



**INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE
MONTERREY**

Aplicación de métodos multivariados en ciencia de datos

Blanca Rosa Ruiz Hernández

Adriana Reyna Lara

Proyecto integrador

Etapas 1. Conociendo el negocio

Integrantes:

Carola Vázquez Arjona A01174562

Paola Michelle Martínez Galeazzi A00839082

Yamil Elias Del Blanco Chávez A00838610

Introducción

La calidad del aire constituye uno de los principales retos ambientales y de salud pública en las ciudades contemporáneas, ya que su deterioro por la presencia de contaminantes tiene un efecto altamente dañino en la salud humana y del medio ambiente. Según estudios realizados en distintas ciudades del país, existen relaciones directas entre el incremento en la concentración de contaminantes en el aire y enfermedades respiratorias y cardiovasculares (Gobierno de la Ciudad de México, s. f.). Este hecho ha motivado guías nacionales con límites cada vez más estrictos para los contaminantes, al reconocerse que el daño puede ocurrir incluso a niveles más bajos que los considerados seguros hace una década (OMS, 2021). México unificó la comunicación del riesgo a través del Índice de Aire y Salud (IAS), estandarizado a nivel nacional por la NOM-172-SEMARNAT-2023 (SEMARNAT, 2024). El IAS se calcula obteniendo valores representativos de cada contaminante en todas las estaciones y por períodos de promediación definidos dependiendo del contaminante. Esto se realiza con la finalidad de categorizar cada contaminante en una escala común de cinco categorías que van desde “Buena” hasta “Extremadamente mala” (INECC, 2025). Por disposición normativa, la condición que se difunde al público por estación es la más riesgosa entre sus contaminantes. Esto permite que el índice sea comparable entre ciudades y que todas las redes comuniquen las mismas categorías y mensajes de riesgo (SEMARNAT, 2024).

En Nuevo León, el Sistema Integral de Monitoreo Ambiental (SIMA), un organismo de la Secretaría de Medio Ambiente, opera la red metropolitana que mide en automático los contaminantes criterio, las variables meteorológicas, y publica estos resultados en tiempo real a través de aire.nl.gob.mx (SIMA, s. f.). Más que un sistema de monitoreo, SIMA representa un esfuerzo institucional por garantizar transparencia, acceso a la información ambiental para todos los ciudadanos, generando evidencia para orientar políticas públicas, diseñar planes de acción y proteger la salud de los habitantes del área metropolitana. El análisis de la calidad del aire es importante para la ciudadanía porque permite convertir mediciones técnicas en señales de riesgo accionables para ajustar actividades al aire libre, activar protocolos escolares/deportivos y orientar acciones de contingencia cuando las condiciones lo ameritan (INECC, 2025). En regiones como la Zona Metropolitana de Monterrey, un contexto urbano e industrial donde convergen actividad económica, tráfico vehicular intenso, pedreras, refinerías y una compleja topografía montañosa, el aire que se respira no siempre cumple con los estándares que garantizan la protección de la salud. El análisis de la calidad del aire debe

ser, por lo tanto, una prioridad no solo para las autoridades, sino también para la ciudadanía en general, donde una buena traducción de la medición técnica a una señal de riesgo puede significar la prevención de enfermedades y la reducción de la exposición de poblaciones vulnerables.

Un buen análisis de la calidad del aire involucra muchos factores, primeramente los contaminantes criterio: $PM_{2.5}/PM_{10}$, partículas asociadas a mortalidad y morbilidad cardiovascular y respiratoria; O_3 , oxidante secundario vinculado a exacerbaciones asmáticas y hospitalizaciones; NO_2 , irritante y precursor de O_3 /partículas; SO_2 , irritante de fuentes industriales; y CO , extremadamente tóxico en altas concentraciones, aunque usualmente bajo en aire exterior urbano (OMS, 2021). Para realizar el análisis de contaminantes, el IAS se apoya en límites sanitarios y reglas de comunicación. En México, los límites para contaminantes criterio están establecidos en Normas Oficiales Mexicanas (NOM) específicas. Cada NOM es la regla sanitaria nacional que fija cómo evaluar y decidir si el aire cumple o no con un nivel de protección a la salud. Para cada especie, la NOM indica que cantidad de tiempo promediar, cuál debe ser el valor límite de concentración y los criterios de validación de los datos. La NOM-025-SSA1-2021 fija ventanas de 24 horas y anual para $PM_{2.5}$, y PM_{10} , ya que, las partículas finas y gruesas se relacionan con eventos agudos (24 h) y eventos crónicos (anuales). La NOM-020-SSA1-2021 fija 1 hora y un promedio móvil de 8 horas para O_3 , ya que es fotoquímico y tiene picos diurnos; el promedio de 8 h capta la exposición sub diaria típica, y la 1 h controla picos muy altos de concentración. La NOM-023-SSA1-2021 fija ventanas de 1 hora y anual para el NO_2 , ya que tiene impactos agudos e impactos crónicos. La NOM-022-SSA1-2019 fija ventanas de 1 hora y 24 horas para el SO_2 , puesto que puede provocar síntomas respiratorios incluso a exposiciones cortas; pero también se controlan episodios sostenidos. Por último, la NOM-021-SSA1-2021 fija ventanas de 1 hora y un promedio móvil de 8 horas para el CO , donde la ventana de 8 h refleja exposición subaguda típica en zonas de tránsito, y la ventana de 1 h limita picos (SSA, 2019, 2021a, 2021b, 2021c, 2021d). A nivel internacional, las guías de calidad del aire de la OMS (2021) recomendaron límites más estrictos, subrayando que no hay un umbral seguro claro para limitar los riesgos en la salud (OMS, 2021). En Estados Unidos, los Estándares Nacionales de Calidad del Aire (NAAQS) establecidos por la Environmental Protection Agency (EPA) sirven como otro punto de comparación. Con la finalidad de realizar estas comparativas analizaremos los 2 contaminantes con concentraciones más elevadas y persistentes en la Zona Metropolitana de Monterrey. Para O_3 la guía sanitaria de la OMS

(2021) recomienda que el promedio en 8 horas sea de 0.050 ppm, una cantidad más estricta que la establecida por la EPA de 0.070 ppm, por su parte, México ya incluye una trayectoria de endurecimiento del promedio de 8h (0.065→0.060→0.051 ppm), lo que acerca el estándar nacional a la OMS en el horizonte de implementación. En $PM_{2.5}$, la OMS recomienda que el promedio anual sea menor o igual a $5 \mu\text{g}/\text{m}^3$, lo cual es más exigente que la EPA ($9 \mu\text{g}/\text{m}^3$) (EPA, 2024). En México, el IAS traduce la situación horaria a clasificaciones nacionales: con un promedio móvil de 12 h, la clasificación Buena es < 15 , Aceptable $> 15\text{--}33$, Mala $> 33\text{--}79$, Muy mala $> 79\text{--}130$ y Extremadamente mala $> 130 \mu\text{g}/\text{m}^3$ (INECC, 2025). Esta información nos ayuda a establecer un marco de referencia para analizar correctamente la Zona Metropolitana de Monterrey, donde podremos ubicar cada lectura tanto en la escala operativa del IAS como frente a las metas sanitarias internacionales OMS/EPA.

Es importante considerar que además de los contaminantes, existen otros aspectos relevantes en el análisis de la calidad del aire. También se considera el aspecto meteorológico, es decir, el viento, la temperatura, humedad, radiación, presión, y todos los factores que rigen la dispersión, transformación y acumulación de contaminantes. Además, se consideran las fuentes y usos de suelo, ya que el tráfico, la industria, las pedreras, la resuspensión de polvo y la cobertura vegetal son aspectos muy importantes que modifican la cantidad y comportamiento de los contaminantes. Por último, también se considera el entorno urbano y topográfico, lo cual es un aspecto importante a considerar en el estudio de la calidad del aire en la Zona Metropolitana de Monterrey, un valle rodeado por sierras que favorece los estancamientos e inversiones térmicas, y las recirculaciones de gases que amplifican picos locales de contaminación. Este tipo de situaciones, las limitaciones de representatividad espacial específica, generan que un punto de medición pueda no ser capturado por completo, y representan, junto con los datos faltantes por interferencias instrumentales, las dificultades principales que se enfrentan en una correcta medición de la calidad del aire.

Descripción del Problema

La Zona Metropolitana de Monterrey combina fuentes urbanas e industriales con una topografía de valle que favorece estancamientos, recirculaciones e inversiones térmicas. Aunque México comunica el riesgo con un índice nacional unificado (IAS), persiste la duda de qué tan bien refleja ese índice la exposición real y los patrones espacio-temporales de Nuevo León. Este contexto motiva la pregunta central de este proyecto:

¿El IAS refleja de forma efectiva el riesgo real en la Zona Metropolitana de Monterrey dada su topografía y meteorología particulares?

Objetivo general

Evaluar la efectividad y representatividad del IAS en la Zona Metropolitana de Monterrey en su dimensión espacial, temporal y sanitaria, y proponer mejoras operativas de comunicación compatibles con la NOM-172.

Justificación del objetivo

Nuevo León presenta episodios recurrentes de contaminantes que afectan grupos sensibles, por lo tanto, evaluar cómo el IAS comunica el riesgo en microambientes del valle permite reducir exposición y orientar medidas preventivas. El IAS ofrece comparabilidad en México, pero validar su representatividad espacial evita sub o sobreestimaciones de riesgo en zonas específicas, sin perder la estandarización. Por último, se cuenta con datos históricos por estación, metadatos y variables meteorológicas de SIMA, por lo tanto, es reproducible y transferible a la operación cotidiana.

Descripción de las fuentes de información:

Para la realización del análisis se consultaron distintas bases de datos provenientes de fuentes oficiales con el objetivo de responder las preguntas pertinentes. Las fuentes de datos con las que se trabajaron fueron las siguientes:

- SIMA: Series horarios de contaminantes (PM₁₀, PM_{2.5}, O₃, NO₂/NO_x, SO₂, CO) y variables meteorológicas (viento, T, RH, presión, radiación, precipitación). Base oficial del ICAS en Nuevo León.
- Copernicus: Modelo digital de elevación de 30 m para pendiente, aspecto, Sky View Factor, profundidad de valle, rugosidad y bloqueo orográfico.
- INEGI/DENUE: Corredores industriales, clústeres y coberturas de uso de suelo en torno a estaciones.
- Red vial (INEGI): Vialidades por clase como proxy de tráfico vehicular.
- Vegetación (NDVI / coberturas): Índices NDVI y áreas verdes en buffers para evaluar efecto de vegetación.

Impacto social principal:

El impacto social principal de este estudio radica en la posibilidad de la necesidad de reevaluar el Índice de Aire y Salud (IAS), ya que en su forma actual no considera la influencia de la topografía urbana en la acumulación y dispersión de contaminantes. En ciudades como Monterrey, rodeadas de sierras y valles, las condiciones geográficas pueden favorecer la concentración de partículas y gases aun cuando los valores promedio de contaminantes no superen los umbrales normativos, generando una exposición mayor a la población de lo que reflejan los indicadores oficiales. Por ello, resulta de alta relevancia analizar la interacción entre factores ambientales como la dirección e intensidad del viento, la altura de la capa de mezcla atmosférica, la humedad relativa o la radiación solar con la estructura topográfica, ya que estas condiciones modulan los episodios de contaminación y determinan riesgos diferenciales para las comunidades. Al incorporar estas dimensiones se fortalece la capacidad de los sistemas de monitoreo y de las políticas públicas para emitir alertas más realistas, diseñar estrategias de mitigación focalizadas, y proteger la salud de la población en contextos urbanos complejos, como el de la Zona Metropolitana de Monterrey.

Fuentes Bibliográficas y de datos

Gobierno de la Ciudad de México. (s. f.). El monitoreo de la calidad del aire.

<http://www.aire.cdmx.gob.mx/default.php?opc=%27ZaBhnmI=%27>

EPA. (2024). *Final rule to strengthen the National Air Quality Health Standard for PM_{2.5} (Fact sheet)*. U.S. Environmental Protection Agency.

<https://www.epa.gov/system/files/documents/2024-02/pm-naaqs-overview.pdf> US EPA

INECC. (2025). *Índice Aire y Salud (IAS): características y aplicación* (presentación/guía, feb. 2025). Gobierno de México.

https://www.gob.mx/cms/uploads/attachment/file/982375/Indice_AIREYSALUD__febrero2025.pdf
Gobierno de México

Organización Mundial de la Salud. (2021). *WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*.

<https://www.who.int/publications/i/item/9789240034228>

SEMARNAT. (2024). *NOM-172-SEMARNAT-2023. Lineamientos para la obtención y comunicación del índice de calidad del aire y riesgos a la salud*. Diario Oficial de la Federación (25 de enero de 2024). https://www.dof.gob.mx/nota_detalle.php?codigo=5715154&fecha=25/01/2024 [Diario Oficial de la Federación](#)

SIMA (Gobierno de Nuevo León). (s. f.). *Portal del Sistema Integral de Monitoreo Ambiental (SIMA)*. <https://aire.nl.gob.mx/> aire.nl.gob.mx

SSA. (2019). *NOM-022-SSA1-2019. Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al dióxido de azufre (SO₂)*. Diario Oficial de la Federación (20 de agosto de

2019). <https://rama.edomex.gob.mx/sites/rama.edomex.gob.mx/files/files/NOM-022-SSA1-2019.pdf>
rama.edomex.gob.mx

SSA. (2021a). *NOM-020-SSA1-2021. Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al ozono (O_3)*. Diario Oficial de la Federación (28 de octubre de 2021). https://www.dof.gob.mx/nota_detalle.php?codigo=5633956&fecha=28/10/2021 [Diario Oficial de la Federación](#)

SSA. (2021b). *NOM-021-SSA1-2021. Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al monóxido de carbono (CO)*. Diario Oficial de la Federación (29 de octubre de 2021). https://www.dof.gob.mx/normasOficiales/8781/salud_3_C/salud_3_C.html [Diario Oficial de la Federación](#)

SSA. (2021c). *NOM-023-SSA1-2021. Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al dióxido de nitrógeno (NO_2)*. Diario Oficial de la Federación (27 de octubre de 2021). <https://rama.edomex.gob.mx/sites/rama.edomex.gob.mx/files/files/NOM-023-SSA1-2021.pdf>
rama.edomex.gob.mx

SSA. (2021d). *NOM-025-SSA1-2021. Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto a las partículas suspendidas PM_{10} y $PM_{2.5}$* . Diario Oficial de la Federación (27 de octubre de 2021). https://dof.gob.mx/nota_detalle.php?codigo=5633855&fecha=27/10/2021

Liga a GitHub con base de datos limpia:

<https://github.com/carolavazquez/RETO-MA2003B>

Comprensión de los datos

En primer lugar, se llevó a cabo una exploración inicial del conjunto de datos. La base de datos original utilizada corresponde al periodo 2023-2024, con un tamaño total de 13,872 registros y 240 columnas. La información está concentrada en una sola hoja; cada fila representa una marca de tiempo común a todas las mediciones y las columnas están organizadas por pares estación-variable. Las estaciones incluidas son: Sureste, Noreste, Centro, Noroeste, Suroeste, Noroeste 2, Norte, Noreste 2, Sureste 2, Suroeste 2, Sur, Norte 2 y Sureste 3. En cada estación se reporta el mismo conjunto de variables de medición: CO (monóxido de carbono), NO (óxido nítrico), NO_2 (dióxido de nitrógeno), NOX (óxidos de nitrógeno totales), O_3 (ozono), PM_{10} (partículas $\leq 10 \mu m$), $PM_{2.5}$ (partículas $\leq 2.5 \mu m$), PRS (presión atmosférica), $RAIN$ (precipitación), RH (humedad relativa), SO_2 (dióxido de azufre), SR (radiación solar), $TOUT$ (temperatura), WSR (velocidad del viento) y WDV (dirección del viento). Todas son variables numéricas continuas, WDV es un ángulo en grados de 0 a 360 y, aunque se almacena como número, se interpreta como variable circular. La hoja incluye además una columna de tiempo que sirve como referencia para alinear todas las series.

Se revisó la calidad de la información con tres focos: valores faltantes, valores fuera de rango físico y valores falsos. Los faltantes se presentan de forma desigual por estación y variable, incluyendo tramos con vacíos prolongados asociados a paros de sensor o mantenimiento. En los contaminantes se detectaron registros negativos, que no son físicamente posibles, y picos aislados mucho más altos de lo esperable en un entorno urbano. En meteorología se observaron inconsistencias como humedades relativas por encima de 100%, velocidades de viento negativas o direcciones fuera del rango 0-360. Estas observaciones guiaron las reglas de depuración posteriores.

Preparación de los datos

La selección del conjunto de datos de 2023-2024 fue priorizada porque el enfoque del estudio es topográfico y requiere cobertura espacial completa: usar todas las estaciones activas al mismo tiempo es más informativo para contrastar efectos de relieve y ubicación que acumular más años con posibles diferencias de cobertura o configuración de la red. Por ello se incluyeron todas las estaciones numeradas y se trabajó con el total de variables disponibles por estación.

Las reglas de depuración aplicadas fueron conservadoras y transparentes. Primero, no se realizaron imputaciones con promedios ni se eliminaron filas completas; los faltantes se conservaron salvo en dos casos de limpieza puntual. En contaminantes se recodificaron como valores faltantes los picos imposibles, tales como PM₁₀ por arriba de 1000 $\mu\text{g}/\text{m}^3$ y O₃ por arriba de 400 ppb, con el fin de evitar que outliers instrumentales distorsionaran los resultados. Por otro lado, se eliminaron del dataset aquellas columnas que presentaban más de 70% de datos faltantes, manteniendo siempre la columna de tiempo. Una vez depurados estos extremos se procedió a la imputación de vacíos cortos dentro de cada serie temporal empleando ventanas de tiempo. Para huecos de hasta seis horas se utilizó interpolación lineal, aprovechando que en ese rango la evolución de los contaminantes suele ser continua y relativamente suave. Posteriormente, en los tramos de hasta veinticuatro horas de duración que permanecían sin datos se aplicó un filtro de Kalman con modelo estructural (StructTS), que permite reconstruir valores considerando la dinámica latente de la serie. Los intervalos mayores a un día no fueron imputados, de manera que se conservaron como ausencias. Este esquema escalonado aseguró que los datos finales tuvieran continuidad suficiente para el análisis sin introducir valores artificiosos en periodos prolongados de vacío.

