche.

Presión atmosférica

Muestra una ligera caída entre las 8 h y 16 h, coincidiendo con el aumento térmico. Este comportamiento es coherente ya que el aire caliente se expande y reduce la presión local. La presión vuelve a aumentar en la noche, cuando la atmósfera se enfría y se densifica.

Velocidad del viento

Presenta un mínimo temprano (6–8 h) y un máximo en la tarde (14–18 h). Esto coincide con la mayor convección y calentamiento solar, que generan movimientos verticales y horizontales del aire.

3. 2 Identificación y tratamiento de valores atípicos

Todos los datos detectados como atípicos se ubican dentro del grupo de eventos de alta polución. Si se consideran los valores máximos registrados durante el “Airpocalypse” como se puede mostrar en la Figura 5. Un análisis epidemiológico basado en los datos de la Embajada de EE. UU. en Pekín reportó una concentración diaria de $PM_{2.5}$ de hasta $569 \mu g/m^3$ (Chen et al., 2020), por lo que es posible contextualizar los valores registrados de estos episodios. En este contexto, resulta razonable asumir que las concentraciones detectadas por encima del umbral considerado en este estudio corresponden a errores de medición.

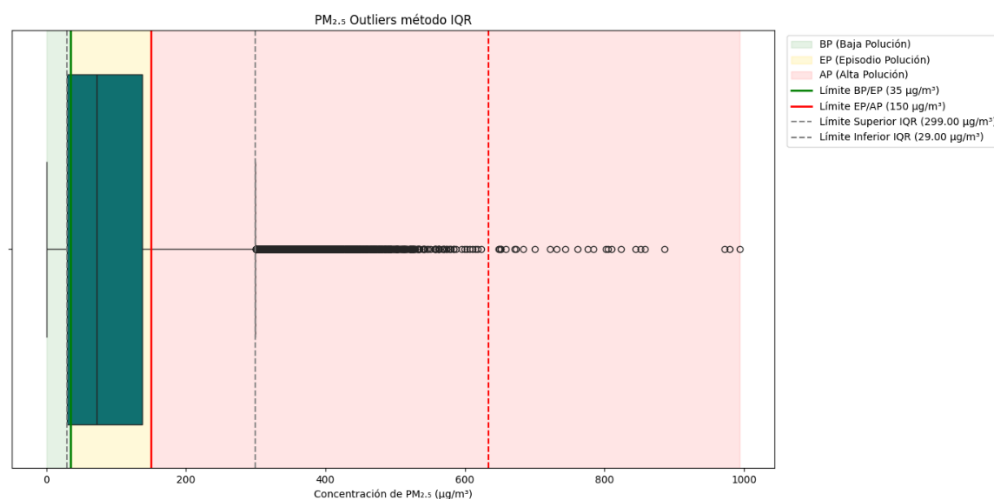


Figura 5. Detección de outliers usando el método del rango intercuartílico (IQR)

Los valores atípicos deben analizarse considerando el contexto físico y ambiental de las variables antes de decidir su exclusión del conjunto de datos. En las gráficas que involucran la variable $PM_{2.5}$, la presencia de datos atípicos resulta más evidente, dado que algunos valores se alejan considerablemente del conjunto principal de observaciones. No obstante, no todos los puntos identificados por el modelo deben considerarse anómalos, ya que varios se mantienen dentro de un rango aceptable que no altera significativamente el

comportamiento general de la variable. Sin embargo, aquellos valores que coinciden con los detectados mediante el método del rango intercuartílico (IQR) sí pueden considerarse candidatos a ser excluidos del análisis, ya que coinciden con los puntos extremadamente alejados que se halló usando este método.

En las gráficas que relacionan las variables temperatura, punto de rocío y presión atmosférica, los valores identificados como atípicos se encuentran, en su mayoría, dentro de una nube de puntos densamente agrupada. Por ello, se considera que estos no representan verdaderas anomalías y, en consecuencia, no serán tratados como outliers. Finalmente, el mayor número de valores detectados como atípicos se observa en las gráficas que incluyen la variable potencia del viento acumulada. Dado que esta variable representa un valor acumulativo, es probable que el algoritmo Isolation Forest identifique como anómalos aquellos casos en los que la potencia del viento alcanza valores elevados. Esto puede deberse a que los episodios en los cuales la dirección del viento se mantiene constante son menos frecuentes, generando registros poco comunes. Sin embargo, estos valores no deben eliminarse, ya que el comportamiento del viento constituye una de las variables de mayor interés dentro del presente estudio.

3.3 Relación entre variables

A partir del pairplot utilizado en la aplicación de Isolation Forest en la Figura 6., junto con el mapa de calor de correlaciones, se analizó la relación entre $PM_{2.5}$ y cada una de las variables numéricas. Este análisis permitió identificar las siguientes relaciones:

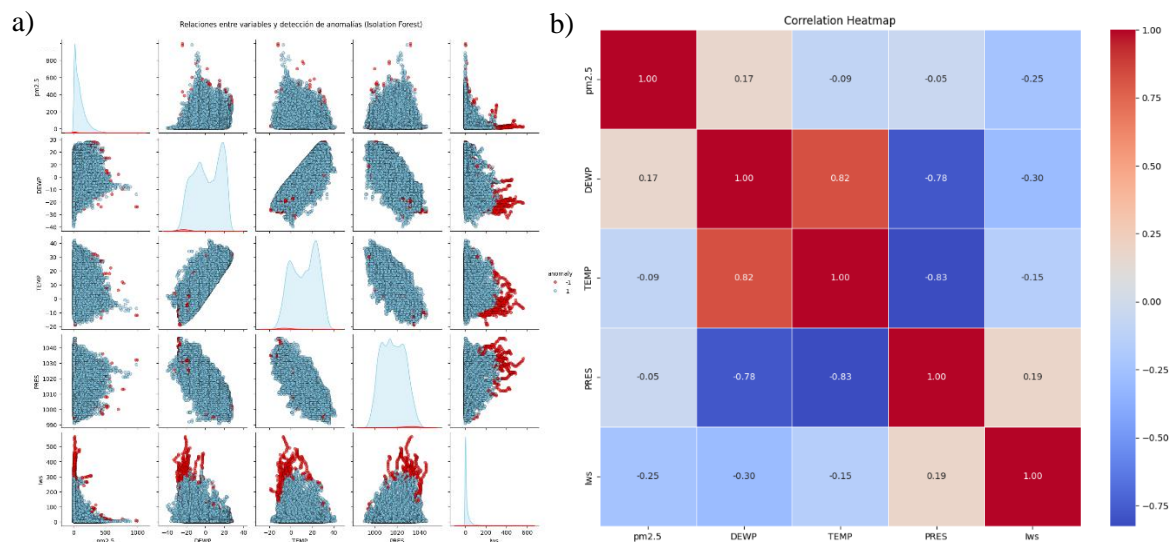


Figura 6. Relaciones entre variables. a) Método de Isolation forest en Pairplots de variables numéricas para identificar valores atípicos. b) Mapa de calor entre variables numéricas

$PM_{2.5}$ vs Temperatura

No se evidencia una correlación fuerte entre la concentración de $PM_{2.5}$ y la temperatura; sin embargo, pueden identificarse valores más altos de $PM_{2.5}$ en el rango de temperatura entre $-10^{\circ}C$ y $0^{\circ}C$, lo cual podría estar asociado a condiciones invernales. Este comportamiento

sugiere analizar con mayor detalle las temporadas frías, particularmente en relación con la ocurrencia de precipitación en forma de nieve que puede limitar la dispersión de contaminantes. También se observa una ligera disminución en los valores de $PM_{2.5}$ a medida que la temperatura aumenta, lo que podría indicar que durante los periodos cálidos o de verano, la mezcla vertical y la mayor turbulencia atmosférica favorecen la dispersión del material particulado y reducen su acumulación en las capas bajas de la atmósfera.

$PM_{2.5}$ vs Punto de rocío

El patrón de $PM_{2.5}$ frente al punto de rocío es similar al observado con la temperatura. El punto de rocío representa la temperatura a la cual el aire debe enfriarse para que el vapor de agua que contiene se condense, por lo tanto, está ligado a la temperatura y a la humedad relativa. A medida que la temperatura aumenta, el aire puede contener más vapor de agua, y si la humedad se mantiene constante, también aumenta el punto de rocío. Esto explica la tendencia lineal ascendente observada entre la temperatura y el punto de rocío, reflejando la coherencia entre ambas variables. En relación con el $PM_{2.5}$, los valores más altos suelen presentarse en condiciones de baja humedad (puntos de rocío bajos), lo que puede favorecer la acumulación de partículas secas en suspensión, mientras que una mayor humedad podría contribuir a la coagulación o sedimentación del material particulado.

$PM_{2.5}$ vs Presión atmosférica

La presión atmosférica muestra una relación ligeramente positiva con el $PM_{2.5}$, evidenciándose un pequeño incremento de las concentraciones entre 1020 y 1040 hPa. Esto puede deberse a que las altas presiones se asocian con condiciones anticiclónicas, caracterizadas por aire descendente, estabilidad y baja ventilación, lo cual limita la dispersión de contaminantes. Asimismo, la relación entre la presión y la temperatura/punto de rocío es negativa: cuando la presión atmosférica disminuye, el aire tiende a ascender, expandirse y enfriarse, aumentando la humedad relativa y el punto de rocío. En cambio, con presiones altas, el aire desciende, se comprime y calienta, reduciendo la humedad y, por ende, el punto de rocío. Este comportamiento puede sugerir que la presión actúa como un indicador indirecto de la estabilidad atmosférica y puede influir en la acumulación o dispersión de material particulado.

$PM_{2.5}$ vs Potencia del viento

A bajas velocidades del viento, las concentraciones de $PM_{2.5}$ son más altas y dispersas, mientras que a mayores velocidades, las concentraciones disminuyen. Esto puede sugerir que el viento actúa como un agente dispersor del material particulado, favoreciendo el transporte del contaminante hacia otras zonas. En condiciones de calma atmosférica, en cambio, el $PM_{2.5}$ tiende a acumularse cerca del suelo debido a la falta de movimiento del aire, lo que eleva los niveles de contaminación. Por otra parte, la relación de la potencia del viento con la temperatura y el punto de rocío es ligeramente descendente, mientras que con la presión atmosférica es ascendente. Este comportamiento es consistente con la dinámica atmosférica (presiones bajas se asocia a movimiento del aire, mientras que altas presiones y temperaturas elevadas favorecen la estabilidad y reducen la ventilación).

3.3.1 Estacionalidad, dirección y velocidad del viento

Durante los eventos clasificados como de Baja Polución, la dirección predominante del viento corresponde al noroeste (NW). En contraste, durante los episodios de Polución y los eventos de Alta Polución, predomina el flujo proveniente del sureste (SE) como se ve en la Figura 7a. En estos últimos casos, la distribución de direcciones del viento presenta proporciones similares; sin embargo, se observa un incremento significativo en la frecuencia de condiciones de viento en calma o estacionario, en comparación con los periodos de baja polución, donde esta condición es mínima. Este comportamiento indica que los estados atmosféricos caracterizados por baja velocidad del viento favorecen la acumulación de material particulado fino, al limitar la dispersión y la renovación del aire en la capa baja de la atmósfera

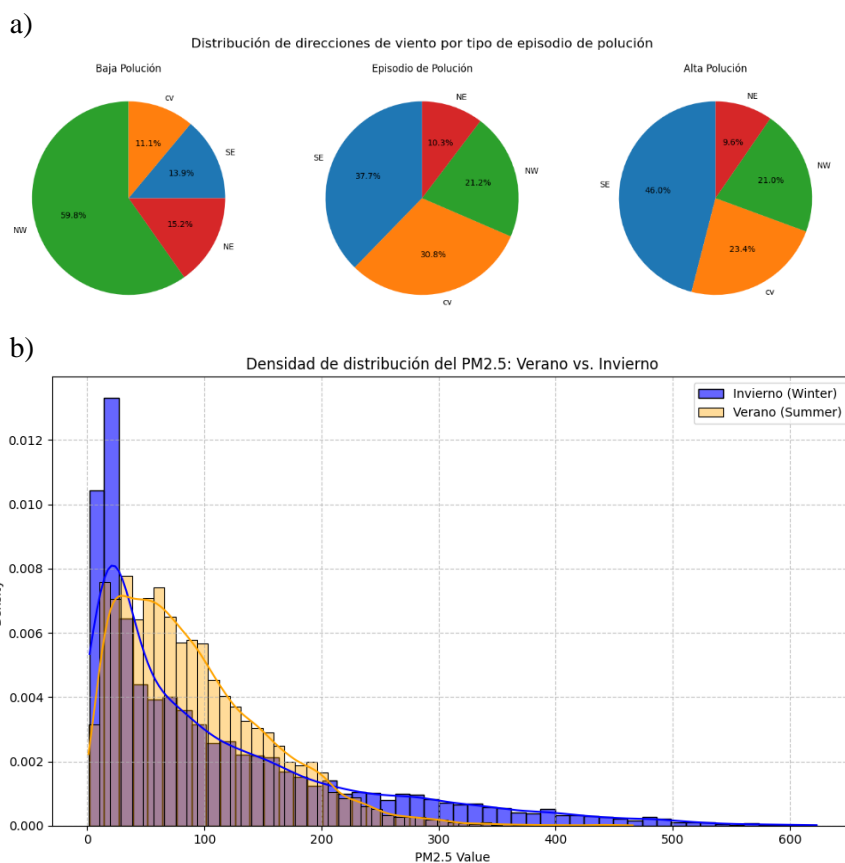


Figura 7. a) Distribución de las direcciones del viento por tipo de episodio de polución b) Densidad de distribución del PM2.5 en verano y en invierno.

Desde una perspectiva geomorfológica, la ubicación de Pekín está un entorno parcialmente rodeado por montañas tiene un papel determinante en la dinámica de dispersión de contaminantes. Hacia el noroeste (NW) se presentan elevaciones pronunciadas del terreno, asociadas a los sistemas montañosos Yan y Taihang, ubicados en la periferia de la ciudad, así como a las montañas Jundu y Xi dentro del área urbana. Esta configuración topográfica actúa como una barrera orográfica que restringe el intercambio de masas de aire. En consecuencia, cuando los vientos provienen del sureste (SE), las masas de aire contaminadas tienden a acumularse sobre la ciudad al encontrarse con las elevaciones del noroeste, generando condiciones de estancamiento atmosférico. Por el contrario, bajo vientos dominantes del noroeste, el flujo de aire se desplaza libremente hacia el sureste, una zona predominantemente plana, lo que facilita la ventilación y dispersión del PM_{2.5} fuera del valle urbano.

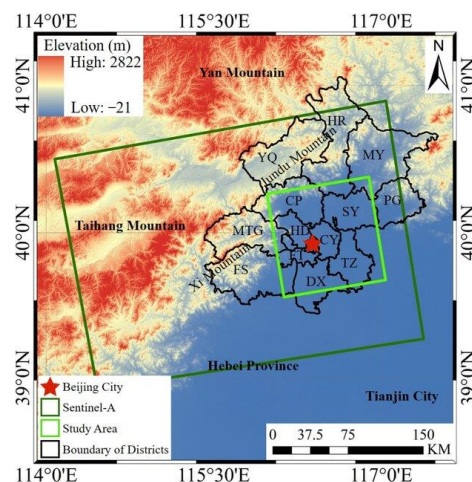


Figura 8. Mapa de modelo de elevación digital (DEM) de Pekín, China (Sheng, H., et al., 2023)

Se observa que durante el verano (Figura 9a y 7b) se presenta un mayor porcentaje de eventos de alta polución en comparación con las demás categorías. De manera opuesta, en invierno predomina la ocurrencia de eventos de baja polución, mostrando un comportamiento estacional contrastante entre ambas estaciones.

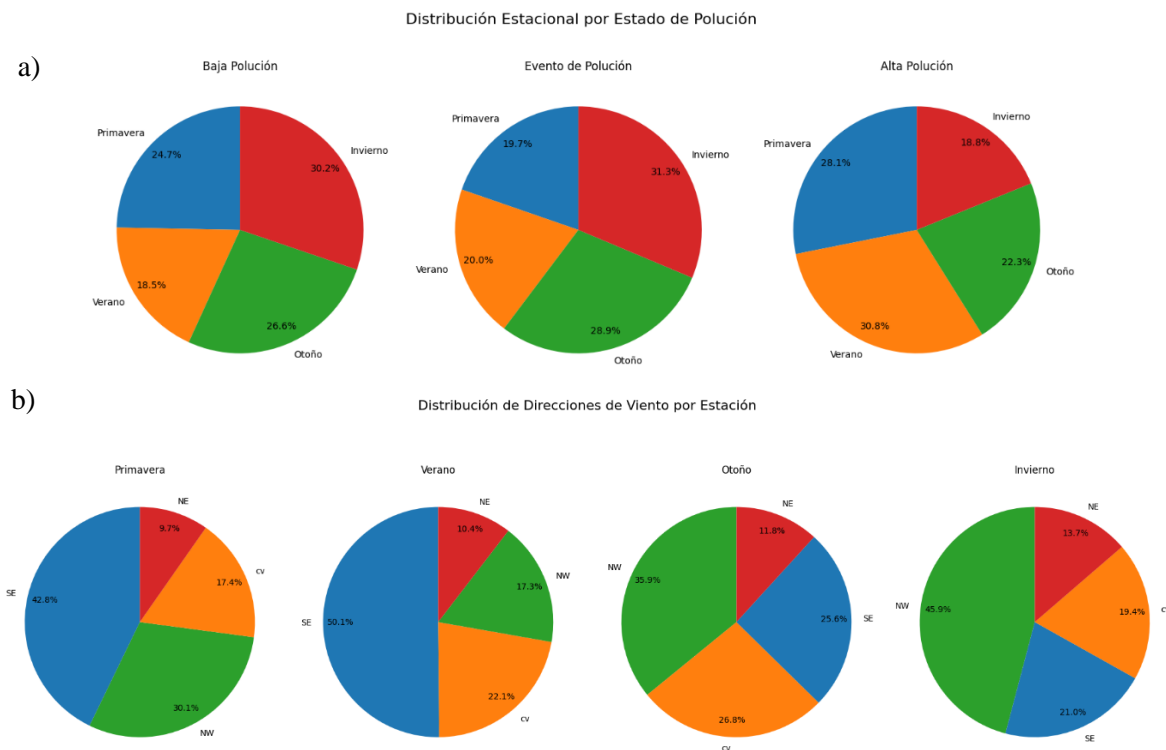


Figura 9. a) Distribución estacional según estado de polución b) Distribución de direcciones de viento por estación

Al analizar la relación entre la dirección del viento y las estaciones del año (figura 7b), se identifica que en invierno predomina el viento proveniente del noroeste (NW), lo cual coincide con la mayor frecuencia de eventos de baja polución asociados a esta misma dirección. En contraste, durante el verano la dirección predominante es sureste (SE), la cual se relaciona principalmente con la ocurrencia de eventos de alta polución, reforzando la existencia de un patrón estacional en la dinámica del material particulado.

Por otro lado, para fortalecer el hallazgo que vincula la dirección NW con los eventos de baja polución, se analizó su relación con las velocidades acumuladas del viento. Se encontró que, efectivamente, los valores más altos de velocidad se presentan cuando el viento proviene del noroeste (Figura 10a), lo que respalda su papel en la dispersión del $PM_{2.5}$. Se hizo un análisis en contraparte también de las características cuando la dirección del viento está en calma o poco variable (Figura 10b), y se encuentra que la mayor frecuencia de condiciones de viento en calma es durante el verano y el otoño, lo que es coherente con los resultados obtenidos anteriormente, donde el verano presenta la mayor proporción de eventos de alta polución.

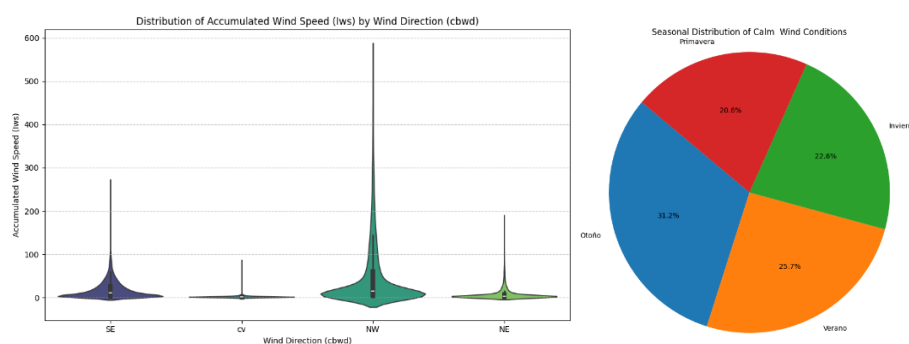


Figura 10. a) Diagramas de violín de potencia acumulada de viento según su dirección. b) Distribución estacional de condiciones calmadas en el viento.

3.3.2 Estacionalidad y condiciones atmosféricas

Dado que en la serie multianual mensual se identificaron febrero y octubre como los meses con los valores más altos de $PM_{2.5}$, y mayo y agosto como los meses con los valores más bajos, se realizó un análisis adicional mediante curvas de regresión entre las concentraciones de $PM_{2.5}$ y dos variables meteorológicas relevantes: el punto de rocío y la presión atmosférica.

En las gráficas, las curvas de regresión indican una tendencia ascendente de la concentración de $PM_{2.5}$ a medida que aumenta el punto de rocío, y una tendencia considerablemente descendente a medida que aumenta la presión, excepto en mayo y agosto, donde se mantienen más estables. Una disminución del punto de rocío y un aumento de la presión suelen ir acompañados de la llegada del viento del norte, que trae aire más seco y fresco.

PM2.5 vs Punto de Rocío y Presión Atmosférica

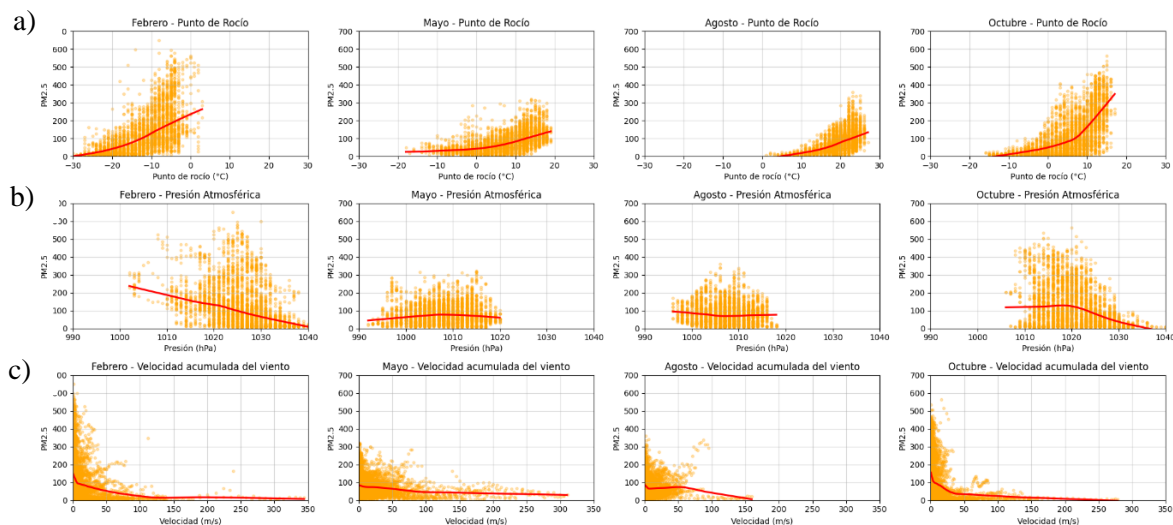


Figura 11. Scatterplots de PM2.5 vs a) punto de rocío, b) presión atmosférica y c) velocidad acumulada del viento en los meses de Febrero, Mayo, Agosto y Octubre

Esto puede relacionarse a que las masas de aire procedentes del norte y noroeste en Pekín suelen tener origen continental, con baja humedad y temperaturas reducidas. Estas condiciones generan un descenso del punto de rocío debido a la menor cantidad de vapor de agua presente en el aire. Al mismo tiempo, el enfriamiento del aire incrementa su densidad, provocando un aumento de la presión atmosférica. Dichas situaciones meteorológicas son típicas de escenarios anticiclónicos, caracterizados por subsidencia del aire y estabilidad atmosférica. Sin embargo, el flujo nortino también puede favorecer la ventilación y dispersión de contaminantes, especialmente cuando se asocia con vientos moderados, contribuyendo temporalmente a la reducción de las concentraciones de PM_{2.5}.

3.4 Casos especiales

3.4.1 2013 versus 2010,2012,2014.

Aunque las variables meteorológicas como temperatura, presión y punto de rocío mostraron diferencias mínimas entre periodos, los cambios más relevantes se dieron en la dinámica del viento: la velocidad media del viento fue menor (21,09 m/s frente a 24,61 m/s en los demás años) y hubo un leve aumento de condiciones de viento en calma o variable. Estos factores probablemente contribuyeron a una menor dispersión de contaminantes.

En el análisis temporal, el episodio prolongado más severo se registró entre el 20 y el 24 de febrero de 2011, con 90 horas consecutivas de alta polución y una concentración media de 439,09 $\mu\text{g}/\text{m}^3$. El valor horario más alto se observó el 9 de abril de 2014, con 580 $\mu\text{g}/\text{m}^3$.

La distribución estacional se mantuvo equilibrada y similar para todos los años, lo que sugiere que las diferencias encontradas en 2013 no se deben a cambios en la composición estacional de los datos. En conjunto, estos resultados indican que las condiciones de menor ventilación atmosférica (especialmente la reducción en la velocidad del viento y el aumento de

condiciones de calma) fueron los factores más influyentes en el incremento de $PM_{2.5}$ durante 2013.

3.4.2 Eventos con precipitación

En cuanto a la posible relación entre las concentraciones de $PM_{2.5}$ y la ocurrencia de precipitaciones, se observa que, en presencia de nieve, los valores de concentración de $PM_{2.5}$ tienden a ser más altos en comparación con los periodos sin precipitación sólida (Figura 12a). Esto podría estar asociado a las condiciones atmosféricas frías y estables típicas durante los eventos de nieve, que limitan la mezcla vertical y favorecen la acumulación de contaminantes en las capas bajas de la atmósfera.

Por otro lado, en los eventos de lluvia, el rango de concentraciones de $PM_{2.5}$ es ligeramente menor, aunque el comportamiento general de la distribución es similar (Figura 12b). Este patrón puede atribuirse al efecto de lavado, por el cual las gotas de lluvia arrastran las partículas en suspensión, reduciendo parcialmente las concentraciones en el aire. Finalmente, se observa un número considerablemente mayor de valores atípicos en los periodos sin precipitación, lo cual se explica por la predominancia de registros horarios sin eventos de lluvia o nieve, que abarcan la mayoría del conjunto de datos y presentan una mayor variabilidad en las concentraciones de $PM_{2.5}$.

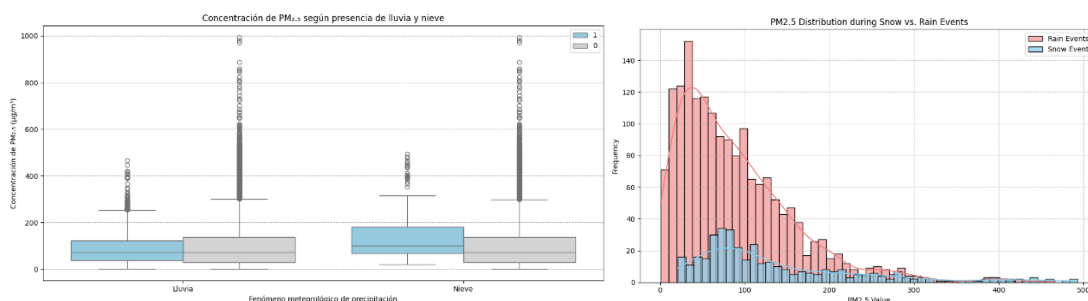


Figura 12. a) Concentración de $PM_{2.5}$ según ocurrencia de lluvia y nieve. b) distribución de $PM_{2.5}$ en eventos donde solo hubo lluvia o solo hubo nieve

El mes donde se registraron más eventos de nieve fue en Febrero y el mes con mayor cantidad de eventos de lluvia fue en Julio. La frecuencia de ocurrencia de los eventos de precipitación es mayor que las de nieve, y puede observarse que para la mayoría de eventos donde hubo lluvia (que en este caso se asocia a la temporada de verano), los niveles de $PM_{2.5}$ eran un poco más bajos que en donde hubo eventos de nieve (asociados a invierno) y donde el viento predominante es en dirección SE que anteriormente se ha asociado a eventos de alta polución, al igual que la dirección predominante en eventos de lluvia es la dirección NW asociada a eventos de menor polución (Figura 13)

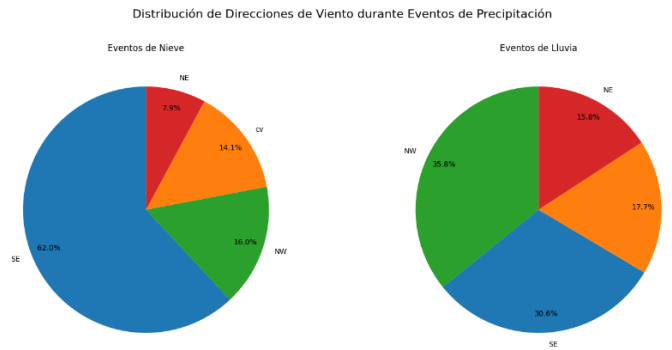


Figura 13. Distribución de direcciones de viento durante eventos de precipitación.

Conclusiones

Los resultados de estacionalidad y dirección del viento muestran que la variabilidad del $PM_{2.5}$ está condicionada en parte por la dirección del viento y la configuración orográfica local. Los eventos de baja polución se asocian con vientos del noroeste, que favorecen la ventilación del valle urbano, mientras que los episodios de alta polución ocurren bajo vientos del sureste y condiciones más frecuentes de calma, que dificultan la dispersión. Esta dinámica coincide con un patrón estacional marcado: en invierno predominan los eventos de baja polución y en verano los de alta polución. Adicionalmente, las relaciones entre $PM_{2.5}$, punto de rocío y presión atmosférica indican que masas de aire seco y de mayor presión (típicamente provenientes del norte) contribuyen a reducir las concentraciones del contaminante.

Los niveles ligeramente más altos de $PM_{2.5}$ en 2013, a pesar de unas condiciones meteorológicas en gran medida similares, sugieren que la reducción de la velocidad del viento y un pequeño aumento de los períodos de calma podrían ser factores importantes en la acumulación de contaminantes.

Se recomienda explorar otros posibles factores contribuyentes, más allá de las variables meteorológicas analizadas (como la actividad industrial, las políticas de emisiones o el transporte atmosférico regional) con el fin de comprender de manera más completa los elementos que pueden influir en el aumento de contaminación.

Referencias

- United States Environmental Protection Agency. (2023). *Particulate Matter (PM) Basics*.
- Marshall, J. (2013). $PM_{2.5}$. *Proceedings of the National Academy of Sciences*, 110(22), 8756-8756.
- Chen, S. (2015). *Beijing $PM_{2.5}$ Data* [Dataset]. UCI Machine Learning Repository.
- Liang, X., Zou, T., Guo, B., Li, S., Zhang, H., Zhang, S., ... & Chen, S. X. (2015). Assessing Beijing's $PM_{2.5}$ pollution: severity, weather impact, APEC and winter heating. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2182), 20150257.

Sheng, H., Zhou, L., Huang, C., Ma, S., Xian, L., Chen, Y., & Yang, F. (2023). Surface Subsidence Characteristics and Causes in Beijing (China) before and after COVID-19 by Sentinel-1A TS-InSAR. *Remote Sensing*, 15(5), 1199.

EPA. (2012). *Air Quality Index (AQI): A guide to air quality and your health*. United States Environmental Protection Agency.

European Commission. (2023). *EU Air Quality Standards*. European Union.

WHO. (2021). *WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization.

Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Wiley.

Ji, D., Li, L., Wang, Y., Zhang, J., Cheng, M., Sun, Y., ... & Miao, H. (2014). The heaviest particulate air-pollution episodes occurred in northern China in January, 2013: Insights gained from observation. *Atmospheric Environment*, 92, 546-556.

Wang, Y., Yao, L., Wang, L., Liu, Z., Ji, D., Tang, G., ... & Xin, J. (2014). Mechanism for the formation of the January 2013 heavy haze pollution episode over central and eastern China. *Science China Earth Sciences*, 57(1), 14-25.