



# **Instituto Politécnico Nacional**

Escuela Superior de Cómputo  
Sección de Estudios de Posgrado e Investigación

## **Análisis de la calidad del aire en la zona norte de la Ciudad de México mediante el empleo de datos de estaciones del SINAICA y sistemas de monitorización móviles y fijos**

TESIS DE MAESTRÍA

Maestría en Ciencias en Sistemas Computacionales Móviles

Presenta

**Ing. Otto Antonio Calzadilla Nápoles**

Directores:

**Dr. Roberto Eswart Zagal Flores**

**Dr. Tonahtiu Arturo Ramírez Romero**



Hoja reservada para el SIP-13

Hoja reservada para el SIP-14

## **PENSAMIENTO**

*A clean and healthy environment is the single most important precondition for  
ensuring good health.*

*Tedros Adhanom Ghebreyesus*

## **DEDICATORIA**

A mis familiares y amigos.

## **AGRADECIMIENTOS**

Quiero agradecer a mis padres y familiares que han posibilitado y garantizado los factores necesarios para mi formación a través de los diferentes niveles de enseñanza.

Deseo extender mi gratitud a todos los profesores de la Escuela Superior de Cómputo y del Centro de Investigación de la Computación del Instituto Politécnico Nacional que han contribuido a mi desarrollo profesional, en especial a mis asesores: Dr. C. Roberto Eswart Zagal Flores y Dr. C. Tonáhtiu Arturo Ramírez Romero, por sus orientaciones precisas y apoyo brindado durante la realización del presente trabajo.

También quiero agradecer a CONAHCYT por otorgarme una beca que constituyó un apoyo fundamental para realizar mis estudios de postgrado.

## RESUMEN

La contaminación del aire constituye el principal riesgo ambiental para la salud. Históricamente en la zona norte de la Ciudad de México (CDMX) se han registrado con frecuencia niveles elevados de concentración de contaminantes atmosféricos que provocan efectos adversos en la salud humana y en los ecosistemas. Las estaciones de monitorización del Sistema Nacional de Información de la Calidad del Aire (SINAICA) que se encuentran en esta región, obtienen valores de concentración de contaminantes atmosféricos de forma espaciada. Para mejorar la evaluación de la calidad del aire en la zona, es fundamental emplear una mayor cantidad de fuentes de datos que describan la distribución de estos contaminantes. El objetivo principal consistió en analizar la calidad del aire en la zona norte de la Ciudad de México mediante el empleo de datos obtenidos de estaciones del SINAICA y sistemas de monitorización móviles y fijos. Como resultado se evaluó el cumplimiento de las Normas Oficiales Mexicanas (NOM) de calidad del aire en la zona de estudio durante los años 2022-2023, donde se emplearon en el análisis los datos de concentración de contaminantes atmosféricos obtenidos por las estaciones Camarones (CAM) y Gustavo A. Madero (GAM). Se obtuvieron tendencias actuales de la concentración de los contaminantes atmosféricos con respecto al tiempo mediante el empleo de software *Microsoft SQL Server*. Se desarrolló un sistema fijo de monitorización remota de la calidad del aire, que presenta un *Raspberry Pi 4 Model B* como dispositivo de adquisición de datos y posibilita la adquisición de los valores de concentración de  $PM_{2.5}$  medidos por un sensor *AM1008W*, y el cálculo del Índice de Aire y Salud para este contaminante. Se obtuvieron modelos de predicción del índice y valor de la concentración de  $O_3$  en la zona de estudio, mediante su entrenamiento con un conjunto de datos de promedios horarios de concentración de  $O_3$ , temperatura ambiente, humedad relativa, dirección y velocidad del viento, obtenidos por la estación GAM en el período 2015 – 2023. El rango de exactitud obtenida por los modelos en la predicción del índice fue de 86.94% - 90.83%, la mayor exactitud correspondió al algoritmo *C4.5*.

## Tabla de contenidos

Capítulo 1. Problemática.....	1
1.1    Introducción .....	1
1.2    Problema científico .....	2
1.3    Objetivo general.....	2
1.4    Objetivos específicos .....	3
1.5    Justificación .....	3
1.6    Estado del arte.....	4
Capítulo 2. Marco teórico .....	8
2.1    La contaminación atmosférica .....	8
2.1.1    Principales contaminantes atmosféricos .....	9
Partículas suspendidas .....	9
Ozono (O <sub>3</sub> ).....	9
Monóxido de carbono (CO).....	10
Dióxido de azufre (SO <sub>2</sub> ) .....	10
Dióxido de nitrógeno (NO <sub>2</sub> ) .....	10
Plomo (Pb) .....	10
2.1.2    Efectos nocivos de la contaminación atmosférica .....	11
Efectos de la contaminación atmosférica sobre la salud.....	11
Efectos de la contaminación del aire en los materiales.....	12
Efectos de la contaminación del aire sobre los ecosistemas y cultivos .....	13



2.2	Normas oficiales mexicanas de calidad del aire .....	13
2.2.1	Índice Aire y Salud .....	15
2.3	Sistemas de monitorización de la calidad del aire .....	16
2.3.1	Clasificación de los SMCA según su principio de operación.....	17
	SMCA Manuales.....	17
	SMCA Automáticos.....	18
	SMCA Mixtos.....	18
2.3.2	Clasificación de los SMCA por su movilidad .....	18
	SMCA Fijos .....	18
	SMCA Móviles .....	18
2.3.3	Sistema de adquisición y transmisión de datos.....	19
2.4	Manejo de datos de la calidad del aire .....	19
2.5	Algoritmos de clasificación .....	21
2.5.1	<i>C4.5</i> .....	21
2.5.2	<i>Naive Bayes</i> .....	21
2.5.3	<i>Bayesian Belief Networks</i> .....	22
2.5.4	<i>Backpropagation</i> .....	22
2.5.5	<i>K nearest-neighbor (Knn)</i> .....	22
2.5.6	<i>Support Vector Machine (SVM)</i> .....	22
2.6	Medidas de error en la comparación de valores .....	23
2.6.1	Error Medio Absoluto (MAE) .....	23
2.6.2	Porcentaje Medio del Error Absoluto (MAPE) .....	23
2.6.3	Raíz del Error Cuadrático Medio (RMSE) .....	23
2.6.4	Coefficiente de correlación .....	23
2.6.5	Coefficiente de determinación .....	24
Capítulo 3. Metodología, preprocesamiento de datos y diseño de la solución de monitoreo		

.....	25
3.1 Estaciones de monitoreo automático atmosférico CAM y GAM.....	25
3.1.1 Preprocesamiento de los datos crudos de contaminantes atmosféricos..	27
Identificación de valores faltantes .....	27
Identificación de valores atípicos .....	27
Análisis de la existencia de valores con ruido .....	29
3.2 Sistema fijo de monitorización remota .....	29
3.2.1 Método de validación.....	31
3.2.2 Preprocesamiento de las bases de datos obtenidas en la Isla de Datos Urbanos	32
Datos faltantes.....	33
Identificación de valores atípicos .....	33
Identificación de valores con ruido.....	34
Integración de los promedios horarios de concentración de PM <sub>2.5</sub> .....	36
Análisis de la relación entre los promedios horarios de concentración de PM <sub>2.5</sub>	36
Calibración del sistema desarrollado mediante el método de regresión lineal ....	37
Evaluación de la efectividad de la calibración.....	39
3.2.3 Visualización remota .....	40
3.3 Sistema monitorización móvil <i>Flatburn</i> .....	41
3.3.1 Implementación del sistema <i>Flatburn</i> .....	41
Preprocesamiento de los datos del Flatburn .....	41
Identificación de datos faltantes .....	42
Identificación de datos atípicos.....	43
Identificación de valores con ruido.....	43
3.4 Metodología de predicción del índice y concentración de O <sub>3</sub> en la zona norte de la CDMX.....	44
3.4.1 Parámetros del conjunto de datos .....	44

3.4.2	Implementación de los modelos de predicción.....	45
3.5	Consideraciones finales del capítulo.....	46
Capítulo 4. Resultados .....		48
4.1	Análisis estadístico valores históricos de las estaciones.....	48
4.1.1	Ozono.....	48
4.1.2	PM <sub>10</sub> y PM <sub>2.5</sub> .....	48
4.1.3	NO <sub>2</sub> .....	50
4.2	Evaluación del cumplimiento de las NOM.....	51
4.2.1	Ozono.....	51
	Datos horarios .....	51
	Promedios horarios de 8h .....	52
4.2.2	PM <sub>10</sub> .....	53
	Promedio de 24 horas .....	53
	Promedio anual .....	53
4.2.3	PM <sub>2.5</sub> .....	54
	Promedio de 24 horas .....	54
	Promedio anual .....	54
4.2.3	NO <sub>2</sub> .....	55
	Datos horarios y promedios anuales .....	55
4.3	Tendencias de contaminantes atmosféricos en el tiempo .....	56
	Ozono.....	56
	PM <sub>10</sub> y PM <sub>2.5</sub> .....	57
	NO <sub>2</sub> .....	59
4.4	Análisis datos obtenidos por el sistema <i>Flatburn</i> .....	60
4.5	Análisis de variabilidad espacial.....	61
4.5.1	Comparación <i>Flatburn</i> – RAMA.....	62

4.5.2	Comparación <i>Davis AirLink</i> – RAMA .....	63
4.6	Predicción del índice y concentración de O <sub>3</sub> en la zona norte de la CDMX ..	63
4.6.1	Predicción del índice de concentración de O <sub>3</sub> .....	64
4.6.2	Predicción del valor de la concentración de O <sub>3</sub> .....	65
	Conclusiones.....	66
	Recomendaciones .....	71
	Referencias bibliográficas.....	72
	Anexos .....	77

## Capítulo 1. Problemática

### 1.1 Introducción

La contaminación del aire constituye el principal riesgo ambiental para la salud. Varios estudios epidemiológicos y experimentales realizados en seres humanos han demostrado que la exposición a elevadas concentraciones de contaminantes atmosféricos se vincula con la existencia de una amplia variedad de efectos adversos, que inciden directamente en la calidad de vida de la población en general (WHO, 2016). Este impacto es más pronunciado en grupos vulnerables, como los niños, mujeres en gestación y adultos mayores, especialmente si padecen de enfermedades preexistentes. Se estima que la exposición a la contaminación del aire causa aproximadamente 7 millones de muertes prematuras cada año (WHO, 2021).

Prácticamente la totalidad de la población mundial (99%) se encuentra expuesta a niveles de contaminación atmosférica que exceden los límites de calidad recomendados por la Organización Mundial de la Salud (OMS) y pone en riesgo su salud (WHO, 2022). Un número sin precedentes de más de 6000 ciudades de 117 países monitorean la calidad del aire; sin embargo, los habitantes de estas localidades son víctimas de niveles insalubres de concentraciones de material particulado (PM) y dióxido de nitrógeno (NO<sub>2</sub>).

Según estudios realizados por el Instituto Nacional de Salud Pública de México, con respecto a la evidencia epidemiológica nacional, se destaca que la contaminación del aire provoca diversos efectos perjudiciales sobre la salud, y afecta significativamente la calidad de vida de las personas que residen en las áreas urbanas de México (SEMARNAT, 2019). En correspondencia con las elevadas concentraciones de contaminantes atmosféricos registradas en diferentes ciudades y áreas metropolitanas, donde se ejecuta el monitoreo de la calidad del aire, se recomienda a los habitantes de estas zonas que realicen cambios sustanciales en sus hábitos para reducir su exposición y minimizar los riesgos para su salud.

De acuerdo con el Informe Nacional de Calidad del Aire del 2020 en México (INECC, 2022), elaborado por el Instituto Nacional de Ecología y Cambio Climático (INECC), el 72 % de los días del año se incumplió al menos una Norma Oficial Mexicana de Calidad del Aire en la Zona Metropolitana del Valle de México (ZMVM). Durante ese año las estaciones de monitoreo pertenecientes al Sistema Nacional de Información de la Calidad del Aire (SINAICA), ubicadas en la zona norte de la Ciudad de México registraron los valores más altos de concentración de  $PM_{10}$  en comparación con las demás estaciones existentes en la capital (SEDEMA, 2023).

El monitoreo de la calidad del aire constituye un aspecto fundamental para evaluar la calidad del aire y sus tendencias. El establecimiento de Sistemas de Monitoreo de la Calidad del Aire (SMCA) ha posibilitado a las autoridades ambientales obtener valores de concentración de contaminantes atmosféricos y formular, con base en los datos obtenidos, estrategias de control y prevención apropiadas para contribuir a la reducción de la contaminación ambiental (Karagulian et al., 2019).

En consideración con lo expuesto anteriormente se plantea el siguiente diseño de investigación:

## **1.2 Problema científico**

Históricamente, en la zona norte de la Ciudad de México se han registrado con elevada frecuencia niveles de concentración de contaminantes atmosféricos que afectan la salud humana y los ecosistemas. Sin embargo, existe una carencia de información detallada en diversas escalas temporales y espaciales, que dificulta la evaluación del grado de exposición y el impacto de la contaminación del aire a nivel urbano. Por lo tanto, se precisa un análisis detallado de la calidad del aire en la región, y desarrollar sistemas de monitoreo que proporcionen datos precisos de contaminación atmosférica y pongan a disposición de la población de forma temprana y precisa esta información, para reducir así el riesgo de afectaciones asociados con la exposición a niveles elevados de concentración de contaminantes atmosféricos.

## **1.3 Objetivo general**

Analizar la calidad del aire en la zona norte de la Ciudad de México mediante el empleo de datos obtenidos de estaciones del SINAICA y sistemas de monitorización móviles y fijos.

## 1.4 Objetivos específicos

- Identificar los contaminantes atmosféricos que han presentado mayor incidencia en la zona norte de la CDMX en los últimos años.
- Analizar estadísticamente los valores de concentración de  $O_3$ ,  $PM_{10}$ ,  $PM_{2.5}$  y  $NO_2$  obtenidos en las estaciones de la Red Automática de Monitoreo Atmosférico (RAMA) ubicadas en la zona norte de la Ciudad de México en los años 2022-2023.
- Evaluar el cumplimiento de las NOM de calidad del aire en la zona de estudio durante los años 2022-2023.
- Obtener tendencias actualizadas del comportamiento de la concentración de los contaminantes atmosféricos con respecto al tiempo en la zona de estudio.
- Desarrollar un sistema fijo de monitorización remota de la calidad del aire.
- Obtener datos de concentración de  $PM_{2.5}$  con el sistema fijo de monitorización remota de la calidad del aire y el dispositivo móvil *Flatburn*, en la zona de estudio.
- Evaluar la variabilidad espacial de la concentración de  $PM_{2.5}$  en la zona de estudio mediante la comparación de los datos obtenidos por los sistemas fijo y móvil, con las estaciones de la RAMA.
- Obtener modelos de predicción del índice y valor de la concentración de  $O_3$  en la zona de estudio.
- Evaluar los modelos de predicción obtenidos mediante el empleo de medidas de error.

## 1.5 Justificación

Históricamente en la zona norte de la Ciudad de México se han registrado con frecuencia niveles elevados de concentración de contaminantes atmosféricos que provocan efectos adversos en la salud humana y en los ecosistemas. Las estaciones de monitorización de la calidad del aire pertenecientes al SINAICA que se encuentran en esta región, obtienen valores de concentración de contaminantes atmosféricos de forma espaciada. Para mejorar la evaluación de la calidad del aire en la zona, es fundamental emplear una mayor cantidad de fuentes de datos que describan la distribución de la concentración de contaminantes atmosféricos. Debido a esto se requiere el desarrollo de sistemas de monitorización de la calidad del aire que obtengan valores de concentración de contaminantes atmosféricos de forma más específica y pongan a disposición de la población de forma temprana y precisa esta información.

## 1.6 Estado del arte

En la actualidad se evidencia un auge en las investigaciones relacionadas con el monitoreo de la calidad del aire y la obtención de modelos de predicción, debido principalmente a la preocupación de las autoridades por el incremento de contaminantes en la atmósfera, y también por la disponibilidad de dispositivos de bajo costo que cumplen con los requisitos establecidos para estas aplicaciones. A continuación, se destacan trabajos relacionados.

Haq (2022) presenta *SMOTEDNN (Synthetic Minority Oversampling Technique with Deep Neural Network)*, un modelo para clasificar la calidad del aire basado en aprendizaje automático. Este algoritmo se implementa en la base de datos *National Air Quality Monitoring Program*, que pertenece al *Central Pollution Control Board (CPCB)* de la India. Se seleccionan nueve parámetros del conjunto de datos: seis de concentración de diferentes contaminantes atmosféricos, el año, mes y el índice de calidad del aire (*AQI*), que es una medida empleada por las autoridades gubernamentales para categorizar la contaminación con respecto de su gravedad. En la evaluación del rendimiento del modelo, se comparan los resultados obtenidos con cuatro algoritmos de clasificación: *XGBoost*, *Random Forest*, *SVM*, *KNN*; para clasificar el índice de calidad del aire (*AQI*) en seis clases. Se realiza un preprocesamiento eficiente de los datos y una optimización rigurosa de los hiperparámetros de cada modelo. Como resultado todos los algoritmos empleados exhiben un rendimiento adecuado con respecto a la exactitud, precisión, sensibilidad y especificidad. Significativamente, con el modelo *SMOTEDNN* se obtiene la mayor exactitud, con un valor de 99,90%.

Yarragunta & Nabi (2021) aplican seis algoritmos de clasificación basados en aprendizaje automático supervisado, a un conjunto de datos de concentración de contaminantes atmosféricos, con el objetivo de pronosticar el índice de calidad del aire en cuatro categorías. Se comparan los resultados obtenidos por los algoritmos: *Logistic Regression*, *Decision Tree*, *Support Vector Machine*, *Random Forest Tree*, *Nave Bayes Theorem* y *K-Nearest Neighbor*. La base de datos pertenece al *Central Pollution Control Board (CPCB)* de la India, esta posee nueve atributos: país, estado, ciudad, ubicación, fecha, valor mínimo, máximo, promedio, y contaminante; se incluyen los contaminantes atmosféricos:  $\text{SO}_2$ ,  $\text{NO}_2$ ,  $\text{PM}_{10}$ ,  $\text{PM}_{2.5}$ ,  $\text{CO}$  y  $\text{O}_3$ . Como resultado el modelo que se obtiene con el algoritmo de *Decision Tree* presenta la mayor exactitud en la predicción del índice de calidad del aire, con un 99.88%.



Al-Janabi et al. (2020) presentan un modelo inteligente de predicción de la calidad del aire (*SAQPM*), basado en técnicas de aprendizaje profundo mediante el empleo de una red neuronal recurrente (RNN), que predice, con dos días de antelación, la concentración de seis contaminantes atmosféricos: partículas  $PM_{2.5}$ ,  $PM_{10}$ , dióxido de nitrógeno ( $NO_2$ ), monóxido de carbono (CO), ozono ( $O_3$ ) y dióxido de azufre ( $SO_2$ ). El modelo consta de cuatro etapas, primeramente, se recupera la información de la base de datos del KDD Cup 2018; luego se preprocesan los datos, mediante el procesamiento de los valores faltantes y la normalización mediante el método *MinMaxScalar* a un rango de 0-1; en la tercera etapa se construye el predictor basado en el método *long short-term memory (LSTM)*, mediante la identificación de la estructura óptima y valores de parámetros (peso, sesgo, número de capas ocultas, número de nodos en cada capa oculta y función de activación) para la red mediante el empleo de un algoritmo de optimización de enjambre de partículas (PSO); finalmente resultados de la evaluación para cada estación se obtienen mediante la lectura de la concentración de cada contaminante cada hora durante un máximo de 30 días y luego se calcula el promedio del error porcentual absoluto medio simétrico (*SMAPE*) durante 25 días. Como resultado el modelo predice la concentración de contaminantes atmosféricos con 48 horas de antelación, con un error porcentual absoluto medio simétrico (*SMAPE*) que oscila entre 4.38 y 13.61, para cada contaminante.

Dhingra et al. (2019) presentan IoT-Mobair, un sistema de monitoreo de la contaminación atmosférica móvil, que emplea tecnología IoT para obtener y analizar valores de calidad del aire. El sistema se compone de sensores de calidad del aire que se integran con dispositivos móviles, una plataforma de procesamiento en la nube y una aplicación móvil para la visualización de datos. Los sensores posibilitan la medición de la concentración de monóxido de carbono (CO) y metano ( $CH_4$ ), y se envían los datos recopilados a través de conexión móvil a la plataforma en la nube. Los resultados de la implementación del sistema en entornos reales demuestran su eficacia para recopilar y analizar valores de concentración de contaminantes atmosféricos.

Gunawan et al. (2018) presentan un sistema portátil de medición de calidad del aire en exteriores. El sistema posee un microcontrolador Arduino para la adquisición y procesamiento de los datos, una interfaz de usuario para visualizar los resultados y sensores de calidad del aire para medir la concentración de contaminantes atmosféricos: monóxido de carbono (CO), ozono ( $O_3$ ),  $PM_{2.5}$  y  $PM_{10}$ . Los resultados experimentales validan la efectividad del sistema para medir parámetros de la calidad del aire y

proporcionar la visualización de los datos recopilados.

Anjomshoaa et al. (2018) presentan City Scanner, una plataforma de monitorización móvil que posee sensores de medición de humedad, temperatura ambiental, y concentración de material particulado: PM<sub>1</sub>, PM<sub>2.5</sub> y PM<sub>10</sub>. Esta se instala en vehículos con el objetivo de recopilar datos en tiempo real en diferentes ubicaciones. El sistema posee una infraestructura de comunicación para transmitir los datos recopilados, y un enfoque de programación que optimiza la ruta y horarios de los vehículos para maximizar la eficiencia en la recolección de datos. Los resultados de la implementación del sistema en la ciudad de Cambridge, Massachusetts, demuestran su eficacia en la obtención de valores asociados a la calidad del aire y variables meteorológicas.

Kumar and Jasuja (2017) presentan un sistema monitorización de calidad del aire basado en IoT. El sistema posee un Raspberry Pi para la adquisición y procesamiento de datos, una interfaz de usuario para visualizar los resultados del monitoreo; y sensores de calidad del aire para medir la concentración de partículas, la temperatura, la humedad y los gases contaminantes. Los resultados de la implementación en entornos reales demuestran la efectividad del sistema para la recolección y procesamiento de datos de calidad del aire.

Gao et al. (2016) presentan Mosaic, un sistema económico de monitoreo de calidad del aire urbano, que emplea dispositivos móviles equipados con sensores de calidad del aire para recolectar valores de concentración de partículas y gases contaminantes. Se destaca la ventaja de su bajo costo y la facilidad de implementación debido al empleo de dispositivos móviles ampliamente disponibles. Los resultados experimentales muestran la eficacia del sistema en la captura de datos de calidad del aire en tiempo real.

Phala et al. (2016) presentan un sistema de monitoreo de calidad del aire basado en los estándares ISO/IEC/IEEE 21451, que definen las interfaces y protocolos para sensores y transductores utilizados en aplicaciones de monitoreo ambiental. El sistema se compone de un nodo de procesamiento de datos, una interfaz de usuario y sensores de calidad del aire que cumplen con los estándares 21451 para medir la concentración de partículas, la temperatura y la humedad. Los resultados de la implementación en un entorno de prueba demuestran la efectividad y precisión del sistema en la recolección y presentación de datos de calidad del aire.

Zheng et al. (2016) presenta un sistema de monitoreo de calidad del aire basado en LPWA (Low-Power Wide-Area) y aborda los desafíos asociados con la supervisión de la calidad

del aire en tiempo real mediante el empleo de tecnologías de comunicación de baja potencia y amplio alcance. El sistema consta de tres componentes principales: sensores de calidad del aire, una red de comunicación LPWA y una plataforma de visualización y análisis de datos; los sensores se despliegan en diferentes ubicaciones para medir parámetros como la concentración de partículas, la temperatura, la humedad y los gases contaminantes; y los datos recopilados por los sensores se envían a través de la red LPWA a una plataforma centralizada para su procesamiento y análisis. El sistema se implementa y se realizan pruebas piloto en un entorno urbano. Los resultados experimentales muestran que el sistema es capaz de recopilar y transmitir datos de calidad del aire de manera efectiva, y la plataforma de visualización proporciona una interfaz intuitiva para el análisis de datos.

Van den Bossche et al. (2015) presentan una metodología para mapear la variación espacial de la calidad del aire urbano mediante el empleo de monitoreo móvil y un conjunto de extenso de datos. Se destacan las limitaciones de los métodos tradicionales de monitoreo fijo de la calidad del aire, que no permiten la captura de la variación espacial detallada en entornos urbanos complejos. En la recolección de los datos se emplearon sensores portátiles y vehículos equipados con instrumentación de medición. Se valida la metodología mediante la ejecución de comparaciones con mediciones de calidad del aire realizadas por estaciones fijas de monitoreo. Los resultados mostraron que el enfoque de monitoreo móvil posibilita registrar la variación espacial detallada en la calidad del aire urbano y proporcionar una representación más precisa de la exposición de las personas a la contaminación del aire.

## Capítulo 2. Marco teórico

En el presente capítulo se abordan los principales contaminantes atmosféricos y sus efectos perjudiciales sobre la salud humana, los ecosistemas y los materiales. Además, se analizan las clasificaciones que poseen los sistemas de monitorización de la calidad del aire (SMCA) con respecto a su principio de operación y su movilidad. Se presentan las Normas Oficiales Mexicanas (NOM) relacionadas con la gestión de la calidad del aire y los aspectos a considerar en el procesamiento de datos obtenidos por estaciones de monitoreo. También, se describen los algoritmos de clasificación que se emplean en la predicción de índice y valor de la concentración de  $O_3$  y se presentan las principales medidas de error.

### 2.1 La contaminación atmosférica

El aire es el fluido que constituye la atmósfera de la Tierra, compuesto principalmente por una mezcla gaseosa que contiene 21% de oxígeno y 78% de nitrógeno, con el resto formado por vapor de agua, gases nobles y dióxido de carbono. El equilibrio de estas concentraciones permite que los seres humanos puedan respirar sin afectar su salud. Sin embargo, este equilibrio puede ser perturbado por la presencia de otros compuestos.

Los contaminantes del aire son sustancias que deterioran la composición y calidad del aire ambiente (INE, 2008). Estos contaminantes pueden ser emitidos de manera natural, a través de procesos como la erosión del suelo, la descomposición de materia orgánica, los incendios forestales y la actividad volcánica. Por otro lado, los contaminantes originados por las actividades humanas se denominan contaminantes antropogénicos, siendo la quema de combustibles fósiles su principal fuente. Este grupo incluye los contaminantes emitidos por fuentes estacionarias, como las industrias; fuentes de área, que comprenden conjuntos de pequeños establecimientos industriales o comerciales; y fuentes móviles, como los vehículos automotores.

Los contaminantes emitidos directamente, ya sea de origen natural o antropogénico, se conocen como contaminantes primarios, como el dióxido de nitrógeno y el monóxido de carbono. Además, existen los contaminantes secundarios, que se forman como resultado de reacciones químicas entre contaminantes primarios y otros componentes del aire. Un ejemplo de contaminante secundario es el ozono, que se genera a partir de la reacción entre el dióxido de nitrógeno y compuestos orgánicos volátiles.

La contaminación atmosférica se refiere a la presencia de contaminantes atmosféricos en la atmósfera terrestre, en concentraciones superiores a las que son naturalmente presentes, y que pueden provocar efectos adversos en la salud humana y los ecosistemas.

### **2.1.1 Principales contaminantes atmosféricos**

Algunos contaminantes del aire han sido regulados debido a sus efectos perjudiciales para la salud de humana, mediante la implementación de límites máximos de concentración en el aire ambiente. Estos contaminantes se conocen como contaminantes criterio, incluyen dióxido de nitrógeno ( $\text{NO}_2$ ), ozono ( $\text{O}_3$ ), dióxido de azufre ( $\text{SO}_2$ ), monóxido de carbono ( $\text{CO}$ ), plomo ( $\text{Pb}$ ) y partículas suspendidas con diámetros aerodinámicos inferiores a 10 y 2.5 micrómetros ( $\text{PM}_{10}$  y  $\text{PM}_{2.5}$ ) (INE, 2008). A continuación, se describen los principales contaminantes atmosféricos.

#### ***Partículas suspendidas***

El material particulado constituye una mezcla compleja de sustancias que se encuentran en estado líquido o sólido y permanecen suspendidas en la atmósfera durante diferentes períodos de tiempo. Según su origen, las partículas pueden clasificarse como primarias o secundarias, y pueden ser tanto de origen natural como antropogénico.

Según su diámetro aerodinámico, las partículas suspendidas se clasifican en tres categorías: menores o iguales a 10 micrómetros ( $\text{PM}_{10}$ ), menores o iguales a 2.5 micrómetros ( $\text{PM}_{2.5}$ ) y menores o iguales a 0.1 micrómetros ( $\text{PM}_{0.1}$ ). El tamaño de estas partículas es un factor crucial para comprender su comportamiento en la atmósfera, la concentración a la que la población puede estar expuesta, y su capacidad para penetrar y permanecer en distintas zonas de las vías respiratorias (Secretaría de Salud, 2014).

#### ***Ozono ( $\text{O}_3$ )***

El ozono constituye un contaminante secundario que se genera mediante una reacción fotoquímica entre óxidos de nitrógeno ( $\text{NO}_x$ ), compuestos orgánicos volátiles (COVs) e

hidrocarburos (HCs), en presencia de luz solar. La formación del ozono también está influenciada por las condiciones geográficas, climatológicas y meteorológicas del entorno. La vida útil del ozono en la atmósfera está determinada por la cantidad y disponibilidad de sus precursores, así como de las condiciones ambientales mencionadas. Las mayores concentraciones de ozono se registran durante las horas del día en las que se alcanzan las temperaturas más altas (Secretaría de Salud, 2014).

### ***Monóxido de carbono (CO)***

El monóxido de carbono (CO) constituye un gas incoloro e inodoro que se genera a partir de la combustión incompleta de compuestos de carbono (Secretaría de Salud, 2021). La operación de vehículos automotores y la actividad industrial son las principales fuentes antropogénicas de CO en la atmósfera. En menor medida, algunos procesos naturales también emiten este contaminante, como los incendios forestales y ciertos procesos biogénicos en los océanos.

### ***Dióxido de azufre (SO<sub>2</sub>)***

El dióxido de azufre (SO<sub>2</sub>) constituye un gas incoloro con un fuerte olor irritante, altamente soluble en agua. Este puede oxidarse para convertirse en trióxido de azufre e iones sulfato, que se combinan para formar sales inorgánicas y ácidos, componentes esenciales de las partículas secundarias. En el aire, el SO<sub>2</sub> es higroscópico, lo que significa que forma aerosoles de ácido sulfúrico y ácido sulfuroso, contribuyendo así a la formación de lluvia ácida. Esta lluvia ácida daña los bosques y acidifica ríos, canales, lagos y suelos. La formación y la duración de los aerosoles están influenciados por las condiciones meteorológicas y la concentración de contaminantes en el aire. Se estima que estos aerosoles permanecen en la atmósfera entre 3 y 5 días (Secretaría de Salud, 2019).

### ***Dióxido de nitrógeno (NO<sub>2</sub>)***

El dióxido de nitrógeno (NO<sub>2</sub>) constituye un contaminante primario originado principalmente de los procesos de combustión, constituyendo la principal fuente de su emisión a la atmósfera. Además de ser un contaminante por sí mismo, el NO<sub>2</sub> contribuye a la generación de otros contaminantes atmosféricos, como las partículas en suspensión (PM<sub>10</sub> y PM<sub>2.5</sub>) y el ozono (O<sub>3</sub>) (Weinmayr et al., 2010).

### ***Plomo (Pb)***

El plomo constituye uno de los metales pesados más comunes en la superficie terrestre,

lo que genera una variabilidad considerable en el riesgo de exposición para la población general. Una de las formas más frecuentes de empleo es como tetraetilo de plomo, que se utiliza como antidetonante de gasolina y constituye su principal fuente de vertimiento a la atmósfera (Khanam et al. 2020).

### 2.1.2 Efectos nocivos de la contaminación atmosférica

La contaminación atmosférica provoca efectos negativos sobre la salud humana, los ecosistemas y los materiales. A continuación, se describen las afectaciones que puede provocar la contaminación atmosférica en estos campos.

#### *Efectos de la contaminación atmosférica sobre la salud*

La contaminación del aire afecta la salud humana de diversas formas. La exposición a corto plazo a los compuestos contaminantes puede causar inflamación, irritación y reacciones alérgicas. A largo plazo, incrementa el riesgo de sufrir asma, cataratas, enfermedades cancerígenas, inflamación de las vías respiratorias y debilitamiento del sistema inmunológico (Abdel-Shafy, 2016). En la Tabla 2.1 se exponen los principales contaminantes y su consecuencia en la salud.

Tabla 2.1 Efectos sobre la salud de los principales contaminantes (Yi et al., 2015).

Contaminante	Efecto sobre la salud
Monóxido de carbono (CO)	Reduce la capacidad de la sangre para transportar oxígeno. En dosis elevadas, puede resultar letal.
Dióxido de nitrógeno (NO <sub>2</sub> )	Afecta significativamente en las vías respiratorias y puede agravar problemas cardíacos preexistentes.
Ozono (O <sub>3</sub> )	Afecta gravemente las vías respiratorias, contribuye a la irritación y daño de los tejidos pulmonares.
Dióxido de azufre (SO <sub>2</sub> )	Aumenta los síntomas del asma y causa broncoconstricción.
Partículas en suspensión	Reducen la esperanza de vida en personas con enfermedades cardíacas. Además, agravan diversas enfermedades y alergias del sistema respiratorio.

Se ha demostrado la existencia de un vínculo directo entre el aumento de la incidencia de enfermedades respiratorias con la exposición inmediata a la contaminación derivada de los gases generados por los motores de combustión interna. Esta tendencia se manifiesta esencialmente en esferas urbanas con elevado tráfico de vehículos. Algunos autores indican que los contaminantes que tienen un mayor impacto negativo son las partículas en suspensión y el ozono (Guo et al., 2016).

La exposición a contaminantes del aire puede ocasionar tanto efectos agudos como crónicos en la salud. Los efectos agudos suelen ser inmediatos y reversibles tras cesar la exposición, incluyen irritación ocular, dolores de cabeza y náuseas (INE, 2008). Por otro lado, los efectos crónicos pueden tardar en manifestarse, son de larga duración y generalmente irreversibles. Entre los efectos crónicos se encuentran la reducción de la capacidad pulmonar y el cáncer de pulmón, debido a la exposición prolongada a contaminantes tóxicos, como el asbesto y el berilio.

Para monitorear los efectos agudos es esencial obtener con mediciones horarias de los contaminantes, lo que permite detectar picos de concentración durante el ciclo diario. Debido a esto, se requiere de métodos automáticos para obtener mediciones precisas y frecuentes. En cuanto a los efectos crónicos, las concentraciones promedio anuales de contaminantes suelen ser suficientes, aunque las variaciones estacionales sugieren que mediciones mensuales ofrecerían una base más descriptiva y útil. Para evaluar el impacto de fuentes específicas de contaminación, es crucial ubicar las estaciones de monitoreo de manera que las concentraciones ambientales medidas sean representativas de las emisiones de dichas fuentes. La ubicación de las estaciones puede determinarse mediante el empleo de un modelo de dispersión adecuado al tipo de contaminante en cuestión.

### ***Efectos de la contaminación del aire en los materiales***

Los materiales que se emplean a frecuentemente en la actividad humana son afectados por la contaminación atmosférica. Entre los compuestos que contribuyen al deterioro de estos materiales se encuentran los derivados del azufre, como el dióxido de azufre, producido principalmente por la combustión derivada del tráfico vehicular.

El incremento en la concentración de dióxido de azufre y iones cloruro acelera la corrosión de materiales como el cobre y el acero. La combinación de sustancias contaminantes como el óxido de azufre, el nitrógeno y el carbono, con el agua presente en la atmosfera contribuye a la formación de las lluvias ácidas. Estas catalizan el deterioro de construcciones fabricadas con materiales metálicos, piedra caliza y otros materiales sensibles al ácido (Tejero-Rivas et al., 2015).

El deterioro de los materiales de construcción, especialmente causado por dióxido de azufre, puede evaluarse mediante muestreos pasivos. Un ejemplo de esto es el uso de bujías de peróxido de plomo, las cuales permiten determinar los índices de sulfatación tras ser expuestas a la atmósfera durante un mes (INE, 2008).



### ***Efectos de la contaminación del aire sobre los ecosistemas y cultivos***

La contaminación del aire afecta tanto a los animales como a los humanos debido a las similitudes biológicas entre ambos. La variación de la acidez del suelo, provocada por los contaminantes, impacta ecosistemas completos, ya que el daño a los árboles afecta a las cadenas tróficas enteras.

Los animales no solo sufren la pérdida de sus hábitats debido a la deforestación, sino que también son impactados por otros contaminantes que se agravan en ambientes ácidos. Los ecosistemas acuáticos también se ven severamente afectados; el aumento de la acidez del agua puede provocar la extinción de especies (Singh & Agrawal, 2007).

Los mantos freáticos son afectados por el incremento de la acidez, que facilita la disolución y asimilación de metales, incluyendo metales pesados que tienen un efecto acumulativo en los organismos y eventualmente se trasladan a los humanos (INE, 2008). El impacto adverso de la contaminación atmosférica en plantas y árboles, tanto en áreas urbanas como en zonas naturales protegidas, puede evaluarse mediante muestreos de 24 horas. Para estudios exhaustivos, es indispensable utilizar equipos de medición continua.

La contaminación atmosférica genera efectos nocivos sobre la salud humana, acelera el deterioro de construcciones y afecta negativamente el equilibrio de varios ecosistemas, por lo que es fundamental realizar el seguimiento continuo de la calidad del aire para disminuir los problemas que se pueden generar.

## **2.2 Normas oficiales mexicanas de calidad del aire**

Con el objetivo de proteger la salud humana, los gobiernos implementan normativas que controlan los niveles de estos contaminantes atmosféricos. Las normas de calidad del aire definen concentraciones admisibles basadas en los riesgos que representan para la salud. Estas normativas establecen las concentraciones permitidas para diferentes períodos de exposición, considerando que una concentración baja durante un tiempo prolongado puede ser tan peligrosa como una alta concentración en un corto período. En México, la Secretaría de Salud es la entidad encargada de evaluar la evidencia sobre los efectos de la contaminación del aire en la salud y de definir los límites permitidos de concentración de los contaminantes (SEMARNAT, 2019). En la Tabla 2.2 se muestran los indicadores que se emplean para evaluar el cumplimiento de las NOM de salud ambiental con respecto a cada uno de los contaminantes criterio.

Tabla 2.2 Indicadores que se emplean para evaluar el cumplimiento de las NOM de salud ambiental con respecto a cada uno de los contaminantes criterio (SEMARNAT, 2019).

Contaminante	NOM	Dato base empleado en la evaluación	Valor límite [µg/m³]
Dióxido de azufre (SO <sub>2</sub> )	NOM-022-SSA1-2019	Promedio horario	196.5
		Promedio 24 horas	104.8
Monóxido de carbono (CO)	NOM-021-SSA1-2021	Promedio horario	30 000
		Promedio móvil de 8 horas	10 000
Dióxido de nitrógeno (NO <sub>2</sub> )	NOM-023-SSA1-2021	Promedio horario	200
		Promedio anual	40
Ozono (O <sub>3</sub> )	NOM-020-SSA1-2021	Promedio horario	176
		Promedio móvil de 8 horas	118
Partículas menores a 10 micras (PM <sub>10</sub> )	NOM-025-SSA1-2021	Promedio 24 horas	60
		Promedio anual	28
Partículas menores a 2.5 micras (PM <sub>2.5</sub> )		Promedio 24 horas	33
		Promedio anual	10
Plomo (Pb)	NOM-026-SSA1-2021	Promedio anual	0.5

El monitoreo de los contaminantes criterio se realiza mediante el empleo de dispositivos que cumplen con las especificaciones requeridas por la Agencia de Protección Ambiental de los Estados Unidos (*EPA*) y la Norma Oficial Mexicana para cada contaminante. Esto asegura que los instrumentos seleccionados proporcionen datos fiables y precisos. Los principios de operación de estos instrumentos de medición se basan en alguna propiedad física o química específica del compuesto a analizar, siendo generalmente métodos específicos con interferencias mínimas. En la Tabla 2.3 se presentan los métodos de medición de la concentración de los contaminantes criterio.

Tabla 2.3 Métodos de medición de la concentración de los contaminantes criterio (SEDEMA, 2024).

Contaminante	Principio de operación del método de medición
Ozono ( $\text{O}_3$ )	Fotometría UV
Monóxido de carbono ( $\text{CO}$ )	Absorción en el infrarrojo
Dióxido de azufre ( $\text{SO}_2$ )	Fluorescencia UV
Dióxido de nitrógeno ( $\text{NO}_2$ )	Quimioluminiscencia
Partículas suspendidas ( $\text{PM}_{10}$ ), ( $\text{PM}_{2.5}$ )	Gravimetría
	Atenuación de radiación beta

### 2.2.1 Índice Aire y Salud

La Norma Oficial Mexicana NOM-172-SEMARNAT-2019 establece las directrices para la obtención y comunicación del Índice de Calidad del Aire y Riesgos a la Salud, con el objetivo de proporcionar información de forma precisa, temprana y continua sobre la calidad del aire, las posibles repercusiones en la salud que provoca y las medidas necesarias para mitigar los efectos adversos (SEMARNAT, 2019). Esta norma presenta validez en la totalidad del territorio nacional de México y es de obligatorio cumplimiento para los gobiernos estatales o municipales encargados de monitorizar la calidad del aire. En esta se define el Índice Aire y Salud que abarca las clasificaciones de calidad del aire y nivel de riesgo asociado en dependencia de los valores de concentración de contaminantes atmosféricos. En la Tabla 2.4 se presenta la clasificación del Índice Aire y Salud en relación con los valores de concentración de  $PM_{10}$ ,  $PM_{2.5}$  y  $NO_2$ .

Tabla 2.4. Clasificación del índice de calidad del aire y salud (SEMARNAT, 2019).

Calidad del aire	Nivel de riesgo asociado	O <sub>3</sub> (Promedio de una hora) [ $\mu\text{g}/\text{m}^3$ ]	PM <sub>10</sub> (Promedio móvil ponderado de 12 h)[ $\mu\text{g}/\text{m}^3$ ]	PM <sub>2.5</sub> (Promedio móvil ponderado de 12 h) [ $\mu\text{g}/\text{m}^3$ ]	NO <sub>2</sub> (promedio de una hora) [ $\mu\text{g}/\text{m}^3$ ]
Buena	Bajo	0 - 100	0 - 50	0 - 25	0 - 201
Aceptable	Moderado	> 100 - 186	> 50 - 75	> 25 - 45	> 201 - 395
Mala	Alto	> 186 - 265	> 75 - 155	> 45 - 79	> 395 - 432
Muy Mala	Muy Alto	> 265 - 343	> 155 - 235	> 79 - 147	> 432 - 470
Extr. Mala	Extr. Alto	> 343	> 235	> 147	> 470

El Índice Aire y Salud se debe publicar en la página electrónica y en cualquier otro medio de difusión que la autoridad a cargo de cada Sistema de Monitoreo de Calidad del Aire determine para este propósito, con el objetivo de informar a la población. Además, se debe difundir el índice que represente el mayor riesgo para la salud humana para cada una de las estaciones que integran el sistema de monitorización de la calidad del aire. En la difusión se asigna un color para caracterizar el Índice Aire y Salud en correspondencia con su clasificación. En el Anexo 1 se presentan las categorías del Índice Aire y Salud.

La divulgación del Índice Aire y Salud debe incluir recomendaciones dirigidas a grupos sensibles y a la población en general, con el propósito de mitigar los efectos perjudiciales

de la exposición a la contaminación atmosférica existente. En el Anexo 2 se presentan los mensajes asociados a las categorías de calidad del aire y riesgos a la salud.

### **2.3 Sistemas de monitorización de la calidad del aire**

Un sistema de medición de la calidad del aire (SMCA) constituye una estructura organizada que combina recursos humanos, técnicos y administrativos destinada a observar y registrar la calidad del aire en una región específica a lo largo del tiempo. Un SMCA debe funcionar de manera eficiente y confiable, alineándose con los objetivos de monitoreo y las necesidades de información establecidas por la normativa federal vigente en materia de calidad del aire (INE, 2008).

Las funciones principales de los SMCA incluyen la medición, registro y procesamiento de los datos sobre la calidad del aire. Los productos generados por estos sistemas abarcan bases de datos, informes, indicadores, y predicciones de contaminación atmosférica. Además, debido a la naturaleza de la información que producen, los SMCA pueden compartir datos con los sistemas de vigilancia epidemiológica operados por el sector salud a nivel local o nacional. En dependencia de los objetivos y alcance del sistema, un SMCA puede medir uno, varios o todos los contaminantes criterio. El objetivo del SMCA es evaluar la calidad del aire o identificar problemas específicos de contaminación atmosférica con la mayor eficiencia y cobertura poblacional posible (SEDEMA, 2022).

El monitoreo constituye una herramienta esencial para identificar y evaluar problemas de calidad del aire. En combinación con modelos de predicción e inventarios de emisiones, el monitoreo es una parte integral de la gestión de la calidad del aire, vinculado con el cumplimiento de objetivos económicos, de evaluación y de regulación. La obtención de información confiable sobre los niveles de contaminación mediante este método es imprescindible para desarrollar planes efectivos de gestión de la calidad del aire.

El objetivo principal del monitoreo de la calidad del aire es proporcionar información esencial a planificadores, científicos y legisladores, para que puedan tomar decisiones apropiadas en la gestión del medio ambiente. El monitoreo proporciona la base científica necesaria para la elaboración de estrategias y políticas, establecer objetivos, evaluar el cumplimiento de metas y la implementación de acciones (Castell et al., 2017).

Para realizar el monitoreo de la calidad del aire, se emplean diversos equipos, tales como analizadores, monitores y sensores. Estas estaciones de monitoreo se agrupan en redes,

que constituyen una parte esencial del SMCA.

El establecimiento de SMCA ha permitido a las autoridades ambientales de la mayoría de las grandes ciudades del mundo enfrentar eficazmente los problemas de contaminación del aire urbano, derivados de las actividades industriales, vehiculares y comerciales necesarias para satisfacer las necesidades de los residentes y proporcionar servicios.

Los SMCA se han convertido en una herramienta fundamental para obtener información fiable sobre la calidad del aire y la concentración de contaminantes específicos. Con base en los datos recopilados, se pueden desarrollar estrategias de control y adoptar medidas oportunas y apropiadas para una gestión ambiental efectiva (Morawska et al., 2018).

La información generada por los SMCA permite adquirir conocimiento y, mediante una interpretación adecuada, entender las tendencias de los contaminantes y su relación con las condiciones meteorológicas existentes. Estos factores son esenciales para determinar la distribución espacial y temporal de los contaminantes, así como las transformaciones que sufren en la atmósfera y los mecanismos de dispersión y transporte. Además, contar con esta información permite realizar predicciones de calidad del aire con anticipación y precisión, y poner en marcha medidas de emergencia incluidas en los planes correspondientes para mitigar los efectos negativos de episodios ambientales por contaminación atmosférica.

### **2.3.1 Clasificación de los SMCA según su principio de operación**

Los sistemas de monitorización de la calidad del aire se clasifican según su tipo de operación en manuales, automáticos y mixtos. A continuación, se describen cada una de las clasificaciones.

#### ***SMCA Manuales***

En los sistemas manuales de medición de calidad del aire, después recolectar los contaminantes, la muestra se envía a un laboratorio para su análisis. Las fases de toma y manipulación de la muestra son esenciales en este tipo de estaciones, ya que de estas dependen la precisión y exactitud de las mediciones. Para garantizar la confiabilidad de los resultados, es crucial que la muestra sea transferida intacta al laboratorio (INE, 2008).

Actualmente, los analizadores automáticos de gases son los únicos métodos validados para medir la concentración de estos contaminantes, mientras que las partículas en suspensión (PST, PM<sub>10</sub> y PM<sub>2.5</sub>) se miden principalmente mediante métodos manuales.

No obstante, los métodos automáticos de medición están experimentando un crecimiento progresivo en su utilización.

### ***SMCA Automáticos***

Los Sistemas de Medición de la Calidad del Aire Automáticos están compuestos por equipos de medición automáticos y continuos. A diferencia de los equipos manuales, estos sistemas realizan el análisis de la información de forma continua mientras se lleva a cabo la toma de la muestra, lo que permite obtener datos en tiempo real (INE, 2008).

### ***SMCA Mixtos***

Los Sistemas de Medición de la Calidad del Aire Mixtos combinan equipos manuales y automáticos. En estos sistemas, la toma de muestras y el análisis de la información se realizan tanto de forma manual como automática. Posteriormente, la información se procesa y se elabora un reporte (INE, 2008).

## **2.3.2 Clasificación de los SMCA por su movilidad**

Los sistemas de monitorización de la calidad del aire se clasifican según su movilidad en fijos o móviles. A continuación, se describen estas clasificaciones.

### ***SMCA Fijos***

Los SMCA se consideran fijos cuando los equipos de medición se instalan en un inmueble de construcción permanente. Estas estaciones se ubican en puntos estratégicos para medir la contaminación ambiental y la calidad del aire de su entorno. Generalmente, las estaciones fijas están integradas en redes de monitorización a nivel local, nacional o continental. Utilizan tecnologías avanzadas con sistemas de calibración que minimizan los errores de lectura (Zamora, 2015).

Estas estaciones son costosas y requieren un espacio considerable para su instalación. Los equipos utilizados suelen ser sensores de gama alta, ya que los sensores de bajo costo presentan fluctuaciones elevadas en las mediciones, por lo que no son frecuentemente empleados en estas instalaciones.

### ***SMCA Móviles***

Los SMCA móviles son sistemas motorizados que se emplean fundamentalmente en proyectos de investigación o en estudios de reconocimiento de los contaminantes atmosféricos en determinadas zonas. Además, se emplean en caso de fallos en

estaciones fijas o durante emergencias atmosféricas (INE, 2008).

En localidades donde no se posee registro de los contaminantes atmosféricos presentes o donde surge problema relacionado con la calidad del aire, se recomienda realizar estudios exploratorios a corto plazo utilizando SMCA móviles, ya que la instalación de SMCA fijos para medir diversos contaminantes constituye una solución significativamente más costosa.

### **2.3.3 Sistema de adquisición y transmisión de datos**

Los datos recopilados por los instrumentos de medición son procesados por el sistema de adquisición de datos y transmitidos al centro de cómputo mediante un sistema de comunicación remota, que puede incluir Internet, microondas, módem, radio, entre otros medios (INE, 2008)

Las redes de medición presentan la capacidad de generar millones de datos anuales, dependiendo de la cantidad de parámetros medidos y la frecuencia de muestreo. Para que los datos sobre la calidad del aire sean válidos, deben estar acompañados de registros sobre el funcionamiento interno de cada instrumento y las condiciones generales del SMCA. Esta información es crucial para garantizar la operación correcta del sistema en su totalidad.

Los SMCA deben garantizar el almacenamiento y respaldo de los datos generados, así como de su información operativa. Es fundamental que cuenten con dispositivos de almacenamiento que permitan la recuperación de la información en caso de fallos en los sistemas de transmisión.

## **2.4 Manejo de datos de la calidad del aire**

Antes de que los datos obtenidos se integren dentro de una base de datos, es necesario ejecutar un proceso de limpieza, verificación y validación. Este proceso puede realizarse de forma manual o por medio de sistemas automatizados de análisis de datos. La Norma Oficial Mexicana referente al establecimiento y operación de sistemas de monitorización de la calidad del aire (SEMARNAT, 2012) establece los siguientes aspectos para el manejo de datos obtenidos por los SMCA:

### **1. Limpieza de datos:**

- a. Definición de banderas de acuerdo con el sistema de adquisición de datos. Se identifican todos los datos con banderas, se asignan fundamentalmente a los

valores que se encuentran fuera del rango de operación y se ajustan a cero los valores negativos dentro del límite de tolerancia establecido por el SMCA.

- b. Conservación de todos los datos crudos y validados, sin eliminar ninguno. Los datos atípicos, se identifican con las banderas correspondientes, y son posteriormente verificados por el equipo técnico, quien también inspecciona el funcionamiento de los equipos.
  - c. Integración de los datos obtenidos por los instrumentos de medición, junto con las banderas asignadas, en una base de datos homologada con la Base de Datos Nacional.
2. Verificación de datos:
- a. Revisión de los registros y/o de las bitácoras de operación de las estaciones.
  - b. Se elaboran estadísticas descriptivas y gráficas para examinar la estructura, las tendencias de los valores y posibles anomalías.
  - c. Verificación de las relaciones existentes entre la concentración de los contaminantes atmosféricos y las variables meteorológicas.
3. Validación de datos:
- a. Ejecución de un análisis de datos conforme a criterios establecidos.
  - b. Se verifica que los conjuntos de datos de variables meteorológicas y contaminantes atmosféricos cumplen con un mínimo del 75% de existencia de datos en el periodo de estudio, para evaluar el cumplimiento de las NOM y obtener tendencias de contaminación atmosférica
  - c. Determinación del cumplimiento de los límites establecidos en las NOM emitidas por la Secretaría de Salud, a través de indicadores de calidad del aire.

En caso de que el interés se centre en obtener datos representativos a nivel de ciudad o zona metropolitana, y dado que las NOM de calidad del aire no especifican métodos específicos para evaluar el cumplimiento de los límites en esa escala, se recomienda seguir las directrices y recomendaciones de la *EPA* para analizar el cumplimiento de sus normas de calidad del aire en regiones específicas (INECC, 2024). Según la *EPA*, la evaluación del cumplimiento de las normativas a nivel metropolitano o de ciudad se realiza mediante el cálculo de un valor representativo basado en los valores obtenidos en cada estación. Este valor representativo es el máximo de los parámetros obtenidos en cada estación. Por lo tanto, si en alguna estación se obtiene un valor de concentración superior a los límites establecidos por las normas de calidad del aire, se considera que en la ciudad



o zona metropolitana también se rebasa este valor límite. El valor representativo del incumplimiento de la normatividad corresponde a las condiciones más desfavorables registradas en la zona.

## 2.5 Algoritmos de clasificación

La clasificación constituye un método de análisis de datos que genera modelos para describir clases de datos relevantes. Estos modelos, conocidos como clasificadores, son capaces de predecir etiquetas de clases categóricas (Han et al., 2012). El proceso de clasificación de datos se divide en dos etapas: la fase de aprendizaje, en la que se construye el modelo de clasificación, y la fase de clasificación, en la que el modelo se utiliza para predecir las etiquetas de clase de datos específicos. A continuación, se detallan los algoritmos de clasificación empleados en la investigación.

### 2.5.1 C4.5

Adopta un enfoque voraz en el que los árboles de decisión se construyen de forma recursiva de arriba hacia abajo, y aplica un enfoque de divide y vencerás. Emplea *gain ratio* como medida de selección de atributos (Han et al., 2012). Esto ayuda a manejar el sesgo hacia atributos con mayor cardinalidad, que constituye una limitación de algoritmo *ID3*. Utiliza técnicas para discretizar atributos continuos, esto permite el manejo de conjuntos de datos que contienen la misma cantidad de atributos categóricos y continuos. Además, incorpora un paso de poda después de la construcción del árbol, que contribuye a evitar el sobreajuste al eliminar nodos que no aportan significativamente a la mejora de la clasificación. También se puede utilizar para problemas de regresión al permitir que las hojas del árbol contengan valores numéricos en lugar de etiquetas de clase. Es más robusto que *ID3* y tiende a generar árboles de decisión más compactos y equilibrados

### 2.5.2 Naive Bayes

Constituye un clasificador estadístico que posibilita predecir probabilidades de pertenencia a una clase, como la probabilidad de que una tupla determinada pertenezca a una clase particular (Han et al., 2012). Este algoritmo asume independencia condicional de clase (el efecto del valor de un atributo en una clase determinada es independiente de los valores del otro atributo). Además, exhibe una alta precisión y velocidad cuando se aplica a bases de datos extensas.

### 2.5.3 *Bayesian Belief Networks*

Constituye un modelo gráfico probabilístico que presenta aplicaciones en la clasificación. Este algoritmo, a diferencia del *Naive Bayesian Classification*, no asume independencia condicional de clase y permite la representación de dependencias entre subconjuntos de atributos. Además, proporciona un modelo gráfico de relaciones causales, sobre el cual se puede efectuar el aprendizaje. Se define por dos componentes: un *directed acyclic graph* y un conjunto de *conditional probability tables*, cada nodo del *directed acyclic graph* representa una variable aleatoria (Han et al., 2012).

### 2.5.4 *Backpropagation*

Constituye un algoritmo de clasificación basado en redes neuronales. El proceso de aprendizaje se realiza a través del procesamiento iterativo de un conjunto de datos de entrenamiento compuesto por tuplas. El valor que predice la red para cada tupla se compara con el valor objetivo conocido. Los pesos se modifican en cada iteración con el objetivo de disminuir el error cuadrático medio existente entre el resultado de la predicción y el valor real. Estos ajustes de la red se ejecutan en la dirección "hacia atrás", propagándose desde las capas ocultas hasta la primera capa (Han et al., 2012).

### 2.5.5 *K nearest-neighbor (Knn)*

Constituye un algoritmo de clasificación basado en el aprendizaje por analogía, mediante la comparación de una tupla de prueba determinada con tuplas de entrenamiento que son similares a esta. Las tuplas de entrenamiento se describen mediante  $n$  atributos, y cada una de estas representa un punto en un espacio de  $n$  dimensiones. Dada una tupla desconocida, el clasificador busca en el espacio del patrón las  $k$  tuplas de entrenamiento más cercanas a la tupla desconocida. Realiza comparaciones basadas en la distancia que asignan intrínsecamente el mismo peso a cada atributo. Por lo tanto, pueden sufrir una precisión deficiente cuando se les asignan atributos ruidosos o irrelevantes. La selección de una métrica de distancia puede ser crítica, se pueden emplear la distancia euclidiana, la distancia de Manhattan u otras medidas de distancia (Han et al., 2012).

### 2.5.6 *Support Vector Machine (SVM)*

Constituye un algoritmo de clasificación de datos lineales y no lineales. En este, se transforman los datos de entrenamiento en una dimensión superior, donde encuentra un hiperplano que separa los datos por clase mediante el empleo de tuplas de entrenamiento esenciales, llamadas vectores de soporte. Presenta un procesamiento lento de los datos de

entrenamiento para bases de datos extensas. Se alcanza una precisión elevada debido a su capacidad para modelar límites de decisión no lineales complejos. La posibilidad de ocurrencia de sobreajuste es significativamente inferior a otros métodos de clasificación. Los vectores de soporte obtenidos también suministran una descripción compacta del modelo aprendido. Este algoritmo se puede emplear en la predicción numérica y en la clasificación (Han et al., 2012).

## 2.6 Medidas de error en la comparación de valores

Cuando se comparan las mediciones entre varios sistemas, es fundamental evaluar las diferencias y errores entre las mediciones para determinar la precisión y exactitud de los sistemas (Han et al., 2012). A continuación, se presentan las ecuaciones de error empleadas en el análisis de los valores de concentración de contaminantes atmosféricos obtenidos:

### 2.6.1 Error Medio Absoluto (MAE)

El error medio absoluto es la media de los errores absolutos.

$$MAE = \frac{1}{n} \sum_{i=1}^n |M_i - R| \quad (\text{Ecuación 1})$$

### 2.6.2 Porcentaje Medio del Error Absoluto (MAPE)

El porcentaje medio del error absoluto es la media de los errores relativos.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left( \frac{|M_i - R|}{R} \right) * 100\% \quad (\text{Ecuación 2})$$

### 2.6.3 Raíz del Error Cuadrático Medio (RMSE)

La raíz del error cuadrático medio es la raíz cuadrada del MSE. Proporciona una medida de la desviación estándar de los errores.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (M_i - R)^2} \quad (\text{Ecuación 3})$$

### 2.6.4 Coeficiente de correlación

El coeficiente de correlación de dos variables aleatorias es una medida de su dependencia lineal (Han et al., 2012). Si cada variable tiene N observaciones escalares, entonces el coeficiente de correlación de Pearson se define como:

$$\rho(A, B) = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{A_i - \mu_A}{\sigma_A} \right) \left( \frac{B_i - \mu_B}{\sigma_B} \right) \quad (\text{Ecuación 4})$$

### 2.6.5 Coeficiente de determinación

Un método para determinar el mejor ajuste es calcular el coeficiente de determinación,  $R^2$ . Este coeficiente mide la capacidad predictiva de un modelo y varía entre 0 y 1. Cuanto más alto sea el valor de  $R^2$ , mejor será el modelo para predecir los datos:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (\text{Ecuación 5})$$

Donde:

$M_i$ : valor medido por el sistema

$R$ : valor de referencia

$n$ : cantidad máxima de observaciones

$i$ : observación actual

$\mu_A$ : media de  $A$

$\sigma_A$ : desviación estándar de  $A$

$\hat{y}_i$ : valores calculados de  $y$

$\bar{y}$ : media de  $y$

## Capítulo 3. Metodología, preprocesamiento de datos y diseño de la solución de monitoreo

En el presente capítulo se describen las fuentes de obtención de datos de concentración de contaminantes atmosféricos en la zona norte de la Ciudad de México: bases de datos de las estaciones de la RAMA: Camarones (CAM) y Gustavo A. Madero (GAM), sistema fijo de monitorización remota, sistema de monitorización móvil *Flatburn*. Se presenta la metodología de análisis de los valores de concentración de contaminantes atmosféricos  $O_3$ ,  $PM_{10}$ ,  $PM_{2.5}$  y  $NO_2$ , obtenidos por las estaciones CAM y GAM en los años 2022 y 2023, que posibilitan realizar un análisis inicial del comportamiento de la calidad del aire en la zona de estudio. En la Figura 3.1 se presenta un diagrama general de la metodología empleada en la investigación.

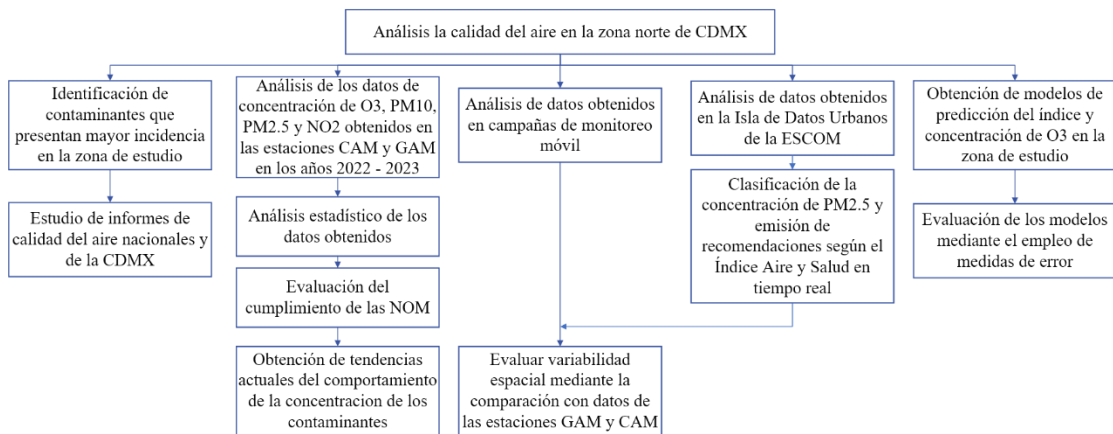


Figura 3.1 Diagrama general de la metodología empleada en la investigación (elaboración propia).

### 3.1 Estaciones de monitoreo automático atmosférico CAM y GAM

Para obtener los valores representativos actuales de la calidad del aire en la zona norte de la Ciudad de México, se recuperaron los valores de concentración de contaminantes atmosféricos  $PM_{10}$ ,  $PM_{2.5}$ ,  $O_3$  y  $NO_2$  de las obtenidos por las estaciones CAM y GAM durante los años 2022 y 2023.

La estación Camarones se encuentra ubicada en la demarcación territorial Azcapotzalco de la Ciudad de México, específicamente en las coordenadas  $19.47^\circ$  de latitud norte y  $-99.17^\circ$  de longitud oeste. Mientras que estación GAM se localiza en la demarcación territorial Gustavo A. Madero de la Ciudad de México, específicamente en las coordenadas  $19.48^\circ$  de latitud norte y  $-99.09^\circ$  de longitud oeste. En la Figura 3.2 se muestran las ubicaciones de estas estaciones (SINAICA, 2023). En los Anexos 3 y 4 se presentan imágenes de estas estaciones.

La obtención de los datos históricos medidos por estas estaciones se realizó a través de la página web del Sistema de Monitoreo Atmosférico (SIMAT). La base de datos se conforma de la integración de los datos de concentración de contaminantes atmosféricos obtenidos por las estaciones Camarones y GAM:  $O_3$ ,  $PM_{10}$ ,  $PM_{2.5}$  y  $NO_2$  en los años 2022 y 2023. A continuación, se describen los atributos que conforman la base de datos:

- Fecha: Se registran los días específicos en los que se midieron valores. Se emplea la siguiente estructura: día/mes/año.
- Hora: Se registran los horarios en los que se midieron valores, en intervalos de una hora.
- Concentración de contaminantes atmosféricos ( $O_3$ ,  $PM_{10}$ ,  $PM_{2.5}$ , y  $NO_2$ ): Atributos numéricos. Se registran los promedios horarios de concentración de estos contaminantes atmosféricos en  $\mu g/m^3$ .

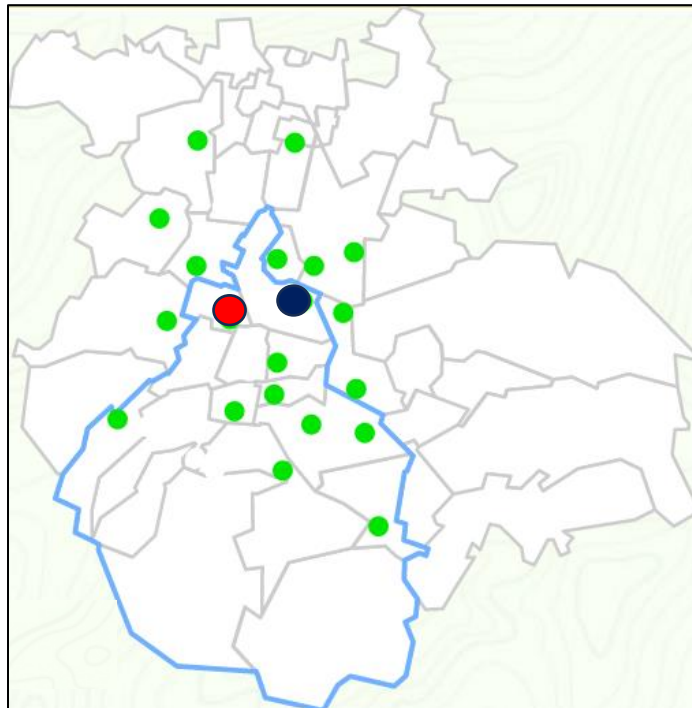


Figura 3.2. Ubicaciones de la estación Camarones (rojo) y Gustavo A Madero (azul) (SEDEMA, 2024).

### 3.1.1 Preprocesamiento de los datos crudos de contaminantes atmosféricos

El proceso de integración y limpieza del conjunto de datos constó de varias etapas, estas se presentan a continuación:

1. Se recuperaron los valores históricos de concentración de contaminantes O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> y NO<sub>2</sub>, medidos por las estaciones en los años 2022-2023 a través de la página web del SIMAT.
2. Se identificaron los datos faltantes en las bases de datos de cada uno de los contaminantes.
3. Se identificaron los datos atípicos en las bases de datos de cada uno de los contaminantes.
4. Se integraron los conjuntos de datos de cada contaminante atmosférico.
5. Se analizó la existencia de datos con ruido en las bases de datos integradas de cada contaminante atmosférico.

#### *Identificación de valores faltantes*

En el análisis de las bases de datos de concentración de O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> y NO<sub>2</sub> se identificaron porcentajes de valores faltantes anuales para cada contaminante, estos se encontraron en un rango de un 6.14% hasta un 75.83%, durante los años 2022 y 2023. En la Tabla 3.1 se presentan los porcentajes de valores faltantes de cada contaminante por año. Estos porcentajes obtenidos se encuentran principalmente influenciados por la inexistencia registros de valores en determinados meses del período de estudio.

Tabla 3.1. Porcentajes de valores faltantes de concentración de O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> y NO<sub>2</sub> en los años 2022-2023 (elaboración propia).

Año	Estación	O <sub>3</sub>		PM <sub>10</sub>		PM <sub>2.5</sub>		NO <sub>2</sub>	
		DF*	%**	DF	%	DF	%	DF	%
2022	CAM	2174	24.82%	4834	55.18%	4834	55.18%	2380	27.17%
	GAM	563	6.43%	6643	75.83%	6643	75.83%	538	6.14%
2023	CAM	2443	27.89%	1815	20.72%	1864	21.28%	2320	26.48%
	GAM	3050	34.83%	3210	36.64%	3293	37.59%	3011	34.37%

\*Datos faltantes; \*\* Porcentaje de datos faltantes con respecto al total de valores posibles.

#### *Identificación de valores atípicos*

Se identificaron los valores atípicos de promedios horarios de concentración de contaminantes atmosféricos en las bases de datos integradas de cada contaminante

atmosférico por estación, mediante el empleo del criterio de encontrarse a una distancia de  $1.5 \times IQR$  (rango inter-cuartil) superior al valor del tercer cuartil.

En la identificación de los datos atípicos de promedio de concentración horaria de cada contaminante, obtenidos en las estaciones CAM y GAM durante los años 2022 y 2023, primeramente, se obtuvieron los diagramas de caja de cada uno de estos parámetros (Figuras 3.3-3.6). En la Tabla 3.2 se presenta la relación de valores atípicos por conjunto de datos.

Tabla 3.2. Cantidad de valores atípicos de concentración de contaminantes atmosféricos de las estaciones CAM y GAM durante los años 2022 y 2023 (elaboración propia).

Período	Estación	Valores atípicos O <sub>3</sub>	Valores atípicos PM <sub>10</sub>	Valores atípicos PM <sub>2.5</sub>	Valores atípicos NO <sub>2</sub>
2022-2023	CAM	387	220	274	344
	GAM	116	100	153	131

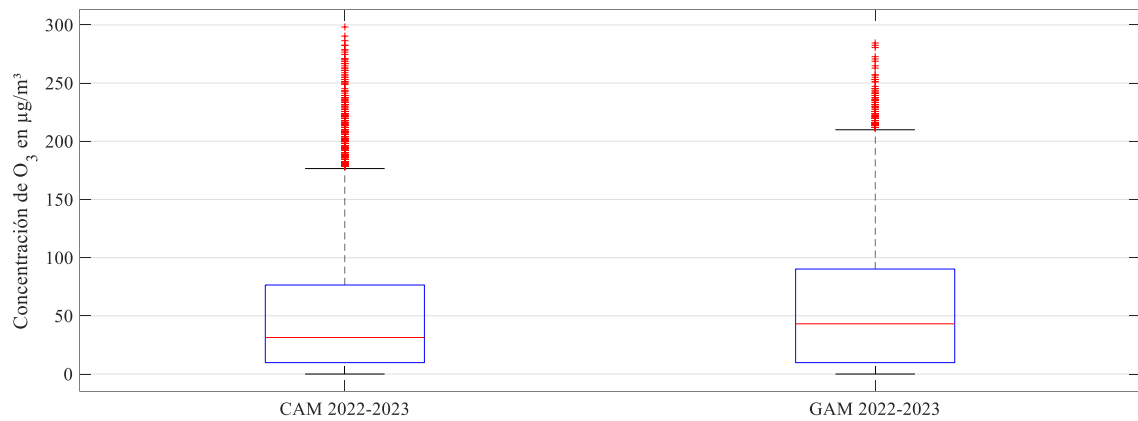


Figura 3.3. Diagrama de caja de los valores de concentración de O<sub>3</sub> de las estaciones CAM y GAM en los años 2022 y 2023 (elaboración propia).

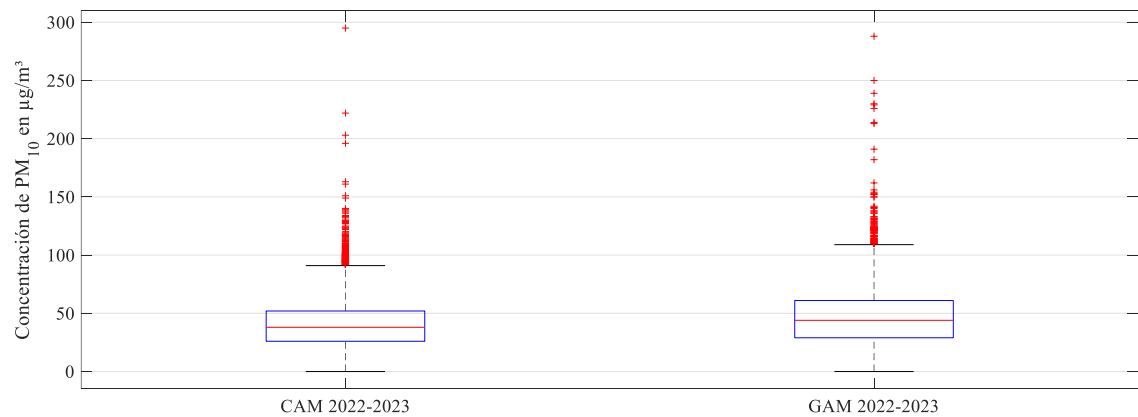


Figura 3.4. Diagrama de caja de los valores de concentración de PM<sub>10</sub> de las estaciones CAM y GAM en los años 2022 y 2023 (elaboración propia).



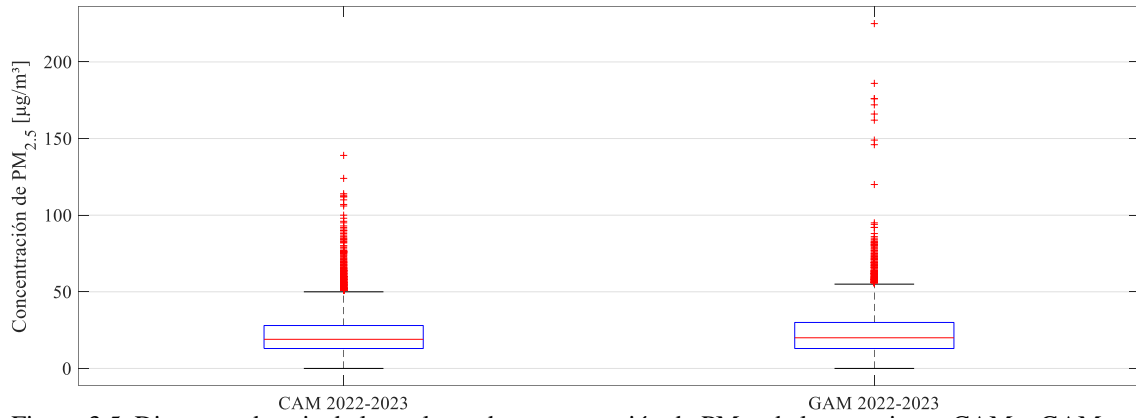


Figura 3.5. Diagrama de caja de los valores de concentración de  $PM_{2.5}$  de las estaciones CAM y GAM en los años 2022 y 2023 (elaboración propia).

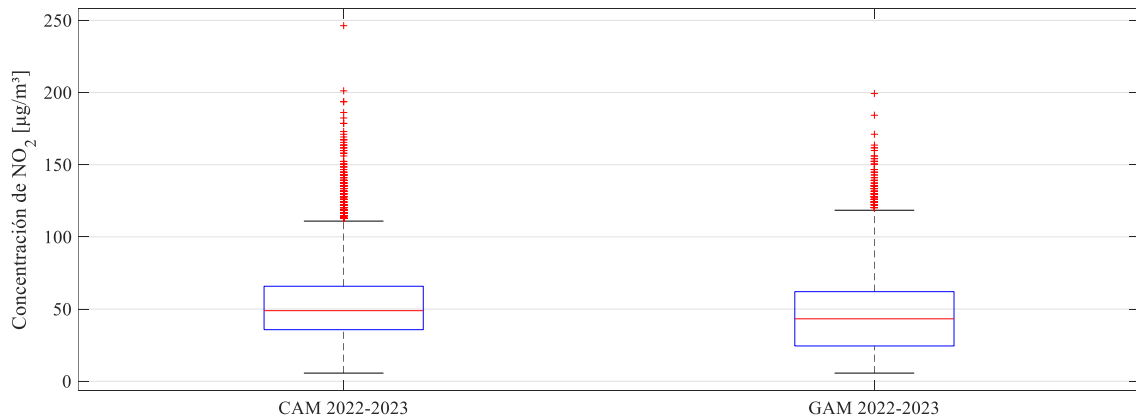


Figura 3.6. Diagrama de caja de los valores de concentración de  $NO_2$  de las estaciones CAM y GAM en los años 2022 y 2023 (elaboración propia).

### ***Análisis de la existencia de valores con ruido***

Para identificar la existencia de valores con ruido, se analizó la validez de los datos atípicos, mediante su comparación con los valores de concentración de cada contaminante obtenidos durante el año 2021 en la región. Estos datos se publicaron en el Informe Anual de Calidad del Aire de Ciudad de México del año 2024 (SEDEMA, 2024), en los Anexos 5-8 se presentan las gráficas de los promedios horarios de cada contaminante obtenidos en la CDMX, durante el año 2021. Como resultado se identificó que durante el año 2021 se registraron valores máximos de concentración de estos contaminantes, superiores a los valores atípicos identificados en los años 2022 y 2023, por lo tanto, se corroboró la validez de los datos atípicos identificados y se descartó la existencia de valores con ruido.

## **3.2 Sistema fijo de monitorización remota**

El sistema de monitorización remota desarrollado posee un *Raspberry Pi 4 Model B* como dispositivo de control y adquisición de datos. Además, presenta un sensor de calidad del aire *AM1008W* que posibilita la obtención de valores de temperatura ambiental, humedad relativa y de concentración de contaminantes atmosféricos:  $CO_2$ ,  $PM_{2.5}$ ,  $PM_{10}$  y

compuestos orgánicos volátiles (VOC). Los valores medidos por el sensor *AM1008W* y los indicadores de funcionamiento del sistema se envían a la plataforma *Initial State*, que posibilita la visualización remota de los valores en tiempo real. Se desarrolló una aplicación en el software *Android Studio* que se conecta con la plataforma *Initial State* y propicia la visualización selectiva de la información. Además, se realizó el registro de los datos obtenidos por el sistema cada 1 minuto en una base de datos almacenada en la tarjeta microSD, que se encuentra conectada al dispositivo *Raspberry Pi*, con el objetivo de posibilitar el análisis de los datos históricos del sistema. En la Figura 3.7, se presenta la arquitectura del sistema de monitorización remota de la calidad del aire desarrollado.

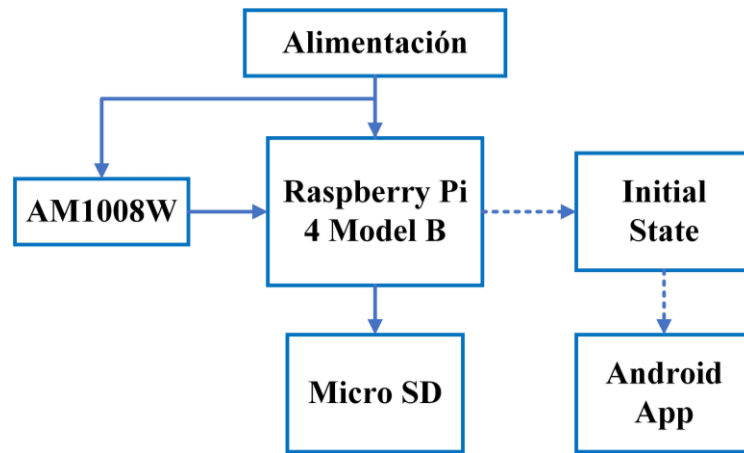


Figura. 3.7. Arquitectura del sistema de monitorización remota de la calidad del aire (elaboración propia). El sensor de calidad del aire para ambientes interiores *AM1008W*, fabricado por la empresa *Cubic Ltd* (Figura 3.8), posee integrados sensores que posibilitan la medición de la concentración de CO<sub>2</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>, VOC, temperatura ambiente y humedad relativa (Cubic Ltd, 2023). En la Tabla 3.3, se presentan las especificaciones técnicas de este dispositivo.



Figura 3.8. Sensor de calidad del aire *AM1008W* (Cubic Ltd, 2023).

Tabla 3.3. Características del sensor de calidad del aire *AM1008W* (Cubic Ltd, 2023).

Características	Especificaciones
Principio de operación	CO <sub>2</sub> : NDIR PM: dispersión láser VOC: semiconductor
Rangos de medición	PM: 0 ~ 1000 µg/m <sup>3</sup> CO <sub>2</sub> : 0 ~ 5000 ppm VOC: 0 ~ 3 level Temperatura: -20°C ~ 70°C RH: 0 - 95%RH
Precisión del sensor de PM <sub>2.5</sub>	0 ~ 35 µg/m <sup>3</sup> , ± 5 µg/m <sup>3</sup> > 35 µg/m <sup>3</sup> , ±15% de lectura 0°C ~ 40°C, 50 ±10%RH Ref. del instrumento: <i>TSI</i>
Precisión del sensor de CO <sub>2</sub>	± (50 ppm + 5% de lectura) a 10°C~50°C, 50±10%RH
Máxima frecuencia de muestreo	1 Hz
Comunicación	UART_TTL (3.3V TTL)
Alimentación	5V DC ± 0.1V

### 3.2.1 Método de validación

Como método de validación de los valores de concentración de contaminantes medido se realizó una comparación con los datos obtenidos por un sensor de calidad del aire *Davis AirLink* (Figura 3.9) que compartió la misma ubicación del sistema desarrollado durante la ejecución de las pruebas experimentales, la Isla de Datos Urbanos de la ESCOM (Anexo 9 y 10). Se analizaron los valores de concentración de PM<sub>2.5</sub> obtenidos por el sistema durante un período de 75 días (15 de marzo-29 de junio del 2023).



Figura 3.9. Sistema de monitorización de calidad del aire *Davis AirLink* (Davis Instruments, 2023).

El dispositivo *Davis AirLink* posee integrado un sensor *PMSA003* que permite la medición de  $PM_{1.0}$ ,  $PM_{2.5}$ ,  $PM_{10}$ . Además, posibilita la medición de la temperatura ambiente y humedad relativa (*Davis Instruments*, 2024). También, permite la visualización en tiempo real de los valores medidos mediante una aplicación móvil. Los datos históricos obtenidos por el sistema se recuperaron mediante empleo del sitio web *WeatherLink*. En la Tabla 3.4, se presentan las características de este dispositivo.

Tabla 3.4. Características del dispositivo *Davis AirLink*. (*Davis Instruments*, 2023).

Características	Especificaciones
Variables	$PM_{1.0}$ , $PM_{2.5}$ y $PM_{10}$ ( $\mu g/m^3$ ), Temperatura ( $^{\circ}F$ ), RH (%)
Sensor de material particulado	Óptico; non-FEM ( <i>PMSA003</i> , <i>Plantower</i> )
Resolución del sensor de material particulado	1 $\mu g/m^3$
Precisión del sensor de material particulado	+/-10 $\mu g/m^3$
Resolución de tiempo	1 min
Comunicación	Wi-Fi
Alimentación de energía	5 VDC AC adaptador
Condiciones ambientales	-10 – 60 $^{\circ}C$ and 0.1 - 100% RH
Dimensiones	2 (L) x 3.5 (W) x 1 (H) para interiores; 4 (L) x 4.5 (W) x 1.5 (H) con protector para exteriores
Peso	3.7 oz (sensor solo); 6.5 oz (con protector para exteriores)

### 3.2.2 Preprocesamiento de las bases de datos obtenidas en la Isla de Datos Urbanos

El preprocesamiento de los conjuntos de datos de concentración de  $PM_{2.5}$  obtenidos en la Isla de Datos Urbanos de la ESCOM constó de varias etapas, estas se presentan a continuación:

1. Identificación de la cantidad de valores faltantes.
2. Identificación de valores atípicos, mediante el empleo del criterio de encontrarse a una distancia de  $1.5 \times IQR$  superior al valor del tercer cuartil.
3. Determinación de la existencia de datos con ruido mediante un algoritmo de análisis temporal, que identifica una diferencia de 5 veces la desviación estándar en las diferencias absolutas existentes entre cada uno de los valores de concentración de  $PM_{2.5}$  con el valor

del promedio móvil correspondiente, se empleó un tamaño de ventana de 5 min.

4. Determinación de la cantidad de horas que cumplieron el criterio de existencia de datos por minuto (75%).
5. Integración de los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por los sistemas de monitorización.
6. Análisis de la relación existente entre los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por los sistemas de monitorización, mediante el empleo de medidas de error.
7. Calibración del sistema desarrollado mediante el método de regresión lineal.
8. Evaluación de la relación de los promedios horarios de concentración de  $PM_{2.5}$  calibrados, obtenidos por el sistema desarrollado, con respecto a los valores obtenidos por el dispositivo *Davis AirLink*.

A continuación, se presenta el proceso de preprocesamiento para los conjuntos de datos de concentración de  $PM_{2.5}$  obtenidos por los sistemas instalados en el Isla de Datos Urbanos de la Escuela Superior de Cómputo.

### ***Datos faltantes***

En el análisis de los conjuntos de datos de concentración de  $PM_{2.5}$  obtenidos en la Isla de Datos Urbanos en el período: 15 de marzo del 2024 - 29 de mayo del 2024, con una resolución de 1 minuto, se identificaron porcentajes de valores faltantes de 11.67% y 10.38% para el sistema desarrollado y el dispositivo *Davis AirLink*, respectivamente. En la Tabla 3.5 se presentan los porcentajes de valores faltantes para cada sistema. Estos porcentajes obtenidos se encuentran principalmente influenciados por períodos de inconsistencias en la conectividad a internet y alimentación de energía eléctrica de los sistemas, durante la etapa de estudio.

Tabla 3.5. Porcentajes de valores faltantes de concentración de  $PM_{2.5}$  en el periodo de estudio (elaboración propia).

Sistema	DF*	DE**	DP***	%DF****	%DE *****
Raspberry	12606	95394	108000	11.67%	88.33%
<i>Davis</i>	11215	96794		10.38%	89.62%

\* Datos faltantes; \*\*Datos existentes; \*\*\*Datos posibles; \*\*\*\*Porcentaje de datos faltantes; \*\*\*\*\* Porcentaje de datos existentes.

### ***Identificación de valores atípicos***

En la identificación de los datos atípicos de concentración de  $PM_{2.5}$  obtenidos por el

sistema desarrollado y el dispositivo *Davis AirLink* en el período de estudio, se obtuvo el diagrama de caja de los valores de estos parámetros (Figuras 3.10 y 3.11), que posibilita la visualización de los valores atípicos. En la Tabla 3.6 se presenta la relación de valores atípicos para cada sistema.



Figura 3.10. Diagrama de caja de los valores de concentración de  $PM_{2.5}$  obtenidos por el sensor *AM1008W* durante el período marzo-mayo del 2024 (elaboración propia).

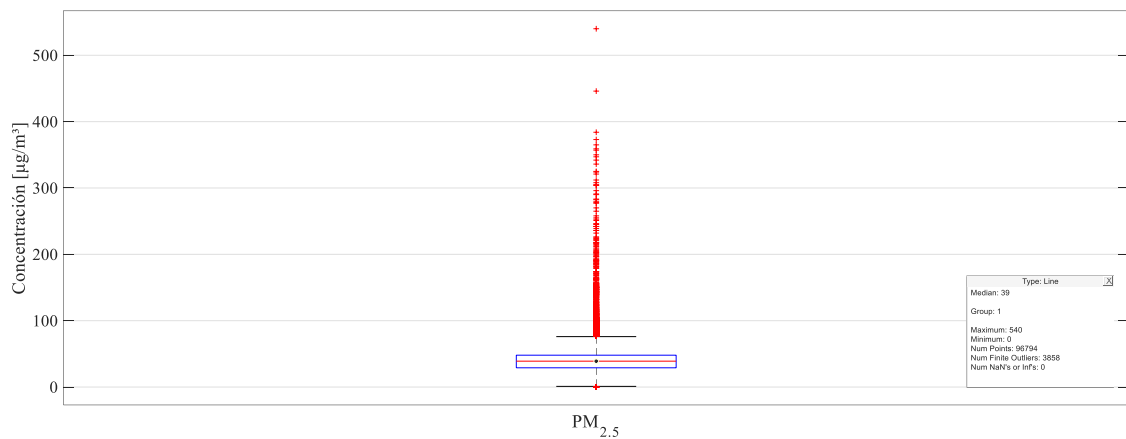


Figura 3.11. Diagrama de caja de los valores de concentración de  $PM_{2.5}$  obtenidos por el dispositivo *Davis AirLink* durante el período marzo-mayo del 2024 (elaboración propia).

Tabla 3.6. Datos atípicos de concentración de  $PM_{2.5}$  obtenidos por el sistema desarrollado y el dispositivo *Davis AirLink* durante el período marzo-mayo del 2024 (elaboración propia).

Método	Valores > 1.5 IQR + Q3 ( <i>GRIMM</i> > 59) ( <i>TSI</i> > 69)	Igual a mínimo ( <i>GRIMM</i> = 0) ( <i>TSI</i> = 1) ( <i>Davis</i> = 0)	Total valores atípicos	Datos existentes	Porcentaje de datos atípicos
<i>GRIMM</i>	4161	1	4162	95394	4.36%
<i>TSI</i>	3670	37	3707		3.89%
<i>Davis</i>	3858	26	3884		4.01%

### Identificación de valores con ruido

Como resultado de la aplicación del algoritmo de análisis temporal de los valores de concentración de  $PM_{2.5}$ , se identificaron 225, 234 y 277 valores con ruido en los conjuntos

de datos obtenidos con el método *GRIMM*, *TSI* y el dispositivo *Davis AirLink*, respectivamente, estos valores resultaron eliminados de la base de datos. En las Figuras 3.12 - 3.14 se presentan los datos con ruido identificados en cada uno de los conjuntos de datos. En la Tabla 3.7. se muestran los resultados de la aplicación del algoritmo de análisis temporal a los conjuntos de datos de concentración de  $PM_{2.5}$  obtenidos en el período de estudio.

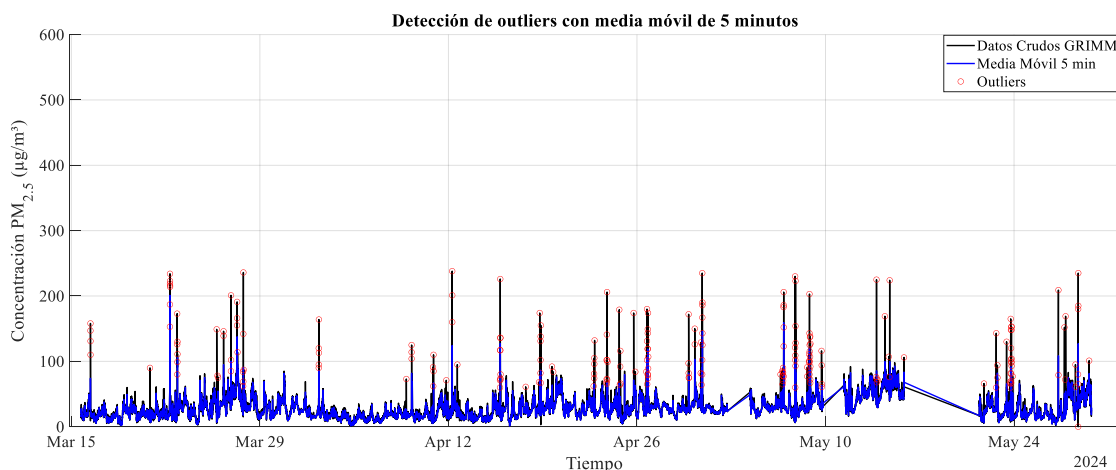


Figura 3.12. Valores de concentración de  $PM_{2.5}$  (*GRIMM*) con ruido identificados mediante el algoritmo de análisis temporal (elaboración propia).

Tabla 3.7. Resultados de la aplicación del algoritmo de análisis temporal a los conjuntos de datos de concentración de  $PM_{2.5}$  obtenidos en el período de estudio (elaboración propia).

Método	Tamaño de ventana (min)	# total de datos	# atípicos	# datos con ruido	# datos válidos	% datos con ruido con respecto a los atípicos
<i>GRIMM</i>	5	95394	4162	225	95169	5.41%
<i>TSI</i>			3670	234	95160	6.38%
<i>Davis</i>		96794	3858	277	96517	7.18%

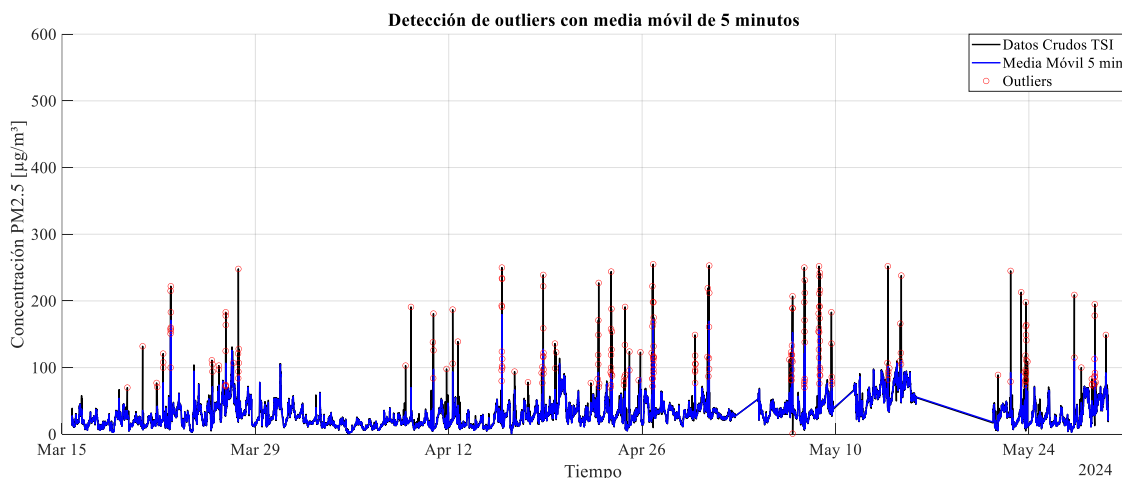


Figura 3.13. Valores de concentración de  $PM_{2.5}$  (*TSI*) con ruido identificados mediante el algoritmo de análisis temporal (elaboración propia).

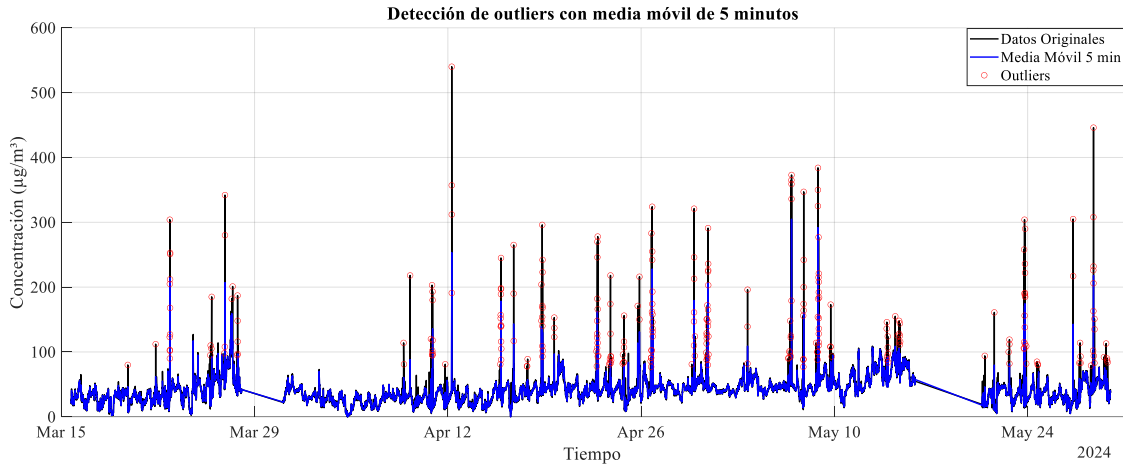


Figura 3.14. Valores de concentración de  $PM_{2.5}$  (Davis AirLink) con ruido identificados mediante el algoritmo de análisis temporal (elaboración propia).

### ***Integración de los promedios horarios de concentración de $PM_{2.5}$***

En la integración de los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por los sistemas instalados en la Isla de Datos Urbanos, primeramente, se verificó la cantidad de horas que cumplían el criterio de existencia de al menos un 75% (45 datos) de los valores por minuto para cada conjunto de datos, luego se calcularon los promedios horarios de las horas que cumplieron el criterio de existencia de datos. Posteriormente se integraron los promedios horarios de cada base de datos, en correspondencia con los datos de fecha y hora. En la Tabla 3.8 se presentan los porcentajes de promedios horarios viables de concentración de  $PM_{2.5}$  en el periodo de estudio.

Tabla 3.8. Porcentajes de valores viables de promedio de concentración horaria de  $PM_{2.5}$  obtenidos en el período de estudio (elaboración propia).

Sistema	Datos faltantes	Datos válidos	Datos posibles	Porcentaje de datos faltantes	Porcentaje de datos válidos
<i>AM1008W-Davis</i>	285	1515	1800	15.83%	84.17%

### ***Análisis de la relación entre los promedios horarios de concentración de $PM_{2.5}$***

En el análisis de los parámetros de relación entre los promedios de horarios de concentración de  $PM_{2.5}$  obtenidos por el dispositivo *Davis AirLink* y cada uno de los métodos de medición del sensor *AM1008W* se obtuvieron valores de MAPE de 33.45% para la relación *GRIMM - Davis* y 28.97% para la relación *TSI - Davis*, que indican la existencia de un error relativo significativo entre los valores obtenidos por los sistemas. En la Figura 3.15 se muestran los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por el dispositivo *Davis AirLink* y los métodos de medición del sensor *AM1008W*. En la Tabla 3.9 se muestran los indicadores de relación obtenidos.



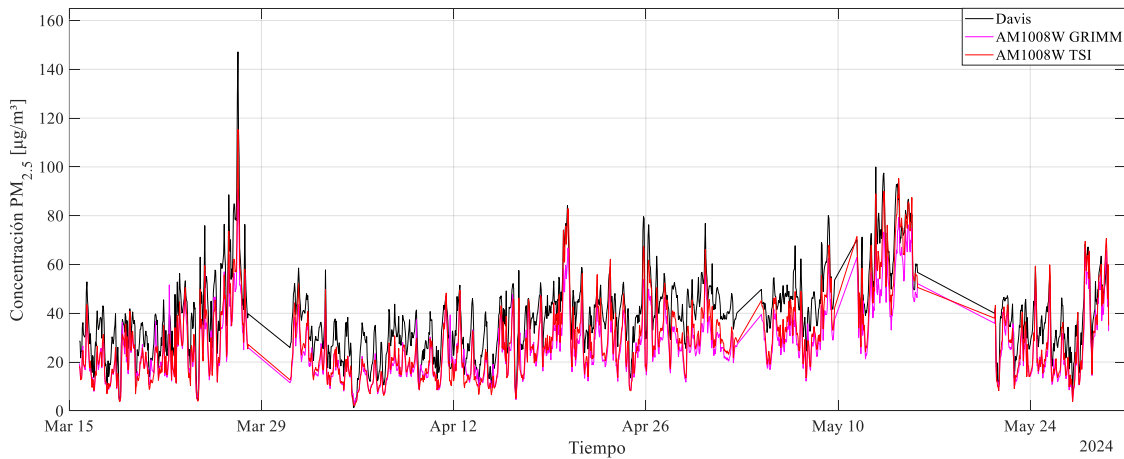


Figura 3.15 Promedios horarios de concentración de  $PM_{2.5}$  obtenidos por dispositivo *Davis AirLink* y los métodos de medición del sensor *AM1008W* (elaboración propia).

Tabla 3.9. Parámetros de evaluación de la relación entre los valores de concentración de  $PM_{2.5}$  obtenidos por los métodos de medición *TSI*, *GRIMM* en el periodo de estudio (elaboración propia).

Relación	Coefficiente de correlación	$R^2$	MAE [ $\mu\text{g}/\text{m}^3$ ]	RMSE [ $\mu\text{g}/\text{m}^3$ ]	MAPE [%]
<i>GRIMM - Davis</i>	0.9357	0.875	12.4763	13.6697	33.45%
<i>TSI - Davis</i>	0.957	0.916	10.1382	11.0166	28.97%

Como existe una elevada correlación lineal en los dos casos, se seleccionó el método de regresión lineal para calibrar el sistema desarrollado y contribuir al aumento de la exactitud en la obtención de valores de concentración de  $PM_{2.5}$  mediante el sensor *AM1008W*.

#### ***Calibración del sistema desarrollado mediante el método de regresión lineal***

En la calibración del sistema desarrollado, se implementó el método de regresión lineal mediante el empleo del software *Matlab*. En el Anexo 11 se presenta el código desarrollado para obtener las ecuaciones de regresión lineal y las medidas de error.

A continuación, se muestran las ecuaciones de ajuste obtenidas para los casos de las concentraciones horarias obtenidas con el método *GRIMM* y *TSI*. En las Figuras 3.16 y 3.17 se presentan los ajustes lineales de los promedios de concentración horaria de  $PM_{2.5}$  obtenidos con los métodos *GRIMM* y *TSI*, respectivamente.

$$\text{Conc. } h \text{ ajustada } PM_{2.5}(GRIMM) = \text{Conc. } h \text{ } PM_{2.5\_GRIMM} * 1.101 + 9.588$$

Donde:

*Conc\_h\_ajustada\_PM<sub>2.5</sub>\_GRIMM*: Concentración horaria de  $PM_{2.5}$  para el método de medición *GRIMM*, ajustada mediante la ecuación de regresión lineal.

*Conc\_h\_PM<sub>2.5</sub>\_GRIMM*: Valores reales de Concentración horaria de  $PM_{2.5}$  obtenidas por el método de medición *GRIMM*.

$$Conc\_h\_ajustada\_PM_{2.5\_TSI} = Conc\_h\_PM_{2.5\_TSI} * 0.937 + 11.744$$

Donde:

*Conc\_h\_ajustada\_PM<sub>2.5</sub>\_TSI*: Concentración horaria de PM<sub>2.5</sub> para el método de medición *TSI*, ajustada mediante la ecuación de regresión lineal.

*Conc\_h\_PM<sub>2.5</sub>\_TSI*: Valores reales de Concentración horaria de PM<sub>2.5</sub> obtenidas por el método de medición *TSI*.

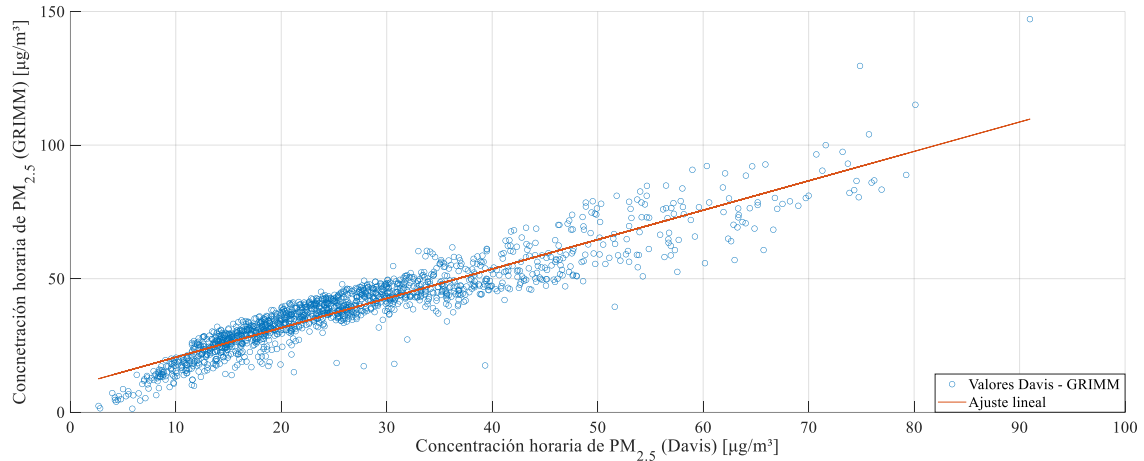


Figura 3.16 Ajuste lineal de los valores de concentración horaria de PM<sub>2.5</sub> obtenidos con el método *GRIMM* (elaboración propia).

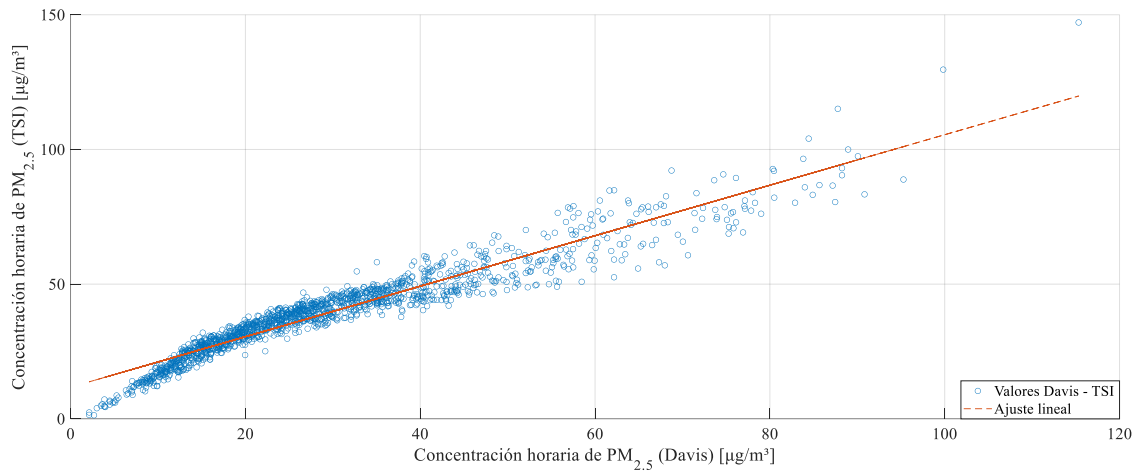


Figura 3.17 Ajuste lineal de los valores de concentración horaria de PM<sub>2.5</sub> obtenidos con el método *TSI* (elaboración propia).

Se aplicaron las ecuaciones de regresión lineal obtenidas para ajustar los valores de concentración de PM<sub>2.5</sub> obtenidos por los métodos *GRIMM* y *TSI* del sensor *AM1008W*. En las Figuras 3.18 y 3.19 se muestran los valores de concentración horaria de PM<sub>2.5</sub> ajustados mediante las ecuaciones de regresión lineal obtenidas, correspondientes a los métodos *GRIMM* y *TSI*, respectivamente.

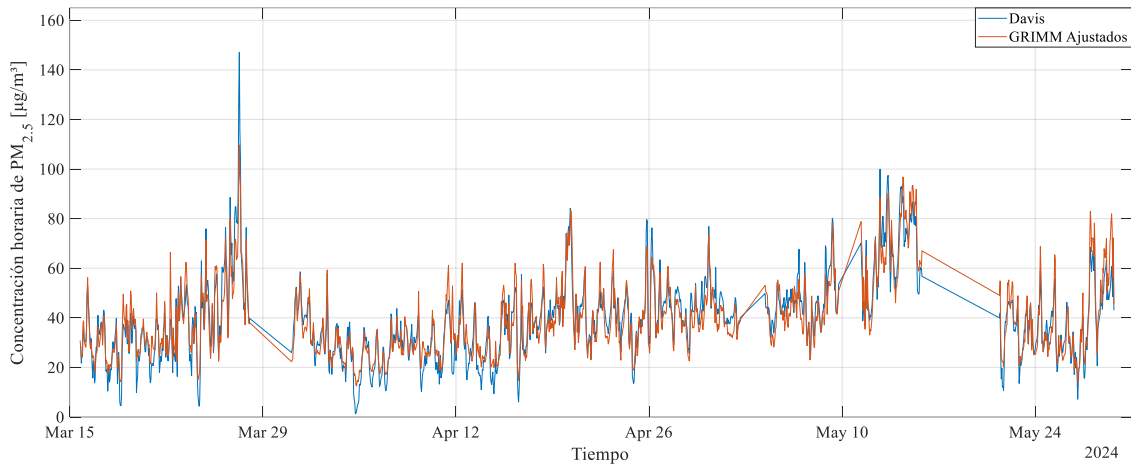


Figura 3.18. Valores de concentración horaria de  $PM_{2.5}$  (GRIMM) ajustados mediante la ecuación de regresión lineal (elaboración propia).

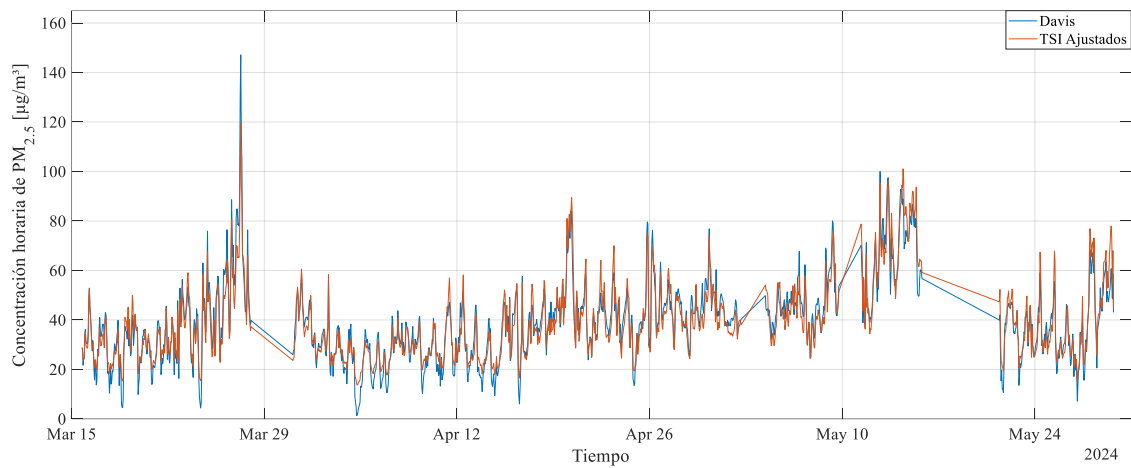


Figura 3.19 Valores de concentración horaria de  $PM_{2.5}$  (TSI) ajustados mediante la ecuación de regresión lineal (elaboración propia).

### Evaluación de la efectividad de la calibración

Se calcularon las medidas de error para evaluar la relación existente entre los valores calibrados de los métodos de medición del sensor *AM1008W*, con respecto a los promedios horarios obtenidos por el dispositivo *Davis AirLink*. En la Tabla 3.10 se muestran los valores de los parámetros de evaluación de la relación.

Tabla 3.10. Parámetros de evaluación de la relación existente entre los valores de concentración de  $PM_{2.5}$  obtenidos por los métodos de medición *TSI*, *GRIMM* en el período de estudio (elaboración propia).

Relación de datos	Coefficiente de correlación	$R^2$	MAE	RMSE	MAPE
<i>GRIMM - Davis</i>	0.936	0.875	4.377	5.765	15.67%
<i>TSI - Davis</i>	0.957	0.916	3.672	4.739	13.53%

Como resultado se identificó que los valores de concentración horaria obtenidos por el método *TSI*, ajustados mediante regresión lineal presentan el menor valor de MAPE, por

lo tanto, estos valores son los que más se asemejan con los valores de concentración horaria obtenidos con el dispositivo *Davis AirLink*. Por lo tanto, se seleccionó este método de medición para mostrar los valores de concentración horaria medidos por el sistema desarrollado y realizar el cálculo del índice de Aire y Salud.

### 3.2.3 Visualización remota

Con el objetivo de posibilitar la visualización remota en tiempo real de los parámetros obtenidos por el sistema: promedio horario de concentración de  $PM_{2.5}$ , clasificación de la calidad del aire y nivel de riesgo asociado según el Índice Aire y Salud para este contaminante, y recomendaciones correspondientes, se envían estos datos a la plataforma *Initial State*. También son transferidos sus indicadores de funcionamiento: temperatura de operación y velocidad de rotación del ventilador del sensor de material particulado. Además, se desarrolló una aplicación Android que se conecta con esta plataforma y permite la visualización selectiva de la información recopilada. En las Figuras 3.20 y Anexo 12, se presenta la interfaz gráfica del sistema, donde se visualizan los valores de concentración de contaminantes medidos y los indicadores de funcionamiento del sistema, respectivamente.



Figura 3.20. Interfaz gráfica del sistema de monitorización de la calidad del aire: valores medidos (elaboración propia).

### 3.3 Sistema monitorización móvil *Flatburn*

El sistema de monitorización móvil *Flatburn* (Figura 3.21) constituye una plataforma de detección modular de código abierto, desarrollado por el *MIT Senseable City Lab*, que permite su implementación en vehículos para medir valores de humedad, temperatura ambiental, concentración de PM<sub>1</sub>, PM<sub>2.5</sub>, PM<sub>10</sub>, NO<sub>2</sub>, CO y SO<sub>2</sub> en movilidad (DeSouza et al., 2020). En la Tabla 3.11 se presentan los sensores que emplea este dispositivo.

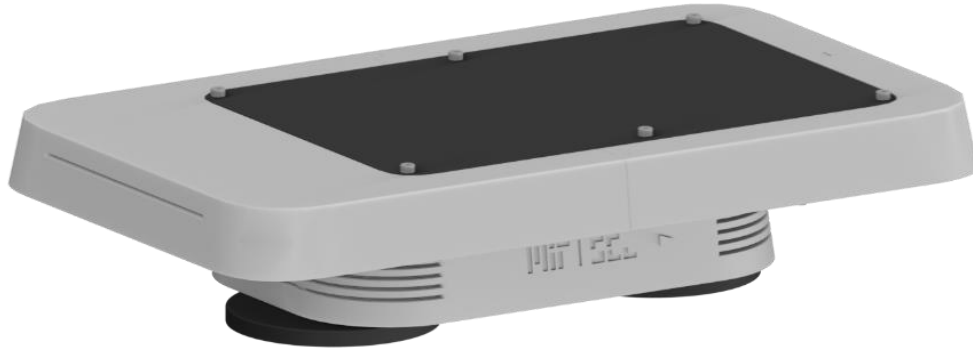


Figura 3.21. Sistema de monitorización móvil *Flatburn* (Senseable City Lab, 2024).

Tabla 3.11. Sensores de calidad del aire del sistema de monitorización *Flatburn* (Senseable City Lab, 2024).

Sensor	Medición de variables	Fabricante
SPS30	PM <sub>1</sub> , PM <sub>2.5</sub> , PM <sub>10</sub>	<i>Sensirion</i>
CO-A4F	CO	<i>Alphasense</i>
NO <sub>2</sub> -A43F	NO <sub>2</sub>	<i>Alphasense</i>
SO <sub>2</sub> -A4	SO <sub>2</sub>	<i>Alphasense</i>
BME280	Humedad, Temperatura	<i>Bosch</i>

#### 3.3.1 Implementación del sistema *Flatburn*

Se realizaron campañas de recolección de datos con el sistema de monitorización móvil de calidad del aire *Flatburn*, durante los meses de abril - noviembre del 2023. El dispositivo se instaló sobre el techo de autobuses de transporte de pasajeros de la ruta La Raza – ESCOM, durante períodos de alrededor de 6 horas (Anexo 13 y 14). En el análisis de los valores de concentración de contaminantes atmosféricos obtenidos se empleó el software *QGIS*.

#### *Preprocesamiento de los datos del Flatburn*

Se recuperaron los datos crudos obtenidos por dispositivo *Flatburn* durante las campañas de monitorización, a través de la tarjeta SD. A partir de estos se filtraron los datos de concentración de PM<sub>2.5</sub> para su posterior preprocesamiento. En la Tabla 3.12 se presentan

los parámetros del conjunto de datos de concentración de PM<sub>2.5</sub> obtenidos del sistema *Flatburn*.

Tabla 3.12. Parámetros del conjunto de datos de concentración de PM<sub>2.5</sub> obtenidos del sistema *Flatburn*. (elaboración propia).

Parámetro	Descripción
Estado	Estado del sensor (0: Registro válido, 1: Registro inválido, 2: Fuera de operación)
ID	Identificador del dispositivo
Hora	Tiempo en formato Unix
Latitud	Latitud en grados con respecto al Ecuador
Longitud	Longitud en grados con respecto al meridiano de Greenwich
PM <sub>2.5</sub>	Concentración de partículas PM <sub>2.5</sub> en µg/m <sup>3</sup> .

En el proceso de preprocesamiento primeramente se convirtieron los datos de la columna “Hora” a formato de hora local de la CDMX. Luego se eliminaron las tuplas que presentaban valores diferentes a 0 en el campo “Estado”. Como resultado se obtuvo una base de datos de 23772 tuplas. Se identificó que la frecuencia de muestreo de registros consecutivos presentó un rango de 1-4 segundos durante períodos de correcto funcionamiento del dispositivo. A continuación, se presentan los procesos de identificación de datos faltantes, valores atípicos y análisis de existencia de datos con ruido.

### ***Identificación de datos faltantes***

Se identificaron un total de 750 valores faltantes de PM<sub>2.5</sub> y 1050 datos faltantes de ubicación en la base de datos, que representan en total un 7.57% del total de tuplas válidas. En la Tabla 3.13 se muestra el desglose de los datos faltantes. Estas tuplas resultaron eliminadas de la base de datos, como resultado se redujo la cantidad de tuplas a un total de 21972.

Tabla 3.13. Relación de datos faltantes de concentración de PM<sub>2.5</sub> obtenidos por el dispositivo *Flatburn* (elaboración propia).

	Datos faltantes	Total de registros válidos	% de registros faltantes con respecto a los validos
Latitud y Longitud(= 0)	1050	23772	4.42%
PM25	750		3.15%
Total	1800		7.57%

### Identificación de datos atípicos

En el proceso de identificación de datos atípicos, se obtuvo el diagrama de caja de los valores de concentración de  $PM_{2.5}$ , para identificar la existencia de los valores atípicos ( $> 44.74 \mu g/m^3$ ). Como resultado se identificó un total de 1195 valores atípicos. En la Figura 3.22 se presenta el diagrama de caja obtenido.

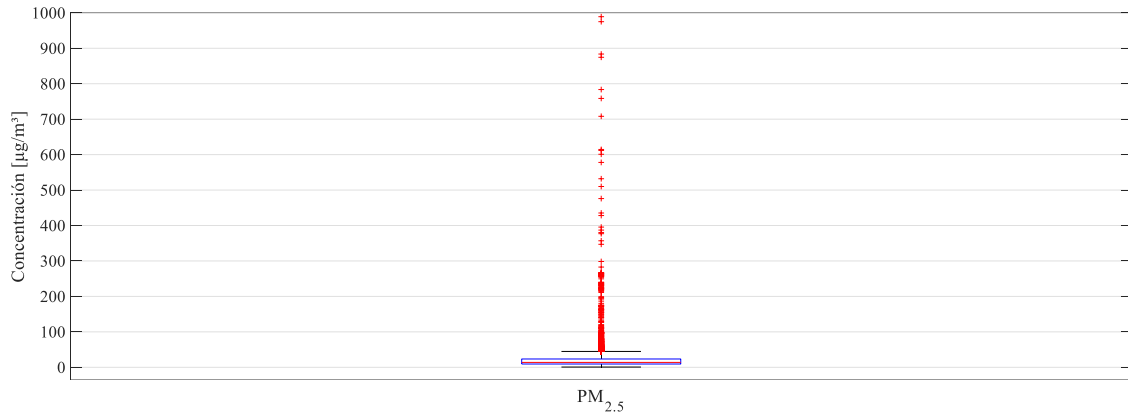


Figura 3.22. Diagrama de caja de los valores de concentración de  $PM_{2.5}$  obtenidos por el dispositivo *Flatburn* durante el período abril-noviembre del 2023 (elaboración propia).

### Identificación de valores con ruido

En la identificación de los datos con ruido de concentración de  $PM_{2.5}$ , primeramente, se eliminaron los valores de  $PM_{2.5}$  iguales a 0, ya que constituye el límite inferior del dispositivo y el dispositivo emite los valores de concentración con una resolución de 2 dígitos decimales y no está respaldado por las condiciones del ambiente de aplicación. También se descartaron los valores de  $PM_{2.5}$  superiores al límite máximo de medición, que es de  $1000 \mu g/m^3$ . Como resultado se redujo la cantidad de tuplas del conjunto de datos a un total de 21268.

Como resultado de la aplicación del algoritmo de análisis temporal, se identificaron 55 valores con ruido que resultaron eliminados del conjunto de datos. Por lo tanto, este se redujo a un total de 21213 tuplas. En la Figura 3.23 se muestran los valores de concentración de  $PM_{2.5}$  con ruido identificados mediante el algoritmo de análisis temporal. En la Tabla 3.14 se presenta la relación de datos con ruido identificados en el conjunto de datos.

Tabla 3.14. Valores de concentración de  $PM_{2.5}$  con ruido, obtenidos por el dispositivo *Flatburn* en el período de estudio (elaboración propia).

Valores = 0	Valores > 1000(rango de medición)	Algoritmo
636	58	55

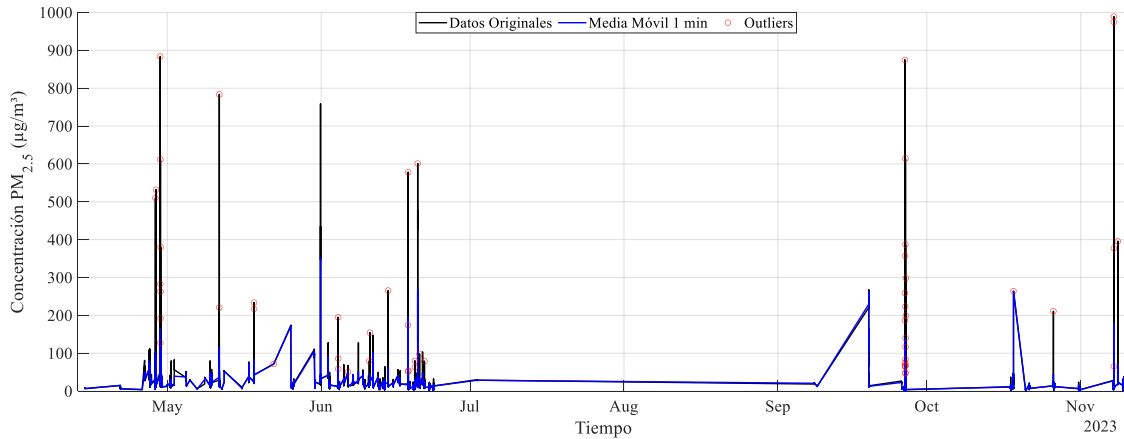


Figura 3.23. Valores de concentración de  $PM_{2.5}$  con ruido identificados mediante el algoritmo de promedio móvil (elaboración propia).

### 3.4 Metodología de predicción del índice y concentración de $O_3$ en la zona norte de la CDMX

El conjunto de datos de concentración de contaminantes atmosféricos y datos meteorológicos empleados en el entrenamiento de los modelos de clasificación se obtuvo de la estación de monitoreo atmosférico GAM, que también presenta una estación de monitoreo meteorológico. La obtención de los datos históricos de concentración de contaminantes atmosféricos medidos por esta estación se realizó a través de la página web del INECC, mientras que los valores de las variables meteorológicas se recuperaron de a través del sitio web del SIMAT.

#### 3.4.1 Parámetros del conjunto de datos

La base de datos se conforma de la integración de datos de concentración horaria de  $O_3$ , variables meteorológicas: temperatura, humedad, dirección y velocidad del viento, obtenidos por la estación GAM, durante los años 2015-2023, y la clasificación de la calidad del aire con respecto al Índice de Aire y Salud para el promedio horario de  $O_3$ . A continuación, se describen los atributos que conforman la base de datos:

- Mes: Atributo nominal. Se registra el mes en que se obtuvo la medición de la concentración de  $O_3$ .
- Día de la semana: Atributo nominal. Se registra el día de la semana en que se obtuvo la medición de la concentración de  $O_3$ .
- Hora: Atributo numérico. Se registra la hora en los que se obtuvo la medición de la concentración de  $O_3$ , con resolución de una hora.



- Temperatura: Atributo numérico. Se registra el promedio horario de temperatura ( $^{\circ}\text{C}$ ) existente en el momento que se obtuvo la medición de la concentración de  $\text{O}_3$ .
- Humedad relativa: Atributo numérico. Se registra el promedio horario de humedad relativa (%), existente en el momento que se obtuvo la medición de la concentración de  $\text{O}_3$ .
- Dirección del viento: Atributo numérico. Se registra el promedio horario de dirección del viento ( $^{\circ}$ ), existente en el momento que se obtuvo la medición de la concentración de  $\text{O}_3$ .
- Velocidad del viento: Atributo numérico. Se registra el promedio horario de velocidad del viento (m/s), existente en el momento que se obtuvo la medición de la concentración de  $\text{O}_3$ .
- Concentración de  $\text{O}_3$ : Atributo numérico. Se registra el promedio horario de concentración de  $\text{O}_3$  en ppb.
- Clasificación de  $\text{O}_3$ : Atributo nominal. Se registra la clasificación del Índice Aire Salud para el promedio de concentración horaria de  $\text{O}_3$ .

### 3.4.2 Implementación de los modelos de predicción

Se aplicaron los algoritmos de clasificación: *C4.5*, *Naive Bayes*, *Bayesian Belief Networks*, *Backpropagation*, *K nearest-neighbor (Knn)*, *Support Vector Machines (SVM)* al conjunto de datos obtenido de la estación GAM, donde se desea predecir el índice y el valor de la concentración de  $\text{O}_3$ , a partir de los valores actuales de: mes, día de la semana, horario, temperatura, humedad relativa, velocidad y dirección del viento.

En la implementación de los algoritmos de clasificación se emplea el software Weka, que presenta un entorno de trabajo que posibilita la implementación de diferentes algoritmos para aprendizaje automático. La versión actual de Weka es software libre, publicado bajo la Licencia Pública General GNU, está implementado en lenguaje Java y soporta aplicaciones para Big Data (Weka, 2023).

En el entrenamiento de los modelos de predicción se empleó un *batch size* = 200. Como esquema de validación se seleccionó el método *Cross Validation*, con una cantidad de divisiones ( $k$ ) igual a 10. Con esta técnica se divide la base de datos en  $k$  conjuntos, para cada conjunto de validación se entrena un modelo con las observaciones del conjunto de entrenamiento, y posteriormente se evalúa el rendimiento del modelo con los datos de validación. Luego se calcula el promedio del error de validación de todas las iteraciones.

Este método proporciona una estimación adecuada de la exactitud en la predicción del modelo final entrenado con todos los datos. Requiere múltiples ajustes, pero realiza un uso eficiente de todos los datos, por lo que se recomienda para conjuntos de datos pequeños (Lei, 2020).

En las Tablas 3.15 se presentan los valores de los hiperparámetros empleados durante el entrenamiento de los modelos mediante el algoritmo *C4.5*, *Bayesian Belief Networks*, *Backpropagation*, *Knn* y *SVM*, respectivamente.

Tabla 3.15. Hiperparámetros seleccionados para el entrenamiento de los modelos (elaboración propia).

Modelos	Parámetros	Valor
<i>C4.5</i>	Factor de confianza	0.25
	Mínimo número de instancias por hoja	2
<i>Bayesian Belief Networks</i>	Estimador	simple
	Algoritmo de búsqueda	<i>K2</i>
<i>Backpropagation</i>	# capas ocultas	(#atributos + #clases)/2
	<i>Momentum</i>	0.9
	<i>Learning rate</i>	0.1
<i>Knn</i>	Algoritmo de búsqueda	<i>KDTree</i>
	<i>DistanceWeighting</i>	<i>Weight by 1-distance</i>
<i>SVM</i>	Calibrador	<i>Logistic</i>
	<i>kernel</i>	<i>PolyKernel</i>

### 3.5 Consideraciones finales del capítulo

1. Se preprocesaron los datos de concentración de O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> y NO<sub>2</sub> obtenidos por la estación Camarones y Gustavo A Madero durante los años 2022-2023, donde se identificó la cantidad de datos faltantes y valores de concentración atípicos de cada contaminante, y se determinó conservar los valores atípicos para el proceso de análisis, al encontrarse en el rango de valores históricos obtenidos en la zona.
2. Se desarrolló un sistema fijo de monitorización remota de la calidad del aire, que presenta un *Raspberry Pi 4 Model B* como dispositivo de control y adquisición de datos, y posibilita la adquisición de los valores de concentración de PM<sub>2.5</sub> medidos por un sensor *AM1008W*, con una resolución de 1 min, y el cálculo del Índice de Aire y Salud para este contaminante. Además, transmite los valores obtenidos a la plataforma *Initial State* en tiempo real.

3. Se corroboró la necesidad de comprobar la precisión de los sensores de calidad del aire de bajo costo, y por lo tanto, la importancia de aplicar métodos de calibración a los sensores de calidad del aire de bajo costo con el objetivo de obtener mediciones precisas de concentración de contaminantes atmosféricos.
4. Se calibró el sistema desarrollado mediante el método de regresión lineal, a partir de los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por el sistema y un dispositivo *Davis AirLink* en la Isla de Datos Urbanos de la ESCOM por un período de 75 días continuos (15 de marzo-29 de junio del 2023). Como resultado se redujo el porcentaje medio de error absoluto (MAPE) de 28.97% a 13.53% entre los promedios horarios obtenidos por el método de medición *TSI* del sensor *AM1008W* y los obtenidos por el dispositivo *Davis AirLink*, que corroboran la elevada relación que existe entre los datos.
5. Se desarrolló una aplicación móvil que posibilita la visualización selectiva y en tiempo real de la concentración de  $PM_{2.5}$ , el promedio horario de concentración de  $PM_{2.5}$  y la clasificación del Índice de Aire y Salud para este contaminante, a partir de su conexión con la plataforma *Initial State*.
6. Se identificaron los parámetros de la base de datos a emplear en el entrenamiento de los modelos de predicción de la clasificación y el valor de la concentración de  $O_3$  en la zona norte de la CDMX.
7. Se establecieron los hiperparámetros de modelos a emplear en la predicción de la clasificación y el valor de la concentración de  $O_3$  en la zona norte de la CDMX.

## Capítulo 4. Resultados

En el presente capítulo se presenta un análisis estadístico de los datos preprocesados de concentración de contaminantes atmosféricos de la estación Camarones y Gustavo A Madero de los años 2022 - 2023 y de los datos obtenidos durante las campañas de monitorización realizadas con el sistema de monitorización móvil *Flatburn*. También se realiza un análisis de la variabilidad espacial de la concentración de  $PM_{2.5}$  en la zona norte de la CDMX, a partir de la comparación de los promedios horarios de concentración de este contaminante obtenidos por el sistema de monitorización fijo desarrollado y el dispositivo *Flatburn*, con las estaciones de la RAMA más cercanas a las ubicaciones donde se realizaron las mediciones.

### 4.1 Análisis estadístico valores históricos de las estaciones

A continuación, se presenta un análisis estadístico de los datos preprocesados de concentración de cada uno de los contaminantes atmosféricos de la estación CAM y GAM en los años 2022 y 2023.

#### 4.1.1 Ozono

En la tabla 4.1 se muestran los parámetros estadísticos obtenidos para los valores de concentración de  $O_3$ , que indican la existencia de un sesgo positivo en los datos. Además, se identificaron los mayores promedios de concentración en la estación GAM en cada año, con valores de 56.99 y 58.87  $\mu g/m^3$  para los años 2022 y 2023 respectivamente. En la Figura 4.1 se muestran los diagramas de caja de los valores de concentración de ozono.

#### 4.1.2 $PM_{10}$ y $PM_{2.5}$

Con respecto al análisis estadístico de los valores de  $PM_{10}$  y  $PM_{2.5}$  se identificaron los mayores promedios de concentración en la estación GAM en cada año, con valores de concentración de  $PM_{10}$  de 50.67 y 45.54  $\mu g/m^3$  para los años 2022 y 2023 respectivamente, y valores de  $PM_{2.5}$  de 23.95 y 21.81  $\mu g/m^3$  para los años 2022 y 2023

respectivamente. Se identificó un descenso en los valores promedios de concentración en las dos estaciones en el año 2023 con respecto a los datos del año anterior. En las Figuras 4.2 y 4.3 se muestran los diagramas de caja de los valores de concentración de  $PM_{10}$  y  $PM_{2.5}$  respectivamente. En las Tablas 4.2 y 4.3 se presentan los parámetros estadísticos de la concentración de  $PM_{10}$  y  $PM_{2.5}$  respectivamente.

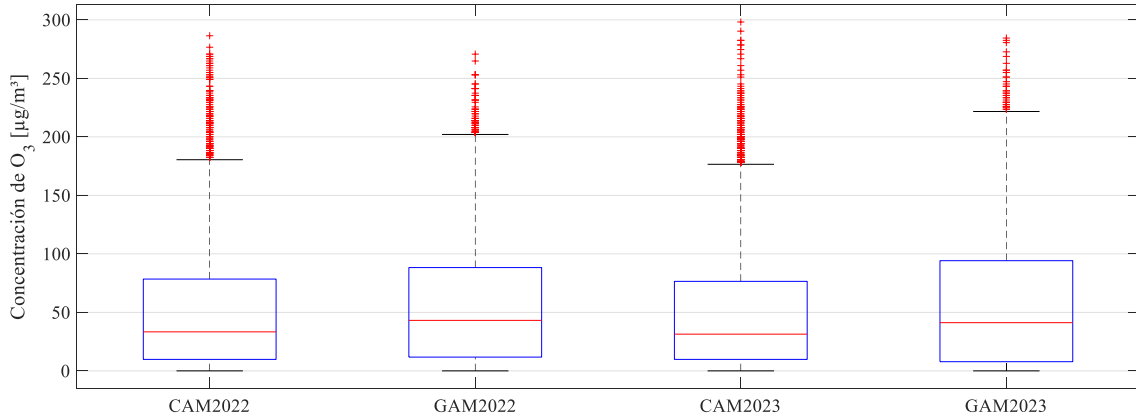


Figura 4.1. Diagrama de caja de los valores de concentración de  $O_3$  de las estaciones CAM y GAM en los años 2022 y 2023 (elaboración propia).

Tabla 4.1 Descripción estadística de los datos preprocesados de  $O_3$  de la estación CAM y GAM en los años 2022 y 2023 (elaboración propia).

Año	Est*	%DV**	Máx [ $\mu g/m^3$ ]	Mín [ $\mu g/m^3$ ]	Prom. [ $\mu g/m^3$ ]	Mediana [ $\mu g/m^3$ ]	Moda [ $\mu g/m^3$ ]	Desv. Est. [ $\mu g/m^3$ ]
2022	CAM	75.18%	286.42	0	51.57	33.35	1.96	52.70
	GAM	93.57%	270.73	0	56.99	43.16	3.92	51.34
2023	CAM	72.11%	298.19	0	50.47	31.38	1.96	52.49
	GAM	65.18%	284.460	0	58.87	41.20	3.92	57.55

\*Estación; \*\* Porcentaje de datos válidos.

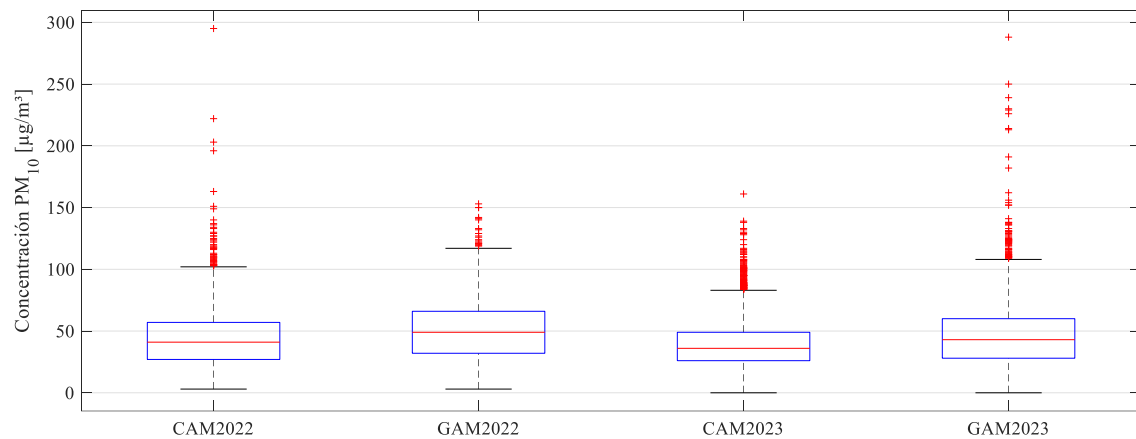


Figura 4.2. Diagrama de caja de los valores de concentración de  $PM_{10}$  de las estaciones CAM y GAM en los años 2022 y 2023 (elaboración propia).

Tabla 4.2 Descripción estadística de los datos preprocesados de  $PM_{10}$  de la estación CAM y GAM en los años 2022 y 2023 (elaboración propia).

Año	Est*	%DV**	Máx [ $\mu\text{g}/\text{m}^3$ ]	Mín [ $\mu\text{g}/\text{m}^3$ ]	Prom [ $\mu\text{g}/\text{m}^3$ ]	Mediana [ $\mu\text{g}/\text{m}^3$ ]	Moda [ $\mu\text{g}/\text{m}^3$ ]	Desv. Est. [ $\mu\text{g}/\text{m}^3$ ]
2022	CAM	44.82%	295	3	44.27	41	42	22.77
	GAM	24.17%	153	3	50.67	49	36	24.07
2023	CAM	79.28%	161	0	38.88	36	28	18.70
	GAM	63.36%	288	0	45.54	43	44	24.66

\*Estación; \*\* Porcentaje de datos válidos.

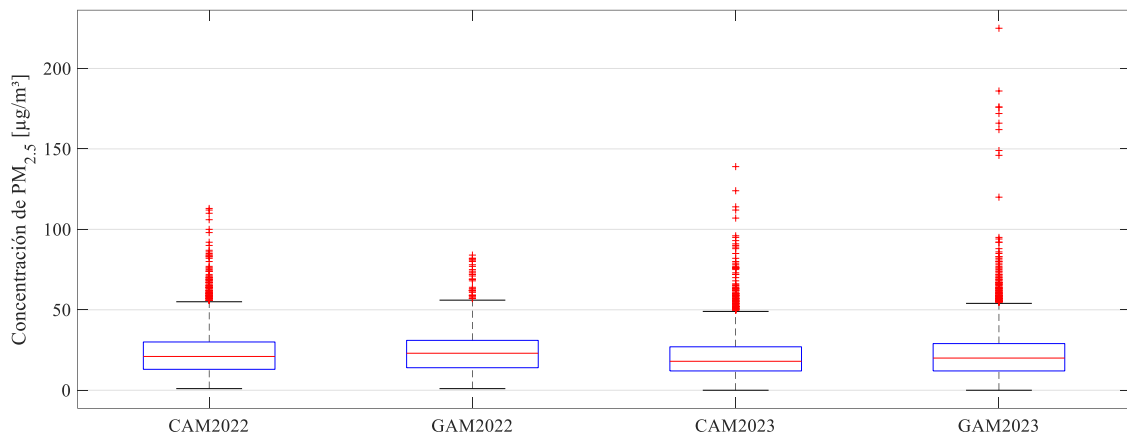


Figura 4.3. Diagrama de caja de los valores de concentración de  $PM_{2.5}$  de las estaciones CAM y GAM en los años 2022 y 2023 (elaboración propia).

Tabla 4.3 Descripción estadística de los datos preprocesados de  $PM_{2.5}$  de la estación CAM y GAM en los años 2022 y 2023 (elaboración propia).

Año	Est*	%DV**	Máx. [ $\mu\text{g}/\text{m}^3$ ]	Mín. [ $\mu\text{g}/\text{m}^3$ ]	Media [ $\mu\text{g}/\text{m}^3$ ]	Mediana [ $\mu\text{g}/\text{m}^3$ ]	Moda [ $\mu\text{g}/\text{m}^3$ ]	Desv. Est. [ $\mu\text{g}/\text{m}^3$ ]
2022	CAM	44.82%	113	1	22.93	21	12	13.64
	GAM	24.17%	84	1	23.95	23	26	13.13
2023	CAM	78.72%	139	0	20.45	18	17	11.69
	GAM	62.41%	225	0	21.83	20	18	14.51

\*Estación; \*\* Porcentaje de datos válidos.

### 4.1.3 $NO_2$

Con respecto al análisis estadístico de los valores de  $NO_2$  se identificaron los mayores promedios de concentración en la estación CAM en cada año, con valores de 51.76 y 54.62  $\mu\text{g}/\text{m}^3$  para los años 2022 y 2023 respectivamente. Se identificó un aumento en los valores promedios de concentración horaria en las dos estaciones en el año 2023 con

respecto a los datos del año anterior. En la Figura 4.4 se muestra el diagrama de caja de los valores de concentración de NO<sub>2</sub>. En la Tabla 4.4 se presentan los parámetros estadísticos de este contaminante.

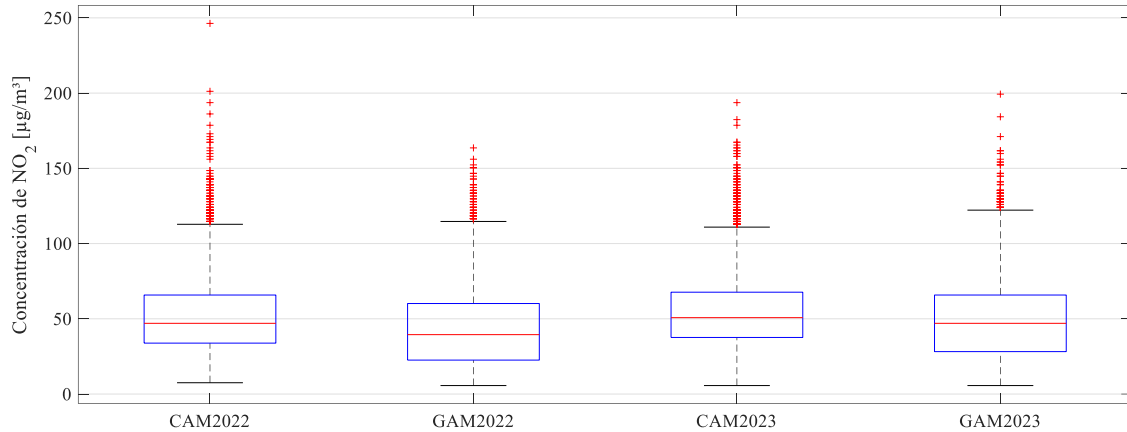


Figura 4.4. Diagrama de caja de los valores de concentración de NO<sub>2</sub> de las estaciones CAM y GAM en los años 2022 y 2023 (elaboración propia).

Tabla 4.4 Descripción estadística de los datos preprocesados de NO<sub>2</sub> de la estación CAM y GAM en los años 2022 y 2023 (elaboración propia).

Año	Est*	%DV**	Máx. [µg/m³]	Mín. [µg/m³]	Media [µg/m³]	Mediana [µg/m³]	Moda [µg/m³]	Desv. Est. [µg/m³]
2022	CAM	72.83%	246.32	7.52	51.76	47.01	43.24	24.74
	GAM	93.86%	163.58	5.64	43.42	39.49	13.16	25.23
2023	CAM	73.52%	193.66	5.64	54.62	50.77	47.01	24.58
	GAM	65.63%	199.31	5.64	49.34	47.01	22.56	26.81

## 4.2 Evaluación del cumplimiento de las NOM

Se evaluó el cumplimiento de las NOM de calidad del aire en la zona norte de la CDMX durante los años 2022-2023, donde se emplearon en el análisis los datos de concentración de contaminantes atmosféricos obtenidos por las estaciones Camarones y Gustavo A Madero en este período y se verificaron los criterios de existencia de datos para cada una de las normas. En el análisis de los conjuntos de datos se empleó el sistema de gestión de base de datos relacional Microsoft SQL Server.

### 4.2.1 Ozono

#### *Datos horarios*

Se analizan los promedios horarios de concentración de O<sub>3</sub> medidos por las estaciones en los años 2022-2023, para evaluar el cumplimiento de la NOM para este indicador. El

valor límite de concentración de  $O_3$  se incumplió en las estaciones de monitoreo en una cantidad de horas que varía de 190 a 271. Se identificó un aumento en el porcentaje de horas donde se incumplió esta norma en el año 2023, en las dos estaciones, en comparación con los datos obtenidos en el año 2022. En la Tabla 4.5 se presenta un análisis de los datos horarios de concentración de  $O_3$  de las estaciones CAM y GAM en los años 2022-2023.

Tabla 4.5. Análisis de datos horarios de concentración de  $O_3$  de las estaciones CAM y GAM en los años 2022-2023 (elaboración propia).

Año	Estación	No. horas con datos válidos	%. Horas con datos válidos	No. Horas > 176 $\mu\text{g}/\text{m}^3$	% Horas > 176 $\mu\text{g}/\text{m}^3$
2022	CAM	6586	75.18%	190	2.88%
	GAM	8197	93.57%	195	2.38%
2023	CAM	6317	72.11%	197	3.12%
	GAM	5710	65.18%	271	4.75%

#### ***Promedios horarios de 8h***

En el análisis de los promedios horarios de 8 horas se obtuvo un mayor porcentaje de días de incumplimiento de la NOM en la estación GAM, con un 32.11% para el año 2022 y 46.18%, para el 2023. Además, se identificó una disminución en el porcentaje de días que se incumplió la NOM en la estación CAM en el año 2023, en comparación con los registros de la estación en el año 2022. Mientras que se identificó un aumento en el porcentaje de días en los que se superó el valor del promedio móvil de  $O_3$  en la estación GAM, en comparación con los registros de la estación en el año 2022. En la Tabla 4.6 se presenta un análisis de los promedios móviles de concentración de  $O_3$  de las estaciones CAM y GAM en los años 2022-2023.

Tabla 4.6 Análisis de los promedios móviles de 8h de  $O_3$  de las estaciones CAM y GAM en los años 2022 – 2023 (elaboración propia).

Año	Estación	No. días con datos válidos	%. Días con datos válidos	No. días ProM8h > 118 $\mu\text{g}/\text{m}^3$	% de días ProM8h > 118 $\mu\text{g}/\text{m}^3$
2022	CAM	287	78.63%	80	27.87%
	GAM	355	97.26%	114	32.11%
2023	CAM	277	75.89%	64	23.10%
	GAM	249	68.22%	115	46.18%

\* Porcentaje de días en los que se supera el límite de promedio móvil de  $O_3$ , con respecto a la cantidad de días con datos válidos.



#### 4.2.2 PM<sub>10</sub>

##### *Promedio de 24 horas*

En la evaluación del cumplimiento de la NOM para el promedio de 24h, primeramente, se verifica el criterio de suficiencia de datos para este indicador. Se identificó un mayor porcentaje de incumplimiento del límite diario en la estación GAM. Se identificó una reducción en el porcentaje de días de incumplimiento del valor límite de 24 h en los años 2022 y 2023 en las dos estaciones. En la Tabla 4.7 se presenta un análisis del cumplimiento del promedio de 24h para la concentración de PM<sub>10</sub> en las estaciones en los años 2022-2023.

Tabla 4.7. Evaluación del cumplimiento del promedio de 24h para la concentración de PM<sub>10</sub> en las estaciones en los años 2022-2023 (elaboración propia).

Año	Estación	No. días con datos válidos	% días con datos válidos	No. Días > 70 $\mu\text{g}/\text{m}^3$	% Días > 70 $\mu\text{g}/\text{m}^3$
2022	CAM	160	43.84%	6	3.75%
	GAM	87	23.84%	14	16.09%
2023	CAM	287	78.63%	3	1.05%
	GAM	230	63.01%	12	5.22%

##### *Promedio anual*

En la revisión del requisito de suficiencia de datos para evaluar NOM anual de PM<sub>10</sub> se identificó en el período de estudio, la única estación que presentó la cantidad de días con datos válidos para evaluar la NOM anual de PM<sub>10</sub> fue la estación CAM en el año 2023, que incumplió la NOM al obtenerse un promedio de valores de concentración anual igual 39 en  $\mu\text{g}/\text{m}^3$ , que es superior al límite anual normado para este contaminante (36  $\mu\text{g}/\text{m}^3$ ). Aunque los datos recuperados por la estación GAM y los obtenidos por la estación GAM en el año 2022 no cumplen el criterio de suficiencia de datos para evaluar la norma anual, se destaca la obtención de promedios anuales superiores al límite normado. En el análisis temporal se apreció una disminución del promedio anual de concentración de PM<sub>10</sub>, medidos por cada una de las estaciones en los años 2022 y 2023. En la Tabla 4.8 se presenta la evaluación de la NOM anual para la concentración de PM<sub>10</sub> en las estaciones en los años 2022-2023.

Tabla 4.8 Evaluación de la NOM anual de PM<sub>10</sub> (elaboración propia).

Año	Est*	Media 1er T [μg/m <sup>3</sup> ]	% ** 1er T [%]	Media 2do T [μg/m <sup>3</sup> ]	% ** 2do T [%]	Media 3er T [μg/m <sup>3</sup> ]	% ** 3er T [%]	Media 4to T [μg/m <sup>3</sup> ]	% ** 4to T [%]	Media anual μg/m <sup>3</sup>
2022	CAM	51	43	51	25	32	59	50	48	44
	GAM	-	0.0	-	0.0	-	0.0	51	95	51
2023	CAM	49	93	39	42	30	90	38	89	39
	GAM	55	99	48	60	25	8	36	86	46

\*Estación; \*\*Porcentaje días que se cumplió el criterio de existencia de datos por trimestre.

### 4.2.3 PM<sub>2.5</sub>

#### *Promedio de 24 horas*

En la evaluación del cumplimiento de la NOM para el promedio de 24 horas, primeramente, se verifica el criterio de suficiencia de datos para este indicador. Se identificó una reducción en el porcentaje de días de incumplimiento del valor límite de 24 horas en los años 2022 y 2023 en las dos estaciones. En la Tabla 4.9 se presenta la evaluación del límite de 24 horas para la concentración de PM<sub>2.5</sub> en las estaciones en los años 2022-2023.

Tabla 4.9. Evaluación del promedio de 24h para el PM<sub>2.5</sub> en las estaciones en los años 2022-2023 (elaboración propia).

Año	Estación	No. días con datos válidos	%. días con datos válidos	No. Días > 41 μg/m <sup>3</sup>	% Días > 41 μg/m <sup>3</sup>
2022	CAM	160	43.84%	3	1.88%
	GAM	87	23.84%	0	0.00%
2023	CAM	286	78.36%	3	1.05%
	GAM	224	61.37%	3	1.34%

#### *Promedio anual*

En la revisión del requisito de suficiencia de datos para evaluar NOM anual de PM<sub>2.5</sub> se identificó en el período de estudio, la única estación que presentó la cantidad de días con datos válidos para evaluar la NOM anual de PM<sub>2.5</sub> fue la estación CAM en el año 2023, que incumplió la NOM al obtenerse un promedio de valores de concentración anual igual 21 en μg/m<sup>3</sup>, que es superior al límite anual normado para este contaminante (10 μg/m<sup>3</sup>). Aunque los datos recuperados por la estación GAM y los obtenidos por la estación GAM en el año 2022 no cumplen el criterio de suficiencia de datos para evaluar la norma anual

(75 % de registros validos), se destaca la obtención de promedios anuales superiores al límite normado. En el análisis temporal se apreció una disminución de los valores de promedio anual de valores de  $PM_{2.5}$ , medidos por cada una de las estaciones en los años 2022 y 2023. En la Tabla 4.10 se presenta la evaluación del límite anual para la concentración de  $PM_{2.5}$  en las estaciones en los años 2022-2023.

Tabla 4.10. Evaluación de la NOM anual de  $PM_{2.5}$  (elaboración propia).

Año	Est*	Media 1er T [ $\mu g/m^3$ ]	% ** 1er T [%]	Media 2do T [ $\mu g/m^3$ ]	% ** 2do T [%]	Media 3er T [ $\mu g/m^3$ ]	% ** 3er T [%]	Media 4to T [ $\mu g/m^3$ ]	% ** 4to T [%]	Media anual $\mu g/m^3$
2022	CAM	22	43	26	25	18	59	28	48	23
	GAM	-	0.0	-	0.0	-	0.0	24	95	24
2023	CAM	24	93	21	42	16	90	22	88	21
	GAM	24	99	24	60	12	8	21	79	22

\*Estación; \*\*Porcentaje días que se cumplió el criterio de existencia de datos por trimestre.

### 4.2.3 $NO_2$

#### *Datos horarios y promedios anuales*

En el análisis de  $NO_2$  se identificó que solamente se incumplió el límite horario de este contaminante en la estación Camarones, en dos ocasiones durante el año 2022. Con respecto al análisis de los promedios anuales, solamente la estación GAM en el año 2022 cumplió con el requisito de suficiencia de datos, y el valor de promedio anual obtenido incumple la norma de concentración límite de este contaminante ( $40 \mu g/m^3$ ). Aunque en los demás casos no se cumplió con el criterio de suficiencia de datos, se destaca que los promedios anuales obtenidos poseen valores superiores al límite de concentración normado de  $NO_2$ . En la Tabla 4.11 se presenta la evaluación del límite anual para la concentración de  $PM_{2.5}$  en las estaciones en los años 2022-2023.

Tabla 4.11. Análisis de datos horarios y promedio anual de  $NO_2$  (elaboración propia).

Año	Estación	No. horas con datos válidos	% . Horas con datos válidos	No. Horas > 200 $\mu g/m^3$	% Horas > 200 $\mu g/m^3$	Promedio anual
2022	CAM	6380	72.83%	2	0.03%	27.5
	GAM	8222	93.86%	0	0.00%	23.1
2023	CAM	6440	73.52%	0	0.00%	29.1
	GAM	5749	65.63%	0	0.00%	26.2

### 4.3 Tendencias de contaminantes atmosféricos en el tiempo

Se obtuvieron tendencias actuales del comportamiento de la concentración de los contaminantes atmosféricos con respecto al tiempo en la zona de estudio, donde se obtuvo el promedio de concentración horaria de  $O_3$ ,  $PM_{10}$ ,  $PM_{2.5}$  y  $NO_2$  por años, meses, días de la semana y horas, mediante el empleo del sistema de gestión de base de datos relacional *Microsoft SQL Server*. Estos valores promedios se compararon con las tendencias históricas de cada una de las estaciones, definidas para los años 2015-2022. A continuación, se presentan los resultados obtenidos.

#### Ozono

Con respecto al análisis de las tendencias de concentración de  $O_3$  en la zona de estudio se identificó que históricamente (2015 - 2022) los promedios anuales obtenidos por la estación GAM son superiores a los medidos por la estación CAM. Además, los meses en los que históricamente se han obtenido los mayores promedios de concentración horaria en la zona de estudio son abril y mayo. Por otro lado, los días en los que históricamente se han obtenido los mayores promedios de concentración horaria son los sábados y domingos. Mientras que las horas en las que históricamente se han obtenido los mayores promedios de concentración horaria son las 2 – 3 pm. Estas tendencias históricas se mantuvieron durante el año 2023. En las Figuras 4.5 – 4.8 se presentan las tendencias de los valores promedios de concentración de  $O_3$  en cada una de las escalas de tiempo, los valores empleados para obtener las gráficas se presentan en los Anexos 15 – 18.

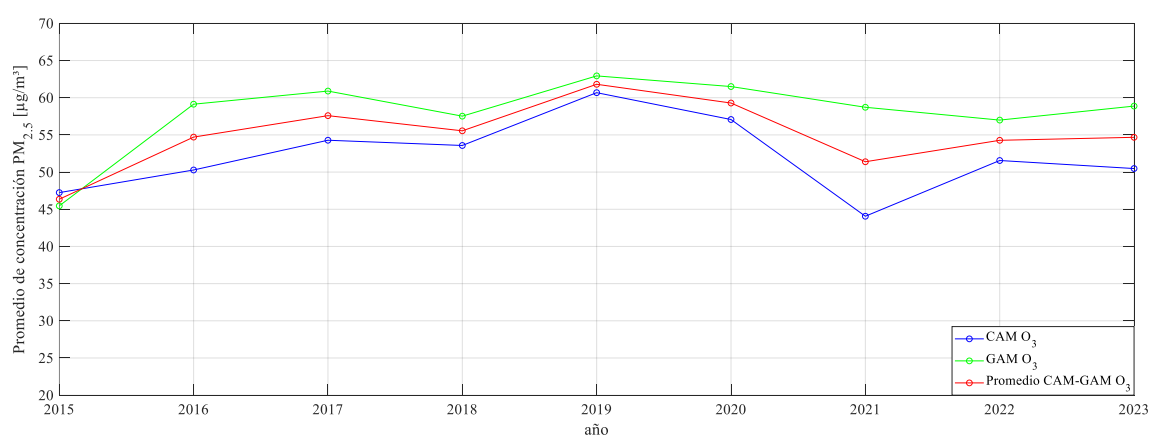


Figura 4.5. Promedios de concentración de  $O_3$  de las estaciones CAM y GAM en el período 2015-2023 (elaboración propia).

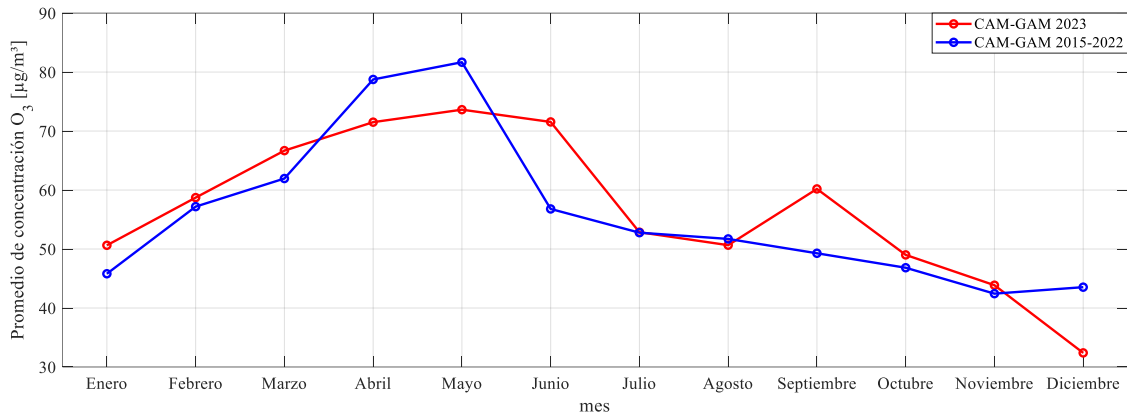


Figura 4.6. Promedio de valores mensuales de concentración de  $O_3$  de las estaciones CAM-GAM en el período 2015-2022 y en el año 2023 (elaboración propia).

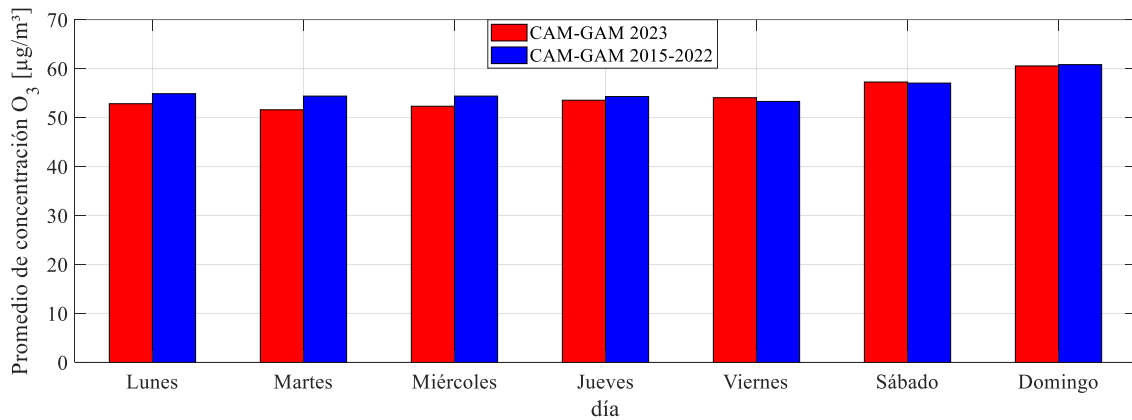


Figura 4.7. Promedio de valores diarios de concentración de  $O_3$  de las estaciones CAM-GAM en el período 2015-2022 y en el año 2023 (elaboración propia).

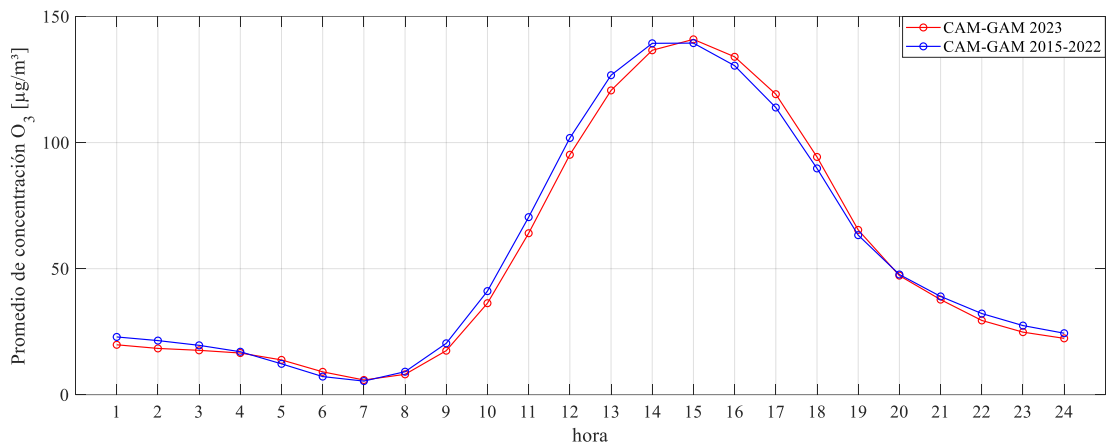


Figura 4.8 Perfil horario de la concentración promedio de  $O_3$  de las estaciones CAM-GAM en el período 2015-2022 y en el año 2023 (elaboración propia).

### **$PM_{10}$ y $PM_{2.5}$**

Con respecto al análisis de las tendencias de concentración de  $PM_{10}$  y  $PM_{2.5}$  en la zona de estudio se identificó que históricamente (2019-2022) los promedios anuales obtenidos por la estación GAM son superiores a los medidos por la estación CAM. Además, los meses en los que históricamente se han obtenido los mayores promedios de concentración horaria en la zona de estudio son diciembre - mayo. Por otro lado, los días en los que

históricamente se han obtenido los mayores promedios de concentración horaria son de martes a sábado. Mientras que las horas en las que históricamente se han obtenido los mayores promedios de concentración horaria son las 9 – 10 am y 10 – 11am, para el caso del  $PM_{10}$  y  $PM_{2.5}$ , respectivamente. Durante el año 2023 se obtuvieron promedios de concentración inferiores para cada una de las escalas temporales en comparación con las tendencias históricas. En las Figuras 4.13 – 4.16 se presentan las tendencias de los valores promedios de concentración de  $PM_{10}$  y  $PM_{2.5}$  en cada una de las escalas de tiempo, los valores empleados para obtener las gráficas se presentan en los Anexos 19 – 26.

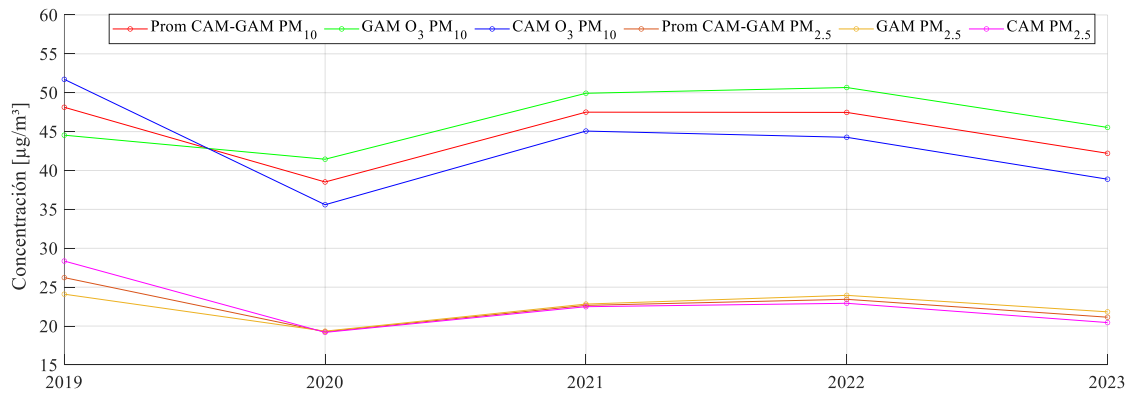


Figura 4.9. Promedios de concentración anual de  $PM_{10}$  y  $PM_{2.5}$  de las estaciones CAM y GAM en el período 2019-2023 (elaboración propia).

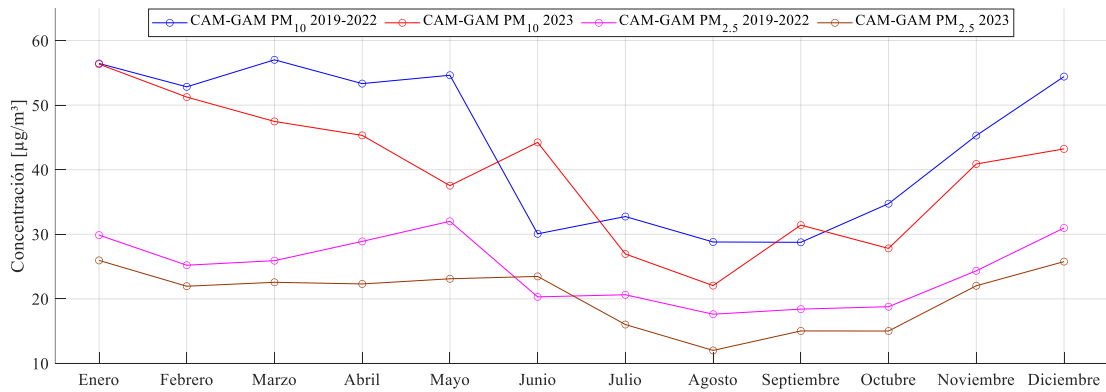


Figura 4.10. Promedio de valores mensuales de concentración de  $PM_{10}$  y  $PM_{2.5}$  de las estaciones CAM-GAM en el período 2019-2022 y en el año 2023 (elaboración propia).

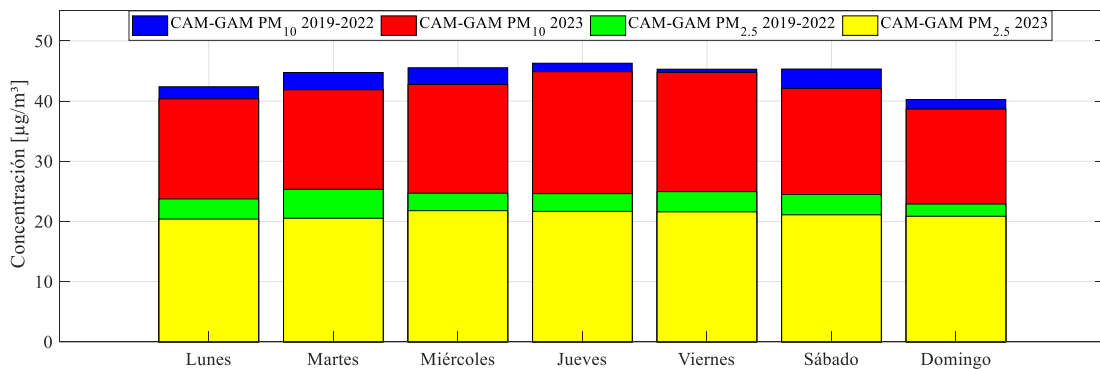


Figura 4.11. Promedio de valores de concentración de  $PM_{10}$  y  $PM_{2.5}$  por días de la semana de las estaciones CAM-GAM en el período 2019-2022 y en el año 2023 (elaboración propia).

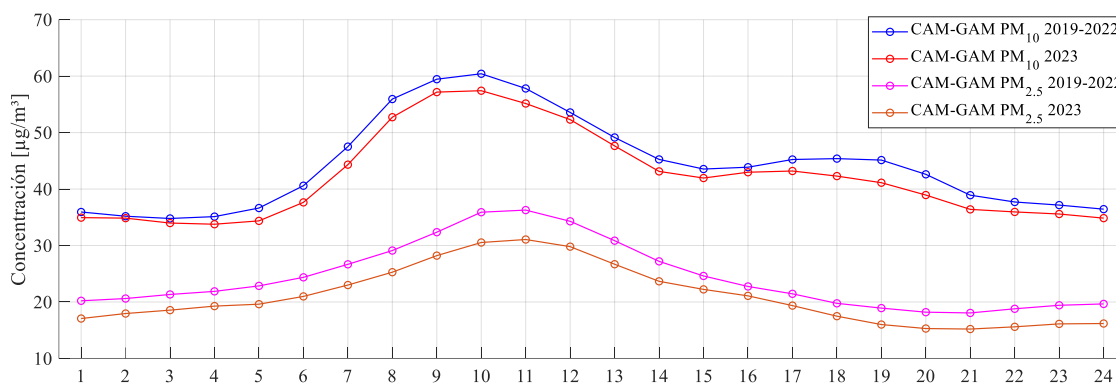


Figura 4.12 Perfil horario de la concentración promedio de  $PM_{10}$  y  $PM_{2.5}$  de las estaciones CAM-GAM en el período 2019-2022 y en el año 2023 (elaboración propia).

### $NO_2$

Con respecto al análisis de las tendencias de concentración de  $NO_2$  en la zona de estudio se identificó que históricamente (2015 - 2022) los promedios anuales obtenidos por la estación CAM son superiores a los medidos por la estación GAM. Además, los meses en los que históricamente se han obtenido los mayores promedios de concentración horaria en la zona de estudio son diciembre y enero. Por otro lado, los días en los que históricamente se han obtenido los mayores promedios de concentración horaria son los días entre semana. Mientras que las horas en las que históricamente se han obtenido los mayores promedios de concentración horaria son las 9 – 10 am. Durante el año 2023 se obtuvieron promedios de concentración superiores para cada una de las escalas temporales en comparación con las tendencias históricas. En las Figuras 4.9 – 4.12 se presentan las tendencias de los valores promedios de concentración de  $PM_{10}$  y  $PM_{2.5}$  en cada una de las escalas de tiempo, los valores empleados para obtener las gráficas se presentan en los Anexos 27 – 30.

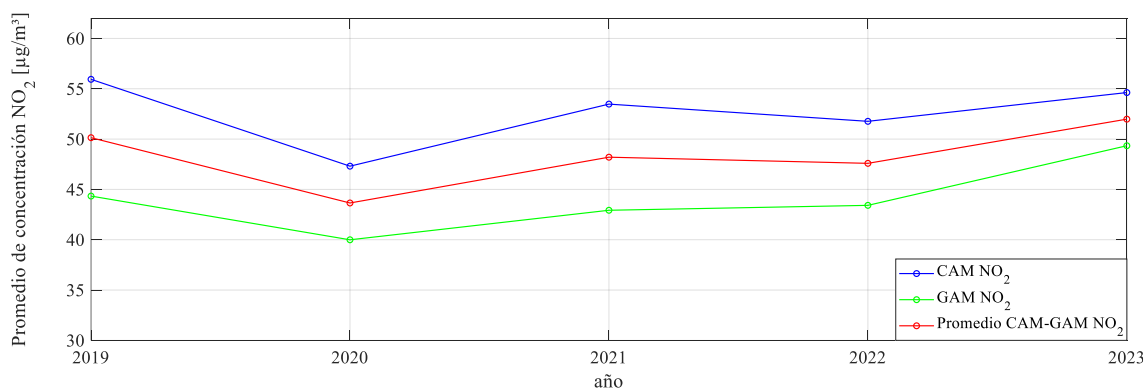


Figura 4.13. Promedios de concentración de  $NO_2$  de las estaciones CAM y GAM en el período 2019-2023 (elaboración propia).

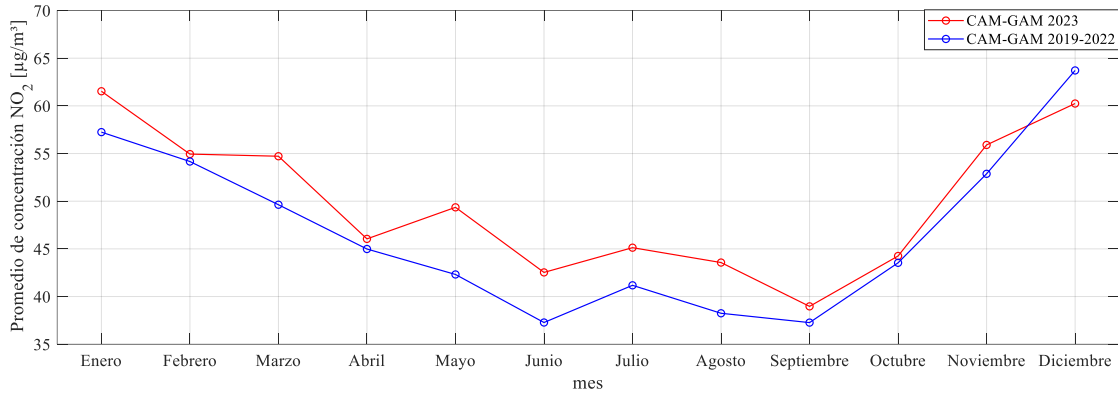


Figura 4.14. Promedio de valores mensuales de concentración de NO<sub>2</sub> de las estaciones CAM-GAM en el período 2019-2022 y en el año 2023 (elaboración propia).

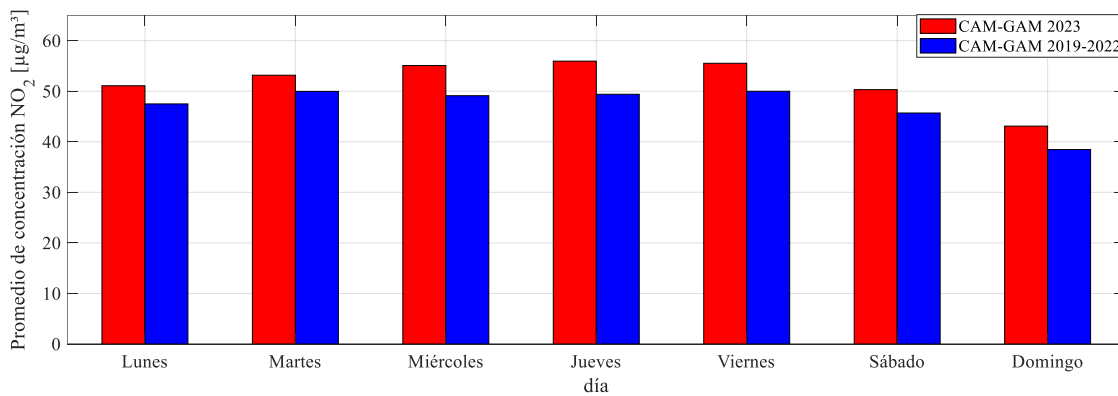


Figura 4.15. Promedio de valores diarios de concentración de O<sub>3</sub> de las estaciones CAM-GAM en el período 2019-2022 y en el año 2023 (elaboración propia).

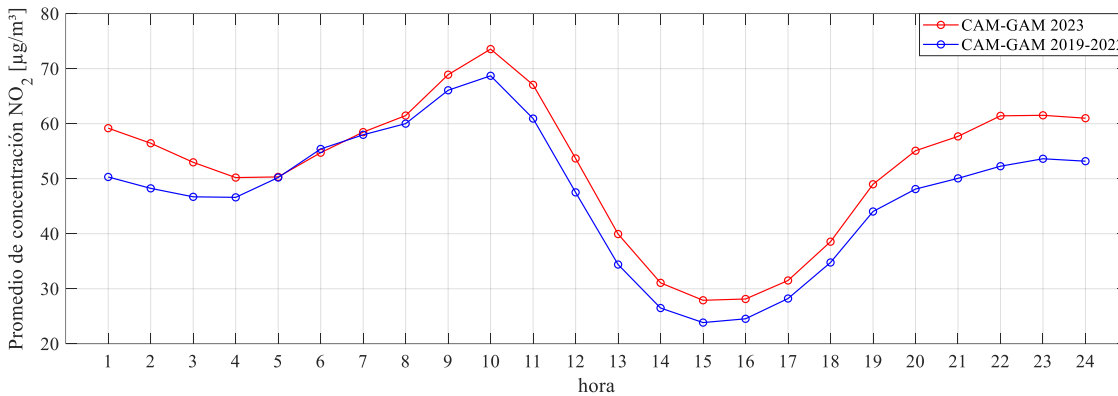


Figura 4.16 Perfil horario de la concentración promedio de NO<sub>2</sub> de las estaciones CAM-GAM en el período 2019-2022 y en el año 2023 (elaboración propia).

#### 4.4 Análisis datos obtenidos por el sistema *Flatburn*

Se clasificaron los valores de concentración de PM<sub>2.5</sub> obtenidos por el dispositivo *Flatburn* durante las campañas de monitorización que se realizaron en la zona de estudio, con respecto al Índice de Aire y Salud. En el análisis de los datos obtenidos se empleó el software QGIS. En la Figura 4.17 se presentan los valores preprocesados de concentración de PM<sub>2.5</sub> obtenidos durante las campañas de monitorización.



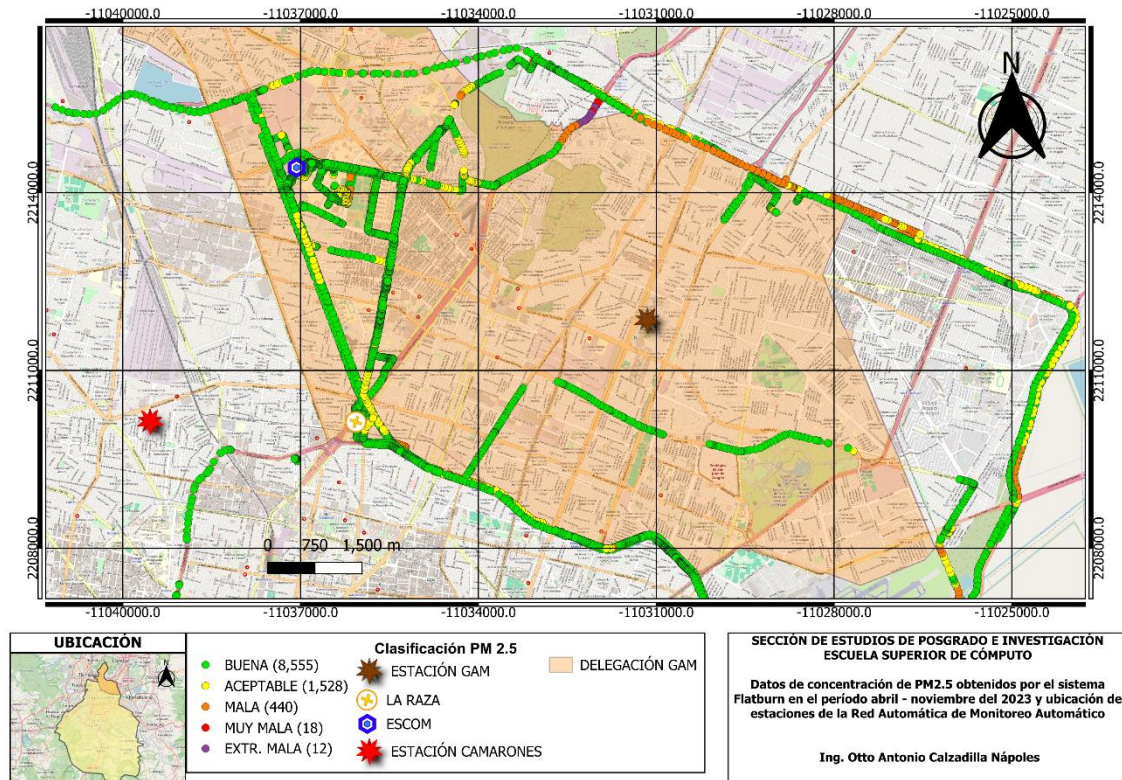


Figura 4.17. Datos preprocesados de concentración de  $PM_{2.5}$  obtenidos por el sistema *Flatburn* durante las campañas de monitorización (elaboración propia).

En el análisis de los datos se identificó que el 18.93% de los valores de concentración de  $PM_{2.5}$  se corresponden niveles de riesgo donde se recomienda a los grupos sensibles y/o población en general ejecutar medidas de reducción a la exposición de la contaminación atmosférica.

Además, se filtraron los valores de concentración de  $PM_{2.5}$  obtenidos en la ruta Raza – ESCOM mediante el empleo del software QGIS. Estos valores se presentan en la Figura 4.18. En el análisis de los datos correspondientes a esta ruta, se identificó que el 12.73% de los valores de concentración de  $PM_{2.5}$  se corresponden niveles de riesgo donde se recomienda a los grupos sensibles y/o población en general ejecutar medidas de reducción a la exposición de la contaminación atmosférica.

#### 4.5 Análisis de variabilidad espacial

Se analizó la variabilidad espacial de la concentración de  $PM_{2.5}$  en la zona de estudio mediante la comparación de los valores medidos con los sistemas fijo y móvil, con los datos obtenidos por las estaciones de la RAMA: CAM y GAM en mismo período.

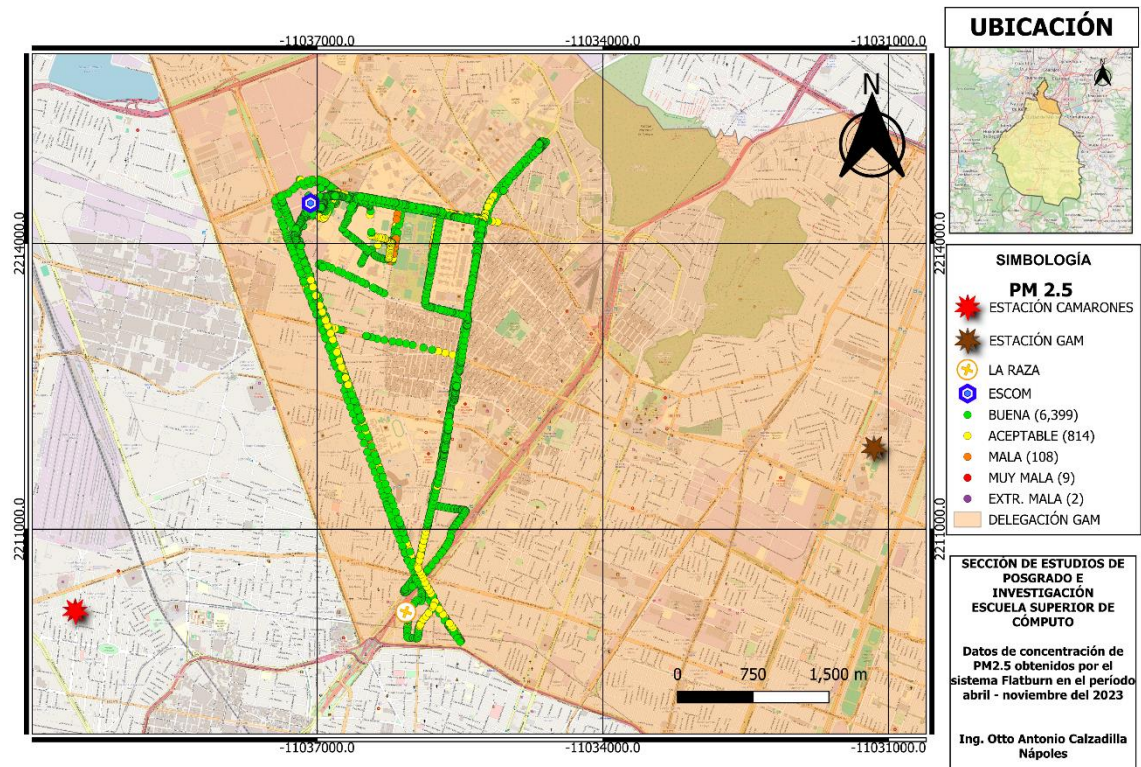


Figura 4.18. Datos preprocesados de concentración de  $PM_{2.5}$  obtenidos por el sistema *Flatburn* durante en la ruta Raza – ESCOM (elaboración propia).

#### 4.5.1 Comparación *Flatburn* – RAMA

En el análisis de la variabilidad espacial a partir de la comparación de los valores medidos por el sistema *Flatburn*, con los datos obtenidos por las estaciones CAM y GAM, se compararon los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por cada sistema durante el período abril – noviembre del 2023.

Como resultado de la comparación se obtuvieron valores de error absoluto medio (MAE) en un rango de 8.54 - 10.76  $\mu g/m^3$ , la mayor diferencia se obtuvo en el análisis *Flatburn* General – GAM. En la Tabla 4.12 se muestran los parámetros considerados en la comparación.

Tabla 4.12 Parámetros de comparación entre los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por el sistema *Flatburn* y las estaciones CAM y GAM en el período abril – noviembre del 2023 (elaboración propia).

Estaciones	Cantidad de registros horarios	Coefficiente de correlación	MAE [ $\mu g/m^3$ ]
<i>Flatburn</i> General - CAM	76	0.5551	8.8154
<i>Flatburn</i> General - GAM	59	0.4524	10.7553
<i>Flatburn</i> Raza_ESCOM - CAM	64	0.6742	8.5394
<i>Flatburn</i> Raza ESCOM - GAM	55	0.4152	10.6267

#### 4.5.2 Comparación *Davis AirLink* – RAMA

En el análisis de la variabilidad espacial a partir de la comparación de los valores medidos por el sistema *Davis AirLink*, con los datos obtenidos por las estaciones CAM y GAM, se compararon los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por cada sistema durante el período 15 de marzo – 29 mayo del 2024. En los Anexos 31 y 32 se muestran el diagrama de caja de los valores obtenidos por cada estación en el período y sus parámetros estadísticos, respectivamente. En la Figura 4.19 se presentan los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por cada estación.

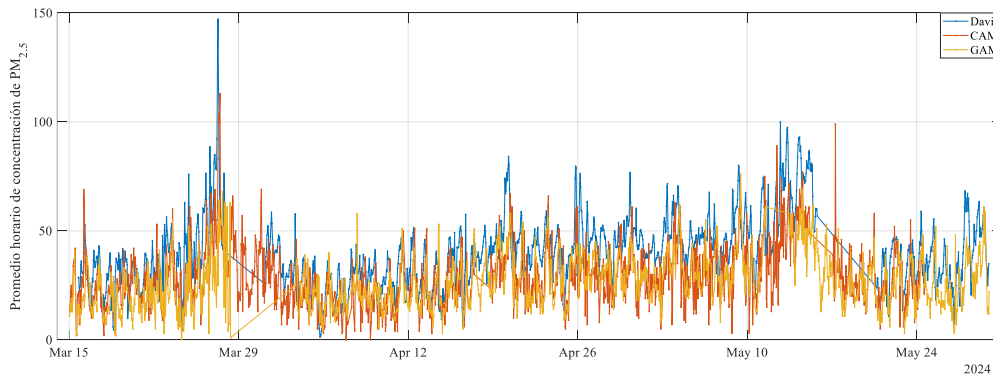


Figura 4.19 Promedios horarios de concentración de  $PM_{2.5}$  obtenidos por las estaciones *Davis*, CAM y GAM en el período 15, marzo – 29, mayo del 2024 (elaboración propia).

Como resultado de la comparación entre los datos obtenidos por las estaciones se obtuvieron valores de error absoluto medio (MAE) de 14.20 y 14.74  $\mu\text{g}/\text{m}^3$ , la mayor diferencia se obtuvo en el análisis *Davis AirLink* – GAM. En la Tabla 4.13 se presentan los parámetros considerados en la comparación.

Tabla 4.13 Parámetros de comparación entre los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por las estaciones *Davis AirLink*, CAM y GAM en el período 15 de marzo – 29 mayo del 2024 (elaboración propia).

Estaciones	# de tuplas	Coefficiente de correlación	MAE [ $\mu\text{g}/\text{m}^3$ ]
<i>Davis AirLink</i> - CAM	1322	0.6262	14.2026
<i>Davis AirLink</i> - GAM	1397	0.5687	14.7433

#### 4.6 Predicción del índice y concentración de $O_3$ en la zona norte de la CDMX

A continuación, se presentan los resultados obtenidos en la predicción del índice y el valor de concentración de  $O_3$  en la zona Norte de la CDMX. Se evaluó el rendimiento de los algoritmos empleados en función de la exactitud, el tiempo de ejecución, el porcentaje de verdaderos positivos, el error absoluto promedio (MAE) y raíz del error cuadrático medio (RMSE).



#### 4.6.1 Predicción del índice de concentración de O<sub>3</sub>

En la Tabla 4.14 se muestran los resultados obtenidos de la aplicación de los algoritmos de clasificación en la base de datos de concentración de O<sub>3</sub> y variables meteorológicas de la Estación GAM, para predecir el índice de concentración de O<sub>3</sub>. En los Anexos 33 - 38 se presentan los resultados obtenidos por cada uno de los algoritmos.

Tabla 4.14. Resultados obtenidos en la predicción del índice de concentración de O<sub>3</sub> (elaboración propia).

Algoritmo	Exactitud [%]	Tiempo de ejecución [s]	Clase	TP rate [%]
<i>C4.5</i>	90.8259	1.16	O3_BUENA	0.960
			O3_ACEPTABLE	0.808
			O3_MALA	0.253
			O3_MUY_MALA	0.041
<i>Naive Bayes</i>	86.46	0.06	O3_BUENA	0.901
			O3_ACEPTABLE	0.794
			O3_MALA	0.404
			O3_MUY_MALA	0.000
<i>Bayesian Belief Networks</i>	86.9375	0.27	O3_BUENA	0.891
			O3_ACEPTABLE	0.872
			O3_MALA	0.323
			O3_MUY_MALA	0.000
<i>Backpropagation</i>	90.2915	100.58	O3_BUENA	0.952
			O3_ACEPTABLE	0.833
			O3_MALA	0.117
			O3_MUY_MALA	0.000
<i>k-Nearest-Neighbor</i>	88.6815	0.08	O3_BUENA	0.949
			O3_ACEPTABLE	0.719
			O3_MALA	0.403
			O3_MUY_MALA	0.216
<i>Support Vector Machine</i>	88.0365	157.57	O3_BUENA	0.952
			O3_ACEPTABLE	0.735
			O3_MALA	0.000
			O3_MUY_MALA	0.000

Se alcanzó la mayor exactitud en la predicción del índice de concentración de O<sub>3</sub> con el algoritmo C4.5, con un 90.83%, y un tiempo de construcción del modelo igual a 1.16

segundos. En la Figura 4.20 se muestran los coeficientes de correlación obtenidos por cada modelo.

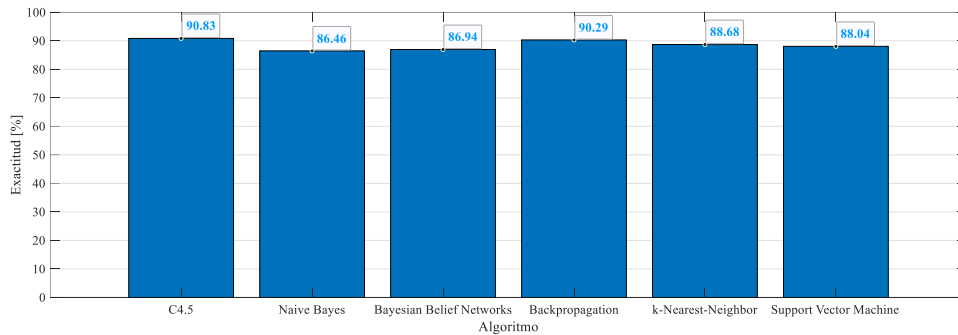


Figura 4.20. Representación gráfica de la exactitud obtenida por los algoritmos de clasificación en la predicción del índice de concentración de  $O_3$  (elaboración propia).

#### 4.6.2 Predicción del valor de la concentración de $O_3$

En la Tabla 4.15 se muestran los resultados obtenidos con la aplicación de algoritmos de clasificación en la base de datos de concentración de  $O_3$  y variables meteorológicas de la Estación GAM, para predecir el valor de la concentración de  $O_3$ .

Tabla 4.15. Resultados obtenidos en la predicción del valor de concentración de  $O_3$  (elaboración propia).

Algoritmo	Tiempo de ejecución [s]	Coefficiente de correlación	MAE [ppb]	RMSE
<i>Backpropagation</i>	73.51	0.7772	14.6494	19.1317
<i>Knn</i>	0.09	0.8925	9.005	13.4286
<i>Support Vector Machine</i>	6521.64	0.8172	12.6995	16.7405

Se alcanzó la mayor exactitud en la predicción del valor de concentración de  $O_3$  con el algoritmo *Knn*, con un valor de error absoluto promedio [MAE] igual a  $9.01 \mu\text{g}/\text{m}^3$  y un coeficiente de correlación de 0.89. En la Figura 4.21 se muestran los coeficientes de correlación obtenidos por cada modelo. En los Anexos 39 - 41 se presentan los resultados obtenidos por cada uno de los algoritmos.

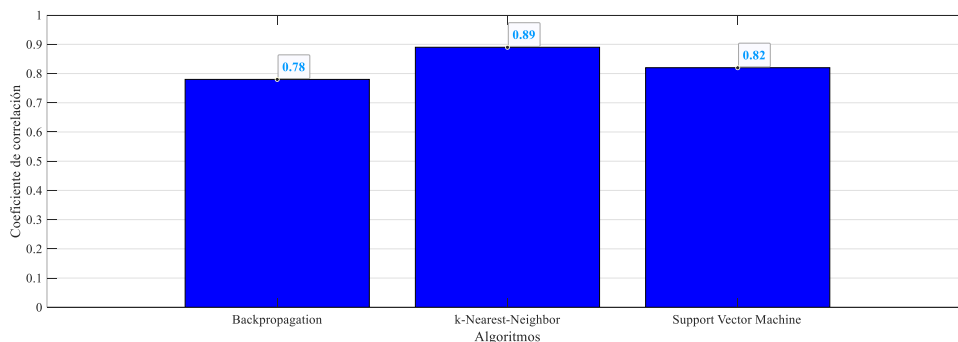


Figura 4.21. Representación gráfica de los coeficientes de correlación obtenidos por los algoritmos de clasificación en la predicción del índice de concentración de  $O_3$  (elaboración propia).

## Conclusiones

1. Se identificaron a los contaminantes atmosféricos criterios:  $O_3$ ,  $PM_{10}$ ,  $PM_{2.5}$  como los contaminantes que mayor incidencia han presentado en la zona norte de la CDMX en los últimos años, mediante el estudio de los informes de calidad del aire nacionales y de la CDMX.
2. Se realizó el análisis estadístico de los datos de concentración de  $O_3$ ,  $PM_{10}$ ,  $PM_{2.5}$  y  $NO_2$  obtenidos por la estación Camarones y Gustavo A Madero durante los años 2022-2023, donde se identificó que los valores de concentración de  $O_3$  y  $NO_2$  presentan un sesgo positivo en las dos estaciones en cada uno de los años de estudio, mientras que el sesgo de los valores de concentración de  $PM_{10}$  y  $PM_{2.5}$  varía según la estación y el año.
3. En el análisis estadístico de los valores de  $O_3$  en la zona norte de la CDMX, se identificaron los mayores promedios de concentración en la estación GAM en cada año, con valores de 56.99 y 58.87  $\mu g/m^3$  para los años 2022 y 2023 respectivamente.
4. En el análisis estadístico de los valores de  $PM_{10}$  y  $PM_{2.5}$  se identificaron los mayores promedios de concentración en la estación GAM en cada año, con valores de concentración de  $PM_{10}$  de 50.67 y 45.54  $\mu g/m^3$  para los años 2022 y 2023 respectivamente, y valores de 23.95 y 21.81  $\mu g/m^3$  para los años 2022 y 2023 respectivamente. Se identificó un descenso en los valores promedios de concentración en las dos estaciones en el año 2023 con respecto a los datos del año anterior.
5. En el análisis estadístico de los valores de  $NO_2$  se identificaron los mayores promedios de concentración en la estación CAM en cada año, con valores de 51.76 y 54.62  $\mu g/m^3$  para los años 2022 y 2023 respectivamente. Se identificó un aumento en los valores promedios de concentración horaria en las dos estaciones en el año 2023 con respecto a los datos del año anterior.
6. Se evaluó el cumplimiento de las NOM de calidad del aire en la zona norte de la CDMX durante los años 2022-2023, donde se emplearon en el análisis los datos de

concentración de contaminantes atmosféricos obtenidos por las estaciones Camarones y Gustavo A Madero en este período y se verificaron los criterios de existencia de datos para cada una de las normas. En el análisis de los conjuntos de datos se empleó el sistema de gestión de base de datos relacional *Microsoft SQL Server*.

7. En la evaluación del cumplimiento de la NOM de  $PM_{10}$ , se identificó el mayor porcentaje de incumplimiento del promedio de 24 horas en la estación GAM con valores de 16.09% y 5.22% para los años 2022 y 2023 respectivamente. Con respecto al promedio anual, solo la estación CAM en el año 2023 cumplió el criterio de existencia de datos para evaluar esta norma, y registró un promedio anual de  $39 \mu g/m^3$ , superior al límite normado; en el resto de los casos no se cumplió el criterio de existencia de datos, pero se identificó que los promedios obtenidos fueron superiores al límite normado.
8. En la evaluación del cumplimiento de la NOM de  $PM_{2.5}$ , se identificaron porcentajes bajos de incumplimiento del promedio de 24 horas durante los años 2022 y 2023, con valores inferiores al 1.88%. Con respecto al promedio anual, solo la estación CAM en el año 2023 cumplió el criterio de existencia de datos para evaluar esta norma, y registró un promedio anual de  $21 \mu g/m^3$ , superior al límite normado; en el resto de los casos no se cumplió el criterio de existencia de datos, pero se identificó que los promedios obtenidos fueron superiores al límite normado.
9. En la evaluación del cumplimiento de la NOM de  $O_3$ , se identificó el incumplimiento del límite horario de concentración en una cantidad de horas que varía de 190 a 271 durante los años 2022 y 2023; además se identificó un aumento en el porcentaje de horas donde se incumplió esta norma en el año 2023, en las dos estaciones, en comparación con los datos obtenidos en el año 2022. Con respecto al promedio horario de 8h se identificaron porcentajes de días de incumplimiento del límite normado que varían de 23.10% a 46.18%, se registraron la mayor cantidad días de incumplimiento de esta norma en la estación GAM.
10. En la evaluación del cumplimiento de la NOM de  $NO_2$ , se identificó que solamente se incumplió el límite horario de este contaminante en la estación Camarones, en dos ocasiones durante el año 2022. Con respecto al análisis del promedio anual, se identificó que solamente la estación GAM en el año 2022 cumplió con el requisito de suficiencia de datos, y el valor de promedio anual obtenido incumple la norma de concentración límite de este contaminante (21 ppb). Aunque en los demás casos no se

cumplió con el criterio de suficiencia de datos, se destaca que los promedios anuales obtenidos poseen valores superiores al límite de concentración normado de  $\text{NO}_2$ .

11. Se obtuvieron tendencias actuales del comportamiento de la concentración de los contaminantes atmosféricos con respecto al tiempo en la zona de estudio, donde se obtuvo el promedio de concentración horaria de  $\text{O}_3$ ,  $\text{PM}_{10}$ ,  $\text{PM}_{2.5}$  y  $\text{NO}_2$  por años, meses, días de la semana y horas, mediante el empleo del sistema de gestión de base de datos relacional Microsoft SQL Server. Estos valores promedios se compararon con las tendencias históricas de cada una de las estaciones, definidas para los años 2015-2022.
12. En el análisis de las tendencias de concentración de  $\text{O}_3$  en la zona de estudio se identificó que históricamente (2015 - 2022) los promedios anuales obtenidos por la estación GAM son superiores a los medidos por la estación CAM. Además, los meses en los que históricamente se han obtenido los promedios de concentración horaria más elevados en la zona de estudio son abril y mayo. Por otro lado, los días en los que históricamente se han obtenido los mayores promedios de concentración horaria son los sábados y domingos. Mientras que las horas en las que históricamente se han obtenido los mayores promedios de concentración horaria son las 2 – 3 pm. Estas tendencias históricas se mantuvieron durante el año 2023.
13. En el análisis de las tendencias de concentración de  $\text{NO}_2$  en la zona de estudio se identificó que históricamente (2015 - 2022) los promedios anuales obtenidos por la estación CAM son superiores a los medidos por la estación GAM. Además, los meses en los que históricamente se han obtenido los promedios de concentración horaria más elevados en la zona de estudio son diciembre y enero. Por otro lado, los días en los que históricamente se han obtenido los mayores promedios de concentración horaria son los días entre semana. Mientras que las horas en las que históricamente se han obtenido los mayores promedios de concentración horaria son las 9 – 10 am. Durante el año 2023 se obtuvieron promedios de concentración superiores para cada una de las escalas temporales en comparación con las tendencias históricas.
14. En el análisis de las tendencias de concentración de  $\text{PM}_{10}$  y  $\text{PM}_{2.5}$  en la zona de estudio se identificó que históricamente (2019-2022) los promedios anuales obtenidos por la estación GAM son superiores a los medidos por la estación CAM. Además, los meses en los que históricamente se han obtenido los promedios de concentración horaria más elevados en la zona de estudio son diciembre - mayo. Por otro lado, los días en los que históricamente se han obtenido los mayores promedios de concentración



- horaria son los días de martes-sábado. Mientras que las horas en las que históricamente se han obtenido los mayores promedios de concentración horaria son las 9 – 10 am y 10 – 11am, para el caso del  $PM_{10}$  y  $PM_{2.5}$ , respectivamente. Durante el año 2023 se obtuvieron promedios de concentración inferiores para cada una de las escalas temporales en comparación con las tendencias históricas.
15. Se desarrolló un sistema fijo de monitorización remota de la calidad del aire, que presenta un *Raspberry Pi 4 Model B* como dispositivo de adquisición de datos. Este sistema posibilita la adquisición de los valores de concentración de  $PM_{2.5}$  medidos por un sensor *AM1008W*, con una resolución de 1 min, y el cálculo del Índice de Aire y Salud para este contaminante. Además, transmite los valores obtenidos a la plataforma *Initial State* en tiempo real.
  16. Se obtuvieron datos de concentración de  $PM_{2.5}$  con el sistema fijo de monitorización remota de la calidad del aire en la Isla de Datos Urbanos de la ESCOM durante el período marzo – mayo del 2024; estos valores se emplearon en la calibración del sistema mediante el método de regresión lineal. Como resultado, se redujo el porcentaje medio de error absoluto (MAPE) de 28.97% a 13.53% entre los promedios horarios obtenidos por el método de medición *TSI* del sensor *AM1008W* y los obtenidos por el dispositivo *Davis AirLink*, que corroboran la elevada relación que existe entre los valores.
  17. En el análisis de variabilidad espacial a partir de los promedios horarios de concentración de  $PM_{2.5}$  registrados por el dispositivo *Davis AirLink* y las estaciones CAM y GAM en el periodo 15, marzo – 29, mayo del 2024, se obtuvieron valores de error absoluto medio (MAE) de 14.20 y 14.74  $\mu g/m^3$ , se obtuvo la mayor diferencia en el análisis Davis – GAM.
  18. Se obtuvieron datos espacio-temporales de concentración de  $PM_{2.5}$  con el dispositivo *Flatburn* en la zona de estudio, en el período abril – noviembre del 2023, mediante su instalación sobre el techo de autobuses de transporte de pasajeros de la ruta La Raza – ESCOM, durante períodos de alrededor de 6 horas, también se recopilaron datos mediante su instalación en vehículos particulares.
  19. En el análisis de los datos preprocesados de concentración de  $PM_{2.5}$  obtenidos durante las campañas de monitorización realizadas con el sistema *Flatburn* se identificó que el 18.93% de los valores de concentración de  $PM_{2.5}$  se corresponden niveles de riesgo donde se recomienda a los grupos sensibles y/o población en general ejecutar medidas de reducción a la exposición de la contaminación atmosférica. Mientras que en los

datos obtenidos en solo la ruta Raza-ESCOM, el 12.73% de los valores obtenidos corresponden a esta clasificación.

20. En el análisis de la variabilidad espacial a partir de los promedios horarios de concentración de  $PM_{2.5}$  registrados por el dispositivo móvil *Flatburn* y las estaciones CAM y GAM en el período abril – noviembre del 2023, se obtuvieron valores de error absoluto medio (MAE) en un rango de 8.54 - 10.76  $\mu g/m^3$ , se obtuvo la mayor diferencia en el análisis *Flatburn* General – GAM.
21. Se obtuvieron modelos de predicción de la clasificación y la concentración de  $O_3$  en la zona de estudio mediante su entrenamiento con un conjunto de datos de promedios horarios de concentración de  $O_3$ , temperatura, humedad relativa, dirección y velocidad del viento, obtenidos por la estación GAM en el período 2015 – 2023. El rango de exactitud obtenida por los modelos de clasificación fue de 86.94% - 90.83%, la mayor exactitud correspondió al algoritmo *C4.5*, con un valor de 90.83%.
22. Se obtuvieron valores de error absoluto promedio en la predicción del promedio horario de concentración de  $O_3$  inferiores a 14.65  $\mu g/m^3$ , la mayor exactitud se obtuvo con el algoritmo *Knn*, con un valor de MAE igual a 9.01  $\mu g/m^3$  y un coeficiente de correlación de 0.89.

## Recomendaciones

Con el objetivo de continuar las investigaciones relacionadas sobre el presente tema se recomienda lo siguiente:

1. Incluir en la calibración de los sensores de calidad del aire de bajo costo, mediante regresión lineal, los parámetros temperatura ambiente y humedad relativa, para contribuir a mejorar la efectividad de la calibración.
2. Incluir una mayor cantidad de parámetros en el entrenamiento de modelos de clasificación, como la presión atmosférica, la radiación solar y la concentración de otros contaminantes atmosféricos como el NO<sub>2</sub>, SO<sub>2</sub>, CO, PM<sub>10</sub>, PM<sub>2.5</sub>.

## Referencias bibliográficas

Abdel-Shafy, H. I., & Mansour, M. S. (2016). A review on polycyclic aromatic hydrocarbons: source, environmental impact, effect on human health and remediation. *Egyptian journal of petroleum*, 25(1), 107-123.

Al-Janabi, S., Mohammad, M., & Al-Sultan, A. (2020). A new method for prediction of air pollution based on intelligent computation. *Soft Computing*, 24(1), 661-680.

Anjomshoaa, A., Duarte, F., Rennings, D., Matarazzo, T. J., deSouza, P., & Ratti, C. (2018). City scanner: Building and scheduling a mobile sensing platform for smart city services. *IEEE Internet of things Journal*, 5(6), 4567-4579.

Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., ... & Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?. *Environment international*, 99, 293-302.

Cubic Sensor and Instrument Co., Ltd. (2023). *Integrated Air Quality Sensor Module AM1008W Technical Specification*. Disponible en : [https://en.gassensor.com.cn/Product\\_files/Specifications/AM1008W Integrated Air Quality Sensor Module Specification.pdf](https://en.gassensor.com.cn/Product_files/Specifications/AM1008W%20Integrated%20Air%20Quality%20Sensor%20Module%20Specification.pdf).

DeSouza, P., Anjomshoaa, A., Duarte, F., Kahn, R., Kumar, P., & Ratti, C. (2020). Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustainable Cities and Society*, 60, 102239.

Dhingra, S., Madda, R. B., Gandomi, A. H., Patan, R., & Daneshmand, M. (2019). Internet of Things mobile-air pollution monitoring system (IoT-Mobair). *IEEE Internet of Things Journal*, 6(3), 5577-5584.

Gao, Y., Dong, W., Guo, K., Liu, X., Chen, Y., Liu, X., ... & Chen, C. (2016, April). Mosaic: A low-cost mobile sensing system for urban air quality monitoring. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer*

*Communications* (pp. 1-9). IEEE.

Gunawan, T. S., Munir, Y. M. S., Kartiwi, M., & Mansor, H. (2018). Design and Implementation of Portable Outdoor Air Quality Measurement System using Arduino. *International Journal of Electrical & Computer Engineering* (2088-8708), 8(1).

Guo, Y., Zeng, H., Zheng, R., Li, S., Barnett, A. G., Zhang, S., ... & Williams, G. (2016). The association between lung cancer incidence and ambient air pollution in China: A spatiotemporal analysis. *Environmental research*, 144, 60-65.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining concepts and techniques*. University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University.

Haq, M. A. (2022). SMOTEDNN: A novel model for air pollution forecasting and AQI classification. *Computers, Materials & Continua*, 71(1).

Instituto Nacional de Ecología (INE). (2008). *Manual 1: Principios de Medición de la Calidad del Aire*. México.

Instituto Nacional de Ecología (INE). (2008). *Manual 2: Sistemas de Medición de la Calidad del Aire*. México.

Instituto Nacional de Ecología (INE). (2008). *Manual 3: Redes Estaciones y Equipos de Medición de la Calidad del Aire*. México.

Instituto Nacional de Ecología y Cambio Climático (INECC). (2022). *Informe Nacional de la Calidad del Aire 2020, México. Ciudad de México: Coordinación General de Contaminación y Salud Ambiental, Dirección de Investigación de Calidad del Aire y Contaminantes Climáticos*. Ciudad de México.

Instituto Nacional de Ecología y Cambio Climático (INECC). (2024). *Informe Nacional de la Calidad del Aire 2021, México. Ciudad de México: Coordinación General de Contaminación y Salud Ambiental, Dirección de Investigación de Calidad del Aire y Contaminantes Climáticos*. Ciudad de México.

Karagulian, F., Barbieri, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., ... & Borowiak, A. (2019). Review of the performance of low-cost sensors for air quality monitoring. *Atmosphere*, 10(9), 506.

Khanam, R., Kumar, A., Nayak, A. K., Shahid, M., Tripathi, R., Vijayakumar, S., Bhaduri, D., Kumar, U., Mohanty, S., & Panneerselvam, P. (2020). Metal (loid)s (As,

Hg, Se, Pb and Cd) in paddy soil: Bioavailability and potential risk to human health. *Science of the Total Environment*, 699, 134330.

Kumar, S., & Jasuja, A. (2017, May). Air quality monitoring system based on IoT using Raspberry Pi. In *2017 International conference on computing, communication and automation (ICCCA)* (pp. 1341-1346). IEEE.

Lei, J. (2020). Cross-validation with confidence. *Journal of the American Statistical Association*, 115(532), 1978-1997.

Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., ... & Williams, R. (2018). Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?. *Environment international*, 116, 286-299.

McKercher, G. R., Salmond, J. A., & Vanos, J. K. (2017). Characteristics and applications of small, portable gaseous air pollution monitors. *Environmental Pollution*, 223, 102-110.

Phala, K. S. E., Kumar, A., & Hancke, G. P. (2016). Air quality monitoring system based on ISO/IEC/IEEE 21451 standards. *IEEE Sensors Journal*, 16(12), 5037-5045.

Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). (2012). *NORMA Oficial Mexicana NOM-156-SEMARNAT-2012, Establecimiento y operación de sistemas de monitoreo de la calidad del aire*. Diario Oficial de la Federación.

Secretaría de Salud. (2019). *NORMA Oficial Mexicana NOM-022-SSA1-2019, Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al dióxido de azufre (SO<sub>2</sub>). Valores normados para la concentración de dióxido de azufre (SO<sub>2</sub>) en el aire ambiente, como medida de protección a la salud de la población*. Diario Oficial de la Federación.

Secretaría de Medio Ambiente y Recursos Naturales (SEMARNAT). (2019). *NORMA Oficial Mexicana NOM-172-SEMARNAT-2019, Lineamientos para la obtención y comunicación del Índice de Calidad del Aire y Riesgos a la Salud*. Diario Oficial de la Federación.

Secretaría de Salud. (2021). *NORMA Oficial Mexicana NOM-020-SSA1-2021, Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al ozono (O<sub>3</sub>). Valores normados para la concentración de ozono (O<sub>3</sub>) en el aire ambiente, como medida de protección a la salud de la población*. Diario Oficial de la Federación.

Secretaría de Salud. (2021). *NORMA Oficial Mexicana NOM-021-SSA1-2021, Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al monóxido de carbono (CO). Valores normados para la concentración de monóxido de carbono (CO) en el aire ambiente, como medida de protección a la salud de la población*. Diario Oficial de la Federación.

Secretaría de Salud. (2021). *NORMA Oficial Mexicana NOM-023-SSA1-2021, Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto al dióxido de nitrógeno (NO<sub>2</sub>). Valores normados para la concentración de dióxido de nitrógeno (NO<sub>2</sub>) en el aire ambiente, como medida de protección a la salud de la población*. Diario Oficial de la Federación.

Secretaría de Salud. (2021). *NORMA Oficial Mexicana NOM-025-SSA1-2021, Salud ambiental. Criterio para evaluar la calidad del aire ambiente, con respecto a las partículas suspendidas PM<sub>10</sub> y PM<sub>2.5</sub>. Valores normados para la concentración de partículas suspendidas PM<sub>10</sub> y PM<sub>2.5</sub> en el aire ambiente, como medida de protección a la salud de la población*. Diario Oficial de la Federación.

Secretaría del Medio Ambiente de la Ciudad de México, (SEDEMA). (2024). *Dirección de Monitoreo Atmosférico*. Disponible en: <http://www.aire.cdmx.gob.mx>

Secretaría del Medio Ambiente de la Ciudad de México (SEDEMA). (2023). *Calidad del aire en la Ciudad de México, Informe anual 2020. Dirección General de Calidad del Aire, Dirección de Monitoreo de Calidad del Aire*. Ciudad de México

Secretaría del Medio Ambiente de la Ciudad de México (SEDEMA). (2022). *Calidad del aire en la Ciudad de México, Informe anual 2019. Dirección General de Calidad del Aire, Dirección de Monitoreo de Calidad del Aire*. Ciudad de México

Senseable City Lab. (2024). Flatburn. MIT Senseable City Lab. <https://senseable.mit.edu/flatburn/>

Singh, A., & Agrawal, M. (2007). Acid rain and its ecological consequences. *Journal of Environmental Biology*, 29(1), 15.

Sistema Nacional de Información de la Calidad del Aire (SINAICA). (2023). *Datos crudos de concentración de contaminantes atmosféricos*. Disponible en: <https://sinaica.inecc.gob.mx/>.

Tejero-Rivas, M. C., Ángel-Meraz, D., Bautista-Margulis, R. G., & Hernández-Morales,

- N. E. (2015). Corrosividad atmosférica del cobre y del acero en dos localidades de Villahermosa, Tabasco. *Ingeniería, investigación y tecnología*, 16(2), 197-206.
- Van den Bossche, J., Peters, J., Verwaeren, J., Botteldooren, D., Theunis, J., & De Baets, B. (2015). Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset. *Atmospheric Environment*, 105, 148-161.
- Weinmayr, G., Romeo, E., De Sario, M., Weiland, S. K., & Forastiere, F. (2010). Short-term effects of PM10 and NO2 on respiratory health among children with asthma or asthma-like symptoms: a systematic review and meta-analysis. *Environmental health perspectives*, 118(4), 449-457.
- Weka, (2023). Weka 3: *Machine Learning Software in Java*. Disponible en: <https://www.cs.waikato.ac.nz/ml/weka/>.
- WORLD HEALTH ORGANIZATION (WHO). (2016). *Ambient air pollution: A global assessment of exposure and burden of disease*. Disponible en <http://apps.who.int/iris/bitstream/handle/10665/250141/9789241511353eng.pdf?sequence=1>.
- WORLD HEALTH ORGANIZATION (WHO). (2021). *New WHO Global Air Quality Guidelines aim to save millions of lives from air pollution*. Disponible en <https://www.who.int/news/item/22-09-2021-new-who-global-air-quality-guidelines-aim-to-save-millions-of-lives-from-air-pollution>.
- WORLD HEALTH ORGANIZATION (WHO). (2022). *Billions of people still breathe unhealthy air: new WHO data*. Disponible en <https://www.who.int/news/item/04-04-2022-billions-of-people-still-breathe-unhealthy-air-new-who-data>.
- Yarragunta, S., & Nabi, M. A. (2021). Prediction of air pollutants using supervised machine learning. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1633-1640). IEEE.
- Zamora Mero, W. J. (2015). Monitorización de los niveles de polución ambiental mediante sensores móviles. Universidad Politécnica de Valencia. España.
- Zheng, K., Zhao, S., Yang, Z., Xiong, X., & Xiang, W. (2016). Design and implementation of LPWA-based air quality monitoring system. *IEEE Access*, 4, 3238-3245.



## Anexos

Anexo 1. Categorías del Índice Aire y Salud (SEMARNAT, 2019).

Calidad del aire	Nivel de riesgo	Descripción del riesgo	Color
Buena	Bajo	Se considera que el riesgo es mínimo o nulo	Verde
Aceptable	Moderado	(PM <sub>10</sub> y PM <sub>2.5</sub> ) Posible agravamiento de enfermedad pulmonar y cardíaca en personas con enfermedad cardiopulmonar y adultos mayores.	Amarillo
Mala	Alto	Para todos los contaminantes criterio existe probabilidad de disminución en la capacidad pulmonar en personas sanas. Incremento en la probabilidad de aparición de síntomas respiratorios en personas sensibles. En personas con enfermedades respiratorias y cardíacas, aumento en la probabilidad de agravamiento y disminución en la tolerancia de la actividad física, así como mayor probabilidad de muertes prematuras en personas con enfermedad cardíaca o pulmonar.	Naranja
Muy Mala	Muy Alto	Para todos los contaminantes criterio, mayor probabilidad de presencia de síntomas respiratorios en población general. Agravamiento de síntomas respiratorios en poblaciones sensibles y en personas con enfermedad pulmonar. Incremento en síntomas cardiovasculares, como dolor precordial, en personas enfermas del corazón, así como mayor probabilidad de muertes prematuras en personas con enfermedad cardíaca o pulmonar.	Rojo
Extr. Mala	Extr. Alto	Para todos los contaminantes criterio, incremento en la probabilidad de síntomas severos respiratorios en población general. Serios efectos respiratorios y agravamiento de síntomas en personas sensibles y en personas con enfermedad pulmonar. Agravamiento de síntomas cardiovasculares en enfermos del corazón e incremento en la probabilidad de muerte prematura en personas con enfermedad pulmonar y cardíaca.	Morado

Anexo 2. Mensajes asociados a las categorías de calidad del aire y riesgos a la salud (SEMARNAT, 2019).

Calidad del aire	Nivel de riesgo asociado	Recomendaciones	
		Grupos sensibles	Toda la población
Buena	Bajo	Disfruta las actividades al aire libre	
Aceptable	Moderado	Considera reducir las actividades físicas vigorosas al aire libre	Disfruta las actividades al aire libre
Mala	Alto	Evita las actividades físicas (tanto moderadas como vigorosas) al aire libre	Reduce las actividades físicas vigorosas al aire libre
Muy Mala	Muy Alto	No realices actividades al aire libre. Acudir al médico si se presentan síntomas respiratorios o cardíacos	Evita las actividades físicas moderadas y vigorosas al aire libre
Extr. Mala	Extr. Alto	Permanece en espacios interiores. Acudir al médico si se presentan síntomas respiratorios o cardíacos	

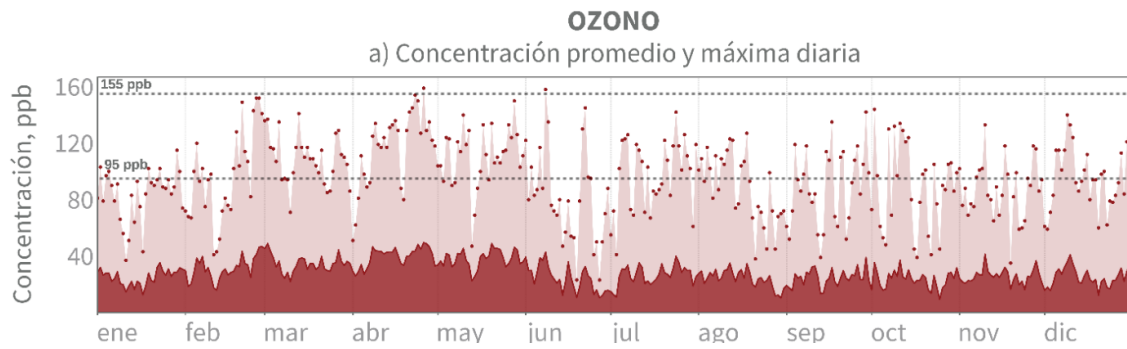
Anexo 3. Estación de monitoreo automático atmosférico Camarones (SINAICA, 2024).



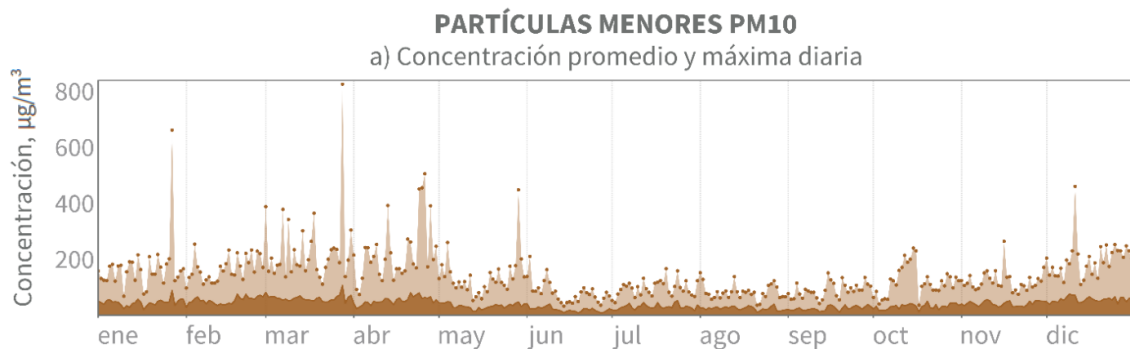
Anexo 4. Estación de monitoreo automático atmosférico Gustavo A Madero (SINAICA, 2024).



Anexo 5. Valores de concentración de  $O_3$  obtenidos en Ciudad de México durante el año 2021 (SEDEMA, 2024).

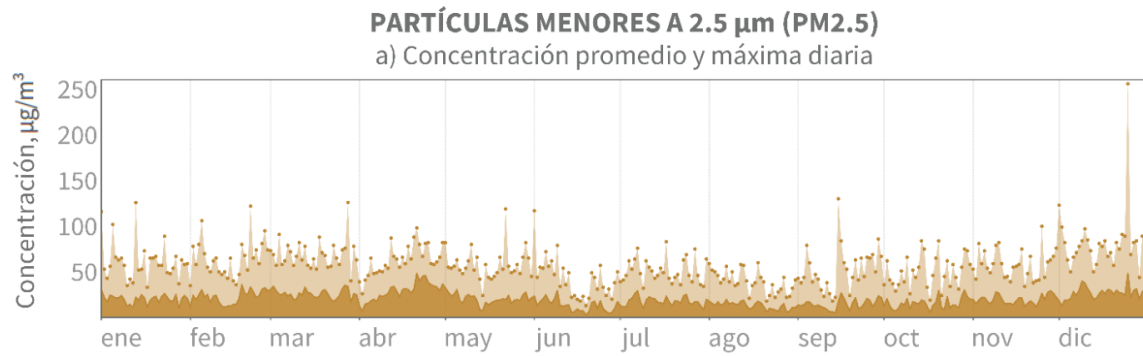


Anexo 6. Valores de concentración de  $O_3$  obtenidos en Ciudad de México durante el año 2021 (SEDEMA, 2024).

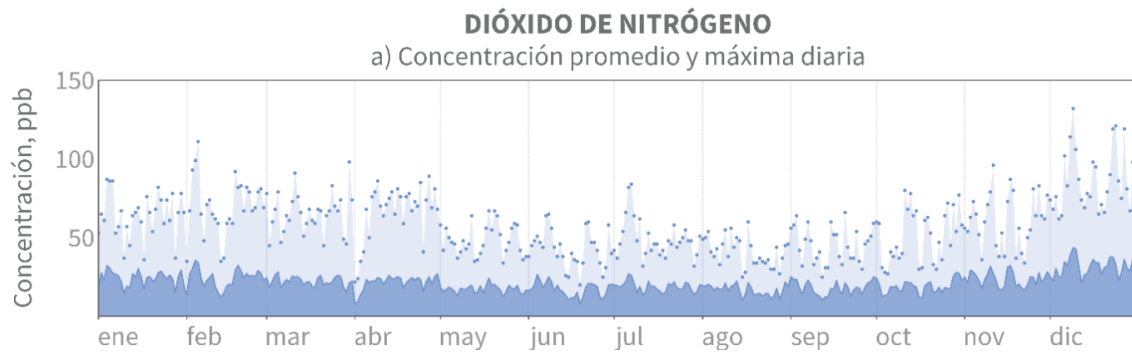




Anexo 7. Valores de concentración de  $\text{PM}_{2.5}$  obtenidos en Ciudad de México durante el año 2021 (SEDEMA, 2024).



Anexo 8. Valores de concentración de  $\text{NO}_2$  obtenidos en Ciudad de México durante el año 2021 (SEDEMA, 2024).



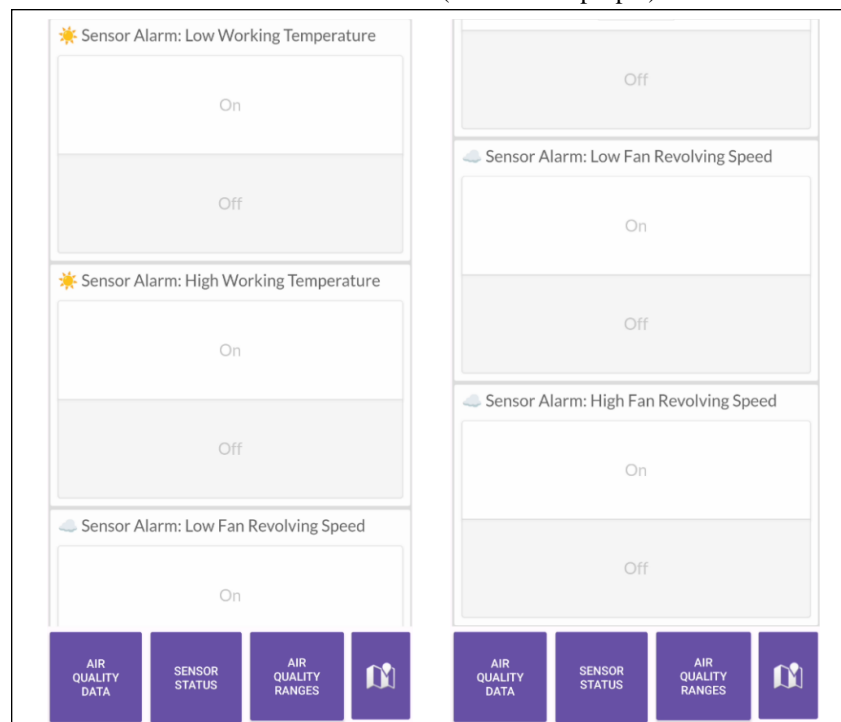
Anexo 9. Isla de Datos Urbanos de la ESCOM.



Anexo 10. Dispositivos de medición de calidad del aire instalados en la Isla de Datos Urbanos de la ESCOM.



Anexo 12. Interfaz gráfica del sistema de monitorización de la calidad del aire: indicadores de funcionamiento del sistema (elaboración propia).



Anexo 13. Camiones empleados en el monitoreo móvil de la calidad del aire.



Anexo 14. Dispositivo *Flatburn* instalado sobre el techo de uno de los camiones de la ruta Raza-ESCOM.





Anexo 15. Promedio de concentración de O<sub>3</sub> por año (elaboración propia).

Año	Promedio de O <sub>3</sub> (CAM) [μg/m <sup>3</sup> ]	Promedio de O <sub>3</sub> (GAM) [μg/m <sup>3</sup> ]	Promedio (CAM- GAM) [μg/m <sup>3</sup> ]
2015	47.24	45.45	46.35
2016	50.28	59.13	54.70
2017	54.28	60.89	57.59
2018	53.58	57.52	55.55
2019	60.68	62.93	61.81
2020	57.07	61.50	59.29
2021	44.06	58.72	51.39
2022	51.56	56.99	54.27
2023	50.48	58.87	54.68

Anexo 16. Promedio de concentración de O<sub>3</sub> por mes (elaboración propia).

Mes	Estación CAM Promedio de O <sub>3</sub> 2023 [μg/m <sup>3</sup> ]	Estación GAM Promedio de O <sub>3</sub> 2023 [μg/m <sup>3</sup> ]	Promedio GAM-CAM 2023 [μg/m <sup>3</sup> ]	Promedio GAM- CAM 2015-2022 [μg/m <sup>3</sup> ]
Enero	45.24	56.01	50.62	45.80
Febrero	53.69	63.74	58.72	57.19
Marzo	64.54	68.82	66.68	61.94
Abril	68.92	74.08	71.50	78.74
Mayo	-	73.63	37.53	81.67
Junio	38.43	104.66	71.55	56.78
Julio	52.85	-	26.94	52.76
Agosto	53.79	47.50	50.64	51.70
Septiembre	58.60	61.78	60.19	49.28
Octubre	47.26	50.75	49.01	46.82
Noviembre	41.94	45.79	43.87	42.43
Diciembre	31.07	33.70	32.39	43.53

Anexo 17. Promedio de concentración de O<sub>3</sub> por día de la semana (elaboración propia).

Día	Estación CAM Promedio de O <sub>3</sub> 2023 [μg/m <sup>3</sup> ]	Estación GAM Promedio de O <sub>3</sub> 2023 [μg/m <sup>3</sup> ]	Promedio de O <sub>3</sub> GAM- CAM 2023 [μg/m <sup>3</sup> ]	Promedio GAM-CAM 2015-2022 [μg/m <sup>3</sup> ]
Lunes	48.52	57.17	52.84	54.89
Martes	48.52	54.69	51.61	54.40
Miércoles	47.50	57.17	52.33	54.40
Jueves	47.18	59.93	53.56	54.30
Viernes	48.79	59.34	54.07	53.31
Sábado	54.07	60.48	57.27	57.06
Domingo	57.91	63.21	60.56	60.83

Anexo 18. Promedio de concentración de O<sub>3</sub> por hora (elaboración propia).

Hora	Estación CAM Promedio de O <sub>3</sub> 2023 [µg/m <sup>3</sup> ]	Estación GAM Promedio de O <sub>3</sub> 2023 [µg/m <sup>3</sup> ]	Promedio de O <sub>3</sub> GAM-CAM 2023 [µg/m <sup>3</sup> ]	Promedio GAM- CAM de O <sub>3</sub> 2015- 2022 [µg/m <sup>3</sup> ]
1	20.07	19.58	19.82	22.93
2	18.60	18.17	18.38	21.49
3	18.09	17.17	17.63	19.61
4	16.85	16.26	16.56	17.07
5	13.61	14.05	13.83	12.27
6	9.40	8.79	9.09	7.22
7	6.02	5.71	5.87	5.42
8	7.51	8.69	8.10	9.15
9	14.81	20.17	17.49	20.40
10	29.53	43.10	36.31	41.11
11	53.85	74.27	64.06	70.44
12	84.32	106.04	95.18	101.83
13	111.88	129.50	120.69	126.72
14	131.28	141.94	136.61	139.37
15	137.88	144.00	140.94	139.45
16	129.85	138.19	134.02	130.50
17	113.14	125.26	119.20	113.91
18	86.63	101.97	94.30	89.76
19	56.74	74.00	65.37	63.30
20	39.79	54.71	47.25	47.69
21	32.06	43.41	37.74	39.00
22	25.07	33.88	29.48	32.21
23	22.29	27.39	24.84	27.45
24	21.93	22.82	22.37	24.40

Anexo 19. Promedio de concentración de PM<sub>10</sub> por año (elaboración propia).

Año	Promedio de PM <sub>10</sub> (CAM) [µg/m <sup>3</sup> ]	Promedio de PM <sub>10</sub> [µg/m <sup>3</sup> ]	Promedio de PM <sub>10</sub> (CAM-GAM) [µg/m <sup>3</sup> ]
2019	51.72	44.54	48.13
2020	35.60	41.45	38.525
2021	45.07	49.93	47.5
2022	44.27	50.67	47.47
2023	38.88	45.54	42.21



Anexo 20. Promedio de concentración de PM<sub>2.5</sub> por año (elaboración propia).

Año	Promedio de PM <sub>25</sub> (CAM) [µg/m <sup>3</sup> ]	Promedio de PM <sub>25</sub> (GAM) [µg/m <sup>3</sup> ]	Promedio de PM <sub>2.5</sub> (CAM-GAM) [µg/m <sup>3</sup> ]
2019	28.36	24.10	26.23
2020	19.19	19.35	19.27
2021	22.49	22.83	22.66
2022	22.93	23.95	23.44
2023	20.45	21.83	21.14

Anexo 21. Promedio de concentración de PM<sub>10</sub> por meses (elaboración propia).

Mes	Estación CAM Promedio de PM <sub>10</sub> 2023 [µg/m <sup>3</sup> ]	Estación GAM Promedio de PM <sub>10</sub> 2023 [µg/m <sup>3</sup> ]	Promedio de PM <sub>10</sub> GAM- CAM 2023 [µg/m <sup>3</sup> ]	Promedio de PM <sub>10</sub> GAM- CAM 2019-2022 [µg/m <sup>3</sup> ]
Enero	50.07	62.63	56.35	56.46
Febrero	49.86	52.64	51.25	52.84
Marzo	46.21	48.74	47.48	57.02
Abril	40.97	49.66	45.32	53.34
Mayo	37.45	47.31	37.53	54.63
Junio	37.55	50.91	44.23	30.07
Julio	30.65	-	26.94	32.74
Agosto	27.69	16.40	22.05	28.80
Septiembre	33.13	29.70	31.42	28.75
Octubre	30.19	25.42	27.81	34.73
Noviembre	38.96	42.81	40.89	45.29
Diciembre	44.67	41.77	43.22	54.42

Anexo 22. Promedio de concentración de PM<sub>2.5</sub> por meses (elaboración propia).

Mes	Estación CAM Promedio de PM <sub>2.5</sub> 2023 [µg/m <sup>3</sup> ]	Estación GAM Promedio de PM <sub>2.5</sub> 2023 [µg/m <sup>3</sup> ]	Promedio de PM <sub>2.5</sub> GAM- CAM 2023 [µg/m <sup>3</sup> ]	Promedio de PM <sub>2.5</sub> GAM- CAM 2015- 2022 [µg/m <sup>3</sup> ]
Enero	24.95	26.94	25.95	29.87
Febrero	22.16	21.73	21.95	25.20
Marzo	23.79	21.32	22.56	25.91
Abril	22.75	21.85	22.30	28.89
Mayo	21.38	24.82	23.10	32.01
Junio	19.64	27.29	23.47	20.30
Julio	16.01	-	16.01	20.63
Agosto	14.32	9.68	12.00	17.61
Septiembre	17.20	12.84	15.02	18.40
Octubre	16.57	13.44	15.01	18.78
Noviembre	21.96	22.08	22.02	24.34
Diciembre	26.61	24.91	25.76	30.97

Anexo 23. Promedio de concentración de PM<sub>10</sub> por día de la semana (elaboración propia).

Día	Estación CAM Promedio de PM <sub>10</sub> 2023 [μg/m <sup>3</sup> ]	Estación GAM Promedio de PM <sub>10</sub> 2023 [μg/m <sup>3</sup> ]	Promedio de PM <sub>10</sub> GAM- CAM 2023 [μg/m <sup>3</sup> ]	Promedio GAM-CAM PM <sub>10</sub> 2015- 2022 [μg/m <sup>3</sup> ]
Lunes	38.24	42.53	40.39	42.38
Martes	38.37	45.40	41.89	44.75
Miércoles	39.70	45.84	42.77	45.54
Jueves	40.73	49.04	44.89	46.30
Viernes	41.70	47.84	44.77	45.30
Sábado	39.32	44.84	42.08	45.33
Domingo	34.31	43.05	38.68	40.26

Anexo 24. Promedio de concentración de PM<sub>2.5</sub> por día de la semana (elaboración propia).

Día	Estación CAM Promedio de PM <sub>2.5</sub> 2023 [μg/m <sup>3</sup> ]	Estación GAM Promedio de PM <sub>2.5</sub> 2023 [μg/m <sup>3</sup> ]	Promedio GAM-CAM PM <sub>2.5</sub> 2023 [μg/m <sup>3</sup> ]	Promedio GAM-CAM PM <sub>2.5</sub> 2015- 2022 [μg/m <sup>3</sup> ]
Lunes	20.09	20.70	20.40	23.74
Martes	19.71	21.35	20.53	25.35
Miércoles	20.99	22.59	21.79	24.70
Jueves	20.47	22.88	21.68	24.63
Viernes	21.13	22.06	21.60	24.96
Sábado	20.94	21.26	21.10	24.48
Domingo	19.82	21.91	20.87	22.90

Anexo 25. Promedio de concentración de PM<sub>10</sub> por hora (elaboración propia).

Hora	Estación CAM Promedio de PM <sub>10</sub> 2023 [μg/m <sup>3</sup> ]	Estación GAM Promedio de PM <sub>10</sub> 2023 [μg/m <sup>3</sup> ]	Promedio de PM <sub>10</sub> GAM- CAM de PM <sub>10</sub> 2023 [μg/m <sup>3</sup> ]	Promedio de PM <sub>10</sub> GAM-CAM de PM <sub>10</sub> 2019-2022 [μg/m <sup>3</sup> ]
1	32.57	37.31	34.94	35.94
2	31.77	37.93	34.85	35.19
3	30.87	37.09	33.98	34.79
4	30.68	36.86	33.77	35.12
5	30.34	38.39	34.37	36.64
6	33.73	41.55	37.64	40.60
7	40.39	48.25	44.32	47.52
8	48.68	56.77	52.73	55.94
9	51.47	62.87	57.17	59.46
10	51.95	62.89	57.42	60.42
11	51.33	58.97	55.15	57.82
12	48.21	56.39	52.30	53.57
13	43.70	51.58	47.64	49.14
14	39.54	46.71	43.13	45.25
15	38.52	45.38	41.95	43.55
16	38.98	46.98	42.98	43.87
17	40.43	45.96	43.20	45.24

18	38.83	45.76	42.30	45.39
19	37.46	44.79	41.13	45.14
20	35.89	41.99	38.94	42.61
21	34.95	37.85	36.40	38.92
22	35.25	36.65	35.95	37.71
23	34.30	36.86	35.58	37.16
24	32.88	36.82	34.85	36.44

Anexo 26. Promedio de concentración de PM<sub>2.5</sub> por hora (elaboración propia).

Hora	Estación CAM Promedio de PM <sub>2.5</sub> 2023 [µg/m³]	Estación GAM Promedio de PM <sub>2.5</sub> 2023 [µg/m³]	Promedio de PM <sub>2.5</sub> GAM- CAM 2023 [µg/m³]	Promedio de PM <sub>2.5</sub> GAM-CAM 2015- 2022 [µg/m³]
1	17.36	16.78	17.07	20.21
2	17.86	18.05	17.96	20.61
3	18.20	18.91	18.56	21.33
4	18.73	19.81	19.27	21.88
5	18.45	20.77	19.61	22.85
6	19.51	22.46	20.99	24.37
7	21.78	24.22	23.00	26.67
8	24.63	25.92	25.28	29.11
9	26.63	29.79	28.21	32.36
10	28.46	32.61	30.54	35.90
11	29.35	32.78	31.07	36.28
12	27.98	31.62	29.80	34.29
13	24.88	28.49	26.69	30.86
14	22.19	25.13	23.66	27.20
15	21.22	23.24	22.23	24.61
16	19.80	22.37	21.09	22.75
17	19.14	19.58	19.36	21.45
18	17.27	17.70	17.49	19.76
19	15.83	16.18	16.01	18.92
20	15.47	15.12	15.30	18.20
21	15.71	14.70	15.21	18.06
22	16.51	14.69	15.60	18.80
23	16.41	15.83	16.12	19.42
24	16.47	15.90	16.19	19.66

Anexo 27. Promedio de concentración de NO<sub>2</sub> por año (elaboración propia).

Año	Promedio de NO <sub>2</sub> [µg/m³] (CAM)	Promedio de NO <sub>2</sub> [µg/m³] (GAM)	Promedio de NO <sub>2</sub> [µg/m³] (CAM-GAM)
2019	55.94	44.34	50.1377
2020	47.31	39.99	43.6507
2021	53.48	42.93	48.201
2022	51.7641	43.4157	47.5899
2023	54.6221	49.3385	51.9803

Anexo 28. Promedio de concentración de NO<sub>2</sub> por mes (elaboración propia).

Mes	Estación CAM Promedio de NO <sub>2</sub> 2023 [µg/m <sup>3</sup> ]	Estación GAM Promedio de NO <sub>2</sub> 2023 [µg/m <sup>3</sup> ]	Promedio NO <sub>2</sub> GAM- CAM 2023 [µg/m <sup>3</sup> ]	Promedio NO <sub>2</sub> GAM-CAM 2019-2022 [µg/m <sup>3</sup> ]
Enero	64.04	59.02	61.53	57.25
Febrero	58.89	50.99	54.94	54.16
Marzo	58.53	50.90	54.72	49.63
Abril	46.91	45.18	46.05	44.99
Mayo	-	49.36	49.36	42.32
Junio	43.70	41.37	42.53	37.28
Julio	45.13	-	45.13	41.18
Agosto	46.65	40.48	43.57	38.24
Septiembre	46.25	31.68	38.97	37.27
Octubre	49.66	38.85	44.25	43.54
Noviembre	62.20	49.60	55.90	52.87
Diciembre	64.74	55.75	60.24	63.71

Anexo 29. Promedio de concentración de NO<sub>2</sub> por día de la semana (elaboración propia).

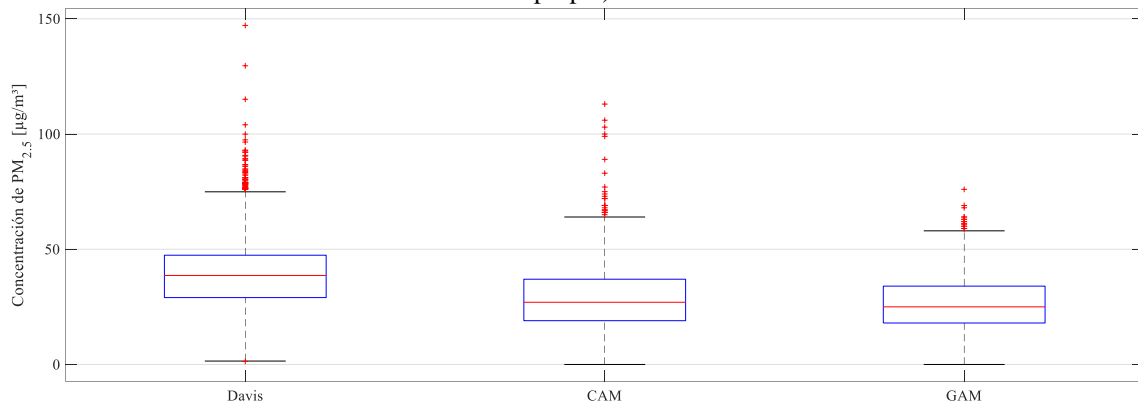
Día	Estación CAM Promedio de NO <sub>2</sub> 2023 [µg/m <sup>3</sup> ]	Estación GAM Promedio de NO <sub>2</sub> 2023 [µg/m <sup>3</sup> ]	Promedio GAM-CAM de NO <sub>2</sub> 2023 [µg/m <sup>3</sup> ]	Promedio GAM-CAM de NO <sub>2</sub> 2019-2022 [µg/m <sup>3</sup> ]
Lunes	53.61	48.55	51.08	47.49
Martes	55.45	50.88	53.16	49.97
Miércoles	57.99	52.16	55.07	49.11
Jueves	58.82	53.08	55.95	49.39
Viernes	59.00	52.05	55.52	49.99
Sábado	53.31	47.31	50.31	45.69
Domingo	44.69	41.50	43.10	38.47

Anexo 30. Promedio de concentración de NO<sub>2</sub> por hora (elaboración propia).

Hora	Estación CAM Promedio de NO <sub>2</sub> 2023 [µg/m <sup>3</sup> ]	Estación GAM Promedio de NO <sub>2</sub> 2023 [µg/m <sup>3</sup> ]	Promedio GAM-CAM de NO <sub>2</sub> 2023 [µg/m <sup>3</sup> ]	Promedio GAM-CAM de NO <sub>2</sub> 2023 [µg/m <sup>3</sup> ]
1	55.90	62.46	59.18	50.32
2	53.64	59.25	56.45	48.25
3	50.75	55.21	52.98	46.71
4	49.28	51.12	50.20	46.61
5	50.32	50.34	50.33	50.21
6	54.66	54.79	54.73	55.40
7	58.55	58.38	58.47	57.98
8	61.94	61.02	61.48	60.01
9	68.86	68.95	68.90	66.07
10	75.36	71.75	73.56	68.71
11	72.62	61.50	67.06	60.93
12	60.39	46.97	53.68	47.51

13	48.38	31.51	39.95	34.40
14	39.35	22.79	31.07	26.49
15	35.26	20.55	27.90	23.85
16	35.05	21.23	28.14	24.54
17	38.09	24.91	31.50	28.22
18	45.11	32.00	38.56	34.78
19	55.04	42.95	48.99	44.04
20	59.87	50.30	55.08	48.12
21	60.23	55.15	57.69	50.06
22	62.74	60.11	61.43	52.27
23	60.56	62.48	61.52	53.63
24	57.76	64.23	61.00	53.18

Anexo 31. Diagramas de caja de los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por las estaciones *Davis AirLink*, CAM y GAM en el periodo 15, marzo – 29, mayo del 2024 (elaboración propia).



Anexo 32. Parámetros estadísticos de los promedios horarios de concentración de  $PM_{2.5}$  obtenidos por las estaciones *Davis AirLink*, CAM y GAM en el periodo 15, marzo – 29, mayo del 2024 (elaboración propia).

Estación	% Datos válidos	Máx. [ $\mu\text{g}/\text{m}^3$ ]	Mín. [ $\mu\text{g}/\text{m}^3$ ]	Media [ $\mu\text{g}/\text{m}^3$ ]	Mediana [ $\mu\text{g}/\text{m}^3$ ]	Moda [ $\mu\text{g}/\text{m}^3$ ]	Desviación estandar [ $\mu\text{g}/\text{m}^3$ ]
<i>Davis</i>	89.56%	147.13	1.33	39.79	38.63	33.45	16.31
CAM	86.78%	113	0	29.08	27	30	14.24
GAM	88.67%	76	0	26.78	25	21	12.02

Anexo 33. Resultados de la aplicación del algoritmo *C4.5* para predecir el índice de concentración de  $O_3$  (elaboración propia).

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.960	0.144	0.958	0.960	0.959	0.817	0.959	0.974	Buena
0.808	0.066	0.749	0.808	0.778	0.722	0.927	0.726	Aceptable
0.253	0.006	0.567	0.253	0.350	0.367	0.917	0.345	Mala
0.041	0.000	0.143	0.041	0.063	0.076	0.776	0.058	Muy Mala
0.908	0.125	0.904	0.908	0.904	0.785	0.951	0.906	

Anexo 34. Resultados de la aplicación del algoritmo *Naive Bayes* para predecir el índice de concentración de O<sub>3</sub> (elaboración propia).

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.901	0.124	0.961	0.901	0.930	0.729	0.958	0.988	Buena
0.794	0.116	0.625	0.794	0.699	0.622	0.919	0.619	Aceptable
0.404	0.014	0.461	0.404	0.431	0.416	0.961	0.398	Mala
0.000	0.000	?	0.000	?	?	0.982	0.052	Muy Mala
0.865	0.119	?	0.865	?	?	0.950	0.897	

Anexo 35. Resultados de la aplicación del algoritmo *Bayesian Belief Networks* para predecir el índice de concentración de O<sub>3</sub> (elaboración propia).

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.891	0.075	0.976	0.891	0.931	0.750	0.968	0.991	Buena
0.872	0.130	0.621	0.872	0.726	0.660	0.927	0.629	Aceptable
0.323	0.010	0.502	0.323	0.393	0.388	0.961	0.405	Mala
0.000	0.000	?	0.000	?	?	0.978	0.037	Muy Mala
0.869	0.083	?	0.869	?	?	0.959	0.902	

Anexo 36. Resultados de la aplicación del algoritmo *Backpropagation* para predecir el índice de concentración de O<sub>3</sub> (elaboración propia).

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.952	0.131	0.961	0.952	0.956	0.811	0.974	0.992	Buena
0.833	0.079	0.720	0.833	0.772	0.715	0.951	0.761	Aceptable
0.117	0.004	0.494	0.117	0.189	0.230	0.950	0.358	Mala
0.000	0.000	?	0.000	?	?	0.964	0.034	Muy Mala
0.903	0.117	?	0.903	?	?	0.969	0.927	

Anexo 37. Resultados de la aplicación del algoritmo *k-Nearest-Neighbor* para predecir el índice de concentración de O<sub>3</sub> (elaboración propia).

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.949	0.171	0.950	0.949	0.949	0.777	0.899	0.950	Buena
0.719	0.069	0.717	0.719	0.718	0.649	0.828	0.626	Aceptable
0.403	0.018	0.395	0.403	0.399	0.381	0.649	0.192	Mala
0.216	0.001	0.232	0.216	0.224	0.223	0.577	0.102	Muy Mala
0.887	0.146	0.887	0.887	0.887	0.740	0.877	0.863	

Anexo 38. Resultados de la aplicación del algoritmo *Support Vector Machine* para predecir el índice de concentración de O<sub>3</sub> (elaboración propia).

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.952	0.233	0.933	0.952	0.942	0.737	0.865	0.928	Buena
0.735	0.083	0.684	0.735	0.709	0.635	0.826	0.555	Aceptable
0.000	0.000	?	0.000	?	?	0.824	0.198	Mala
0.000	0.000	?	0.000	?	?	0.500	0.001	Muy Mala
0.880	0.197	?	0.880	?	?	0.856	0.832	

Anexo 39. Resultados de la aplicación del algoritmo *Backpropagation* para predecir el promedio horario de la concentración de O<sub>3</sub> (elaboración propia).

Correlation coefficient	0.7772
Mean absolute error	14.6494
Root mean squared error	19.1317
Relative absolute error	62.1252 %
Root relative squared error	66.3648 %
Total Number of Instances	59690

Anexo 40. Resultados de la aplicación del algoritmo *k-Nearest-Neighbor* para predecir el promedio horario de la concentración de O<sub>3</sub> (elaboración propia).

Correlation coefficient	0.8925
Mean absolute error	9.005
Root mean squared error	13.4286
Relative absolute error	38.1885 %
Root relative squared error	46.5815 %
Total Number of Instances	59690

Anexo 41. Resultados de la aplicación del algoritmo *Support Vector Machine* para predecir el promedio horario de la concentración de O<sub>3</sub> (elaboración propia).

Correlation coefficient	0.8172
Mean absolute error	12.6995
Root mean squared error	16.7405
Relative absolute error	53.8561 %
Root relative squared error	58.0699 %
Total Number of Instances	59690