

# Evidencia del Reto: Reporte Individual

Juan José de Jesús Hernández Beltrán (A00836747)

## 1. Introducción

La contaminación del aire se ha vuelto un tema cada vez más preocupante en México, sobre todo en las grandes ciudades como Guadalajara, Ciudad de México, Toluca o la propia Monterrey, pues su incremento ha llegado a puntos en los que la salud de la población puede verse comprometida. Instituciones como el Sistema Integral de Monitoreo Ambiental [SIMA] Nuevo León se encargan de estudiar la calidad del aire para entenderlos y buscar soluciones a los problemas actuales y futuros de contaminación.

En esta línea, un estudio de carácter matemático-estadístico podría ser una herramienta especialmente útil para conocer de manera más exacta las relaciones entre las variables contaminantes y climatológicas y cómo estas influyen en la calidad del aire de nuestra ciudad. El objetivo de nuestra investigación es conocer cómo el viento (dirección y velocidad) influye en la concentración de los contaminantes, así como las relaciones entre ellos; y proveer al socio formador herramientas de análisis inferencial que consideren la información espacial con la que contamos para realizar descubrimientos que permiten una mejor toma de decisiones para la generación de nuevas políticas públicas que promuevan la mejora de la calidad del aire y el cuidado ambiental en nuestra ciudad.

## 2. Metodología

SIMA cuenta con 15 estaciones de monitoreo ambiental distribuidas a través del área metropolitana de Monterrey que capturan lecturas sobre los contaminantes (CO, NO, NO<sub>2</sub>, NOX, O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>) y las condiciones meteorológicas (presión atmosférica, precipitación, humedad relativa, radiación solar, temperatura, velocidad del viento, dirección del viento). Se nos entregaron cinco bases de datos con datos correspondientes a los años 2020 a 2024, cuyas lecturas estaban promediadas para ajustarse a una frecuencia horaria.

### 2.1. Limpieza y Preprocesamiento de Datos

Se decidió trabajar únicamente con la base de datos más reciente: 2023 - 2024. A partir de ella, se separaron 15 bases de datos individuales, una para cada estación, que se trabajaron de forma paralela. De ellas, se seleccionaron únicamente las variables de interés correspondientes a los contaminantes más la dirección y velocidad del viento.

Con un pequeño análisis exploratorio fue evidente que había varios datos nulos y algunos otros datos que parecían poco realistas. Por ejemplo, había períodos anormalmente largos (meses) en los que una variable que se sabía periódica o caótica, mantenía un valor casi constante; así como valores excesivamente atípicos de concentración de contaminantes en momentos puntuales. Respecto a estos últimos, se cuestionó su validez directamente con los socios formadores, quienes confirmaron que se trataba de datos válidos consecuencia de actividades humanas que no debían corregirse. El resto de datos se imputaron/corrigieron utilizando interpolación basada en el contexto.

Originalmente se pensó en una interpolación lineal basada en los datos existentes adyacentes pero, si bien resultaba muy funcional cuando los datos nulos estaban aislados entre sí, otorgaba resultados muy poco realistas cuando había cantidades considerables de datos consecutivos a interpolar. Para solucionar esto se optó por una interpolación spline, que ajusta polinomios de cúbicos por partes entre cada par de puntos de datos, lo que garantiza una curva suave manteniendo la continuidad tanto en la primera como en la segunda derivada en los puntos de conexión, conocidos como nudos. La cantidad de datos imputados fueron 39376, es decir, el 3.64% de los datos totales.

### 2.2. Análisis Direccional

Se realizó una investigación bibliográfica sobre técnicas existentes para el análisis del viento y la dispersión de contaminantes en el aire. A partir de ella se determinaron algunos métodos que podían aplicarse sobre los datos para relacionar el viento con la concentración de los contaminantes, e incluso, con las relaciones entre contaminantes. A continuación se describen dichas técnicas.

El análisis direccional fue nuestro principal marco de trabajo. Este tipo de análisis tiene su diferenciador en su capacidad de entender la información del viento (dirección y velocidad) no solo como un par de variables numéricas más, sino como información con un contexto espacial que puede brindar mucha más información. Nuestro objetivo con esto era poder establecer relaciones geográficas visuales sobre los contaminantes, no solo en cada estación de forma aislada, sino aprovechando la información de todas.

### 2.2.1. Gráficos Polares de Concentración de Contaminantes

Estos gráficos permiten visualizar la relación entre la velocidad del viento ( $r$ ), su dirección ( $\theta$ ) y la concentración de un determinado contaminante (color) a través de un mapa de calor en coordenadas polares, donde el rojo representa mayor concentración, al contrario del color azul.

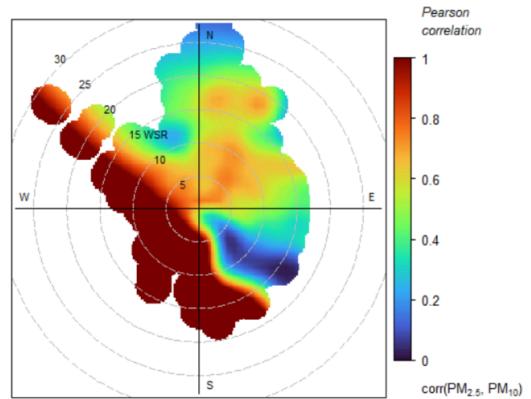
Además de conocer información general sobre el comportamiento del viento (por ejemplo, en la gráfica de la derecha, donde se observa que el viento alcanza mayores velocidades cuando viaja en dirección noroeste o sureste), este tipo de gráficos permite hacer inferencias sobre el efecto local del viento en los contaminantes, como si son arrastrados por el viento (como en la gráfica de la izquierda, donde las concentraciones aumentan con altas concentraciones) o bien si son dispersados por este (como en la gráfica derecha, donde la concentración disminuye a medida que las velocidades del viento aumentan).

Estas gráficas representan una superficie continua definida por la intensidad de los contaminantes en cada una de las coordenadas polares del viento que se tienen disponibles en la base de datos. Este resultado se logra a través de modelación matemática, utilizando también técnicas de smoothing. La primera de estas técnicas es formar pequeños bins de entre  $5^\circ$  y  $10^\circ$  en la dirección y 40 intervalos para la velocidad del viento, pues partir en estos conjuntos (cuyas dimensiones tienen antecedente bibliográfico) se puede reducir la cantidad de datos, capturando con suficiente detalle el comportamiento de estas variables, sin perder fidelidad.

Utilizando las medias de cada bin, se aplica una transformación de los datos, pasando de coordenadas polares (dirección, velocidad) a componentes rectangulares –para aprovechar las bondades de trabajar con una función isotrópica–. Con estos componentes se ajusta un Generalized Additive Model (GAM) que permite relacionarlos con la raíz cuadrada de las concentraciones, pues esta transformación suele generar datos con distribuciones más amigables (como vimos en clase) que facilitan una interpolación precisa. Trabajar con GAMs es conveniente en estudios relativos a la contaminación del aire porque las relaciones no suelen ser lineales y la interacción entre variables toma un rol significativo. Con este modelo ajustado correctamente, que representa una superficie continua, se pueden calcular y graficar las concentraciones asociadas a una coordenada polar incluso en los "huecos" (observaciones con cierta velocidad y dirección que no se tienen) que hubiera en el espacio de los componentes rectangulares.

### 2.2.2. Pairwise Statistics y Gráficos Polares de Correlación

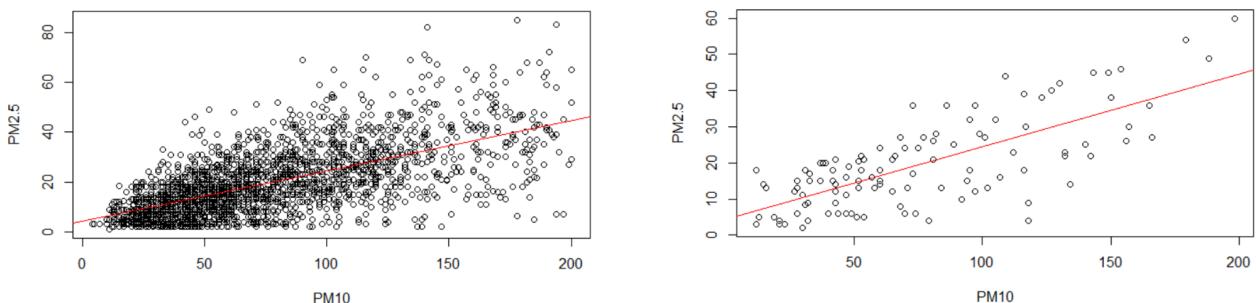
Tomando como base la teoría detrás del gráfico polar de concentración, se plantea un gráfico polar del coeficiente de correlación de Pearson de un par de contaminantes. Pairwise statistics aprovecha el proceso de categorización/binarización de las observaciones muestrales y en vez de usar el modelo aditivo generalizado (GAMs) del PolarPlot convencional debido a que los bins generados podrían tener pocos pares de observaciones y asignar el mismo peso a cada par en la prueba no genera resultados acertados, utiliza un kernel Gaussiano para asignar el peso a cada par de variables en base a su proximidad con el centroide del bin, asegurando a su vez que cada par de variables esté estandarizado y tenga significación en el cálculo de la regresión.



Se observó que existen pares de contaminantes que cambian notablemente su correlación en función de la dirección del viento. El caso de la gráfica mostrada arriba resultó especialmente curioso por lo radical que es el cambio, pues resulta que cuando el viento sopla entre  $160^\circ$  y  $290^\circ$  la correlación entre PM2.5 y PM10 resulta ser muy alta, a diferencia del resto de direcciones. Estos hallazgos hacen pensar que crear modelos que relacionen un par de variables contaminantes puede resultar completamente inviable en algunos comportamientos del viento, pero muy alcanzable en otros.

### 2.3. Regresión Lineal

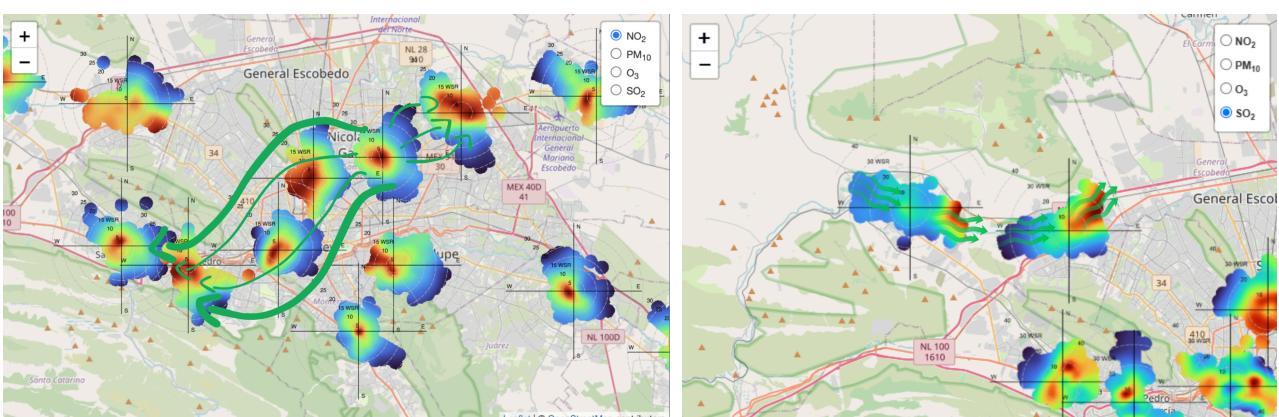
Considerando la información obtenida con pairwise statistics y los gráficos polares de correlación, se planteó hacer un modelo de regresión que permitiera explicar la relación de PM2.5 con PM10 que, se espera, debería existir y ser “sencilla”. Para valorar la influencia del viento en la viabilidad para implementar un modelo de este tipo se contrasta un modelo ajustado con la totalidad de los datos (excluyendo atípicos e influyentes) con otro modelo ajustado únicamente con datos filtrados para una estación del año determinada y condiciones de viento que propician una alta correlación según pairwise statistics (excluyendo atípicos e influyentes).



En estos gráficos de dispersión puede observarse que los datos de la gráfica izquierda (con un ajuste  $r^2$  de 0.45), si bien tienen una tendencia que podría decirse lineal, la dispersión y heterocedasticidad vuelven complicado crear un modelo robusto, pues de los cuatro supuestos (residuos con media cero, normalidad, homocedasticidad e independencia) ninguno se cumple. Una vez que se filtran los datos, el ajuste  $r^2$  pasa a ser de 0.74 y los residuos cumplen ahora con media cero y normalidad. Aún con este progreso, no recomendaría utilizar un modelo así, pues no cumple con la totalidad de los supuestos. Valdría la pena probar con métodos distintos, como regresión cuantílica o regresión robusta. Queda como trabajo a futuro.

### 3. Resultados

A partir de los procesos descritos anteriormente se generaron mapas interactivos que permiten visualizar las gráficas polares en contexto geográfico para cada estación, facilitando el entendimiento de los contaminantes a través del viento sobre toda el área metropolitana.



En la gráfica izquierda puede observarse un corredor de viento que arrastra ligeramente contaminantes entre estaciones. En la gráfica de la derecha se observa un pequeño caso de flujo de contaminantes en el viento. Los mapas interactivos están disponibles en nuestro [repositorio](#).

### 4. Conclusión

En mi opinión se cumplió con el objetivo que nos propusimos en un inicio: brindarle al socio formador una herramienta de trasfondo estadístico-matemático para conocer más sobre el comportamiento de los contaminantes en toda el área metropolitana. Si bien, nosotros no poseemos el conocimiento teórico de química y meteorología que se requiere para inferir adecuadamente a partir de la información visual que generamos, estamos seguros de que nuestros socios formadores sí lo tienen y pueden aprovechar de gran manera nuestro trabajo para encontrar insights que les permitan promover iniciativas que mejoren la situación de la ciudad en materia de contaminación del aire.

Algunas cosas que podrían implementarse a futuro serían pruebas estadísticas para confirmar, bajo cierto nivel de confianza, algunos fenómenos identificables en las gráficas, como los corredores de viento, el flujo, concentración y dispersión de contaminantes por la acción del viento, entre muchos otros. Para ello deberán contrastarse las lecturas de distintas estaciones de monitoreo que estén relacionadas por la acción del viento, en tiempos con un ligero desfase (lag).

## Referencias

1. OMS (2024) Contaminación del aire ambiente (exterior) y salud. Organización Mundial de la Salud. Centro de Prensa. Notas Descriptivas. [https://www.who.int/es/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/es/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)
2. Secretaría de Medio Ambiente de Nuevo León (2022) PIGECA: Plan Integral para la Gestión Estratégica de la Calidad del Aire.
3. EPA. (2021). National Air Quality: Status and Trends of Key Air Pollutants. United States Environmental Protection Agency.
4. World Health Organization. (2021). Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide.
5. OMS. (2021). WHO global air quality guidelines: particulate matter (PM<sub>2.5</sub> and PM<sub>10</sub>), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Organización Mundial de la Salud.
6. Gobierno de México(2023). Índice AIRE Y SALUD, herramienta para informar sobre la calidad del aire y las recomendaciones para proteger la salud. <https://www.gob.mx/comisionambiental/documentos/presentacion-de-la-conferencia-de-prensa>
7. Zhang, X., & Ma, L. (2024). Impact of commuting on mental well-being: Using time-stamped subjective and objective data. *Transportation Research Part F: Psychology and Behaviour*, 107, 395–412.<https://doi.org/10.1016/j.trf.2024.09.009>
8. Carslaw, D. C. (2024). Polar plots. In D. C. Carslaw & J. Davison, *The openair book: A guide to the analysis of air pollution data*. Retrieved November 21, 2024, from [https://bookdown.org/david\\_carslaw/openair/sections/directional-analysis/polar-plots.html](https://bookdown.org/david_carslaw/openair/sections/directional-analysis/polar-plots.html)
9. Wang, X., Zhang, Y., Chen, H., Yang, X., & Chen, J. (2018). Relationships between fine particle pollution and meteorological parameters in winter in Beijing, Tianjin and Shijiazhuang, China. *Atmospheric Environment*, 173, 210-220.
10. Stull, R. B. (2012). *An Introduction to Boundary Layer Meteorology* (Vol. 13). Springer Science & Business Media.
11. Kumar, R., Barth, M. C., Pfister, G. G., Naja, M., & Brasseur, G. P. (2015). WRF-Chem simulations of a typical pre-monsoon dust storm in northern India: influences on aerosol optical properties and radiation budget. *Atmospheric Research*, 164-165, 142-155.
12. Zhang, Y. L., & Cao, F. (2015). Fine particulate matter (PM<sub>2.5</sub>) in China at a city level. *Scientific Reports*, 5, 9238.
13. Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (3rd ed.). John Wiley & Sons.
14. Meng, X., Liu, C., Zhang, L., Wang, W., Stowell, J., Kan, H., & Liu, Y. (2020). Estimating PM<sub>2.5</sub> concentrations in Northeastern China with full spatiotemporal coverage, 2016-2018. *Environmental Pollution*, 266, 115284.
15. Sosa-Echeverría, R., Alarcón-Jiménez, A. L., Torres-Barrera, M. del C., Jaimes-Palomera, M., Retama-Hernández, A., Sánchez-Álvarez, P., Granados Hernández, E., & Bravo-Álvarez, H. (2018). Spatial and temporal variation of acid rain in the Mexico City Metropolitan Zone. *Atmósfera*, 32(1), 71–85. <https://doi.org/10.20937/ATM.2019.32.01.05>