

Calibración de un sensor de bajo costo

Jhonatan Stiven Blanco Melo
Angie Yuliana Sandoval Reyes

*

Universidad Industrial de Santander
CL 9 Cra. 27, Bucaramanga, Santander

6 de diciembre de 2022

Índice

1. Introducción	2
2. Calibración sin media móvil	2
3. Calibración con media móvil	5
4. Conclusiones	9
5. Referencias	10

Resumen

En el siguiente artículo se realizó la calibración de un sensor de bajo costo de $PM_{2.5}$. Para realizar la calibración se tomaron datos de un sensor de mayor calidad como referencia. En la primera parte, a partir de la distancia entre los datos se halló una función de calibración para luego proyectarlos sobre una función lineal (el mejor ajuste), de modo que al aplicar esta función a los datos se evidenció una reducción en la distancia de los datos, siendo que, el error de los datos sin calibrar fue de 12,8%, luego de esta calibración se redujo hasta un 7,44%. La segunda parte consistió en realizar una calibración utilizando el concepto de media móvil, además de un solapamiento en los intervalos. También se empleó el concepto de promedios móviles exponencialmente ponderados (EWMA), de modo que, se aplicó mayor peso a los datos sin solapamiento. Mediante el EWMA se halló que la ventana que proporciona el error mínimo para construir el modelo corresponde a un 1/4,64 de los datos, ya con esta porción se realizó el mismo procedimiento de la primera parte obteniendo una nueva función de calibración con un error del 3,98 %.

* e-mail: jhonatan2211497@correo.uis.edu.co y angie2210728@correo.uis.edu.co

1. Introducción

Junto con el avance de nuevas tecnologías se han desarrollado una serie de sensores de todo tipo, los sensores de bajo costo vienen acompañados de lo que hoy se conoce como IoT (internet de las cosas) que se refiere a una red colectiva con tecnologías implementadas para el procesamiento de datos que facilita la comunicación entre los dispositivos y sistemas a través de redes de comunicación, para favorecer la masificación se busca ahorrar en los costes lo que afecta notablemente la precisión de estos sensores, esto puede ser contrarrestado en gran medida por medio de una calibración inteligente teniendo como referencia datos medidos con sensores de mayor calidad para que de esta forma las lecturas del sensor de bajo costo sean mucho más acertadas.

Para esta calibración es necesario conocer el error de medición del sensor de bajo costo y la comparación de los datos se llevara a cabo con el criterio de media móvil, la cual consiste en analizar los datos promedio locales de ambos conjuntos de datos. Se emplearon datos extraídos de un dataset de concentración $PM_{2,5}$ que se refiere a la materia particulada (por sus siglas en inglés, particulate matter) que tiene al menos $2,5\mu m$ de diámetro, partículas que se generan como residuos de todo tipo de combustión, según estudios, generan contaminación que puede llegar a ser nociva para la salud.

2. Calibración sin media móvil

Los datos de referencia se emplearon para calibrar los datos tomados por el sensor de bajo costo, la figura 1 enseña como se presentan inicialmente los datos.

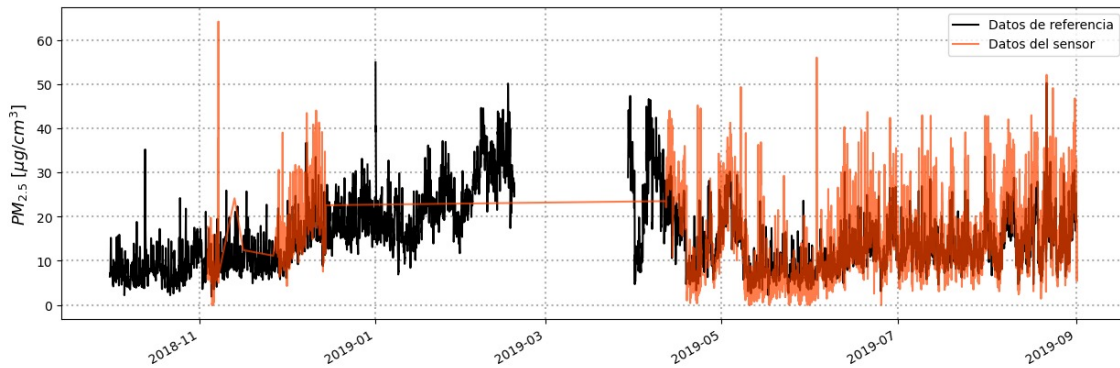


Figura 1: Datos $PM_{2,5}$ de referencia junto con los datos del sensor de bajo costo.

Como primera observación se evidencia una discontinuidad en los datos tomados, en la primera parte solo una pequeña porción de los datos fue medida por el sensor de bajo costo, para efectos prácticos se realizó el calibre y el análisis de datos recortando esta porción de la información debido a esas irregularidades, en rasgos generales se evidencia que existe una correlación entre los datos de referencia (Respuesta esperada) figura 1.

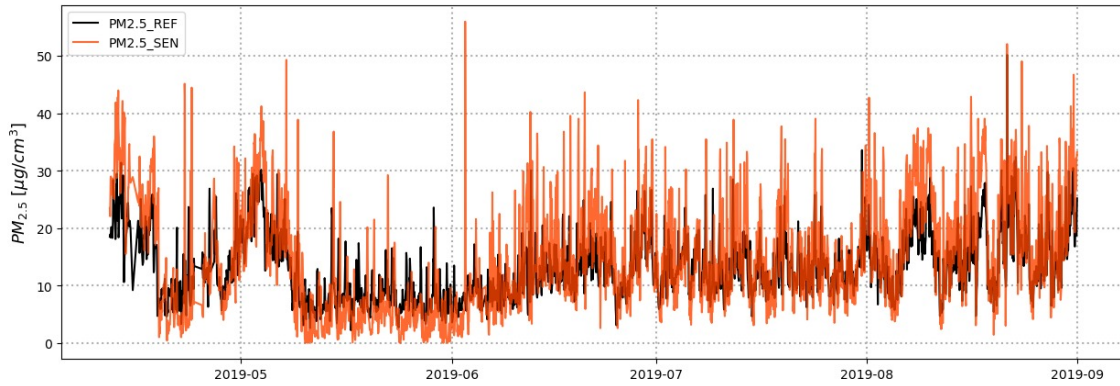


Figura 2: Recorte de los datos en la figura 1.

A pesar de la correlación observada se ve que los datos no coinciden del todo, el objetivo es que los datos tomados por el sensor de bajo costo sean lo más parecido posible a los datos de referencia que es nuestra respuesta esperada, la distancia entre lecturas esta dada por:

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

donde \mathbb{R} son los datos de referencia (PM2.5-REF) y $\hat{\mathbb{R}}i$ son los datos del sensor (PM2.5-SEN). La distancia relativa \mathcal{D}/N (N representa el tamaño de la muestra o lo que es lo mismo, el número de lecturas) entre estos dos conjuntos de datos es de 12.929 %.

Una forma de reducir la distancia entre los datos de referencias y los medidos por el sensor es mediante una calibración. La calibración se puede ver como una función que recibe los datos del sensor (Input) para dar como resultado una respuesta cercada a la esperada esperada, esto es:

$$C(\hat{\mathbb{R}}) = \hat{y} \approx \mathbb{R}. \quad (2)$$

Para encontrar la función de calibración se analizaron los datos de referencia respecto a los datos del sensor (eje x y eje y respectivamente de la figura 3) para luego proyectarlos sobre el subespacio de un tipo de funciones en este caso se eligió una lineal debido a que se comportaba de buena manera para estos datos.

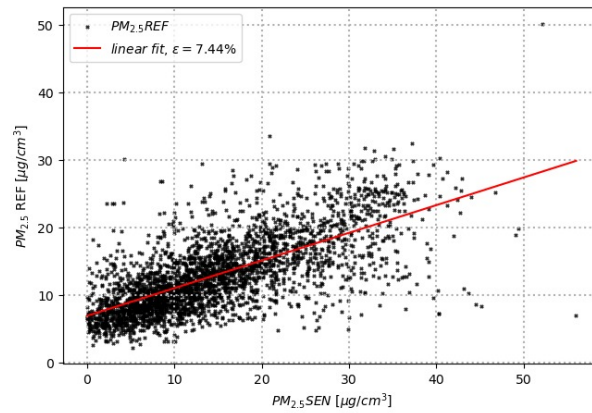


Figura 3: Modelo lineal de PM2.5-SEN vs PM2.5-REF con un error del 7.44 %

La función de calibración obtenida mediante el procedimiento anteriormente mencionada fue la siguiente:

$$C(x) = 6,9 + 0,41x, \quad (3)$$

Luego de aplicar la calibración a los datos del sensor se obtuvo un comportamiento más cercano a los datos de referencia, el distancia respecto a los datos sin calibrar, sin calibración se tenía un error del 12,8 %, luego de esta calibración el error obtenido se redujo hasta un 7,44 %.

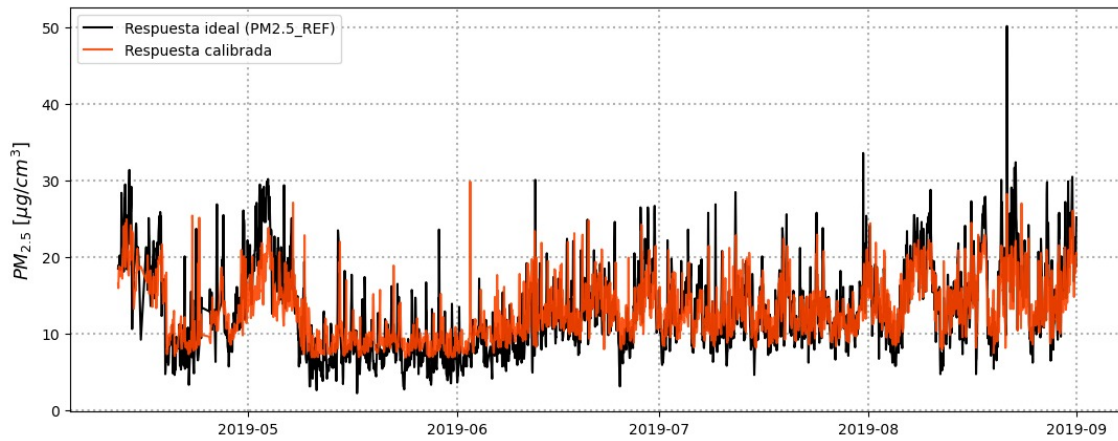


Figura 4: Datos calibrados a partir de la ecuación (3).

El error se redujo significativamente, sin embargo fue posible reducirlo aún más usando el concepto de la media móvil.

3. Calibración con media móvil

La media móvil consiste en definir un intervalo de datos (la ventana) y promediar esos intervalos, estas ventanas pueden tener un solapamiento, o sea, para cada intervalo se tomaran datos del intervalo anterior. Además se empleara el método de promedios móviles exponencialmente ponderados (EWMA), el cual consiste en aplicar ponderaciones a los valores de una serie temporal para dar mayor importancia a los más recientes en el cálculo del promedio móvil y menos a los datos anteriores por lo que el solapamiento tiene menor peso. Esto permite una mayor capacidad de reacción del sensor a cambios en el valor medido.

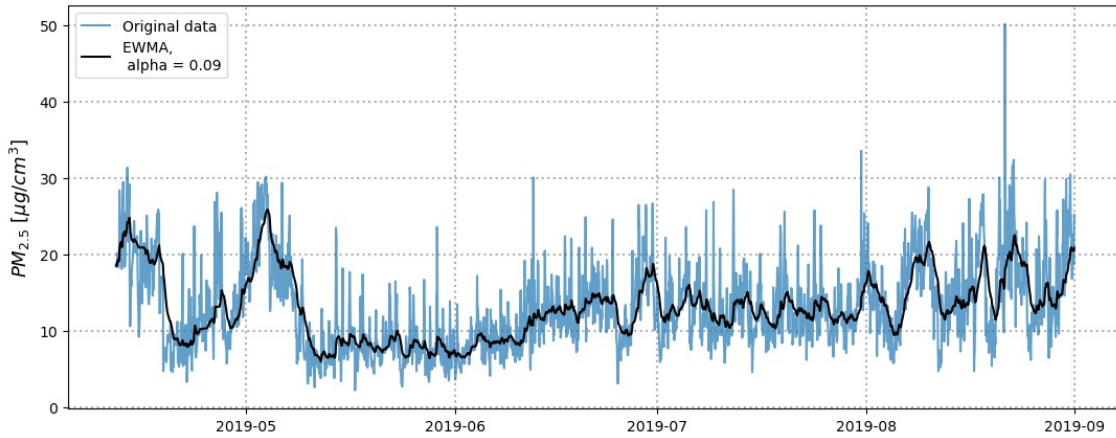


Figura 5: Ejemplo de como actua el EMWA en los datos

El uso de la media móvil en la calibración de un sensor es útil para estabilizar y suavizar la señal medida, lo que puede ayudar a reducir el ruido y a mejorar la precisión y la sensibilidad del sensor. Además, el uso del método EWMA permite una mayor flexibilidad y adaptabilidad del sensor a diferentes entornos y condiciones de medición ya reducir el ruido implica también reducir las lecturas que responden a situaciones particulares del entorno y del momento en el que se midieron los datos. El que los datos solapados tengan menor peso hace que el promedio responda más rápidamente a cambios en los datos, lo que la hace más sensible a las variaciones, la sensibilidad puede ser ajustado mediante el parámetro alpha de la ecuación 5.

En general los promedios móviles ponderados obedecen a la siguiente expresión, donde w representa el peso aplicado:

$$y_t = \frac{\sum_{i=0}^t w_i x_{t-i}}{\sum_{i=0}^t w_i}. \quad (4)$$

En particular para el EWMA los pesos se calculan con la siguiente formula, donde α representa el factor de suavizado:

$$\alpha = \frac{1}{(1 + w)}, w \geq 0. \quad (5)$$

$$y_t = \frac{x_t + (1 - \alpha)x_{t-1} + (1 - \alpha)^2x_{t-2} + \dots + (1 - \alpha)^tx_0}{1 + (1 - \alpha) + (1 - \alpha)^2 + \dots + (1 - \alpha)^t}. \quad (6)$$

Este parámetro α es el responsable del grado de suavizado de la EWMA. Cuanto mayor sea el valor de α , más peso se le dará a los valores más recientes en la serie de datos, lo que resultará en una EWMA más sensible a los cambios en los datos. Por otro lado, si el valor de α es menor, se le dará menos peso a los valores más recientes, lo que resultará en una EWMA menos sensible a los cambios en los datos.

Para una mejor calibración es importante elegir el parámetro α que mejor se ajuste a la respuesta esperada, también es importante hallar la porción de los datos usados para hallar la función de calibración, para esto se vario el parámetro α en las distintas porciones de los datos de referencia y así se encontró la ventana con el parámetro α que proporciona el error minimo, se encontró que la mejor porción de los datos para construir el modelo es la que corresponde a un 1/4,64 de los datos, particularmente los datos que van hasta las 9:00 a.m del día 12 de mayo del 2019 con $sm = 9,92$, donde $sm = \alpha^{-1} - 1$, así que el parametro alpha que corresponde a este sm es $\alpha = 0,092$.

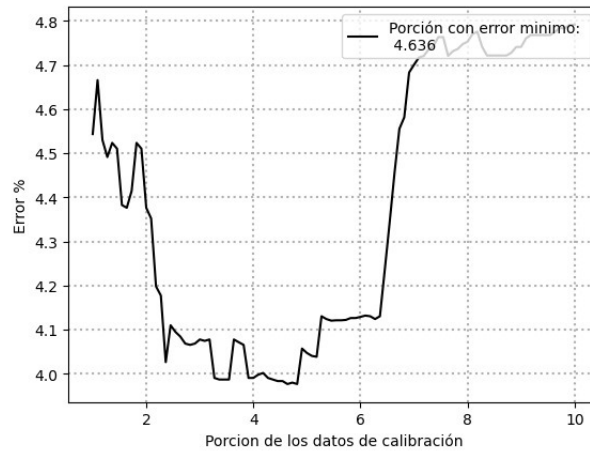


Figura 6: Error en función de la porción de datos tomada en la construcción de la función de calibración

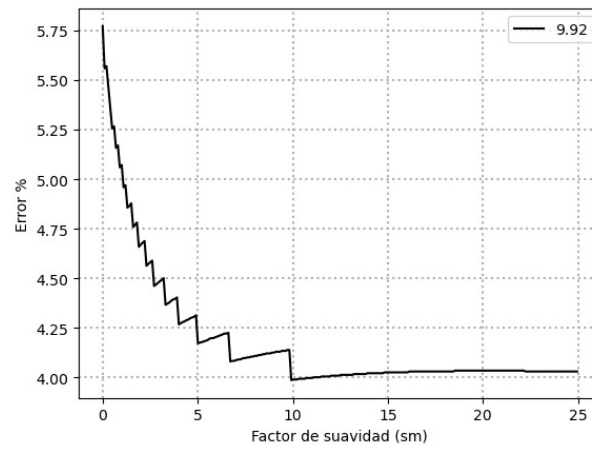


Figura 7: Error en función del factor de suavidad

El mejor factor de suavidad y la mejor porción de datos calculada para hallar el EWMA se visualiza en la figura 8

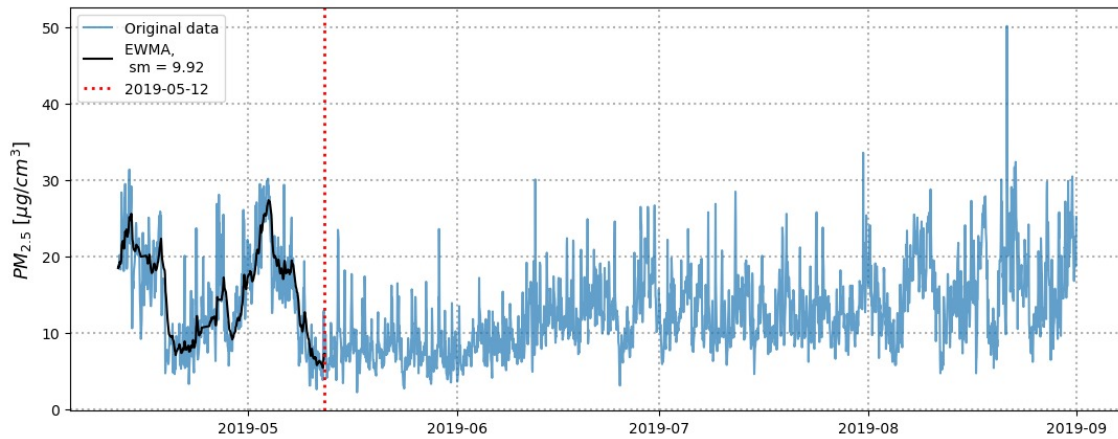


Figura 8: Caption

Con la porción y el factor sm determinado, se hizo el mismo procedimiento mostrado en la sección anterior, se hizo la regresión lineal de la que se obtuvo la función de calibración:

$$C(x) = 4,43 + 0,59x \quad (7)$$

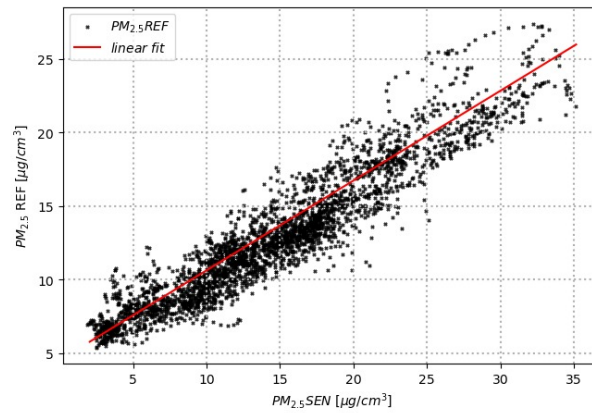


Figura 9: Modelo lineal usando EWMA

Como se observa en la figura 9 los datos están más confinados en una recta, es decir, menos dispersos que cuando no se usa EWMA (véase la figura 3).

Como se muestra en la figura 8 el EWMA y por tanto la función de calibración fue hallado teniendo en cuenta los datos que van hasta el 2019-05-12, de modo que los datos posteriores sirvieron para evaluar como se comporta nuestro modelo para lecturas futuras.

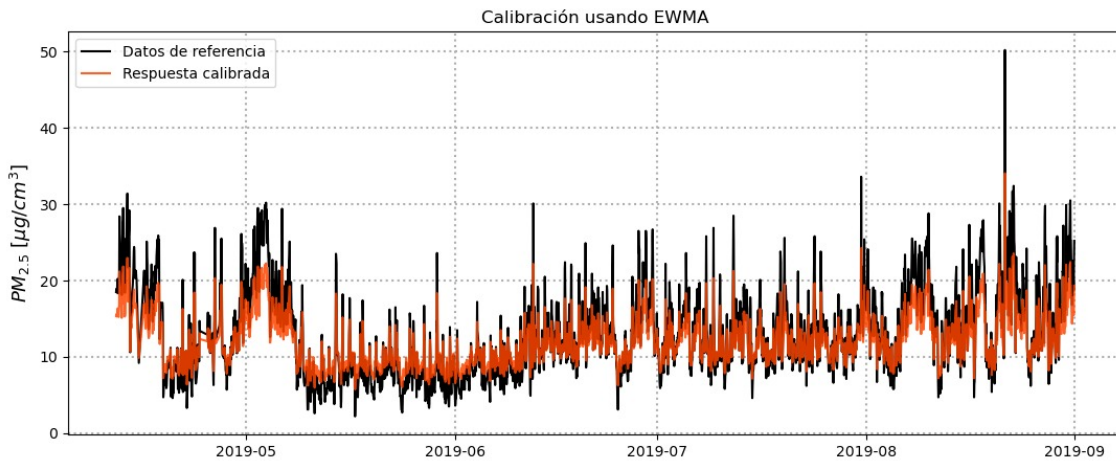


Figura 10: Datos de referencia vs datos calibrados con la ecuación 7

Se obtuvo un mejor resultado que el obtenido sin EWMA, el error de la calibración fue del 3.98 % respecto a los datos originales $PM_{2.5}$ REF, una mejora significativa respecto 7.44 % mostrado en la sección anterior.

Si se compara la respuesta calibrada con los datos de referencia con EWMA se obtiene un error del 3.28 % y se puede ver de mejor manera el comportamiento de la calibración.

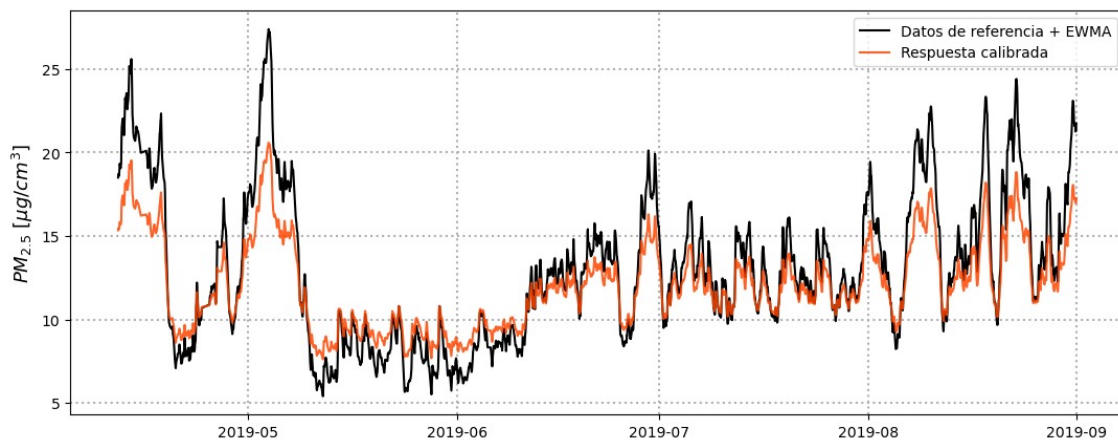


Figura 11: Datos de referencia + EWMA vs datos calibrados con la ecuación 7

Se observa como el modelo tiene un comportamiento excelente la mayoría del tiempo, los problemas aparecen cuando hay cambios bruscos.

4. Conclusiones

Teniendo en cuenta los resultados y las calibraciones realizadas, se llegó a las siguientes conclusiones:

- Se puede obtener una calibración bastante precisa para datos obtenidos por sensores de bajo costo mediante el uso de media móvil.
- Cuantificar el error de medición de un sensor y hallar una calibración para el mismo, lleva al concepto de métrica en espacios vectoriales y en general aproximaciones.
- La calibración empleando el EWMA evidencia una mejora significativa en la precisión de los datos respecto a los datos calibrados sin EWMA.
- La ventana que proporciona mayor precisión en la calibración de los datos, construyendo la función de calibración, es de un $1/4,64$ de los datos.

De forma general, la física computacional y la ciencia de datos desempeñan un papel importante en la solución de problemas inclusive a nivel cotidiano que le conciernen a la población.

5. Referencias

Referencias

- [1] RADEČIĆ, D., *Time Series From Scratch - Exponentially Weighted Moving Averages (EWMA) Theory and Implementation*: Towardsdatascience, 2021, Agosto 9.
<https://towardsdatascience.com/time-series-from-scratch-exponentially-weighted-moving-averages/>
- [2] NÚÑEZ L., *Métricas, datos y calibración inteligente*: Git-Hub, 2020. <https://github.com/nunezluis/MisCursos/blob/main/MisMateriales/Asignaciones/TallerDistancias/TallerDistancias.pdf>

