Calibración de Sensores de Bajo Costo

Samuel Dario Rico Hurtado - 2230653

September 2025

1 Introducción

Los sensores son dispositivos que están presentes en la mayoría de los dispositivos en la actualidad y es que son tan útiles y versátiles que son indispensables en la actualidad. Sin embargo estos aparatos no son perfectos y muchas veces tienen errores en la toma de información.
Por lo que en el presente capitulo vamos a comparar los datos o el patrón tomado por el sensor, con los datos "reales" los cuales llamamos patrón de referencia y trataremos por medio del método de aproximación de mínimos cuadrados, calibrar el sensor para reducir el error entre los dos patrones de datos .

2 Metodología

El objetivo de este informe es comparar los datos tomados con el sensor con el patrón de referencia, medir el error entre ambos y calibrar el sensor para minimizar el error, por lo tanto, mostraremos que métodos matemáticos y computacionales utilizamos para llevar a cabo estos objetivos.
Los datos del patrón de referencia los obtenemos por medio del repositorio del profesor Luis Núñez en el archivo llamado "Datos Estaciones AMB.xlsx".
Para este informe tomaremos unicamente las mediciones del sensor que van del 2018 - 11- 1 al 2018 - 11- 30, que se encuentran en el mismo repositorio del patrón de referencia.
Luego de obtener el conjunto de datos que vamos a comparar, procedemos a la parte computacional. Vamos a elaborar un código en Google collab para poder procesar la gran cantidad de datos que poseemos.

2.1 Error de medición

Lo primero que haremos es calcular el error de medición de los sensores de bajo costo, para esto debemos hallar la distancia euclidiana entre las mediciones del sensor y el patrón de referencia. La distancia euclidiana la obtenemos de la siguiente manera:

$$D(D,\hat{D}) = \sqrt{\sum_j (f(\epsilon_j) - \hat{f}(\epsilon_j))^2} \tag{1}$$

Siendo $f(\epsilon_j)$ los datos del patrón de referencia para un tiempo ϵ_j y siendo $\hat{f}(\epsilon_j)$ los datos de los sensores para el tiempo ϵ_j.
Por lo que debemos emparejar los datos que tienen tiempos iguales, restarlos, elevarlos al cuadrados, sumar los cuadrados y aplicarles raíz para hallar la distancia euclidiana. Sin embargo los dos conjuntos de datos difícilmente coinciden en tiempo, por lo que debemos precisar una ventana de tiempo para obtener el promedio de las mediciones de los dos conjuntos de datos en esa ventana y emparejar esos promedios de las mediciones.
Luego de especificar la ventana de tiempo y sacar los promedios para los dos conjuntos de datos, podemos hallar el error de medición o la distancia euclidiana para esa ventana de tiempo.
En el código encontramos el error de medición para varias ventanas de tiempo distintas para observar en que ventana se asemejan mas ambos conjuntos de datos. En el código de python encontramos la distancia euclidiana para ventanas de 1 hora, 3 horas, 6 horas, 12 horas y 24 horas.
Ademas de hallar la distancia euclidiana que nos da el error de medición de los sensores, también hallaremos otro tipo de valores como el error cuadrático medio(RMSE) Y el error absoluto medio.
El error cuadrático medio(RMSE) estará dado por:

$$RMSE = \sqrt{\frac{1}{N}\sum_j (f(\epsilon_j) - \hat{f}(\epsilon_j))^2} \tag{2}$$

Donde la N representa el numero total de pares de datos que estamos comparando. Por otro lado el error absoluto medio esta dado por:

$$MAE = \frac{1}{N}\left|\sum_j f(\epsilon_j) - \hat{f}(\epsilon_j)\right| \tag{3}$$

2.2 Calibración

Ya teniendo la distancia o el error que hay entre las mediciones del sensor y las mediciones del patrón de referencia, podemos tratar de hacer un ajuste a las mediciones del sensor para disminuir este error.
El ajuste se hace por medio de aproximación de funciones por el método de mínimos cuadrados. Estableceremos parejas coordenadas de la forma (x,y) = $(f(\epsilon), \hat{f}(\epsilon))$, entonces deberemos hallar una función lineal que mejor se aproxime a todos los puntos coordenados.

$$f(\epsilon_j) = a * \hat{f}(\epsilon_j) + b \tag{4}$$

Con el método de mínimos cuadrados, hallamos el par de valores a y b tal que las mediciones del sensor se ajusten de la mejor manera al patrón de referencia. Luego haremos el mismo proceso de calibrar el sensor hallando la función lineal que mejor aproxime los conjuntos de datos, pero lo realizaremos con la mitad de los datos. Esto para poder aplicar nuestra calibración a la otra mitad de los datos y poder observar que tan bien funciona nuestro ajuste o calibración.

3 Tratamiento de datos

Una pequeña parte de los datos del patrón de referencia es:

```
print(patron_referencia.head())

             Date&Time    PM10   PM2.5 NO2 ppb  O3 ppb Temp_Aire  Lluvia  \
0            Date&Time   ug/m3   ug/m3   ug/m3   ug/m3         C      mm
1  2018-10-01 00:00:00  NoData  NoData  NoData  NoData    NoData  NoData
2  2018-10-01 01:00:00  NoData  NoData  NoData  NoData    NoData  NoData
3  2018-10-01 02:00:00  NoData  NoData  NoData  NoData    NoData  NoData
4  2018-10-01 03:00:00  NoData  NoData  NoData  NoData    NoData  NoData

  Humedad_Relativa      WD      WS R_Solar
0                %  grados     m/s    W/m2
1           NoData  NoData  NoData  NoData
2           NoData  NoData  NoData  NoData
3           NoData  NoData  NoData  NoData
4           NoData  NoData  NoData  NoData
```

Figure 1: Patrón de Referencia

Podemos notar que cada dato esta separado por 1 hora, por lo que en la medición de los sensores establecemos una ventana de 1 hora, resultando los datos de la siguiente manera:

	pm25
fecha	
2018-11-03 23:00:00	8.680555
2018-11-04 00:00:00	8.472222
2018-11-04 01:00:00	6.909722
2018-11-04 02:00:00	6.347222
2018-11-04 03:00:00	7.833333

Figure 2: Medición de los sensores

Emparejando cada dato de los conjuntos por sus tiempos y calculando la distancia euclidiana, el error y el error cuadrático medio obtenemos:

```
Distancia euclídea: 74.58884133836176
RMSE: 6.237432260489162
                     pm25_sensor  pm25_ref     error
2018-11-03 23:00:00     8.680555      14.9 -6.219445
2018-11-04 00:00:00     8.472222       9.7 -1.227778
2018-11-04 01:00:00     6.909722      11.1 -4.190278
2018-11-04 02:00:00     6.347222      12.2 -5.852778
2018-11-04 03:00:00     7.833333       9.1 -1.266667
```

Figure 3: Error de medición - ventana:1H

Graficando los resultados de la anterior, podemos observar:

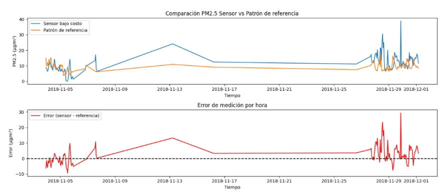

Figure 4: Gráfica error de medición - ventana:1H

Ahora hallamos la distancia euclidiana, el error cuadrático medio y el error absoluto entre los dos conjuntos de datos para varias ventanas de tiempo. Los resultados se muestran en la siguiente tabla:

	Ventana	Distancia	RMSE	MAE
0	1H	74.588841	6.237432	4.382950
1	3H	39.960373	5.488979	4.125986
2	6H	27.459717	5.099141	3.753528
3	12H	18.423541	4.342470	3.252399
4	24H	15.714400	4.358390	3.017743

Figure 5: Tabla de errores para varias ventanas

La gráfica que relaciona las ventanas y el error se muestra a continuación:

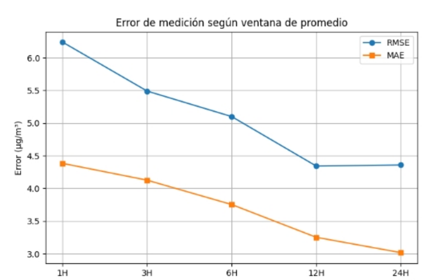

Figure 6: Gráfica Error - Ventanas

Ahora procedemos con la calibración. Hallando la función lineal por el método de mínimos cuadrados para la mitad de los datos, nos resulta la siguiente función con el nuevo error correspondiente:

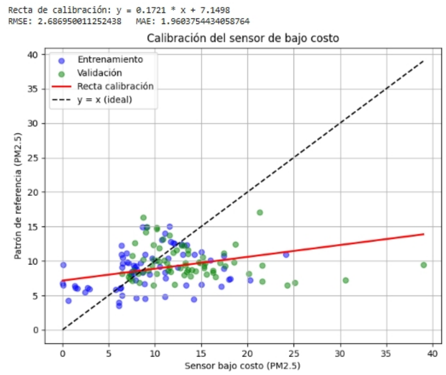

Figure 7: Calibración Función Lineal

Ahora extrapolando nuestra calibración a la otra mitad de los datos y graficando nos resulta de la siguiente manera:

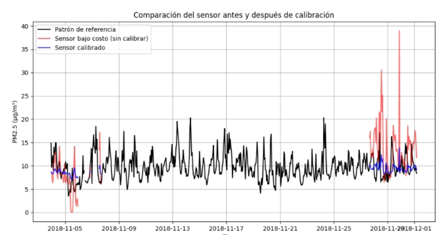

Figure 8: Gráfica calibración

4 Análisis de Datos

Observando detenidamente los datos proporcionados en la sección anterior, podemos notar que el error de medición es cada vez menor cuanto mayor es la ventana que consideramos. Esto nos indica que en un periodo largo de tiempo, los promedios de los conjuntos de datos son muy parecidos.
Si ponemos nuestra atención en la gráfica 4 de error de medición, podemos examinar que hay un gran agujero de datos correspondiente a la medición de los sensores, sugiriendo que los sensores dejan de tomar datos o fallan durante largos periodos de tiempo.
Pasando a la calibración de nuestro sensor, podemos notar como las mediciones

se aproximan mas al patrón de referencia gracias a nuestra calibración. La diferencia entre nuestras mediciones normales y las mediciones calibradas es abismal, pero esta brecha abismal es positiva ya que nos permite aproximarnos mucho mas al patrón de referencia. Estos buenos resultados en la calibración solo indican el poder que tiene el álgebra de espacios vectoriales en la vida cotidiana.

5 Conclusión

En este trabajo se comparó el comportamiento de un sensor de bajo costo frente a un patrón de referencia, evaluando sus errores de medición y aplicando una calibración con el método de mínimos cuadrados. Se observó que, aunque el sensor presenta diferencias notables con respecto al patrón en intervalos cortos, al aumentar la ventana de promediado esas diferencias se reducen.

La calibración lineal permitió acercar las mediciones del sensor a las del patrón, logrando una disminución tanto del error cuadrático medio como del error absoluto medio. Esto demuestra que los sensores de bajo costo pueden ser una alternativa viable, siempre y cuando se realice un proceso de ajuste adecuado y se establezcan los límites en los que sus datos calibrados son confiables. En resumen, la calibración y el análisis estadístico son herramientas clave para mejorar la utilidad de sensores más económicos.