

Métricas, datos y calibración inteligente

E.D Fonseca*

E.Paez**

Universidad Industrial de Santander

Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones

Versión 1 9 de Sep 2021

Índice

1. Introducción	2
2. Metodología	2
3. Resultados	9
4. Conclusiones	10
5. Referencias	10

Resumen

El propósito de estudio consistió en el análisis de las métricas de un sensor de bajo costo (IoT) y su calibración respecto a las mediciones de la estación de monitoreo del Sistema de Vigilancia de Calidad del Aire, en Santander Colombia. La comparación de los datos del sensor IoT generados instalados en paralelo con la estación de referencia, contribuye a la evaluación del rendimiento y la precisión de los sistemas de sensores lowcost en un contexto del mundo real, y respalda su calibración y desarrollo posterior.

A continuación se presenta el proceso de desarrollo de un algoritmo de calibración para dicho sensor, como anexo se presenta el código completo de Python, que permite su reproducibilidad ajustando las mediciones tomadas por el sensor IoT con las mediciones de referencia de la estación AMB.

* e-mail:duvanfonsecabril@gmail.com

** e-mail: edwin2218421@correo.uis.edu.co

1. Introducción

En el año 2011 la firma CISCO (Internet Business Solutions Group) señalaba que para el 2020, estarían conectados a internet 50 mil millones de dispositivos, cada uno con múltiples sensores, no obstante, dichas estimaciones no tenían en cuenta el aumento exponencial de las tecnológicas, avances en IoT, macrodatos (bigdata) entre otros, tal cantidad de dispositivos y sensores existentes, suponen una serie de retos, como lo son la energía que se debe suministrar, la calibración/actualización de los mismos y el aprovechamiento de los datos obtenidos.

En el informe del 2018 la agencia Europea de Medio Ambiente exponían sus resultados basados en sensórica de bajo coste, entre los principales problemas observados se expresaba una preocupación de la escasa fiabilidad de los datos medidos por dichos sistemas, además, uno de los mayores retos en este ámbito corresponde a la dificultad de conciliar las medidas realizadas por las redes de bajo coste con los equipos de monitorización ambiental certificados. Generalmente los datos obtenidos por los sensores lowcost se comparan con los datos de referencia, para obtener un algoritmo de calibración (Borrego et al., 2016)(Buehler et al., 2020)(Cross et al., 2017)(Spinelle et al., 2017) y dada la variabilidad de las condiciones ambientales y el rendimiento de los sensores de bajo coste, este tipo de comparación se debe realizar de manera regular y tomar entre 2 y 3 semanas de datos a comparar (Spinelle et al., 2017). Por lo tanto, la caracterización cuidadosa del sensor, la calibración y el procesamiento de datos son importantes para garantizar la precisión de la medición.

En este estudio se realizó la cuantificación del error en la medición de un sensor de bajo costo respecto a las mediciones realizadas por la estación AMB - Normal en Santander, Colombia, que mide la cantidad de partículas atmosféricas (PM) que tienen un diámetro de menos de 2.5 micrómetros suspendidas en el aire (PM2.5). En la primera sección se especifican los procesos llevados a cabo para la resolución del problema desde el análisis de los datos

2. Metodología

En una primera instancia se ordenaron los datos de la estación AMB - NORMAL y los datos obtenidos por los sensores IoT de acuerdo con la fecha y hora en la cual fueron obtenidos. Para el análisis de datos se procedió a seleccionar los datos que se obtuvieron en las mismas fechas y de los cuales se contarán con al menos 24 registros diarios, como se muestra en la Figura 1, existen espacios de tiempos donde no se cuentan con registros para la estación AMB, el sensor IoT o para ambos simultáneamente, esto dio como resultado una lista de datos desde el 7/06/2019 hasta el 31/08/2019 (Ver Cuadro 1), espacio de tiempo donde se cuenta con una continuidad en los registros de la estación AMB y el sensor IoT.

El análisis, manipulación de datos y los diferentes cálculos se efectuó haciendo uso de las siguientes librerías de Python:

- pandas as pd
- matplotlib.pyplot as mpl
- matplotlib.ticker as tk

- seaborn as sns
- pylab as pl
- numpy as np
- from scipy.special import legendre
- math
- xlrd



Figura 1: Diagramas de dispersión - Estación AMB Normal y Sensor IoT

Cuadro 1: Comparación de datos de la Estación AMB - Normal y Sensores IoT, ordenados en orden cronológico desde el 7 de junio hasta el 31 de agosto

AMB			IoT	
<i>Fecha</i>	<i>Hora</i>	<i>Valor</i>	<i>Hora</i>	<i>Valor</i>
7/06/2019	0:00:00	7.7	0:02:49	9.916666
	1:00:00	8.2	1:03:00	10.999999
	2:00:00	11.3	2:02:50	21.25
31/08/2019	21:00:00	22.8	21:02:49	32.41667
	22:00:00	25.2	22:02:49	33.25
	23:00:00	18	23:03:00	6.916666

La figura 2 muestra el grafico de líneas para los datos ordenados desde el 7/06/2019 hasta el 31/08/2019, en total se obtuvieron 2063 datos continuos, 24 registros diarios para el sensor IoT y la estación AMB.

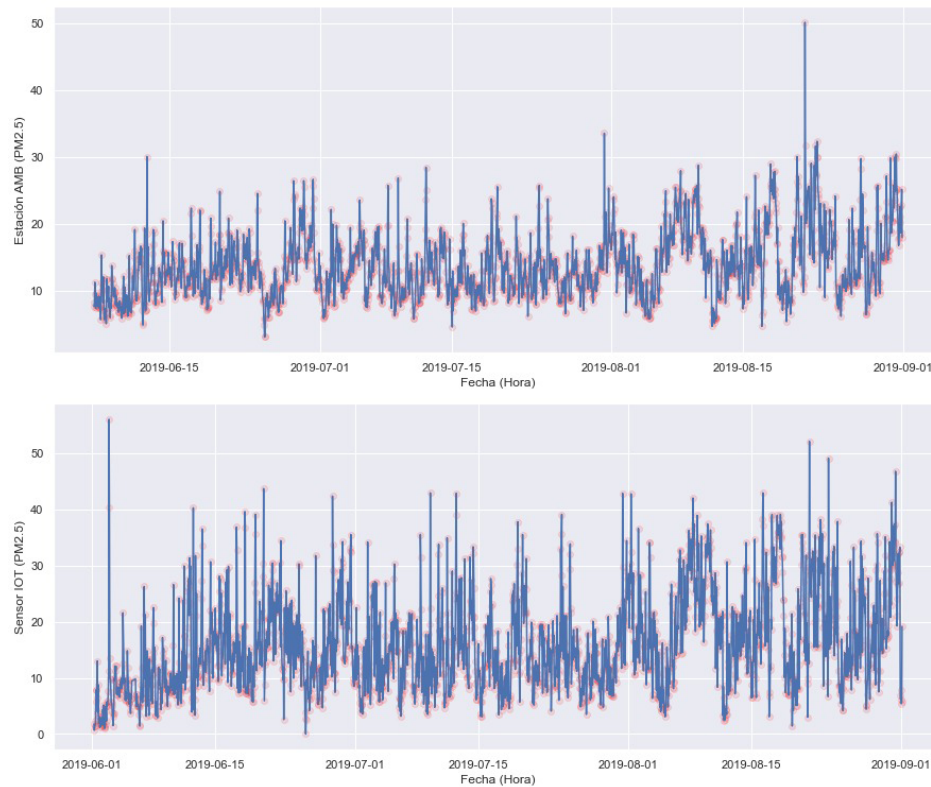


Figura 2: Representación Grafica, datos obtenidos por la estación y el sensor IoT por hora desde el 7/06/2019 hasta el 31/08/2019

Si bien el espacio de tiempo corresponde al mismo para los datos de la estación AMB como para el Sensor IoT, las mediciones se realizaron en diferentes horas, por lo cual fue necesario obtener el promedio diario para dicho espacio de tiempo, dando como resultado un total de 86 registros, estos datos fueron tabulados (ver Cuadro 2) y en la Figura 3 se muestra la comparación grafica de los datos promedios.

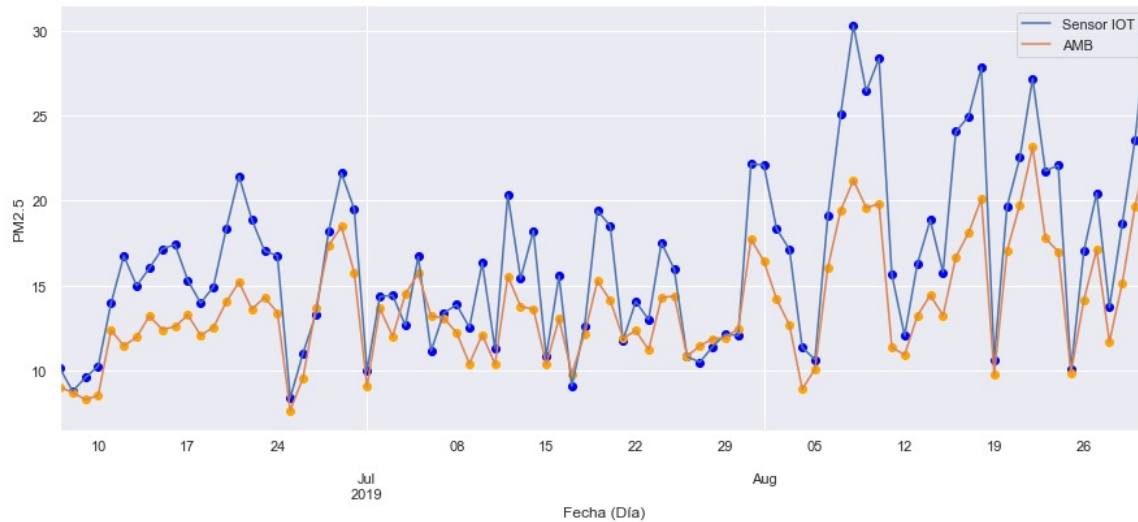


Figura 3: Promedios de datos diarios desde el 7/06/2019 hasta el 31/08/2019 para estación AMB y Sensor IoT

Cuadro 2: Tabulación valores promedios diarios para la estación AMB y sensor IoT

Valores		
<i>Fecha</i>	<i>AMB</i>	<i>IoT</i>
7/06/2019	9,029166	10,131944
8/06/2019	8,708333	8,784722
9/06/2019	8,283333	9,586805
29/08/2019	15,129166	18,611111
30/08/2019	19,65	23,572916
31/08/2019	22,869565	30,347222

De esta manera se obtuvieron mediciones comparables y una vez ordenados y analizados se procedió a encontrar la distancia entre los puntos, en primera instancia se reescalaron los "X" para poder encontrar el polinomio de interpolación como combinación lineal de los polinomios de Legendre, desde 1 a 86 que corresponden al primer y último día, hasta -1 a 1, mediante el uso del algoritmo que se muestra en la figura 4:

```
# Para el calculo de Las nuevas X
X = []
for i in np.arange(-1, 1+float(2/85), float(2/85)):
    X.append(round(i, 6))
```

Figura 4: Algoritmo para cambio de variable

Los nuevos X se tabularon y a partir de ellos se generaron 3 ecuaciones correspondientes a los primeros 3 datos (7/09/2019 - 10/06/2019), para ello se calcularon los coeficientes de la serie, para plantear un sistema de ecuaciones inicial (Ver Figura 5) correspondiente a la primer parte de la función a trazos que definen los valores de los promedios diarios.

```
# Para el cálculo de los coeficientes para hallar los polinomios con las nuevas X
P = []
for i in np.arange(0, len(Comparar['AMB'])):
    fila = []
    for j in np.arange(0, 3):
        pol = legendre(j)
        fila.append(round(pol(X[i]), 6))
    P.append(fila)

# Resuelve el sistema de ecuaciones para hallar C0, C1 y C2
inicio = 0
final = 7
A = np.matrix([[2, 3], [1, -2]])
b = np.matrix([[8], [-10]])
x = (A**-1)*b
resultado = []
for i in range(0, 100):
    A = np.matrix([P[inicio], P[inicio+1], P[inicio+2]]), P[inicio+3], P[inicio+4], P[inicio+5], P[inicio+6])
    B = np.matrix([Comparar['AMB'][inicio], Comparar['AMB'][inicio+1], Comparar['AMB'][inicio+2]]), Comparar['AMB'][inicio+3], Comparar['AMB'][inicio+4], Comparar['AMB'][inicio+5], Comparar['AMB'][inicio+6])
    C = (A**-1)*B
    inicio = inicio + 1
    if inicio > 83:
        inicio = 83
        A = np.matrix([P[inicio], P[inicio+1], P[inicio+2]]), P[inicio+3], P[inicio+4], P[inicio+5], P[inicio+6])
        B = np.matrix([Comparar['AMB'][inicio], Comparar['AMB'][inicio+1], Comparar['AMB'][inicio+2]]), Comparar['AMB'][inicio+3], Comparar['AMB'][inicio+4], Comparar['AMB'][inicio+5], Comparar['AMB'][inicio+6])
        C = (A**-1)*B
        resultado.append(C)
        break
    else:
        resultado.append(C)
```

Figura 5: Lineas de Código Python para la generación del primer sistema de ecuaciones

Para la generación de la siguiente ecuación del siguiente grupo de datos, se tomaron los últimos dos datos de la anterior ventana y se sumó el siguiente dato es decir los datos desde el 8/06/2019 hasta el 9/09/2019 como se observa en la figura 6, la tercera ecuación desde el 9/06/2019 hasta el 11/06/2019 y de esta manera hasta completar el ultimo día.

		P1	Fecha	AMB	Sensor IoT
	P2		7/06/2019	9,03	10,13
P3			8/06/2019	8,71	8,78
			9/06/2019	8,28	9,59
			10/06/2019	8,54	10,26
			11/06/2019	12,37	13,95

Figura 6: Ventana móvil

En total se hallaron 84 polinomios de orden 2 que definen una función a trazos para los datos de la estación AMB y 84 para el Sensor IoT, el proceso se llevo a cabo mediante el uso de las siguientes líneas de código (Ver figura 7) En la parte inferior se aprecian los primeros 3 polinomios.

```

Y = []
for i in np.arange(0, len(Comparar['AMB'])-2):
    pol = resultado[i].tolist()[0][0]*legendre(0) + resultado[i].tolist()[1][0]*legendre(1) + resultado[i].tolist()[2][0]*legendre(2)
    if i == 84:
        n = 2
    else:
        n = 1
    for j in np.arange(X[i], X[i+n], 0.002352941176):
        Y.append(round(pol(j), 6))
    print(pol)

2
-94 x - 199.4 x - 96.4
2
613.6 x + 1166 x + 562.1
2
3233 x + 6096 x + 2882
2
-4289 x - 7708 x - 3451

```

Figura 7: Calculo de los polinomios

Una vez calculados los diferentes polinomios fue posible obtener una gráfica para el sensor IoT y la Estación AMB, las cuales se aprecian en la figura 8, en ella se puede evidenciar una curva suavizada a lo largo de los puntos experimentales diferente a la curva inicial (Figura 3). A partir de los polinomios encontrados es posible el cálculo de la distancia euclídea entre las dos mediciones, haciendo uso de la ecuación general.

Para el proceso de calibración del sensor, se encontró la diferencia entre los valores del Sensor IoT y los valores de la estación AMB (Distancia), en la figura 9 se pueda observar la diferencia punto a punto entre los valores, en el eje X se muestran los valores del Sensor IoT y en el eje Y la diferencia correspondiente a los valores de la estación AMB, es decir si los datos fuesen iguales los valores en Y serían 0. La línea azul corresponde a los promedios de 8 ventanas de valores, datos con los cuales se obtuvo una función polinómica de grado 7 por medio de los polinomios de Lagrange. La línea de código correspondiente a este proceso se muestra en la figura 10.



Figura 8: Graficas polinomios

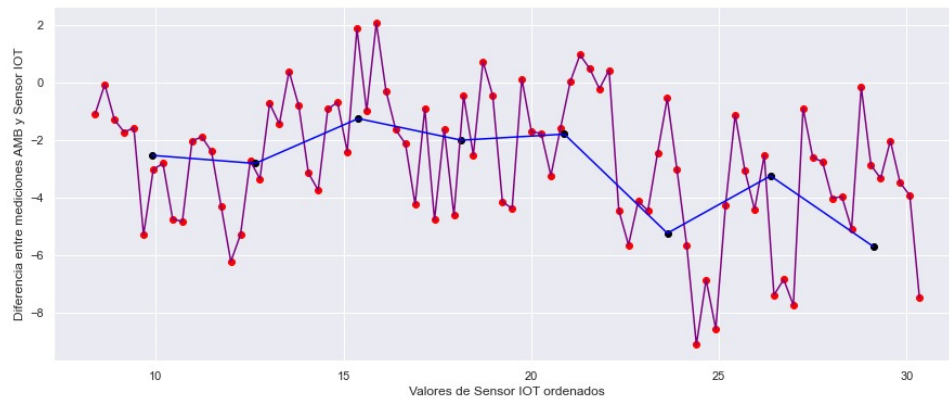


Figura 9: Diferencia punto a punto entre los valores del sensor IoT y la estación AMB, con valores promedio para 8 ventanas de datos


```

Polinomios = {}
px_0 = []
P0 = lagrange(prom_X, prom_D)
print(P0)
for i in np.arange(prom_X[0], prom_X[5], 0.20110347263812983):
    px_0.append(P0(i))

# Para la gráfica
fig, ax = plt.subplots(figsize=(15,6))
plt.xlabel('Valores de Sensor IOT ordenados')
plt.ylabel('Diferencia entre mediciones AMB y Sensor IOT')
plt.scatter(np.arange(X[0], X[len(X)-1], 0.2584150279), D, color = 'red', alpha = 1)
plt.plot(np.arange(X[0], X[len(X)-1], 0.2584150279), D, color = 'purple')
plt.scatter(np.arange(9.902870346539533, 30.321180393298434, 2.745659672), prom_D, color = 'black', alpha = 1)
plt.plot(np.arange(9.902870346539533, 30.321180393298434, 2.745659672), prom_D, color = 'blue')
plt.plot(np.arange(prom_X[0], prom_X[5], 0.20110347263812983), px_0, color = 'green')
plt.show()

```

Figura 10: Línea de python para la obtención del polinomio y su respectiva graficación

3. Resultados

Por medio del cálculo de las distancias fue posible generar un algoritmo de calibración, en el cual se introducen los valores medidos por el sensor y mediante la evaluación en un polinomio de grado 7:

$$-3,052e^{-5}x^7 + 0,003883x^6 - 0,2071x^5 + 6,005x^4 - 102,3x^3 + 1023x^2 - 5573x + 1,274e^4$$

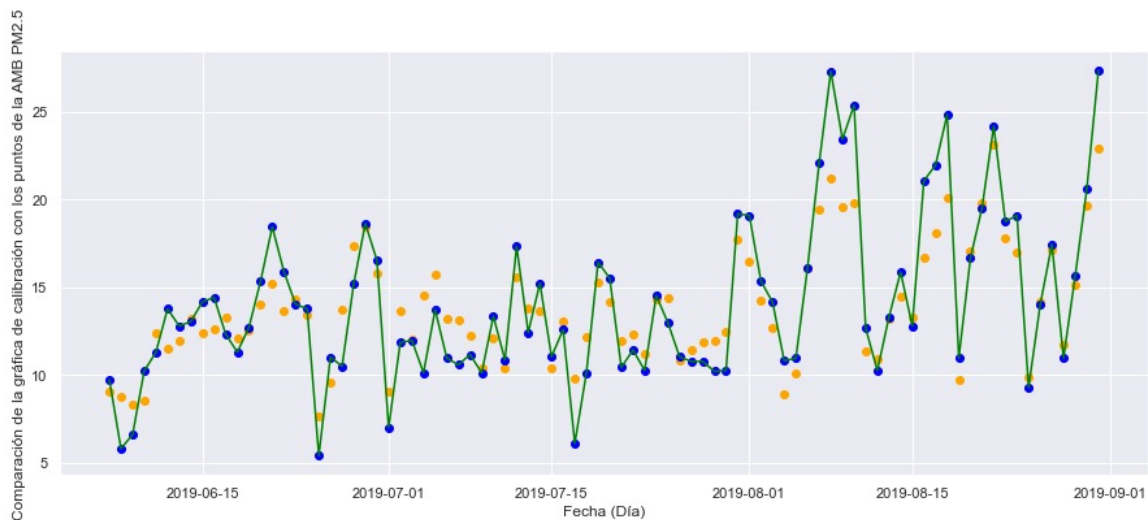


Figura 11: Comparación datos ajustados con el algoritmo de calibración y datos estación AMB

Es posible ajustarlos a los valores proporcionados por la estación de la AMB, se observa que éste ajuste es mayormente válido para los valores comprendidos entre 9.90 y 19.96 aproximadamente, es decir, que para los valores que arroje el sensor IOT para éste intervalo, se pasan por el polinomio de orden 7 anterior. En la figura 11 se aprecian los datos del sensor IoT después de ajustarse con el algoritmo. Los puntos amarillos corresponden a los datos de la estación AMB y en ella se aprecia una disminución en la distancia.

4. Conclusiones

El algoritmo de calibración otorga una disminución del error aproximadamente del 21.6 % si bien los valores contribuyen al mejoramiento en la toma y medición de datos, es posible ajustar aún más la curva de calibración iterando diferentes ventanas de tiempo y aumentando el grado del polinomio.

5. Referencias

- Borrego, C., Costa, A., Ginja, J., Amorim, M., Coutinho, M., Karatzas, K., Sioumis, T., Katsifarakis, N., Konstantinidis, K., De Vito, S., Esposito, E., Smith, P., André, N., Gérard, P., Francis, L., Castell, N., Schneider, P., Viana, M., Minguillón, M., Reimringer, W., Otjes, R., von Sicard, O., Pohle, R., Elen, B., Suriano, D., Pfister, V., Prato, M., Dipinto, S. and Penza, M., 2016. Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise. *Atmospheric Environment*, 147, pp.246-263.
- Buehler, C., Xiong, F., Zamora, M., Skog, K., Kohrman-Glaser, J., Colton, S., McNamara, M., Ryan, K., Redlich, C., Bartos, M., Wong, B., Kerkez, B., Koehler, K. and Gentner, D., 2020. *Stationary and Portable Multipollutant Monitors for High Spatiotemporal Resolution Air Quality Studies including Online Calibration*.
- Cross, E., Williams, L., Lewis, D., Magoon, G., Onasch, T., Kaminsky, M., Worsnop, D. and Jayne, J., 2017. Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. *Atmospheric Measurement Techniques* 10(9), pp.3575-3588
- Spinelle, L., Gerboles, M., Villani, M., Aleixandre, M. and Bonavitacola, F., 2017. Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO₂. *Sensors and Actuators B: Chemical*, 238, pp.706-715.