

Universidad Industrial de Santander
Facultad de Ingeniería de Sistemas e
Informática

MATEMÁTICAS AVANZADAS

Métricas, Datos y Calibración Inteligente

Autor:
Jhon Edinson López Durán

Profesor:
Luis Alberto Nuñez

Septiembre del 2020

1. Resumen

El material particulado fino ambiental ($PM_{2.5}$) es uno de los contaminantes atmosféricos más estudiados en el mundo, este se define como el conjunto de partículas sólidas y/o líquidas (a excepción del agua pura) presentes en suspensión en la atmósfera [1], que se originan a partir de una gran variedad de fuentes naturales o antropogénicas y poseen un amplio rango de propiedades morfológicas, físicas, químicas y termodinámicas. La presencia en la atmósfera de este contaminante ocasiona variedad de impactos a la vegetación, a los materiales y al hombre, entre ellos, la disminución visual en la atmósfera, causada por la absorción y dispersión de la luz. Además, la presencia del material particulado está asociada con el incremento del riesgo de muerte por causas cardiopulmonares en muestras de adultos [2]. Es necesario, además de realizar mediciones de la concentración de este contaminante, evaluar su comportamiento en el espacio y el tiempo, asociándolo con los fenómenos meteorológicos, composición química y origen, los cuales permitan orientar estrategias de control y realizar seguimiento por parte de las autoridades ambientales interesadas. Debido a que los sensores utilizados hoy en día para la medición del $PM_{2.5}$ son costosos, presentaremos el uso de sensores de bajo costo (**IoT**, internet of things) para la medición de estas partículas y análisis de calibración respecto a sensores de alto costo utilizados por la AMB (Acueducto Metropolitano de Bucaramanga). Esto con el fin de minimizar el costo de llevar registros de este contaminante en nuestro ambiente.

2. Introducción

El material particulado (MP) es un conjunto de partículas sólidas y líquidas emitidas directamente al aire, tales como el hollín de diesel, polvo de vías, el polvo de la agricultura y las partículas resultantes de procesos productivos [3]. Según la normatividad colombiana, el MP no se sedimenta en períodos cortos sino que permanece suspendido en el aire debido a su tamaño y densidad. Estas partículas en suspensión son una compleja mezcla de productos químicos y/o elementos biológicos, como metales, sales, materiales carbonosos, orgánicos volátiles, compuestos volátiles, hidrocarburos aromáticos policíclicos y endotoxinas que pueden interactuar entre sí formando otros compuestos [4]. Debido a que son de tamaño, forma y composición variada, para su identificación se han clasificado en términos de su diámetro aerodinámico que corresponde al diámetro de una esfera uniforme en unidad de densidad que alcanza la misma velocidad terminal de asentamiento que la partícula de interés y que está determinado por la forma y densidad de la partícula. De acuerdo a esto, pueden ser clasificadas como finas y gruesas [?]. Otros autores afirman que el diámetro de las partículas atmosféricas en suspensión, varía desde nanómetros hasta decenas de micras. Generalmente se identifican diferentes rangos de tamaños de partícula denominados “modas”, que están relacionados en su mayoría con el mecanismo de formación de las partículas: nucleación, Aitken, acumulación y moda gruesa [?].

Estas partículas pueden variar su tamaño y composición a través de varios procesos como el vapor de condensación, la evaporación, la coagulación a través de las colisiones del movimiento browniano por la hidrodinámica y fuerzas gravitacionales o eléctricas [?]. El material particulado posee una amplia gama de propiedades morfológicas, químicas, físicas y termodinámicas. Este es emitido a la atmósfera producto de diferentes actividades, tanto naturales como antropogénicas. Entre las actividades naturales se encuentran las emisiones volcánicas y las partículas arrastradas por el viento, como el aerosol marino [?]. Dentro de las actividades antropogénicas se encuentran las emisiones de gases de escape y partículas procedentes de los buques de navegación marítima. Además, algunos investigadores han divulgado que la emisión de contaminantes de origen vehicular, constituye una de las causas más relevantes en el deterioro de la calidad del aire, lo que afecta la composición química de la atmósfera a nivel local y regional. La extracción de la atmósfera del material particulado en suspensión se produce por diversos mecanismos, bien sea por deposición seca o por deposición húmeda, cuya eficacia depende esencialmente del diámetro, composición química y propiedades termodinámicas de las partículas. El material particulado está asociado con compuestos con conocida actividad genotóxica, mutagénica o carcinogénica.

Según el Departamento de Protección Ambiental de Hong Kong, el escape de los vehículos diesel es la principal causa del alto nivel de MP en la zonas urbanas. Los principales elementos encontrados por la quema de combustibles fósiles, transformación de metales e incineración de residuos son el **Pb, Cd, Zn, Cu, Ni, V, Sb**. Los diferentes elementos presentes en las muestras del material particulado pueden estar asociados con materia mineral, aerosol marino, compuestos de azufre, compuestos de nitrógeno y compuestos de carbono. Conocer la composición química del material particulado, tiene relevancia no solo desde el punto de vista de la química de la atmósfera, sino también sobre la calidad del aire que se respira en las ciudades. Muchas propiedades de las partículas pueden influir de manera importante en su comportamiento en el aire y sus efectos sobre el ambiente y la salud. Los modelos de contaminación del aire juegan un papel importante en la ciencia, debido a su capacidad para investigar la importancia de procesos relevantes, al igual que un mayor rol en la aplicación. Son los únicos métodos que pueden cuantificar la relación entre emisiones y concentración-deposición, incluyendo las consecuencias de escenarios futuros y las estrategias de disminución. Estos modelos han sido usados para calcular el impacto de las emisiones volcánicas y modelos productivos.

Para analizar y mostrar la calibración de este material se tiene dos conjuntos de datos que capturan la concentración de material $PM_{2.5}$ para dos estaciones, una estación perteneciente a la AMB y otra estación de IoT. El objetivo es calcular un modelo que calibre las mediciones de la estación IoT con respecto a la estación de AMB. Estos datos al igual que muchos en la realidad no se pueden modelar en crudo, para esto es necesario realizarle un preprocesamiento y ajustarlos de tal manera que puedan ser usados por algún tipo de clasificador

al igual que se tiene que definir que es lo que se quiere predecir.

3. Metodología

Inicialmente realizamos un preprocesamiento de los datos, el cual consiste en eliminar los datos vacíos y crear columnas que almacenen la fecha, hora y día de la toma de la muestra para realizar las comparaciones necesarias. Luego del tratamiento de los elementos nulos, se observa que el número de datos de la estación IoT es de 5008 y es menor a la cantidad de la estación de la AMB 7920, además de que algunas muestras fueron tomadas en días diferentes y algunas horas tienen mas de una muestra, para esto lo que se hace es tomar el dataset con menor numero de ejemplos, en este caso la estación IoT y promediar las horas por día, esto con el fin de que cada día tenga 24 muestras y luego tomar los datos que fueron muestreados en el mismo tiempo para la estación de la AMB esta da como resultado en ambos conjuntos de datos con 2994 muestras.

En el segundo paso se realiza una exploración de los datos para observar como se distribuye uno respecto al otro para esto se tomaron todos los datos y se promediaron sus horas, después se calculo la distancia euclidiana (1) como se muestra en la figura 1, donde se puede observar que se obtiene una mayor distancia entre los datos durante las horas pico, y esto se puede observar en los promedios de cada estación donde se ve que normalmente durante las horas pico la estación de la AMB se incrementa el numero de partículas y en la IoT disminuye.

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

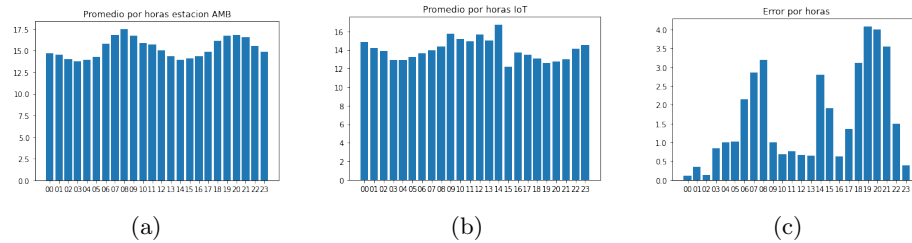


Figura 1: (a) y (b) Promedio por horas para ambas estaciones, (c) Error entre las estaciones

También se realizó un promedio por meses como se observa en la figura 2, en este caso se observa que la distancia entre los meses es mucho mas grande, pero esto se debe a que para algunos meses se cuenta con mayor cantidad de datos que para otros. También en la figura 3 se observa como se distribuyen los datos de la estación IoT respecto a los de la AMB donde se observa que existen

algunas mediciones poco comunes (outlayers) que pueden afectar el modelo y las predicciones.

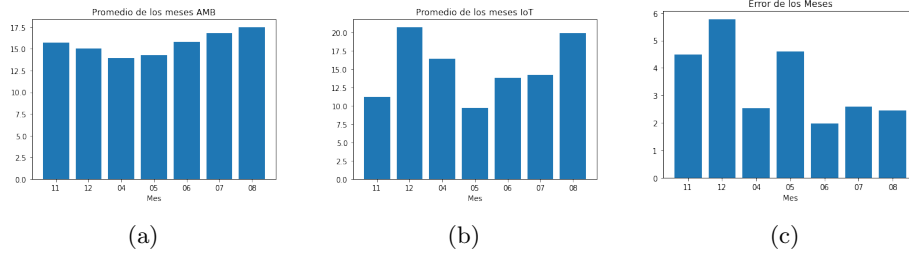


Figura 2: (a) y (b) Promedio por meses para ambas estaciones, (c) Error entre las estaciones.

Como tercer paso se realizo una ventana móvil para suavizar los datos y reducir los outliers que se presentan, para medir la precision de estas ventanas se uso la distancia euclidiana respecto a los datos de la AMB, el resultado de esta distancia y de las ventanas se muestra en la tabla ?? donde se observa que la menor distancia se obtiene al no usar la ventana móvil, esto se debe a que como se observa en la figura 1 la media de ambos datos es diferente y al promediar los datos, esta media se hace un poco mas grande aumentando el error. Adicionalmente se muestra cada una de estas distribuciones en la figura 3.

Ancho Ventana	Distancia
1	0.1648
2	0.2183
3	0.2579
4	0.2207

Cuadro 1: Distancia para los diferentes valores de la ventana móvil.

Una vez se definido nuestro conjunto de datos, se procede a entrenar un modelo para generar un modelo lineal tal que los datos de la estación IoT sean lo mas cercanos a los datos de la estación de la AMB, esto significa determinar un factor α que calibre nuestros datos de la estación IoT y este α correspondería a la pendiente de nuestro modelo. para este caso se obtuvo que el factor $\alpha = 0,30292034$ utilizando el 70 % de los datos. En la figura 4 se muestra la distribución de los datos y la recta que mejor se ajusta a estos.

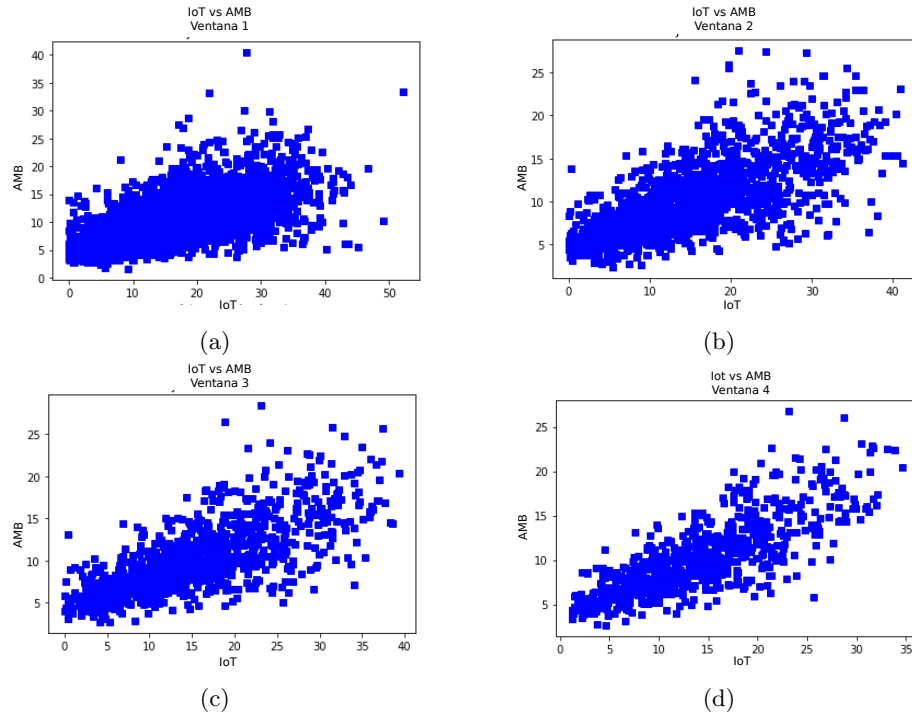


Figura 3: Distribución de los datos para diferentes valores de la ventana móvil.

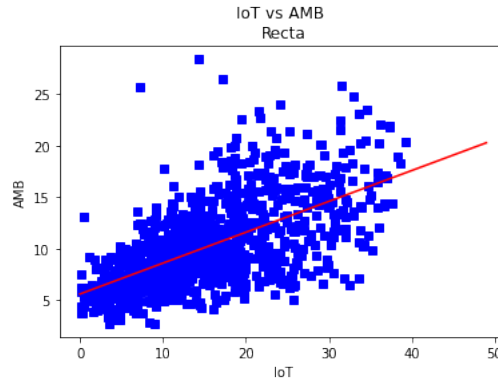


Figura 4: Distribución de los datos sin ventana móvil y la recta que mejor se ajusta a estos datos con un $\alpha = 0,30292034$.

Para determinar el alcance del modelo y el conjunto mínimo de datos para generar el mismo modelo, se vario el porcentaje de datos para el entrenamiento y testeo. Se calculo la distancia respecto a los datos de la AMB, el resultado

de estas distancias se muestra en la tabla ??, donde se observa que apartir de usar mas del 50 % de los datos la distancia aumentó, lo que significa que con este modelo se puede predecir hasta un 50 % mas de muestras sin aumentar el error significativamente, esto quiere decir que se puede tolerar hasta un error aproximadamente del 10 %.

% de datos	Distancia
30 %	0.1484
40 %	0.1180
50 %	0.0965
60 %	0.1102

Referencias

- [1] E. Mészáros, *Atmospheric Chemistry*. Elsevier, 1981.
- [2] C. A. Pope Iii, R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston, “Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution,” *Jama*, vol. 287, no. 9, pp. 1132–1141, 2002.
- [3] G.-C. Fang, C.-N. Chang, C.-C. Chu, Y.-S. Wu, P. P.-C. Fu, I.-L. Yang, and M.-H. Chen, “Characterization of particulate, metallic elements of tsp, pm2. 5 and pm2. 5-10 aerosols at a farm sampling site in taiwan, taichung,” *Science of the Total Environment*, vol. 308, no. 1-3, pp. 157–166, 2003.
- [4] S. Billet, G. Garçon, Z. Dagher, A. Verdin, F. Ledoux, F. Cazier, D. Courcot, A. Aboukais, and P. Shirali, “Ambient particulate matter (pm2. 5): physico-chemical characterization and metabolic activation of the organic fraction in human lung epithelial cells (a549),” *Environmental research*, vol. 105, no. 2, pp. 212–223, 2007.
- [5] F. F. García, R. A. Agudelo, and K. M. Jiménez, “Distribución espacial y temporal de la concentración de material particulado en santa marta, colombia,” *Revista Facultad Nacional de Salud Pública*, vol. 24, no. 2, pp. 73–82, 2006.
- [6] H. U. Roll, *Physics of the Marine Atmosphere: International Geophysics Series, Vol. 7*. Elsevier, 2016, vol. 7.
- [7] G.-C. Fang, Y.-S. Wu, S.-Y. Chang, J.-Y. Rau, S.-H. Huang, and C.-K. Lin, “Characteristic study of ionic species in nano, ultrafine, fine and coarse particle size mode at a traffic sampling site,” *Toxicology and industrial health*, vol. 22, no. 1, pp. 27–37, 2006.
- [8] A. Juárez, C. GAY, and Y. Flores, “Impact of the popocatepetl’s volcanic activity on the air quality of puebla city, méxico,” *Atmósfera*, vol. 18, no. 1, pp. 57–69, 2005.