Chicos, la idea de estas tareas es que sean informes que paso a paso los lleven a conclusiones acerca de sus datos, no que sean una receta y ya. Cuidado con la forma de presentar el trabajo y con las conclusiones. Deben ir además perfeccionando su forma de escribir, lo ideal es ir concluyendo a lo largo del desarrollo del informe

3.9

Universidad Nacional de Colombia, sede Medellín
Facultad de Minas

Juan Pablo Martínez Betancur
Joseph Alberto Fuentes Pineda

Análisis de Datos Ambientales
Tarea 1

Introducción

Esta primer tarea está basada en los datos descargados a través del SIATA (Sistema de Alerta Temprana de Medellín y el Valle de Aburrá, ver *Anexo 1*), los cuales corresponden al material particulado de 2,5 micras (PM2,5) en la estación de medida: Universidad Nacional, El Volador. Cabe señalar que el período de tiempo analizado va desde el primero de enero del año 2014 hasta el 31 de enero del año 2017; además, se resalta que las unidades de medida del PM2,5 son de microgramos por metro cúbico (ug/m^3).

El fin de la toma de estos datos es poder avanzar en el desarrollo del trabajo final, el cual va encaminado a determinar la correlación o no entre la mala calidad del aire en el AMVA (Área Metropolitana del Valle de Aburrá) y el aumento de enfermedades cardiorrespiratorias en las últimas décadas, siendo esto un tema de suma importancia para este territorio, ya que las crisis de material particulado en el aire son más recurrentes y de mayor magnitud. De tal manera, se pretende obtener un mejor entendimiento del fenómeno, sus causas y consecuencias, y el posible acercamiento a soluciones del problema.

Punto 1: Conseguir una serie temporal de más de mil datos

Como se mencionó, los datos fueron obtenidos a través de la página web del SIATA. Su resolución temporal es horaria, pero fueron promediados los datos cada 24 horas a fin de obtener una media diaria de PM2,5. Para esto, se usaron herramientas básicas de Excel como filtrado y tablas dinámicas, y también fue necesario programar en VBA (Visual Basic para Aplicaciones) para ordenar los datos (ver código realizado en *Anexo 2*).

Punto 2 y 3: Lectura de los datos, lectura y gráfico de la serie de tiempo

Una vez organizados los datos se pasaron a un archivo “.txt”, por lo que fue fácil la lectura desde Python, como se observa en la *Imagen 1*.

```
#Lectura de serie de tiempo
fileData="pm25m.txt"
```

```
fileData
```

```
'pm25m.txt'
```

```
Datos=np.genfromtxt(fileData,delimiter=",", dtype="str")
Datos
```

```
array([['D-M-A', 'pm25'],
       ['1-1-2014', '35.9565217391304'],
       ['2-1-2014', '37.625'],
       ...,
       ['29-1-2017', '24.2083333333333'],
       ['30-1-2017', '24.7083333333333'],
       ['31-1-2017', '33.2083333333333']], dtype='|S16')
```

```
PM25=np.array(Datos[1:,1]).astype(float)
```

Imagen 1. Lectura de los datos

Se continuó construyendo un vector de Fechas y otro de PM2,5, que fuesen correspondientes, y luego se dio paso a la creación de un "DataFrame" para realizar la lectura y gráfico de la serie a lo largo del tiempo (*Imagen 2*).

```
Fechas=[]
for i in range(len(PM25)):
    Fechas.append(datetime.strptime(Datos[i+1,0],'%d-%m-%Y'))
Fechas=np.array(Fechas)
Fechas[1100]
#plt.plot(Fechas,PM25)
```

```
datetime.datetime(2017, 1, 31, 0, 0)
```

```
#Serie de datos
Serie_Datos=pd.DataFrame(PM25,Fechas)
Serie_Datos
```

```
fig, ax = plt.subplots()
ax.plot(Serie_Datos.index,Serie_Datos[0])
fig.autofmt_xdate()
plt.ylabel('PM2,5 [ug/m^3]')
plt.xlabel('Tiempo')
plt.title("Serie de tiempo de los datos: PM 2,5")
```

Imagen 2. Lectura de la serie de PM2,5

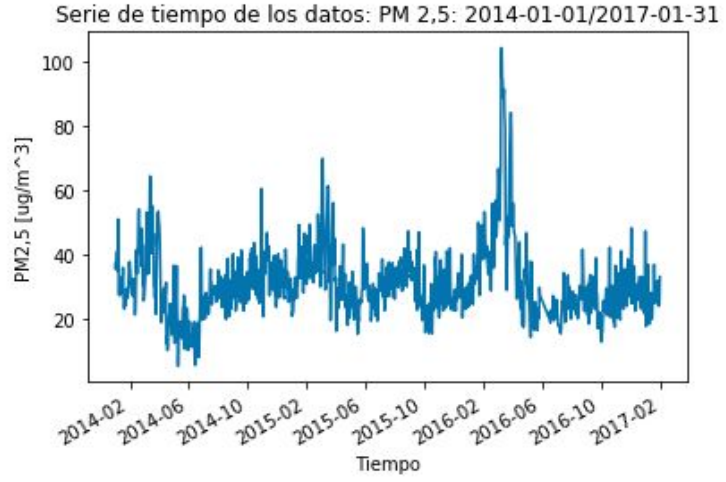

Gráfico 1. Serie de tiempo de PM2,5 en el periodo 2014-01-01/2017-01-31.

Punto 4 y 5: Estimar FDA (Función de Distribución Acumulada), fdp (función de distribución de probabilidad), percentiles (P) e índices estadísticos:

Se creó un vector de percentiles y se graficó posteriormente, obteniendo así la FDA (ver *Gráfico 2*). Posteriormente, a través de las funciones del paquete "numpy" de Python, se procedió a determinar el histograma, con un total de intervalos de 50; con el cual se calculó la fdp, que se puede ver en el *Gráfico 3* con la respectiva media y desviación estándar. Otros datos calculados para la serie de PM2,5 como cuartiles, intercuartil (IQR), varianza (var), desviación estándar (std), coeficientes de Kurtosis (Kurt), asimetría (asim) y Yule-Kendall, son presentados en la *Tabla 1*.

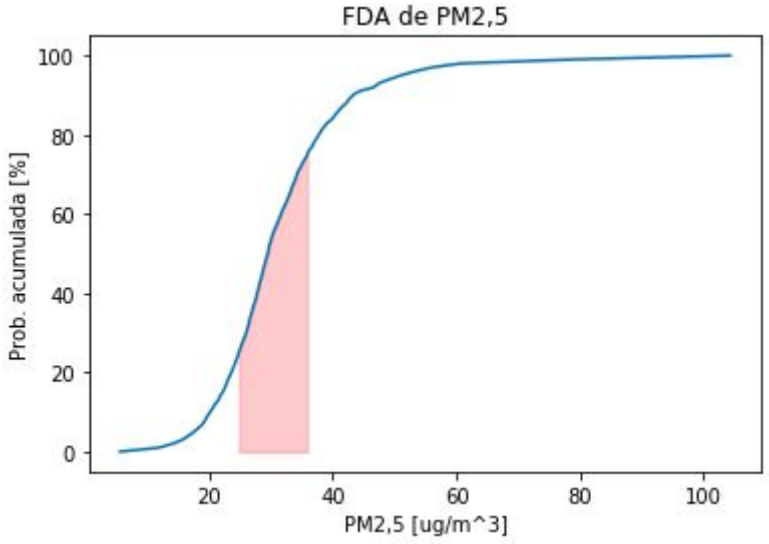

Gráfico 2. FDA para la serie de tiempo de PM2,5. La zona sombreada en color rosa, corresponde al rango del IQR.

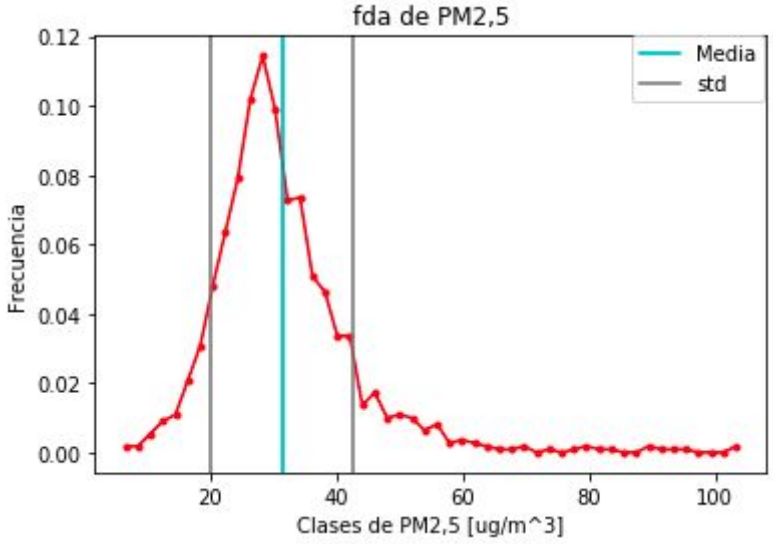

Gráfico 3. fda de la serie de tiempo de PM2,5 con su respectiva media y desviación estándar (std).

Media	31,426
var	128,161
std	11,321
P25	24,792
P50/Mediana	29,500
P75	35,875
IQR	11,083
Kurt	7,648
asim	1,952
YK	0,150

Tabla 1. Resumen de los estadísticos/índices de la serie de tiempo de PM2,5.

Punto 6 y 7: ¿Son estacionarios los histogramas/fdp's e índices en las series móviles?:

Una vez obtenido lo anterior, se procedió a hacer un análisis por tramos o períodos consecutivos de la serie de PM2,5, con el fin de lograr una información más certera y precisa. Para ello se estableció una "ventana" de observación de 90 días, es decir, 3 meses, siendo que esta ventana recoge un conjunto de datos de la serie original de PM2,5 y se les hace la respectiva estimación de los histogramas. La ventana es móvil, lo cual indica que se va moviendo este rango de 90 días a lo largo de toda las fechas originales. Así, resultan entonces un total de 1011 series que son analizadas por separado (todos sus histogramas e índices). El periodo de observación va entonces de 2014-01-01 a 2016-11-03 y los gráficos que resumen la información se muestran a continuación:

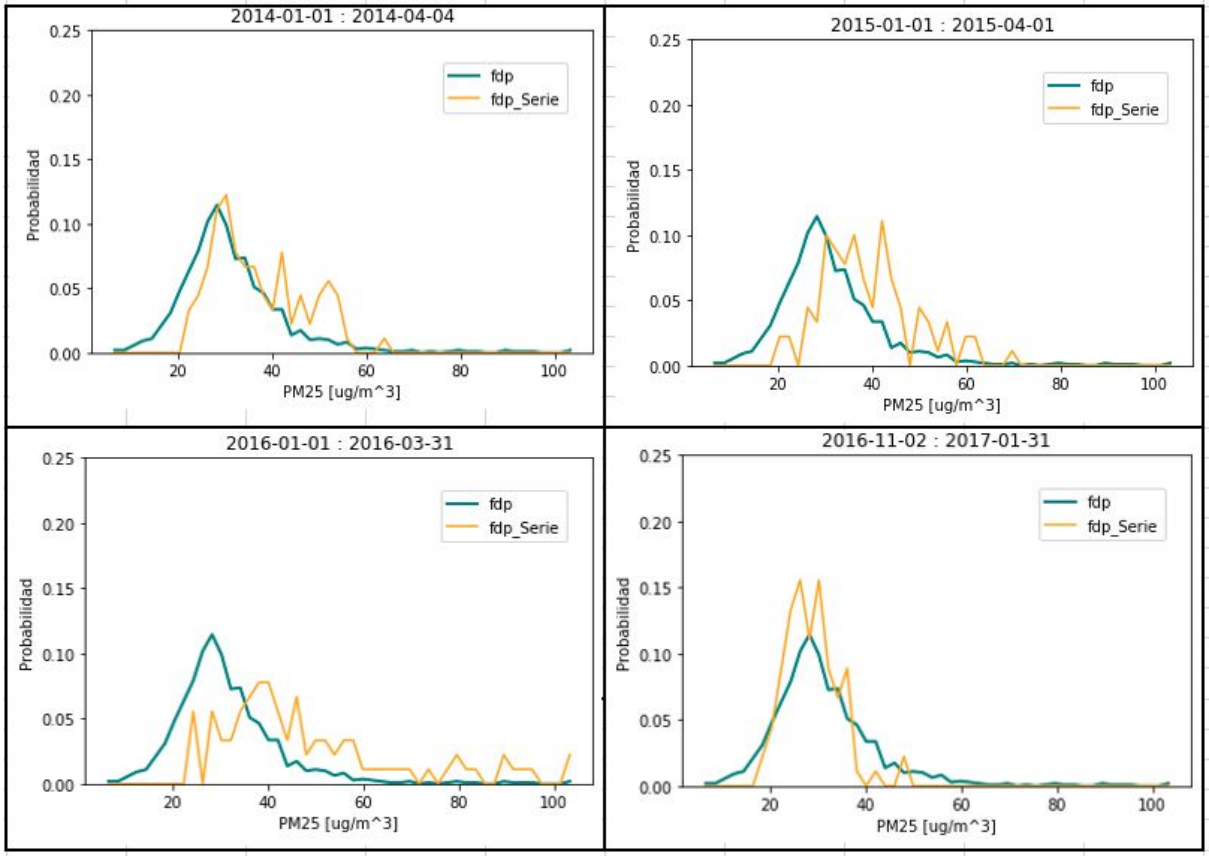

Gráfico 4. Comparación de la fdp general con ciertas fdp de series móviles, para PM2,5.

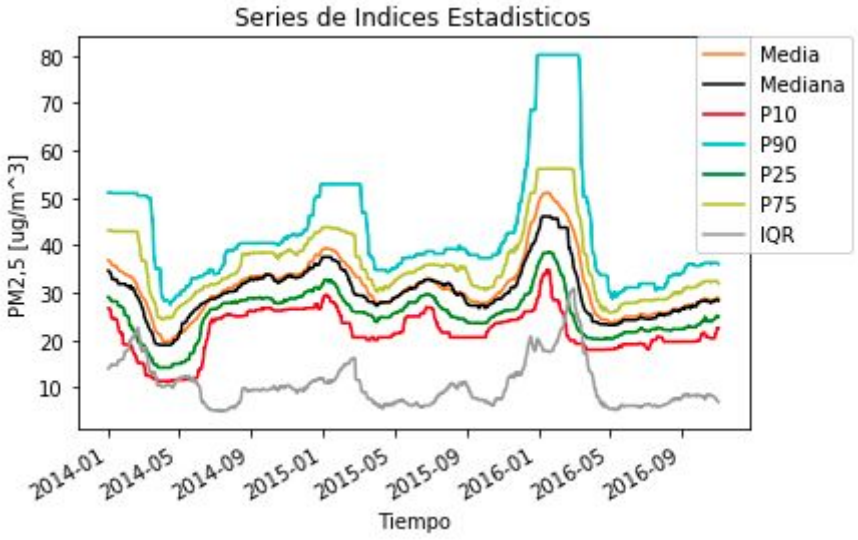

Gráfico 5. Muestra de los diferentes índices para las series móviles, con PM2,5.

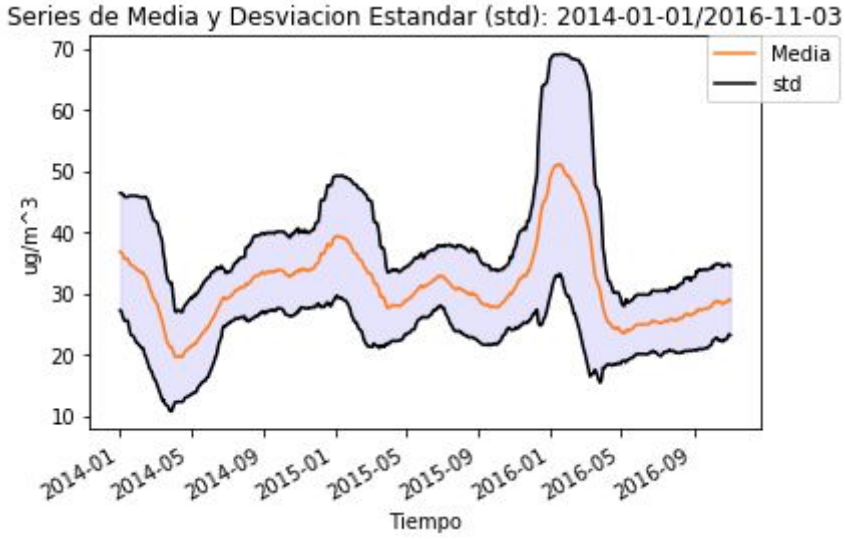

Gráfico 6. Medias y el rango de la respectivas std´s en las series móviles, con PM2,5.

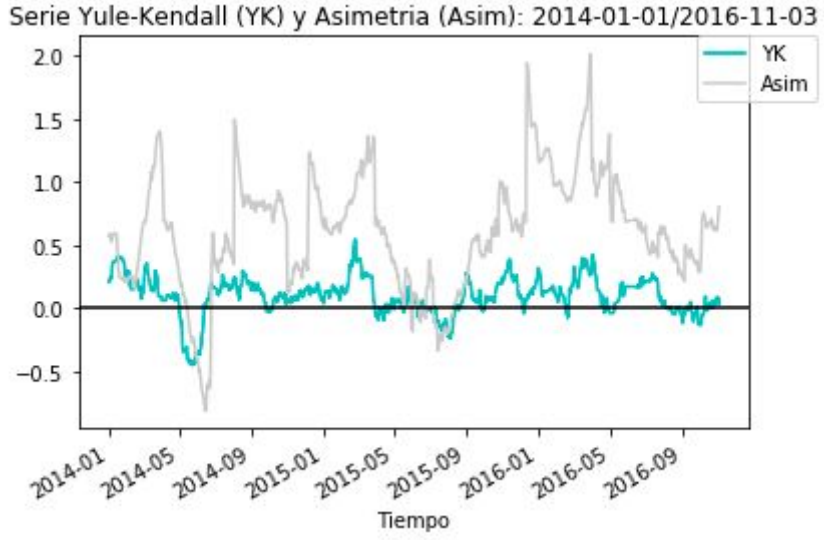

Gráfico 7. Coeficiente de YK y coeficiente de asimetría, con PM2,5.

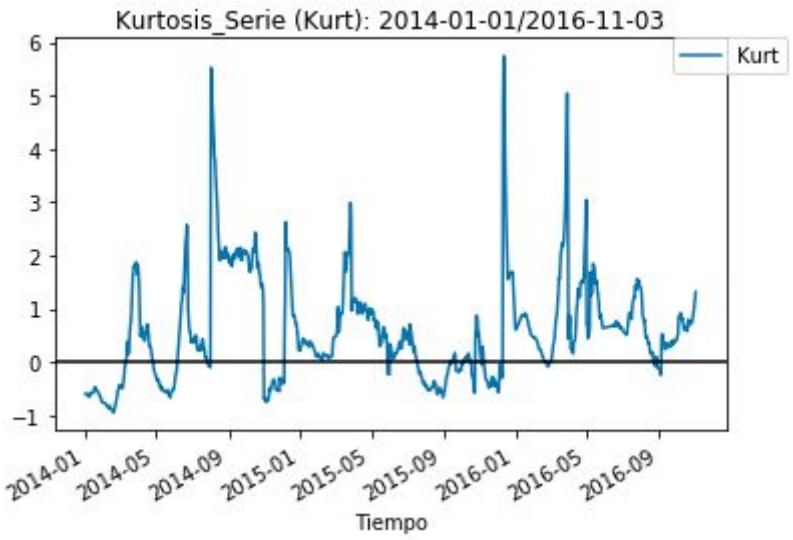

Gráfico 8. Coeficiente de Kurtosis con PM2,5.

Punto 8: Análisis de la tendencia en la serie

Se desarrolló una función que trabaja con la ecuación de Mann-Kendall [1], resultando entonces las variables de la *Tabla 2*. Es importante señalar que por tratarse de un test de doble cola pueden existir tanto valores por encima como por debajo de un valor de referencia (la media), el alpha de interés en este caso se debe dividir entre dos.

S	-28108,00
var	148494316,67
Z	-2,31

Y la comparación con la distribución normal? las conclusiones? Los demás estadísticos?

Tabla 2. Resumen de los datos entregados por la función de Mann-Kendall

Punto 9: Discusión de los resultados/Conclusiones

Lo primero que hay que resaltar es que no se trata de una serie de muchos datos (1101 en total), debido a que no se tomó un registro muy lejano porque es en los últimos 5 años donde se han venido presentando las contingencias ambientales por la mala calidad del aire en el AMVA, así que en un principio se quería analizar un periodo corto de tiempo para una sola estación. Es evidente que la serie de tiempo original de PM 2,5 tiene gran variabilidad (ver *Gráfico 1*), y se notan picos drásticos en los que el material particulado sobrepaso los 100ug/m^3 en AM (abril, mayo) aproximadamente de 2016. Esto tiene sentido siendo que en estos meses es cuando se presenta la primer época de invierno del año, entonces es donde existen las mayores alertas por calidad del aire al presentarse el fenómeno de inversión térmica, que deja los contaminantes “atrapados” en el Valle de Aburrá. La fdp y los índices de esta serie general (*Gráfico 3 y Tabla 1*), concuerdan. El coeficiente de asimetría siendo mayor a cero indica que su cola derecha está alargada, mientras que el de Kurtosis al ser también mayor a cero, indica que los datos se concentran sobre la media, y se encuentra más espigada la distribución, como se puede observar en el respectivo gráfico. Cabe señalar que la std es de considerable magnitud, sabiendo que la media es de 31,42ug/m^3 y la OMS establece que en un dia la concentración no debe superar los 25ug/m^3 para no afectar la salud.

Ojo con esas conclusiones! Asegúrense de investigar correctamente el fundamento físico de su problema.. Para empezar los países tropicales no tienen estaciones.

Ahora, se estudian las series móviles. Es fácil darse cuenta del *Gráfico 4* que las fdp´s móviles o sus respectivos histogramas no son estacionarios, y pro el contrario muy variables en el tiempo (esto se puede confirmar con el gif anexado). Profundizando en los índices o estadísticos móviles, que se encuentran resumidos en el *Gráfico 5*, también es fácil ver que no estacionarios, y que tanto en este punto como en la serie general, no se logra ver una tendencia marcada, pero si un cierto indicio de ciclos, que podrían deberse al propio sistema bimodal de lluvias de la región andina de Colombia, que es este periodo de transición entre temporada seca a húmeda lo que causa el fenómeno de la inversión térmica, y por tanto, de las contingencias ambientales por mala calidad del aire.

En el *Gráfico 6*, tanto la media como la std, también tienden a mostrar una especie de ciclo de la concentración del material particulado, donde los picos ocurren precisamente alrededor de los meses de MAM y SON (marzo, abril mayo y septiembre, octubre y noviembre), periodos donde la radiación solar es bajar y ocurre lo contrario en las demás épocas, afirmando que la concentración de PM2,5 está muy marcada por la radiación del momento sobre la superficie terrestre. En cuanto a los Gráficos 7 y 8, existe también una concordancia con las fdp´s obtenidas, ya que en la mayoría del tiempo kurt, asim y YK son mayores a 1 indicando que es espingada la distribución (los datos se concentran alrededor de la media) y que tiene su cola derecha estrecha y alargada (es decir, el PM2,5 aumenta su concentración drásticamente en ciertos periodos de tiempo).

Los Gráficos 9 y 10 afirman que la serie general no es realmente estacionaria, y nuevamente se pueden observar ciertas zonas muy marcadas en concentración de PM2,5 y que parecen tener cierto ciclo, lo que resulta de un alto interés de análisis para las tareas posteriores.

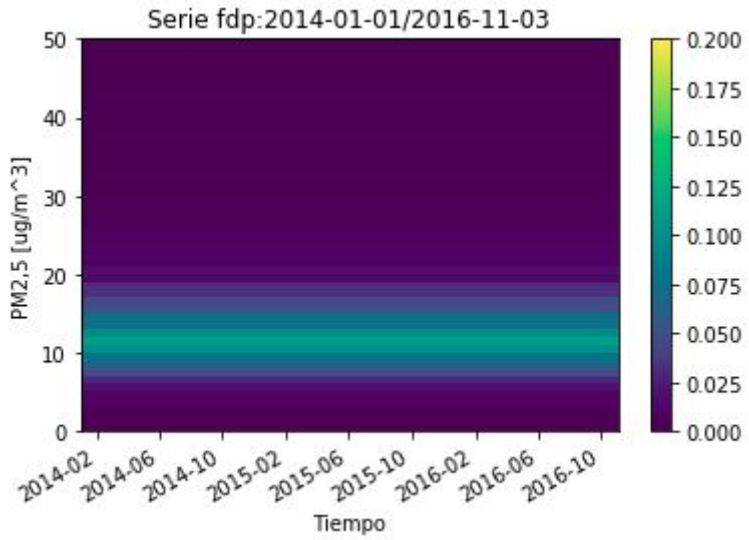

Gráfico 9. Matriz de la serie fdp original.

Ojo, esto no se debe mostrar, podría inducir a conclusiones erróneas, la siguiente es la adecuada…. en el código esto únicamente hacía parte de una demostración

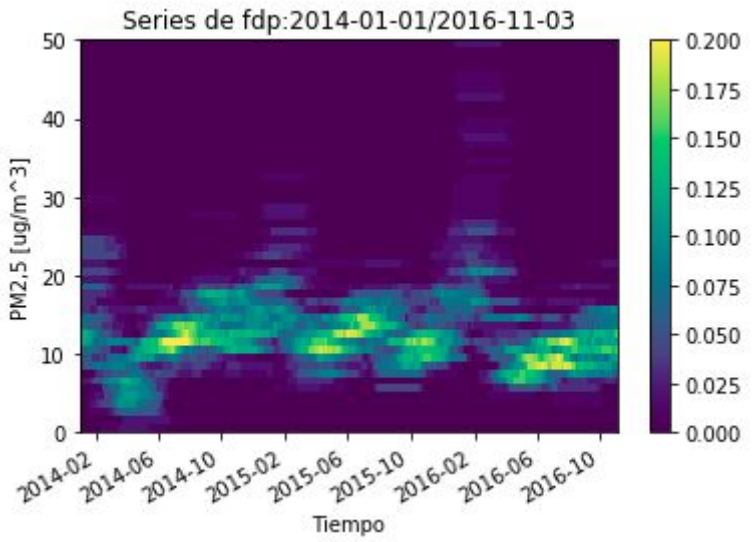

Gráfico 10. Matriz de la series fdp's móviles.

Bibliografía consultada

- [1] Cantor, D. EVALUACIÓN Y ANÁLISIS ESPACIOTEMPORAL DE TENDENCIAS DE LARGO PLAZO EN LA HIDROCLIMATOLOGÍA COLOMBIANA. Universidad Nacional de Colombia, Sede Medellín. Tesis de maestría en Recursos Hidráulicos. pp. 12-13.
- Wilks, D. Statistical Methods in the Atmospheric Sciences, second Edition. Department of Earth and Atmospheric Sciences. Cornell University. Chapter 3: Empirical Distributions and Exploratory Data Analysis. pp. 23-69.

Anexos

1- Página usada para la descarga de datos de PM2,5:
https://siata.gov.co/descarga_siata/index.php/index2/login

2- Códigos en VBA para organizar los datos de PM2,5:

```
Private Sub CommandButton1_Click()
b = 1118
For i = 41 To 41
    For j = 6 To 36
        a = Hoja1.Cells(j, i)
        Hoja5.Cells(b, 7) = a
        b = b + 1
    Next j
Next i


cont2 = 1
For k = 1 To 37
    cont1 = 1
    For m = 1 To 31
        c = Hoja1.Cells(cont1 + 5, 1)
        Hoja5.Cells(cont2 + 1, 1) = c
        cont1 = cont1 + 1
        cont2 = cont2 + 1
    Next m
Next k


contador = 0
e = 0
r = 0
x = 0
,
For n = 1 To 37
    x = x + 1
    If x <= 12 Then
        For p = 1 To 31
        contador = contador + 1
        Hoja5.Cells(contador + 1, 2) = x
        Next p
    ElseIf x > 12 And x <= 24 Then
        e = e + 1
        For t = 1 To 31
        contador = contador + 1
        Hoja5.Cells(contador + 1, 2) = x * e / x
        Next t
    ElseIf x > 24 And x <= 36 Then
        r = r + 1
        For y = 1 To 31
        contador = contador + 1
        Hoja5.Cells(contador + 1, 2) = x * r / x
        Next y
    Else
        For w = 1 To 31
        contador = contador + 1
        Hoja5.Cells(contador + 1, 2) = x / x
       Next w
    End If
```

- Se anexa al correo el “gif” que muestra la variabilidad o no estacionariedad de los histogramas o fdp.
- Se anexa al correo el archivo excel con macros que muestra el código para organizar los datos.