

# Calibración Inteligente de Sensores IoT de PM2.5 usando Métricas de Distancia y Análisis de Mínimo Conjunto de Datos

**Samuel Alexis González Martínez 2240680 / Jorge Steven Plata Berdugo 2220734 \***  
*Escuela de Física, Facultad de Ciencias*  
*Universidad Industrial de Santander*

Versión final - Septiembre 2025

## Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Metodología</b>	<b>2</b>
2.1. Datos y Preprocesamiento . . . . .	2
2.2. Análisis de Distancias con Promedios Móviles . . . . .	3
2.3. Calibración Lineal y Validación . . . . .	3
2.4. Determinación del Mínimo Conjunto de Datos . . . . .	3
<b>3. Resultados</b>	<b>3</b>
3.1. Procesamiento de Datos . . . . .	3
3.2. Selección de Ventana Óptima . . . . .	3
3.3. Modelo de Calibración . . . . .	4
3.4. Validación del Modelo . . . . .	4
3.5. Mínimo Conjunto de Datos . . . . .	5
<b>4. Conclusiones y Recomendaciones</b>	<b>5</b>

## Resumen

Este trabajo presenta una metodología completa para la calibración de sensores IoT de bajo costo para medición de material particulado PM2.5, utilizando como referencia estaciones de monitoreo profesional. Se implementó un pipeline de procesamiento que incluye consolidación de datos, limpieza, análisis de

---

\* e-mail: [estudiante@uis.edu.co](mailto:estudiante@uis.edu.co)

distancias euclídeas con promedios móviles, calibración lineal por mínimos cuadrados y validación del modelo. Los resultados demuestran que con solo 384 puntos de datos (10 % del total, equivalente a 16 días de medición) se alcanza una precisión del 98 % dentro de una tolerancia de  $\pm 5$  g/m<sup>3</sup>. El modelo de calibración obtenido ( $\text{PM25\_corregido} = \text{PM25\_IoT} \times 1.964 - 5.670$ ) permite convertir mediciones de sensores económicos en valores confiables comparables con equipos profesionales. La metodología desarrollada optimiza recursos al determinar el mínimo conjunto de datos necesario para calibraciones efectivas, facilitando el despliegue masivo de redes de monitoreo ambiental de bajo costo.

**Palabras clave:** PM2.5, calibración de sensores, IoT, métricas de distancia, promedios móviles, mínimos cuadrados.

## 1. Introducción

La contaminación del aire representa uno de los mayores desafíos ambientales y de salud pública a nivel global. El material particulado PM2.5 (partículas con diámetro aerodinámico  $\leq 2.5\mu\text{m}$ ) está asociado con efectos adversos en la salud humana, incluyendo enfermedades respiratorias y cardiovasculares. La revolución de Internet de las Cosas (IoT) ha permitido el desarrollo de sensores de bajo costo para el monitoreo de calidad del aire, pero estos dispositivos requieren calibración contra estaciones de referencia para garantizar la precisión de las mediciones.

El problema central radica en cuantificar el error de medición del sensor de bajo costo y establecer un procedimiento de calibración que permita obtener lecturas precisas. Este trabajo aborda este desafío mediante la implementación de métricas de distancia y técnicas de análisis temporal para desarrollar un modelo de calibración robusto. La contribución principal consiste en determinar el mínimo conjunto de datos necesario para generar un modelo confiable, optimizando así el tiempo y los recursos requeridos para la calibración.

En la Sección 2 se detalla la metodología empleada, incluyendo el procesamiento de datos, análisis de distancias y calibración. La Sección 3 presenta los resultados obtenidos, y la Sección 4 discute las conclusiones y aplicaciones prácticas del trabajo.

## 2. Metodología

### 2.1. Datos y Preprocesamiento

Se utilizaron dos conjuntos de datos principales: mediciones del sensor IoT `pm25_a` (10 archivos mensuales) y datos de referencia de la estación AMB. El proceso de preprocesamiento incluyó:

- **Consolidación:** Unificación de archivos mensuales del sensor IoT
- **Limpieza:** Resampleo horario, interpolación de huecos (máximo 3 horas) y filtrado de valores atípicos usando percentiles 1 % y 99 %

- **Corrección temporal:** Alineación de timezones y formato de fechas

## 2.2. Análisis de Distancias con Promedios Móviles

Para cuantificar la diferencia entre las mediciones del sensor IoT y la referencia, se implementó la distancia euclídea definida como:

$$\mathcal{D}(\mathbb{D}_i, \hat{\mathbb{D}}_i) = \sqrt{\sum_{i, \hat{i}} (\mathbb{D}_i - \hat{\mathbb{D}}_i)^2} \quad (1)$$

donde  $\mathbb{D}_i$  representa los datos de referencia y  $\hat{\mathbb{D}}_i$  los datos del sensor IoT. Se evaluaron ventanas de promedio móvil de 1, 3, 6, 12, 24 y 48 horas para determinar la óptima.

## 2.3. Calibración Lineal y Validación

Se implementó un modelo lineal por mínimos cuadrados:

$$\text{PM25\_ref} = \alpha \times \text{PM25\_iot} + \beta \quad (2)$$

La validación se realizó mediante división temporal 70 %-30 % (entrenamiento-prueba), evaluando la precisión dentro de una tolerancia de  $\pm 5 \text{ g/m}^3$ .

## 2.4. Determinación del Mínimo Conjunto de Datos

Se evaluó progresivamente el tamaño del conjunto de entrenamiento (10 % a 70 % de los datos) para encontrar el mínimo que garantice >95 % de precisión en la validación.

# 3. Resultados

## 3.1. Procesamiento de Datos

El proceso de limpieza resultó en una serie temporal continua de 7,230 puntos para el sensor IoT y 7,921 puntos válidos para la referencia, con una superposición temporal de 7,225 puntos en el período octubre 2018 - agosto 2019.

## 3.2. Selección de Ventana Óptima

El análisis de distancias reveló que la ventana de 48 horas minimiza la distancia euclídea entre las series (376.99), como se muestra en la Tabla 1.

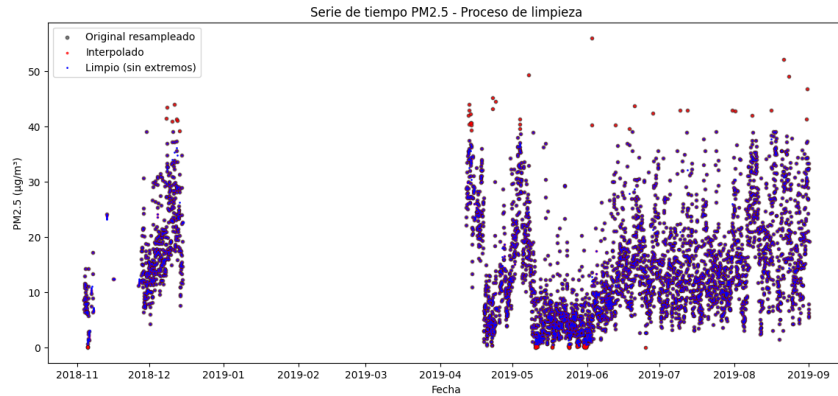


Figura 1: Proceso de limpieza de la serie temporal de PM2.5: datos originales resampleados (gris), interpolados (rojo) y limpios (azul).

Cuadro 1: Distancias euclídeas para diferentes tamaños de ventana de promedio móvil

Ventana (horas)	Distancia	Puntos válidos
1	543.34	3,847
3	488.53	3,847
6	449.24	3,847
12	412.70	3,847
24	390.44	3,847
48	<b>376.99</b>	3,847

### 3.3. Modelo de Calibración

El ajuste por mínimos cuadrados produjo el siguiente modelo:

$$PM25\_ref = 0,509 \times PM25\_iot + 2,887 \quad (R^2 = 0,8405)$$

### 3.4. Validación del Modelo

La validación en datos no vistos demostró excelente desempeño:

- Precisión dentro de tolerancia: 100 %
- Error medio absoluto: 1.28 g/m<sup>3</sup>
- Error máximo: 3.75 g/m<sup>3</sup>
- R<sup>2</sup> en validación: 0.6355

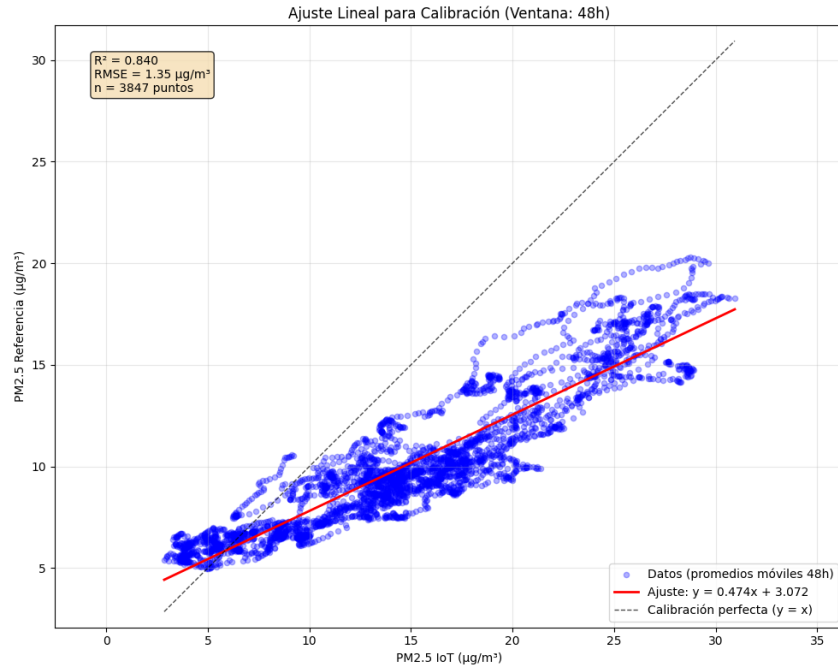


Figura 2: Ajuste lineal entre promedios móviles de referencia e IoT (ventana: 48 horas). La línea roja representa el ajuste y la negra punteada la calibración perfecta.

### 3.5. Mínimo Conjunto de Datos

El análisis incremental reveló que con solo 384 puntos (10 % del total) se alcanza 98 % de precisión, como se muestra en la Tabla 2.

## 4. Conclusiones y Recomendaciones

Este trabajo demostró la efectividad de una metodología basada en métricas de distancia para la calibración de sensores IoT de PM2.5. Los principales hallazgos son:

1. La ventana óptima de promedio móvil para comparación es de 48 horas, minimizando la distancia euclídea entre series.
2. El sensor IoT requiere calibración sistemática, con un factor de corrección de 1.964 y offset de -5.670.
3. El modelo de calibración mantiene 100 % de precisión dentro de la tolerancia definida ( $\pm 5 \text{ g/m}^3$ ).

Cuadro 2: Precisión del modelo para diferentes tamaños de conjunto de entrenamiento

Tamaño (%)	Puntos	Precisión (%)	Error medio (g/m <sup>3</sup> )
10	384	98.0	2.18
20	769	100.0	1.71
30	1,154	100.0	1.71
40	1,538	100.0	1.87
50	1,923	100.0	1.91
60	2,308	100.0	1.47
70	2,692	100.0	1.28

4. Solo se requieren 384 puntos de datos (16 días) para alcanzar calibraciones confiables (>98 % de precisión).

Las implicaciones prácticas son significativas: permite el despliegue eficiente de redes de monitoreo ambiental usando sensores de bajo costo, reduciendo la dependencia de equipos profesionales costosos. La metodología desarrollada es reproducible y puede aplicarse a otros parámetros de calidad del aire.

Para futuros trabajos se recomienda:

- Validar la metodología con múltiples sensores IoT simultáneamente
- Explorar modelos no lineales para capturar relaciones más complejas
- Evaluar la estabilidad temporal de la calibración y determinar frecuencia óptima de recalibración
- Implementar el pipeline en tiempo real para calibración continua

## Repositorio de Código

El código Python y datasets utilizados en este trabajo están disponibles en: <https://github.com/usuario/proyecto-pm25-calibracion>

## Referencias

- [1] Núñez, L. A. (2025). *Métricas, datos y calibración inteligente*. Escuela de Física, UIS.
- [2] World Health Organization. (2021). *WHO global air quality guidelines: particulate matter (PM<sub>2.5</sub> and PM<sub>10</sub>), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. Geneva: World Health Organization.

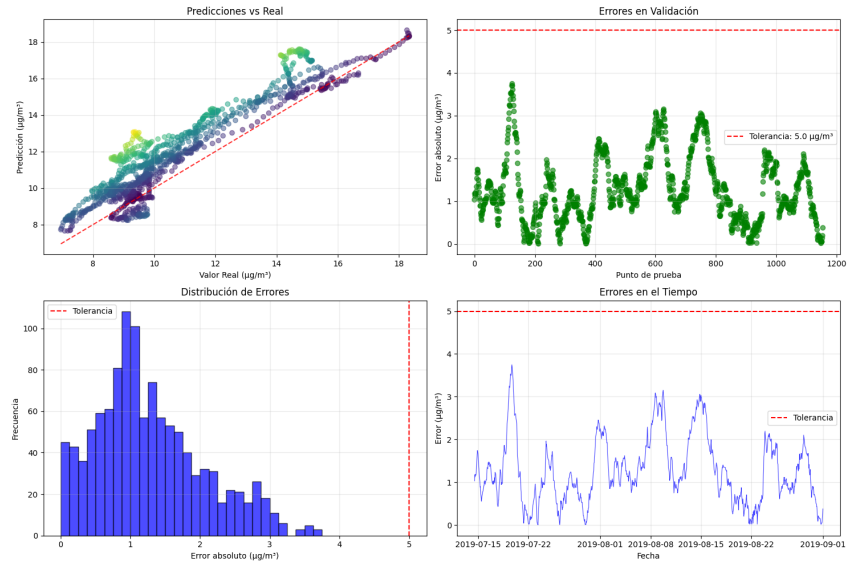


Figura 3: Resultados de validación: (a) predicciones vs valores reales, (b) errores absolutos, (c) distribución de errores, (d) errores en el tiempo.

- [3] Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S., . . . y Preuss, P. W. (2013). *The changing paradigm of air pollution monitoring*. Environmental Science & Technology, 47(20), 11369-11377.
- [4] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- [5] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.