



Tecnológico de Monterrey

Etapa 2. Conociendo los datos

Aplicación de métodos multivariados en ciencia de datos MA2003B (Gpo 201)

Equipo 5

Profesores:

Blanca Rosa Ruiz Hernández

Jesús José Armenta Segura

Integrantes:

Gisel Regina Benítez Calvillo | A00228137

Julio Avelino Amador Fernández | A01276513

Luis Eduardo Martínez Aguirre | A00840037

Miguel Angel González Gutiérrez | A01198604

Jueves, 09 de Octubre de 2025

Objetivo Principal

Detectar y explicar patrones temporales, geográficos y ambientales modulados por fenómenos como inversión térmica, vientos débiles y lluvia, y los contrastes invierno-verano, identificando como variables a contaminantes atmosféricos; cuantificar y generar modelos predictivos para anticipar y mitigar episodios críticos, mejorando la gestión y políticas públicas relacionadas con la salud en las áreas con mayor densidad poblacional dentro del Área Metropolitana de Monterrey.

Se decidió enfocar el objetivo a las políticas públicas respecto a la salud ya que se consideró que el problema de la salud por efectos climatológicos y contaminantes atmosféricos se ha vuelto muy relevante en Monterrey. Es por ello que se decidió usar solamente las estaciones ubicadas en zonas con mayor densidad poblacional, excluyendo a aquellas lejos de zonas concurridas y ciudades grandes, en particular las estaciones NO₂, NO₃, NE₃, SE₂ y SE₃.

Es importante destacar que también se incluyeron sólo las variables relevantes respecto al objetivo. En particular, se excluyeron aquellas que no aportan información sobre fenómenos como inversión térmica, vientos débiles y lluvia, y contrastes invierno-verano o que resultan ser teóricamente redundantes y no funcionan como indicadores. Con estos motivos, se mantuvieron las variables O₃, NO₂, PM_{2.5}, PM₁₀, TOUT, RH, RAINF, SR, WSR, WDR.

Técnicas estadísticas que contestarán tu objetivo

Las técnicas estadísticas que permitirán contestar el objetivo principal del estudio, analizar la estacionalidad y variación temporal de los contaminantes atmosféricos en Monterrey, así como sus combinaciones y determinantes, se basan en métodos de análisis de series temporales, modelado multivariante y regresión. Estas herramientas posibilitan identificar patrones diarios y estacionales, evaluar la influencia de las condiciones meteorológicas y desarrollar modelos predictivos que anticipen episodios críticos de contaminación.

El análisis de series temporales, como el modelo ARMA estacional, permite detectar tendencias y ciclos en las concentraciones de contaminantes a lo largo del tiempo. Este tipo de modelo resulta útil para captar los patrones estacionales característicos del invierno y el verano en Monterrey, y para generar predicciones a corto y mediano plazo del Índice de

Calidad del Aire. Además, se propone aplicar variantes más avanzadas como los modelos SARIMA, que permiten ajustar mejor las fluctuaciones por estaciones y eventos meteorológicos específicos como la inversión térmica o los vientos débiles.

Por otro lado, el uso de modelos multivariantes, como los modelos espaciales gaussianos implementados en estudios previos en Europa, permitirá analizar la relación simultánea entre distintos contaminantes ($PM_{2.5}$, PM_{10} , O_3 , NO_2 , SO_2) y sus interdependencias. Este enfoque ayuda a identificar combinaciones de contaminantes que, al coexistir, potencian los efectos negativos sobre la calidad del aire. También permite incorporar variables meteorológicas (temperatura, lluvia, velocidad del viento) como factores explicativos en la variación de las concentraciones, mejorando la interpretación de los resultados.

Asimismo, se plantea el uso de modelos de regresión lineal múltiple o regresión no lineal para cuantificar la relación entre las variables meteorológicas y los niveles de contaminantes. Estas técnicas permitirán estimar la magnitud y dirección de los efectos de cada variable ambiental, además de facilitar la construcción de modelos predictivos simples que complementen los pronósticos más complejos del sistema actual.

En conjunto, las técnicas de análisis de series de tiempo, modelado multivariante y regresión estadística constituyen un marco robusto para cumplir los objetivos del proyecto. Estas herramientas no solo ayudarán a describir y entender los patrones de contaminación en Monterrey, sino que también servirán para desarrollar indicadores predictivos y apoyar la formulación de políticas públicas basadas en evidencia científica.

Descripción y justificación de variables escogidas

Variable	Descripción	Tipo	Unidades	Valores posibles	Valores atípicos	Valores nulos
NO_2 – Dióxido de nitrógeno	Gas contaminante primario generado principalmente por la combustión vehicular e industrial. Es precursor del ozono troposférico y su concentración	Numérico	ppb	0 – 175 (2025)	4,611 – 1.36%	14,651 – 4.11%

	aumenta con inversión térmica, baja ventilación y tráfico intenso.					
O ₃ – Ozono	Contaminante secundario formado por reacciones fotoquímicas entre NO _x y COVs en presencia de radiación solar. Altamente dependiente de la meteorología. Es el principal indicador de episodios críticos y contingencias	Numérico	ppb	0 – 185 (2025)	4,460 – 1.31%	17,021 – 5.01%
PM ₁₀ – Material particulado menor a 10 micrómetros	Partículas sólidas o líquidas suspendidas en el aire de origen natural o antropogénico. Afectan la salud respiratoria y su concentración varía con el viento, precipitación y estabilidad atmosférica.	Numérico	µg/m ³	0 – 820 (2025)	4717 – 1.39%	12,487 – 3.67%
PM _{2.5} – Material particulado menor a 2.5 micrómetros	Partículas finas que penetran profundamente en el sistema respiratorio.	Numérico	µg/m ³	0 – 350 (2025)	2441 – 0.72%	100,458 – 29.56%
RAINP – Precipitación	Cantidad de agua caída en forma de lluvia. Tiene un efecto depurador sobre la atmósfera (reduce PMy gases solubles) y se relaciona con el ciclo estacional.	Numérico	mm/hr	0 – 25 (2025)	514 – 0.15%	5,985 – 1.76%
RH – Humedad	Porcentaje de vapor de agua en el aire.	Numérico	%	0 – 100 (2025)	8 – 0.00%	34,606 – 10.18%

relativa	Influye en la formación de aerosoles, deposición húmeda y dispersión de contaminantes..					
SO ₂ – Dióxido de azufre	Gas contaminante primario generado por la quema de combustibles fósiles con contenido de azufre. Aunque sus concentraciones actuales suelen ser bajas, sigue siendo monitoreado por normativa.	Numérico	ppb	0 – 405 (2025)	3908 – 1.15%	22,816 – 6.71%
SR – Radiación solar	Energía solar incidente, esencial para las reacciones fotoquímicas que generan ozono. Varía estacionalmente y es fundamental para explicar los picos de O ₃ .	Numérico	kW/m ²	0 – 1.2 (2025)	6 – 0.00%	15,151 – 4.46%
TOUT – Temperatura ambiente	Temperatura del aire exterior. Influye directamente en la formación de ozono y en el desarrollo de inversiones térmicas.	Numérico	°C	-4.5 – 45 (2025)	4 – 0.00%	15,735 – 4.63%
WSR – Velocidad del viento	Rapidez del viento superficial. Determina la dispersión o acumulación de contaminantes: vientos débiles favorecen estancamiento; vientos fuertes, dispersión.	Numérico	km/hr	0 – 40 (2025)	1315 – 0.39%	9,184 – 2.70%
WDR – Dirección del viento	Indica la procedencia del viento. Es clave para identificar el	Numérico	°	0-360 (2025)	0 – 0.00%	7,711 – 2.27%

	transporte de contaminantes desde distintas fuentes.					
--	--	--	--	--	--	--

Las variables seleccionadas permiten analizar los factores que influyen en la calidad del aire y su efecto en la salud. Los contaminantes O3, NO2, PM2.5 y PM10 se incluyeron por su relevancia sanitaria y su relación con fenómenos atmosféricos; el O3 refleja procesos fotoquímicos ligados a la radiación y la temperatura, el NO2 representa emisiones vehiculares e industriales, y las partículas indican riesgos directos para la salud. Las variables meteorológicas (TOUT, RH, RAINF, SR, WSR, WDR) ayudan a explicar la dispersión, acumulación o eliminación de contaminantes: la temperatura y radiación influyen en la formación de ozono, la lluvia limpia el aire, la humedad afecta la formación de partículas, y el viento determina la dispersión y dirección de transporte. En conjunto, estas variables permiten identificar patrones estacionales, efectos de inversión térmica y episodios críticos de contaminación en zonas pobladas.

Exploración de datos

Variable <chr>	Media <dbl>	Mediana <dbl>	Moda <dbl>	Mínimo <dbl>	Máximo <dbl>	Rango <dbl>	Varianza <dbl>	Desviación <dbl>
O3 (ppb)	26.92021904	23.000	11.00000	-3.098239e+01	212.0677	243.05004	3.590815e+02	18.9494463
NO2 (ppb)	16.73369758	13.700	6.70000	-1.184203e+01	158.0000	169.84203	1.313556e+02	11.4610477
PM2.5 (ug/m3)	19.85616492	16.280	10.00000	-1.286035e+01	999.0000	1011.86035	2.540006e+02	15.9373956
PM10 (ug/m3)	59.46344643	50.000	43.00000	-8.637302e-01	1000.0000	1000.86373	2.068262e+03	45.4781516
TOUT (°C)	-31.75036662	24.390	26.84526	-1.015892e+04	948.3667	11107.29112	4.010011e+05	633.2465149
RH (%)	-28.37353363	56.000	79.00000	-9.999393e+03	298.6795	10298.07282	5.567683e+05	746.1690556
RAINF (mm/h)	0.02345920	0.000	0.00000	-6.114401e-01	83.4500	84.06144	3.947448e-01	0.6282872
SR (kW/m2)	0.09765859	0.002	0.00000	-9.999000e+03	7.7860	10006.78600	5.182254e+02	22.7645639
WSR (km/h)	8.22771031	7.500	5.80000	-2.000544e-01	112.7000	112.90005	2.116554e+01	4.6006024
WDR (azimutal)	146.82992069	126.000	141.00000	1.000000e+00	371.5302	370.53023	9.039454e+03	95.0760419

Figura 1. Medidas de tendencia central y de dispersión para todas las variables.

De las medidas de tendencia central, se observa un comportamiento inusual para *TOUT* y *RH*. Principal insight, es considerar una transformación para mejorar la descripción de estas.

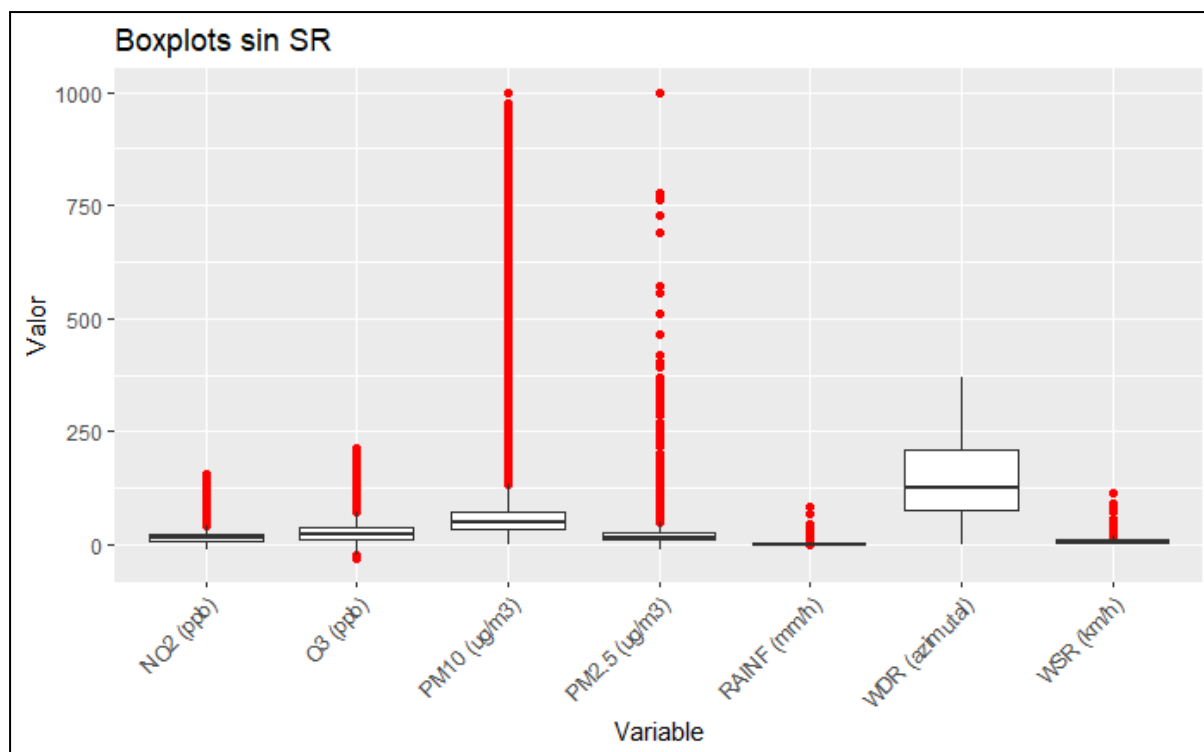


Figura 2. Gráficas de caja de variables, exceptuando RH (%), TOUT (°C) y SR(kW/m2)

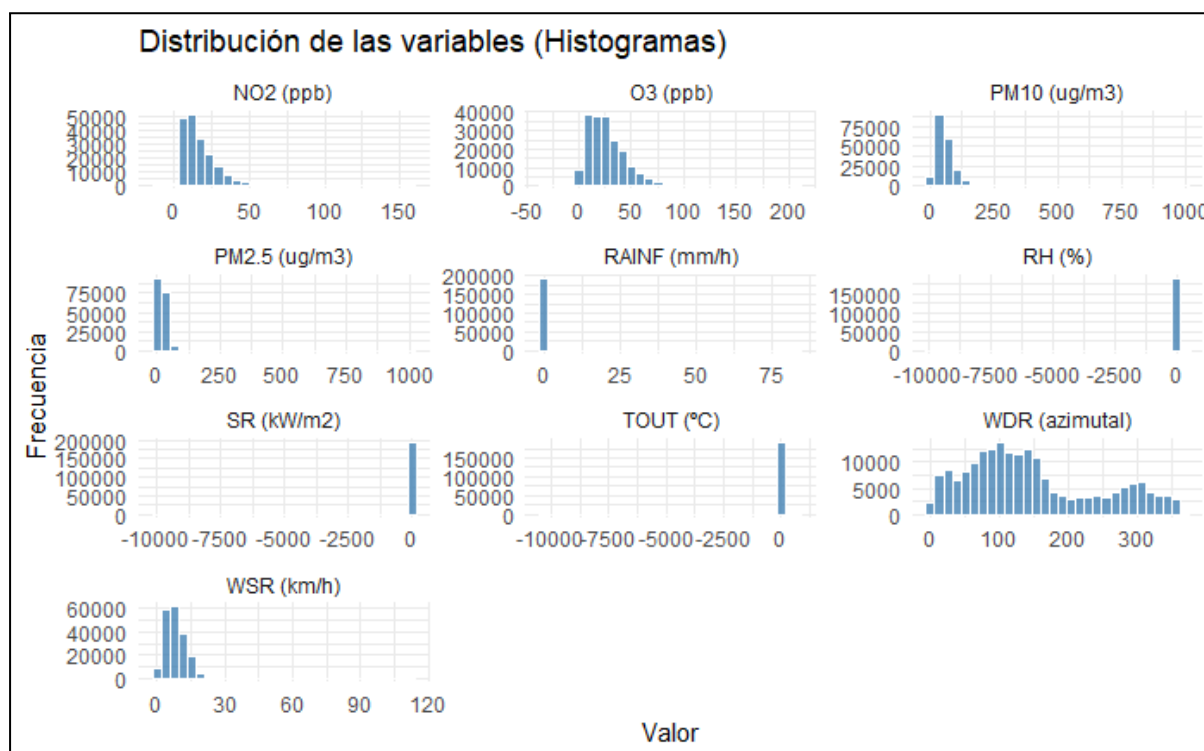


Figura 3. Distribución de histogramas de total de variables.

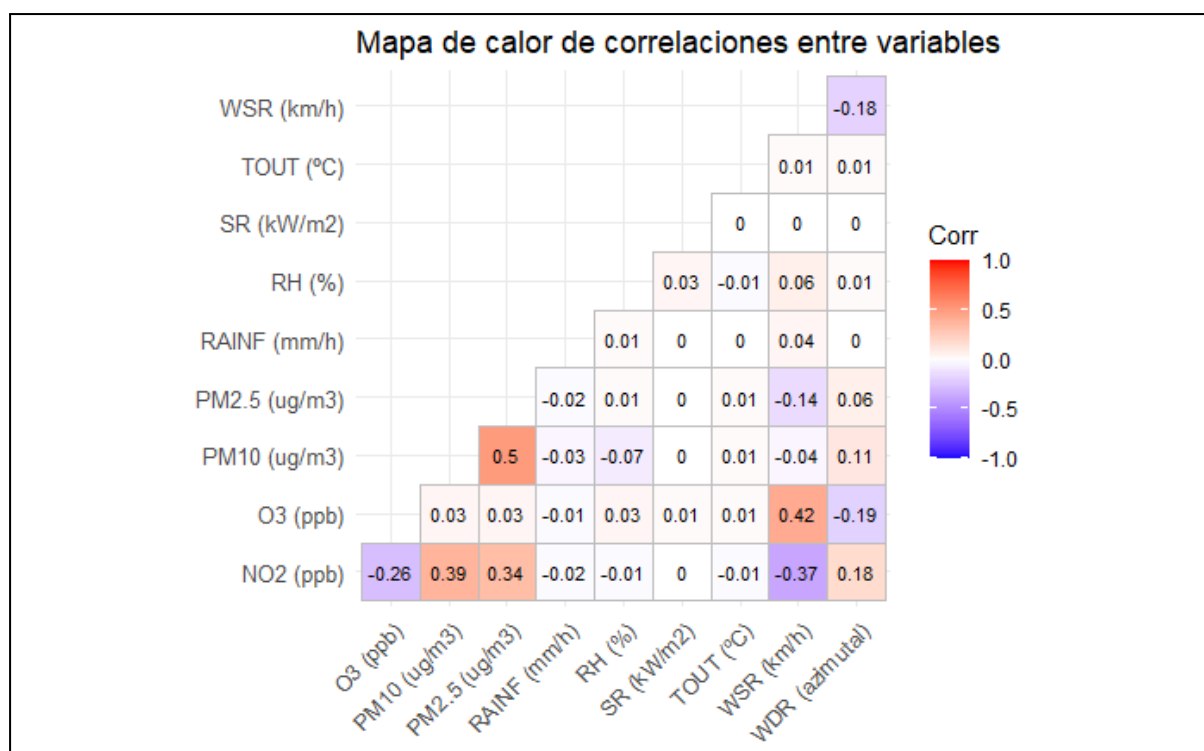


Figura 4. Mapa de calor de correlaciones entre total de variables.

Anexo: <https://github.com/mikeg00/Evidencias-Reto-MA2003B.201>

Referencias bibliográficas

- Zavala-Romero, O., Segura-Chavez, P. A., Camacho-Gonzalez, P., Zavala-Hidalgo, J., Garcia, A. R., Oropeza-Alfaro, P., Romero-Centeno, R., & Gomez-Ramos, O. (2024). Operational Ozone forecasting system in Mexico city: A machine learning framework integrating forecasted weather and historical Ozone data. *Atmospheric Environment*, 121017. <https://doi.org/10.1016/j.atmosenv.2024.121017>
- Ruiz-Villavicencio, E., López-López, M. Á., Cetina-Alcalá, V. M., & Ramírez-Guzmán, M. E. (2020). MODELACIÓN y ESTIMACIÓN DE NO₂ y O₃ EN ZONAS RURALES y SUBURBANAS DEL VALLE DE MÉXICO. *Revista Internacional De Contaminación Ambiental*. <https://doi.org/10.20937/rica.5354>
- Yang, Q., Yuan, Q., Li, T., Shen, H., & Zhang, L. (2017, August 21). *The relationships between PM_{2.5} and meteorological factors in China: Seasonal and regional variations*. arXiv.org. <https://arxiv.org/abs/1708.06072>
- Kirešová, S., & Guzan, M. (2022). Determining the Correlation between Particulate Matter PM₁₀ and Meteorological Factors. *Eng—Advances in Engineering*, 3(3), 343–363. <https://doi.org/10.3390/eng3030025>
- Ramos-Herrera, S., Bautista-Margulis, R., & Valdez-Manzanilla, A. (23 de marzo, 2010). *Estudio estadístico de la correlación entre contaminantes atmosféricos y variables meteorológicas en la zona norte de Chiapas, México*. https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0186-29792010000100005
- ¿Cuáles son las principales fuentes de precursores de ozono? (2020, December 3). PortalAmbiental.com.mx.

<https://www.portalambiental.com.mx/sabias-que/20201201/cuales-son-las-principales-fuentes-de-precursores-de-ozono>