

Métricas, datos y calibración inteligente

Brayan Monroy* Emmanuel Martinez
Universidad Industrial de Santander
Dirección de la Institución

Versión $\alpha\beta$ fecha del documento

Índice

1. Calculo de distancia	1
2. Calibración	3
3. Conclusiones y Recomendaciones	4

Resumen

En el presente informe se hace desarrollo del taller de distancias, partiendo del procesamiento de los datos por medio del calculo de los promedios locales y el posterior calculo de la distancia siguiendo la definición de la distancia euclidiana. Finalmente se realizo la calibración de las medidas obtenidas por sensores de bajo costo dadas unas medidas de referencia mediante un ajuste lineal por medio del método de mínimos cuadrados. El ajuste lineal siguió dos estrategias, la primera sobre todo el conjunto de datos y posteriormente se hizo un análisis de la porción mínima del conjunto de datos para realizar una óptima calibración.

1. Calculo de distancia

Dado el conjunto de datos de referencia $\mathbb{D}_i = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$ y el conjunto de datos a calibrar $\hat{\mathbb{D}}_i = \{(x_1, y_1), (x_2, y_2) \dots (x_m, y_m)\}$, se realizo el calculo del error entre los promedios locales de los datos de referencia y los datos a calibrar, $f(\epsilon_j), \hat{f}(\epsilon_j)$. Empleando una ventana de media móvil definidas por una ventana común $a_i \leq x_i, \hat{x}_i \leq b_j$, donde $e_j = a_j + (b_j - a_j)/2$ con j indica el numero de ventanas y promedios locales a definir en el rango de variación de los datos, experimentalmente se hizo el calculo para los valores de $j = \{2, 4, 15, 21, 28, 34\}$. Posteriormente se realizo el calculo de la distancia entre las medidas de referencias y las medidas a calibrar dada la formula de la distancia euclidiana,

* e-mail: brayan2180032@correo.uis.edu.co

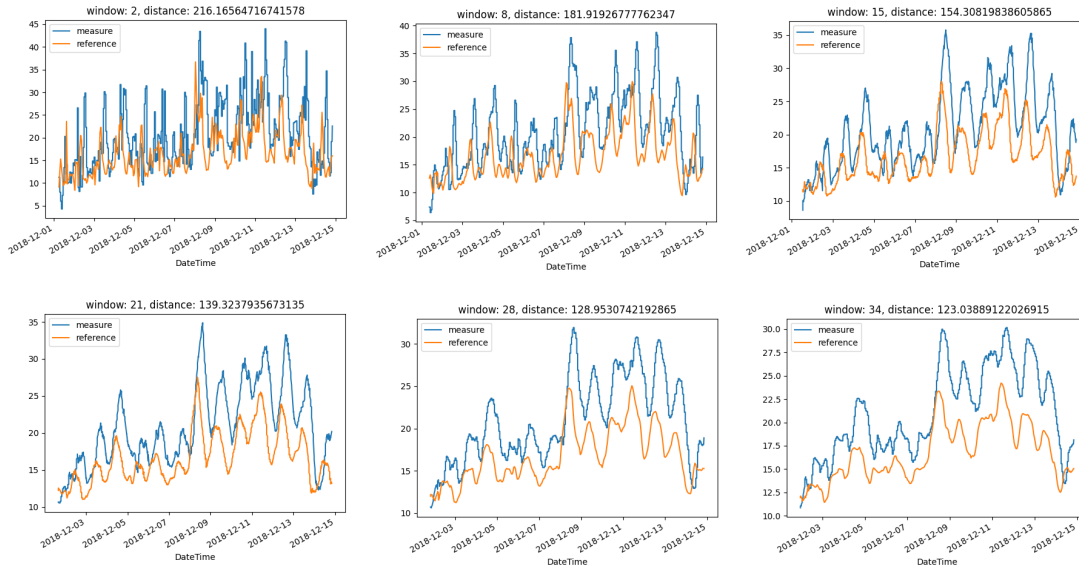


Figura 1: Promedios locales para distintos anchos de ventana móvil.

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

En la Figura 1 se muestran los resultados de los promedios locales calculados como la distancia entre este nuevo ajuste de los datos, como se puede observar conforme se aumenta el tamaño de la ventana, menor es la distancia obtenida, pero su vez hay una perdida en la fidelidad del comportamiento estacional de los datos originales. Para un análisis mas general, se realizo el calculo de la distancia entre las medidas de referencia para unos valores de ventana móvil en el intervalo [8, 654] como se observa en la Figura 2.

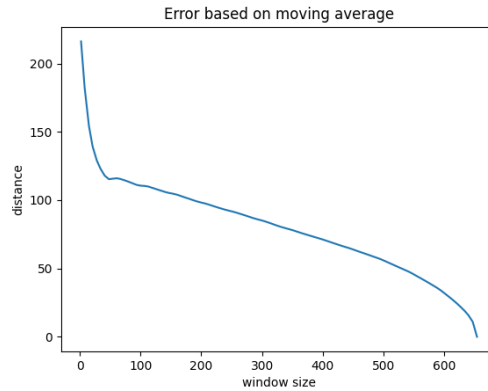


Figura 2: Error para distintos anchos de ventana móvil.

2. Calibración

Para la calibración de las muestras promediadas $(\hat{f}(\epsilon_j), f(\epsilon_j))$, se hizo uso de un ajuste de mínimos cuadrados para determinar un modelo de ajuste lineal, $f(\epsilon_j) = \alpha \hat{f}(\epsilon_j)$. De esta forma, el calculo de α fue determinado como,

$$\alpha = \frac{\langle f(\epsilon) | \hat{f}(\epsilon) \rangle}{\langle \hat{f}(\epsilon) | \hat{f}(\epsilon) \rangle} \quad (2)$$

Como se muestra en la Figura 3(a), se presenta una gran dispersión entre los datos de referencia y los datos a calibrar al no ser evidente un comportamiento lineal de los mismos. Sin embargo, una vez se realiza el ajuste lineal, como podemos observar en la Figura 3(b), los datos de las estaciones de bajo costo se ajusta correctamente a los datos de referencia, presentando un comportamiento similar y una menor distancia en comparación con los datos sin calibrar, diferenciándose por un desfase temporal posiblemente debido a un error en la documentación de la captura de los datos.

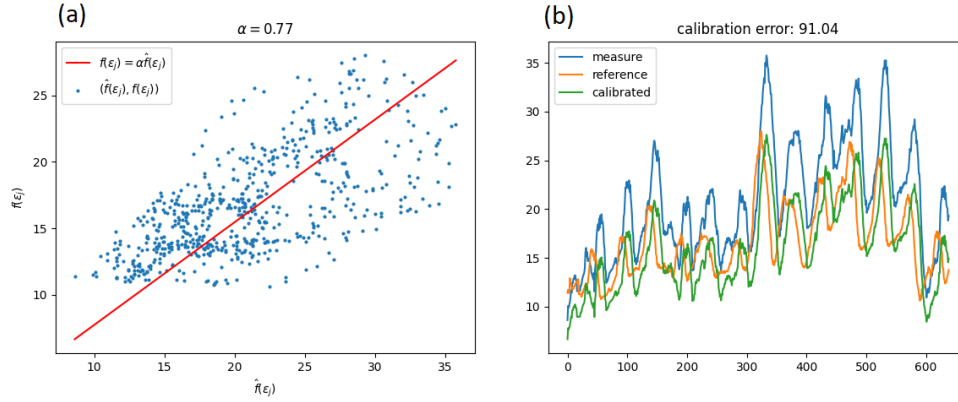


Figura 3: (a) Ajuste lineal por el método de mínimos cuadrados, (b) Calibración de las muestras dado el modelo lineal obtenido.

Posteriormente, se siguió otra alternativa mediante la división de los conjuntos de datos, en un principio se realizó una división equitativa con el 50 % de las muestras como conjunto de calibración y el otro 50 % como conjunto de validación, los resultados se pueden observar en la Figura 4.

Finalmente, se evalúa la estrategia por división de conjunto de datos para distintos porcentajes de división los cuales fueron $\{10, 20, 30, 40, 50, 60, 70, 80\}$ para el conjunto de calibración y su respectivo complemento para el conjunto de validación. Los resultados obtenidos se pueden observar en la Cuadro 2. Como se puede observar, conforme mayor es el porcentaje del conjunto de calibración, menor es la distancia entre el conjunto de validación y de referencia, por lo que podría indicar una mayor precisión en la calibración de las muestras de las estaciones de bajo costo.

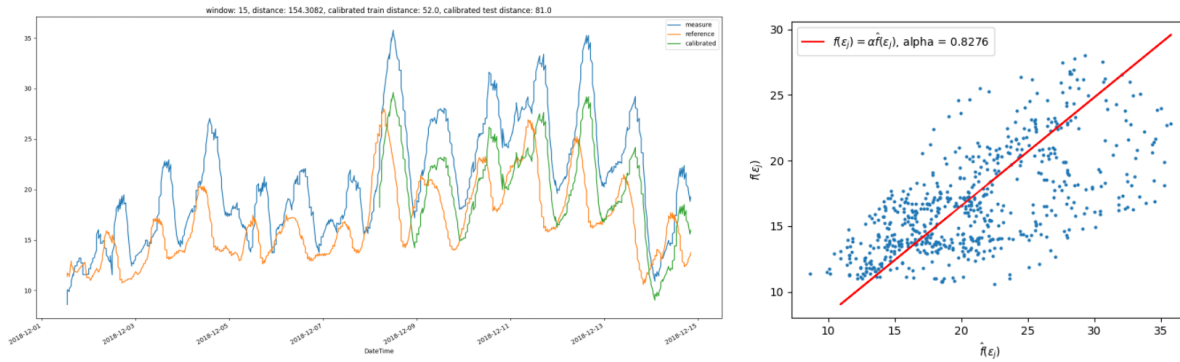


Figura 4: Ajuste lineal empleando el 50 % del conjunto de muestras de referencia y calibración.

porcentaje	D calibración	D validación	alpha
20	30.0	93.0	0.8391
30	40.0	85.0	0.8137
50	52.0	81.0	0.8096
70	70.0	58.0	0.7884
100	91.0	-	0.7727

Cuadro 1: Primera columna, variación del porcentaje de conjuntos para realizar el ajuste lineal de calibración. Segunda y tercera columna, distancia del conjunto de calibración y conjunto de validación. Última columna, factor de ajuste lineal obtenido.

3. Conclusiones y Recomendaciones

En el presente informe se emplearon los conceptos aprendidos de aproximación de funciones y bases ortogonales en el problema de calibración de sensores dados unas muestras de referencia. Durante el desarrollo del informe, se observó un comportamiento inusual al analizar el cálculo de la distancia de los conjuntos de datos en función del tamaño de la ventana móvil, en donde a un mayor tamaño de ventana la distancia era menor. Sin embargo, dada la definición de distancia euclídea utilizada, podemos evidenciar como esta no contempla el tamaño del conjunto de muestras, por lo que los valores de distancia obtenidos pueden estar sesgados a obtener una menor distancia conforme en conjunto de datos disminuye sin poder evidenciar correctamente el error general, una recomendación que sugerimos para un mejor análisis es dividir la distancia obtenida por la cantidad de muestras evaluadas.