

# MÉTRICA, DATOS Y CALIBRACIÓN INTELIGENTE

J. C. Gómez - J. A. Pitre\*

UNIVERSIDAD INDUSTRIAL DE SANTANDER

Bucaramanga, Santander

25 de septiembre de 2025

## Índice

<b>1. Introducción</b>	<b>2</b>
1.1. Objetivo General . . . . .	2
1.2. Objetivos Específicos . . . . .	2
<b>2. Metodología</b>	<b>3</b>
2.1. Carga y Agrupación de Datos . . . . .	3
2.2. Cálculo de Error y Ajuste . . . . .	3
2.3. Evaluación de Ventanas y Análisis Comparativo . . . . .	3
2.4. Evaluación de Alcance y Validez . . . . .	4
<b>3. Resultados</b>	<b>4</b>
<b>4. Conclusiones</b>	<b>5</b>

## Resumen

En nuestra época, la propagación de sensores de bajo costo en aplicaciones de la Internet de las Cosas (IoT) plantea el desafío de garantizar la precisión y exactitud de sus mediciones. Estos dispositivos, aunque accesibles para muchas personas, requieren ser calibrados frente a estaciones de referencia para asegurar su veracidad. Este trabajo aborda el problema de calibración desde la perspectiva matemática, vinculando la idea de *métrica* con la discrepancia entre datos de referencia y lecturas de sensores IoT.

La metodología consistió en cargar y limpiar al código de Python los conjuntos de datos, agruparlos en ventanas de tiempo y calcular métricas de error como la distancia euclídea y el error cuadrático medio (RMSE). Posteriormente, se aplicó un ajuste lineal mediante un factor de calibración  $\alpha$ , entrenado con un subconjunto de datos y evaluado en un conjunto de prueba.

Los resultados muestran que un intervalo óptimo de 30 minutos permite obtener un factor de calibración de  $\alpha = 0,7434$ , reduciendo el RMSE de 6,3175 a 4,2982  $\mu g/m^3$ . Además, se tuvo un

---

\* e-mail: jhoam2231544@correo.uis.edu.co

alcance del 77,5 % dentro de la tolerancia de  $5,0 \mu\text{g}/\text{m}^3$ , con un mínimo de  $N_{\min} = 80$  ventanas de entrenamiento y un horizonte continuo de validez de 7 horas.

De este modo, confirman la importancia de la calibración inteligente como estrategia fundamental para garantizar la validez de sensores IoT en escenarios de monitoreo ambiental y tecnológico.

## 1. Introducción

En la actualidad, vivimos una época caracterizada por un crecimiento acelerado en el número de sensores desplegados en nuestro entorno cotidiano. Estos dispositivos, en su mayoría de bajo costo, forman parte de sistemas de monitoreo ambiental, industrial y biomédico, y constituyen un eje central de la Internet de las Cosas (IoT) por sus siglas en inglés “*Internet of Things*” [1]. Sin embargo, la ventaja económica y la escalabilidad de dichos sensores suelen estar acompañadas de limitaciones en precisión y estabilidad, lo que hace necesaria su calibración frente a un patrón de referencia [2].

Desde la perspectiva de la metrología, la calibración no solo permite garantizar la exactitud de las mediciones, sino también asegurar la veracidad y la evaluación de la incertidumbre asociada [3]. Por otra parte, en redes de sensores distribuidas surge la posibilidad de emplear técnicas de calibración colaborativa o incluso sin referencia, lo cual representa un campo de investigación en expansión [4].

Matemáticamente, la calibración se encuentra íntimamente ligada al concepto de *métrica*: la elección de una función de distancia o error define la forma en que se mide la discrepancia entre las lecturas de los sensores y los valores de referencia. Así, métodos como los mínimos cuadrados se asocian al uso de normas cuadráticas, mientras que el empleo de métricas robustas permite reducir la influencia de valores atípicos o poco comunes [5].

De este modo, la calibración de sensores en el contexto de la IoT no solo representa un problema tecnológico, sino también un escenario donde los métodos matemáticos de estimación, optimización y análisis de métricas cobran relevancia, alineándose con los objetivos de este curso.

Por lo tanto, el problema está en cuantificar cuál es el error de medición del sensor de bajo costo y, cómo calibrarlo para que podamos establecer nuevas lecturas que sean más precisas. En la Sección 2 discutimos la metodología empleada para el desarrollo del problema, mientras que en la Sección 3 se presentan los resultados obtenidos y finalizamos el artículo con las conclusiones en la Sección 4.

### 1.1. Objetivo General

Cuantificar los errores de los sensores y calibrarlos mediante un patrón de referencia.

### 1.2. Objetivos Específicos

- Calcular la distancia euclídea para varios valores de la ventana móvil
- Determinar el alcance del modelo lineal
- Determinar el alcance para las predicciones dentro de la tolerancia dada

- Determinar el mínimo conjunto de datos para generar el modelo y cuál será su máximo alcance para una tolerancia dada.

## 2. Metodología

Para calibrar los datos tomados por los sensores de bajo costo y ajustarlos con respecto a un conjunto de datos de referencia, se utilizó un código en Python <sup>1</sup>, el cual se desglosará por fases en las siguientes secciones.

### 2.1. Carga y Agrupación de Datos

Para empezar se cargaron al código los datos de referencia desde una hoja de excel <sup>2</sup> y los datos tomados por una estación de IoT <sup>3</sup>, dado que existían columnas con datos faltantes para cierta fecha, se realizó una limpieza de esos valores numéricos. Además, se estandarizó el tiempo en el que los datos fueron tomados; una vez que los datos a tratar estén ordenados, se procedió a la agrupación de dichos valores por ventanas de tiempo tanto en los datos de referencia como en los de los sensores. Para este caso, la ventana se tomó en un intervalo de tiempo de 30 minutos; luego, se calculó un promedio para los valores de cada ventana y se tuvieron en cuenta únicamente aquellas en las que el intervalo de tiempo coincidía en ambos archivos de datos.

### 2.2. Cálculo de Error y Ajuste

En segundo lugar, se debe hallar la diferencia bruta entre las ventanas ya alineadas de ambos grupos de datos (los de referencia y los de IoT); esto se hace mediante el cálculo de la distancia euclídea y el error cuadrático medio. De este modo, para realizar el ajuste lineal de los datos se proporcionó un factor de calibración ( $\alpha$ ) que minimiza el error cuadrático entre la serie de referencia y la serie IoT escalada ( $\alpha \times IoT$ ) mediante un enfoque de mínimos cuadrados.

### 2.3. Evaluación de Ventanas y Análisis Comparativo

En esta fase, se consideraron distintos tamaños de ventana para examinar su efecto en el proceso de calibración. Para cada tamaño de ventana se repite el procedimiento de ventaneo, alineación, cálculo de errores y ajuste de  $\alpha$ . Los resultados se resumen en una tabla (ver Tabla 1) que compara los valores del error cuadrático medio antes y después de aplicar la calibración.

Posteriormente, se selecciona el tamaño de ventana que produce el menor RMSE tras la calibración. Para esta ventana óptima, se visualizan tanto la serie temporal original como la calibrada, permitiendo observar el efecto de la corrección sobre la correspondencia entre ambos conjuntos de datos (ver Fig 1a).

<sup>1</sup>[https://colab.research.google.com/drive/1fn8m-CSClqx54aAuGK\\_zzt-by\\_tFjyG\\_](https://colab.research.google.com/drive/1fn8m-CSClqx54aAuGK_zzt-by_tFjyG_)

<sup>2</sup><https://github.com/nunezluis/MisCursos/tree/main/MisMateriales/Asignaciones/TallerDistancias/DatosDistancias>

<sup>3</sup><https://blissair.com/what-is-pm-2-5.htm>

	Tamaño	Número de ventanas	RMSE	Factor de calibración	RMSE después de calibración
0	10	136	6.431119	0.666924	4.696535
1	15	137	6.281412	0.665400	4.618873
2	20	138	6.303331	0.663080	4.617393
3	30	141	6.317478	0.660355	4.582270
4	45	143	6.281477	0.666360	4.618663
5	60	143	6.237432	0.668715	4.598528

Cuadro 1: Resultados de calibración para diferentes tamaños de ventana

## 2.4. Evaluación de Alcance y Validez

En esta última fase, luego de seleccionar la ventana óptima, se dividieron sus datos en conjunto de entrenamiento y prueba, el parámetro de calibración  $\alpha$  se ajusta usando el conjunto de datos de entrenamiento y se evaluó qué tan bien se cumple en la serie IoT calibrada y si coincide con la de referencia en el conjunto de prueba dentro de una tolerancia específica. Se informó el porcentaje de puntos del conjunto de prueba que caen dentro de esa tolerancia y el rango de valores IoT en los que se cumple dicha condición.

Finalmente, se determinó la cantidad mínima de datos de entrenamiento necesaria para alcanzar un nivel de cobertura objetivo en un bloque de prueba fijo. Con base en este ajuste, se calculó también el horizonte temporal máximo en el que la serie calibrada del IoT se mantiene dentro de la tolerancia especificada, lo que permite establecer los límites de confiabilidad del sensor calibrado.

## 3. Resultados

Posteriormente a la selección de la mejor ventana se realizó un gráfico que muestra la comparación entre los datos de referencia y los datos de IoT antes y después del ajuste con el factor  $\alpha$ , (ver Fig 1), en el cual se puede observar que tras el ajuste con el factor lineal  $\alpha$  la discrepancia es en promedio menor que antes. Por lo tanto, al introducir este nuevo factor se permite a los datos tener un mejor ajuste para su comparación, dando los siguientes valores [6]: antes de la calibración el error cuadrático medio era de  $6,3175\mu g/m^3$  y luego de la calibración era de  $4,5823\mu g/m^3$  con un valor de  $\alpha$  ( $\alpha = 0,660355$ ).

Adicionalmente, el procedimiento de calibración lineal aplicado al conjunto de entrenamiento de los datos del sensor IoT frente a la estación de referencia arrojó los siguientes resultados principales. Se determinó que el ventaneo óptimo corresponde a un intervalo de 30 minutos, con un factor de calibración de  $\alpha = 0,7434$ . En el conjunto de entrenamiento, el error cuadrático medio fue de  $4,2982\mu g/m^3$  mejorando el valor obtenido del factor en la ventana más óptima.

En la fase de validación, utilizando una tolerancia absoluta de  $5,0\mu g/m^3$ , se observó que el 77,5% de las mediciones del conjunto de prueba se mantuvo dentro de dicha tolerancia. El rango de validez para las lecturas del sensor IoT se encontró entre 6,429 y  $21,286\mu g/m^3$ , intervalo en el cual la calibración es confiable.

Asimismo, se comprobó que es suficiente utilizar únicamente 80 ventanas de entrenamiento

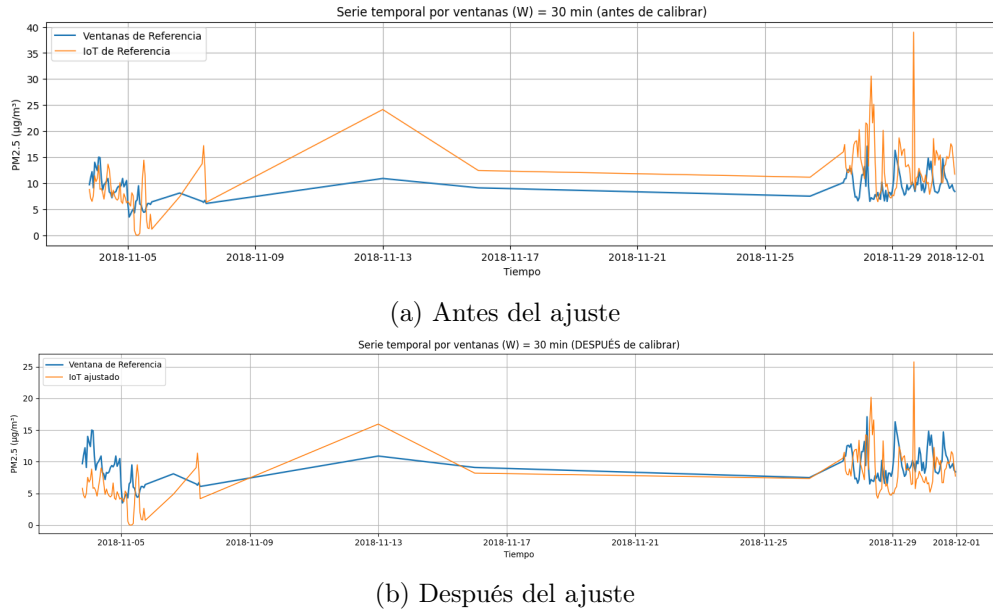


Figura 1: Comparación de los datos de referencia con los datos tomados por la IoT. (a) Muestra la comparación antes del ajuste, (b) Muestra la comparación después del ajuste de  $\alpha$

para alcanzar una cobertura del 85,7 % en el conjunto de prueba, obteniendo un nuevo valor de  $\alpha = 0,6404$ . Finalmente, la calibración se mantuvo válida de forma continua durante un horizonte de aproximadamente 14 ventanas, es decir, cerca de 420 minutos (7 horas), específicamente entre las 11:00 del 28 de noviembre y las 00:00 del 29 de noviembre de 2018 <sup>4</sup>.

## 4. Conclusiones

En este trabajo se ajustaron una serie de datos obtenidos por estaciones *IoT* a un conjunto de datos tomados en estaciones de referencia, utilizando aproximaciones lineales y relacionando la discrepancia entre los conjuntos con la idea de métrica en el espacio (la distancia entre las cosas). Para comprobar la validez del procedimiento, los datos fueron ajustados mediante un factor lineal  $\alpha$ , el cual se “entrenó” con un subconjunto y se evaluó en otro de prueba. Este proceso arrojó un valor de  $\alpha = 0,7434$ , logrando que la calibración de los datos presentara la menor diferencia respecto a las mediciones de las estaciones de referencia (AMB). Asimismo, se determinó la cantidad mínima de ventanas necesarias para la aproximación de los conjuntos ( $N_{min} = 80$ ), con lo cual se obtuvo que el 77,5 % de los datos permaneciera dentro de la tolerancia establecida de  $5,0 \mu\text{g}/\text{m}^3$ .

Dado que vivimos en un mundo en constante cambio y evolución tecnológica, el uso de sensores de bajo costo se ha vuelto indispensable en la cotidianidad, encontrando aplicaciones en la industria, el comercio e incluso en el hogar. Por esta razón, resulta de suma importancia la calibración de dichos

<sup>4</sup>Para más detalles y claridad sobre el código puede dirigirse al colab: [https://colab.research.google.com/drive/1fn8m-CSCIQx54aAuGK\\_zzt-by\\_tFjyG\\_](https://colab.research.google.com/drive/1fn8m-CSCIQx54aAuGK_zzt-by_tFjyG_)

sensores bajo un patrón de referencia preestablecido, evitando así la toma de datos erróneos o poco representativos de la realidad.

## Referencias

- [1] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645–1660, 2013.
- [2] Ali Alhashimi. *Statistical Sensor Calibration Algorithms*. PhD thesis, Linköping University, 2018.
- [3] Joint Committee for Guides in Metrology. Evaluation of measurement data—guide to the expression of uncertainty in measurement (gum 1995 with minor corrections). Technical report, JCGM, 2008.
- [4] Jing Wang, Laura Balzano, and Robert Nowak. Reference-free calibration in sensor networks. *arXiv preprint arXiv:1805.11999*, 2018.
- [5] Peter J Rousseeuw and Annick M Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, 2005.
- [6] J. C. Gómez. Asignación 1, 2025. Código y experimentos implementados en Google Colab.