

* Sobre 4.5

Trata de entregarlo a tiempo la próxima,
en general hiciste un muy buen trabajo...
Ojo con el análisis de tendencia

3.9

Ten cuidado con la redacción y
con el uso de lenguaje científico

TALLER NO 1. ANÁLISIS DE SERIE DE DATOS

ANTONY ESTEFANY HINESTROZA CÓRDOBA

**ESTUDIANTE DE:
ESPECIALIZACIÓN EN APROVECHAMIENTO DE RECURSOS HIDRÁULICOS**

TÓPICOS AVANZADOS EN SISTEMAS HIDROLÓGICOS

**UNIVERSIDAD NACIONAL DE COLOMBIA, SEDE MEDELLÍN
FACULTAD DE MINAS, DEPARTAMENTO DE GEOCIENCIAS Y MEDIO AMBIENTE
MEDELLÍN, SEPTIEMBRE DE 2018**

TABLA DE CONTENIDO

1. INTRODUCCIÓN	3
2. OBJETIVO GENERAL	3
3. SERIE DE DATOS	4
4. LECTURA DE LOS DATOS	4
5. GRÁFICA DE LOS DATOS	4
6. ESTIMACIÓN DE PDF (HISTOGRAMAS)	6
6.1. PERCENTILES	6
6.2. PROBABILIDADES	7
7. ESTIMACIÓN DE ÍNDICES DE PDF	9
7.1. ÍNDICES DE LOCALIZACIÓN	9
7.2. ÍNDICES DE DISPERSIÓN.....	9
7.3. ÍNDICES DE ASIMETRÍA	10
8. SON ESTACIONARIOS LOS ÍNDICES	11
9. TENDENCIA DE LA SERIE	12
10. DISCUSIÓN RESULTADOS ENCONTRADOS	12
11. CONCLUSIONES	13

LISTADO DE TABLAS

Tabla 1. Localización de la Estación Limnigráfica La Pintada.	4
Tabla 2. Estimación de índices de medidas de sumatorias numéricas.....	5

LISTADO DE FIGURAS

Figura 1. Lectura de la serie de datos utilizando el software Python 2.7	4
--	---

LISTADO DE GRÁFICAS

Gráfica 1. Serie de caudales diarios Estación La Pintada (1965 – 2015).	5
Gráfica 2. Histograma de los percentiles	6
Gráfica 3. Intervalos de clase evaluados	7
Gráfica 4. Índice del comportamiento de la probabilidad de ocurrencia de eventos.....	8
Gráfica 5. Vector de la ocurrencia del rango de caudales más presentados en el punto de interés	8
Gráfica 6. Comportamiento de los índices de Localización	9
Gráfica 7. Comportamiento de los índices de Dispersión	10
Gráfica 8. Comportamiento de los índices de Asimetría.....	10
Gráfica 9. Evaluación de estacionariedad de los índices.....	11
Gráfica 10. Contornos de la estacionariedad de los índices	11
Gráfica 11. Tendencias de la serie de los datos de caudales analizados.....	12

1. INTRODUCCIÓN

El análisis de la información hidrometeorológica, hoy día sea convertido en una herramienta fundamental para la toma de decisiones, ya que esta nos ayuda a entender sobre el comportamiento de los fenómenos macroclimáticos, que pueden ocurrir en el entorno de los periodos de tiempo. Como son los comportamientos en la precipitación de algunas zonas, lo que hace parte de la oscilación de los caudales de los ríos en el territorio colombiano.

Por ello se requiere para el análisis estadístico de las series de tiempo que se disponga de herramientas que nos ayuden a minimizar tiempo y agilicen la entrega de información amena y de fácil entendimiento, para quien la lea, por lo que se requiere de involucrarle a la ingeniería herramientas de programación, las cuales complementan y/o son muy importante a la hora de toma de decisiones en los diferentes ámbitos, siendo estos investigativos o laborales.

Algunas herramientas estadísticas como los índices de localización, dispersión y asimétricos, nos ayudan a generar ideas globales de comportamiento de nuestras series de tiempo que se estén evaluando, por lo que se requiere de hacer trabajos con mucha responsabilidad y así ser más acertados a la hora de dar algún diagnóstico.

2. OBJETIVO GENERAL

Realizar el análisis de una serie de datos que comprenda como mínimo 1.000 datos, teniendo en cuenta las herramientas estadísticas tales como los índices de localización, asimetría y dispersión, así como también los percentiles y la probabilidad.

3. SERIE DE DATOS

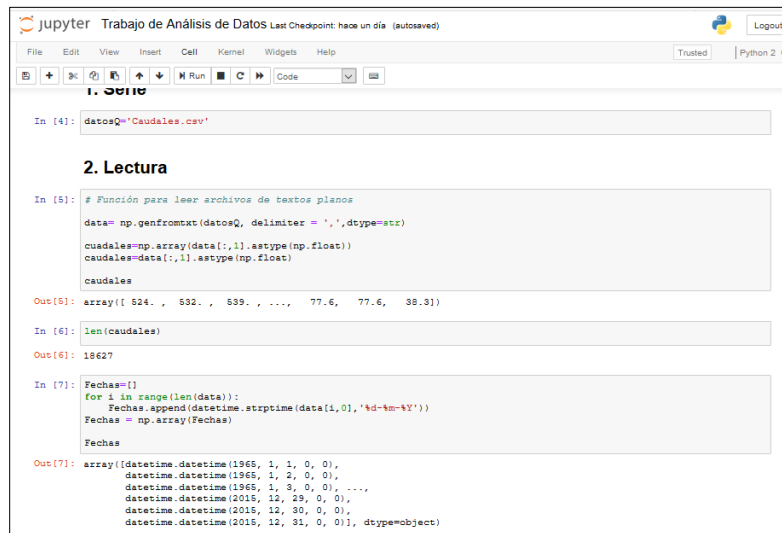
Para este trabajo se toma, la información de los registros históricos de caudales de la estación limnigráfica La Pintada (26187110), se utilizó una serie de datos comprendida entre el año 1965 y 2015 (50 años de registros). Esto último es representativo para el análisis de la influencia de eventos macroclimáticos ya que garantiza al menos la ocurrencia de dos (2) eventos Niños y 2 eventos Niña. En la Tabla 1 se presenta la localización de la estación limnigráfica y los año de inicio y el periodo de medición que se tiene para la serie.

Tabla 1. Localización de la Estación Limnigráfica La Pintada.

Código	Estación	X (m)	Y (m)	Altitud (msnm)	Inicio	Fin
26187110	La Pintada	829.268,76	1.115.330,59	543,0	1.965	2.015

4. LECTURA DE LOS DATOS

Se realiza una lectura de la serie de datos utilizando el software Python 2.7, mediante la plataforma de Anconda-Jupyter notebook, donde se lee la información utilizando algunos de los paquetes de programación que se presentan con este software mencionado (ver Figura 1), que se presenta con el informe.



```

jupyter Trabajo de Análisis de Datos Last Checkpoint: hace un día (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 2.7.0

1. Serie

In [4]: datosQ="Caudales.csv"

2. Lectura

In [5]: # Función para leer archivos de textos planos
data= np.genfromtxt(datosQ, delimiter = ',', dtype=str)
caudales=np.array(data[:,1].astype(np.float))
fechas=data[:,0].astype(np.float)
caudales

Out[5]: array([ 524. ,  532. ,  539. , ...,  77.6,  77.6,  38.3])

In [6]: len(caudales)

Out[6]: 18627

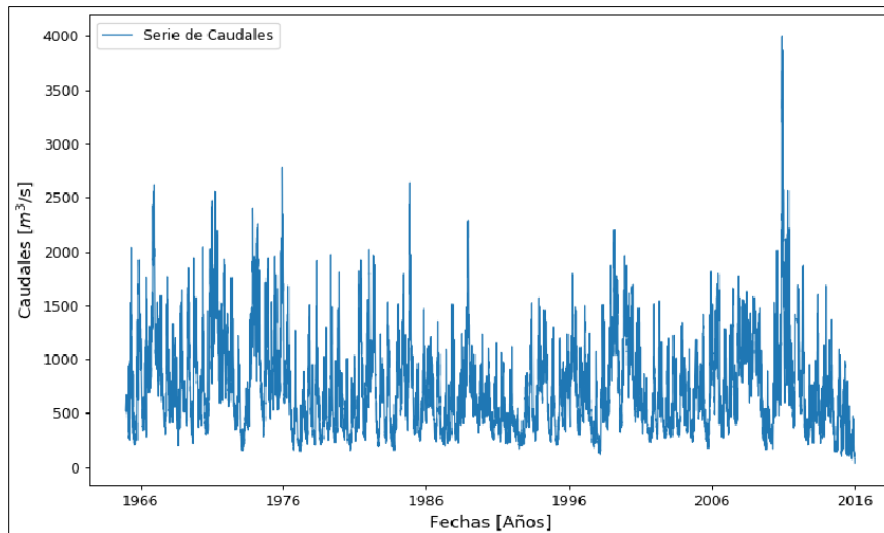
In [7]: Fechas=[]
for i in range(len(data)):
    Fechas.append(datetime.strptime(data[i,0], '%d-%m-%Y'))
Fechas = np.array(Fechas)
Fechas

Out[7]: array([datetime.datetime(1965, 1, 1, 0, 0),
               datetime.datetime(1965, 1, 2, 0, 0),
               datetime.datetime(1965, 1, 3, 0, 0), ...,
               datetime.datetime(2015, 12, 29, 0, 0),
               datetime.datetime(2015, 12, 30, 0, 0),
               datetime.datetime(2015, 12, 31, 0, 0)], dtype=object)
    
```

Figura 1. Lectura de la serie de datos utilizando el software Python 2.7

5. GRÁFICA DE LOS DATOS

Mediante la utilización de software Python 2.7, después de leer la serie de datos se procedió a graficar la información, ya que esto nos permite tener una idea de los registros que se tienen. De la Gráfica 1, se puede observar que existen diferentes comportamientos a lo largo de los 50 años de registros, como también que se presentaron registros que están por encima de los 3.000 m³/s, estos están aproximadamente entre el periodo comprendido en los años de 2.010 y 2.011.



Gráfica 1. Serie de caudales diarios Estación La Pintada (1965 – 2015).

De la serie de datos reportados entre el año 1965 y 2015, se realizaron algunos análisis de índices de localización, dispersión y asimétricos, los cuales se presentan en la Tabla 2, donde la media de caudales fue del orden de 716,23 m³/s, lo que nos indica que esta fuente hídrica en el punto evaluado presentan un promedio de 716,23 m³/s, durante los 50 años que comprende la serie evaluada, variación de caudales. También se observa que los percentiles 90 y 10 analizados se encuentran en tanto por encima como por debajo de la mediana y la media, de lo anterior se puede deducir que estas variables se encuentran afectadas por los diferentes fenómenos del ciclo anual de caudales en el punto de interés, donde se puede presentar un comportamiento bimodal con valores máximos alrededor de mayo y diciembre, lo cual es congruente con el doble paso de la ZCIT (Zona de Convergencia Inter Tropical) sobre Colombia. Los caudales mínimos se presentan durante los meses de febrero y agosto, estando también muy relacionado con los periodos lluviosos en el territorio colombiano. Es de anotar que en el periodo comprendido entre el año 1965 y 1976, Colombia en general, estuvo influenciada por la ocurrencia de una serie de eventos (fenómeno del niño), catalogados como moderado, mientras que en los últimos tiempos se presentaron los eventos más extremos de precipitación comprendidos entre los fenómenos del niño y de la niña (Montealegre & Pabon, 2000), lo cual puede argumentar el incremento de los valores medios.

Tabla 2. Estimación de índices de medidas de sumatorias numéricas

Índices de Localización	
Media	716,23
Mediana	621,00
P90	1.289,00
P10	280,00
Índices de Dispersión	
Rango Intercuartílico (IQR)	417,59
Desviación Estándar	537,00
Índices Asimétrico	
Asimetría	1,3705
Yule-Kendall	0,6308

6. ESTIMACIÓN DE PDF (HISTOGRAMAS)

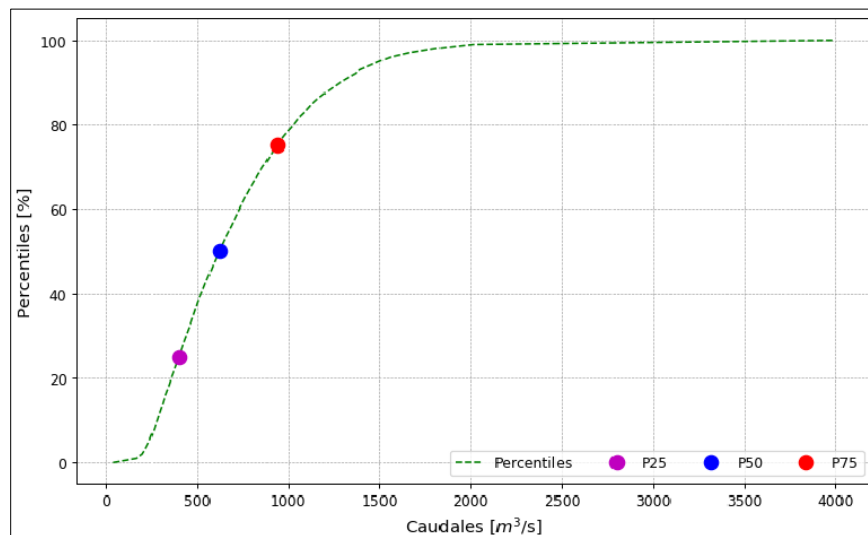
Para el análisis de la información se realizó la estimación de comportamiento de algunos histogramas, generados mediante la evaluación de los percentiles y la probabilidad de ocurrencia de algunos fenómenos de

6.1. PERCENTILES

En la Gráfica 3 se presenta la información de la evaluación de los percentiles, donde observa que el percentil 25 (P_{25}) de los datos evaluados de la serie de caudales de 50 años, se encuentra por el orden de los $399,0 \text{ m}^3/\text{s}$, donde el rango que comprende el 25% de los datos analizados se encuentra de $38,3$ a $399,0 \text{ m}^3/\text{s}$, los cuales se encuentran por debajo de los valores encontrado en la media y mediana de los datos de la serie de caudales.

Según lo analizado en el percentil 75 (P_{75}), se concluye que el 75% de los datos analizados están por debajo de $1.000,0 \text{ m}^3/\text{s}$, sobre la fuente hídrica superficial evaluada (Río Cauca), como también que el 25% restante de los datos superan los $1.000,0 \text{ m}^3/\text{s}$, correspondiente a un rango de datos de 239 datos, de lo cual se puede deducir que los caudales más elevados se presentaron en el año 2.010 donde en Colombia se presentó un temporada invernal muy crítica entre los meses de octubre y diciembre del año en mención, por lo que se puede argumentar que estos valores ya están por encima de los índices de localización como la media y el P_{90} .

De lo anteriormente dicho se puede concluir que el aumento del caudal encontrado en la serie de datos analizados posiblemente se deba a la situación que se originó debido a las abundantes lluvias provocadas por depresiones tropicales; y frentes cálidos propios de la época húmeda, ocasionados por el fenómeno de La Niña, lo que trajo consigo la ocurrencia de fuertes aguaceros que causaron grandes inundaciones en diferentes zonas de Colombia.

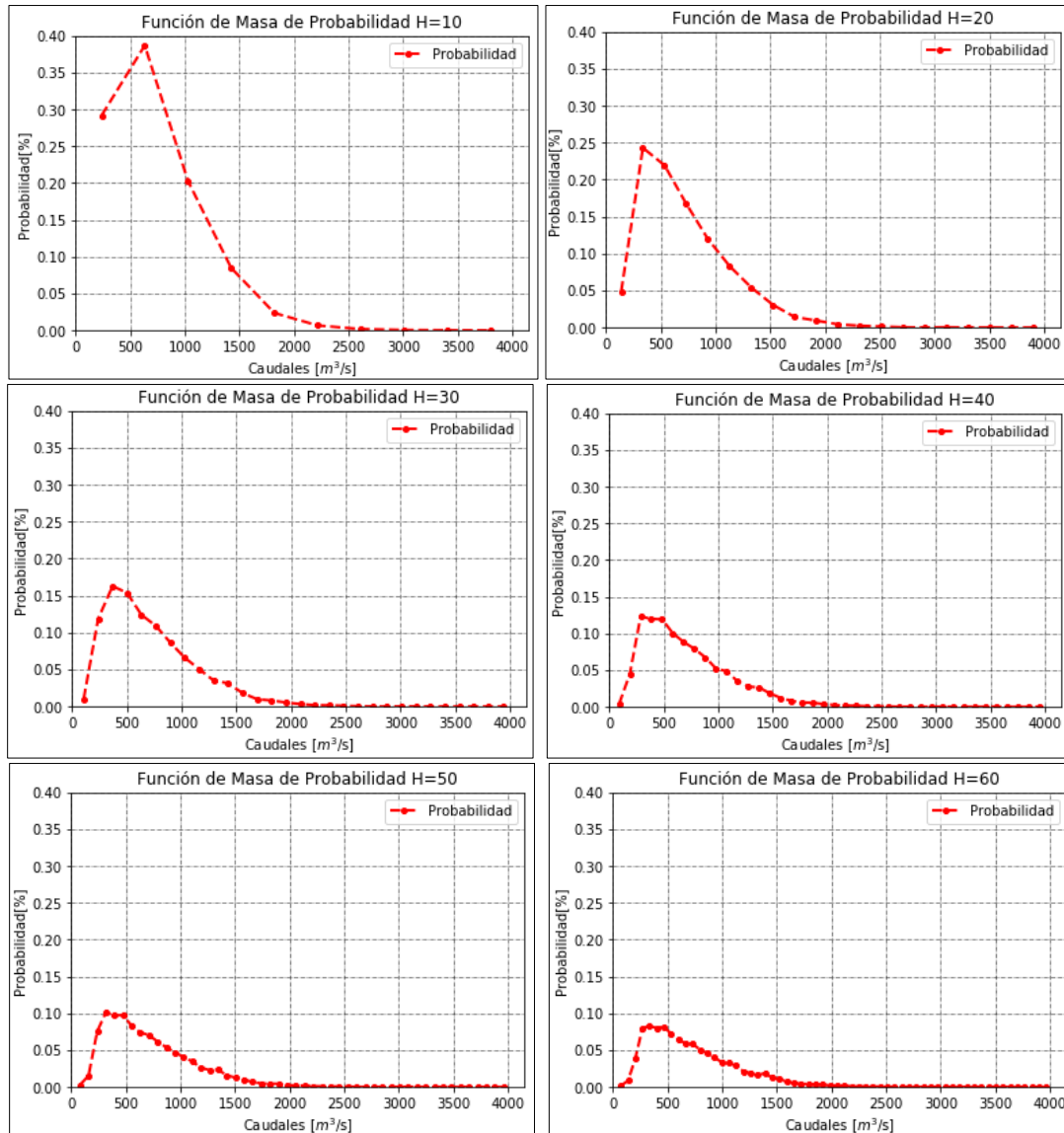


Gráfica 2. Histograma de los percentiles

El análisis del número de bins hace parte del “detrás de cámaras”, en un informe debe aparecer solo el mejor resultado, el que consideraste más adecuado de acuerdo con el objetivo

6.2. PROBABILIDADES

Para determinar el mejor intervalo de clase, para evaluar la probabilidad de ocurrencia de la masa de los datos se hicieron algunas pruebas con diferentes intervalos, donde en la Gráfica 3 se muestra el comportamiento de cada uno de ellos de la serie, de donde se determina que el mejor intervalo para la evaluación de la probabilidad del comportamiento de la mayor masa de datos o donde se presentan los caudales más cercanos a la media de los datos.

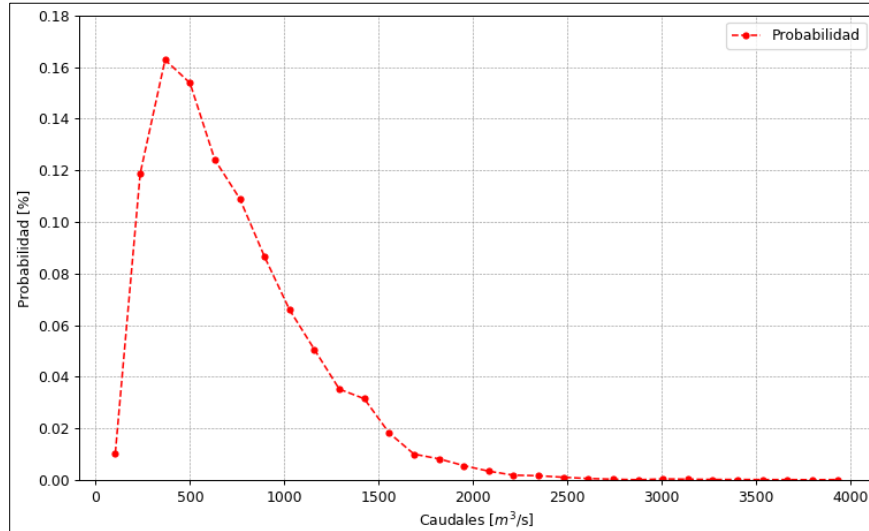


Gráfica 3. Intervalos de clase evaluados

Lo anterior es un punto de partida para determinar el mejor intervalo de clase del análisis de los datos, de donde se definió que el intervalo que mejor representa el comportamiento de los datos es el de 30 el cual muestra mejor el comportamiento de la masa de probabilidad.

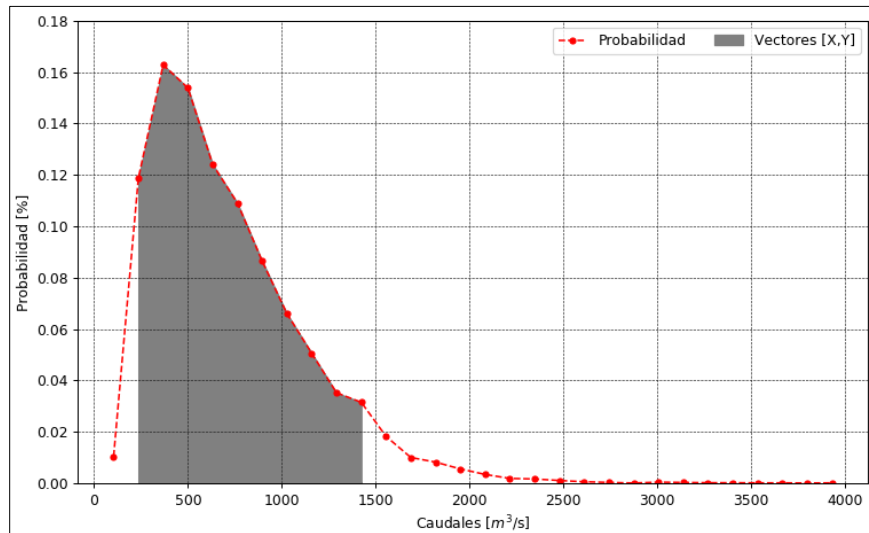
Según el comportamiento de los datos encontrados en la serie analizada, se puede deducir que esta presenta la masa de ocurrencia más representativa de caudales que varían de 38,3 a 1.500 m³/s, donde los valores que tiene una mayor probabilidad de presentarse se encuentran entre los 300 a 600 m³/s. como también podemos concluir que los valores que superan 1.500 m³/s, tienen una menor probabilidad de ocurrencia, por lo que estos

están más asociados a las diferentes anomalías que ocurren por los fenómenos de la niña en diferentes temporadas o ciclos anuales, tal como ocurrió en la temporada invernal de finales del año 2.010 y 2.011, lo que registro datos en esta serie que solo se presentaron una vez durante la longitud de esta serie analizada (ver Gráfica 4).



Gráfica 4. Índice del comportamiento de la probabilidad de ocurrencia de eventos

En la Gráfica 5 se relata el rango de caudales más frecuentes, que se presentaron durante la serie de datos que se está analizando, donde el 16,2%, son los datos que presentan mayor frecuencia de ocurrencia, estado por debajo de la media de localización de la serie y tan solo el 0,1% de los datos muestra una frecuencia de ocurrencia muy mínima durante dicha serie, siendo los valores que están por encima de los 1.500 m^3/s .



Gráfica 5. Vector de la ocurrencia del rango de caudales más presentados en el punto de interés

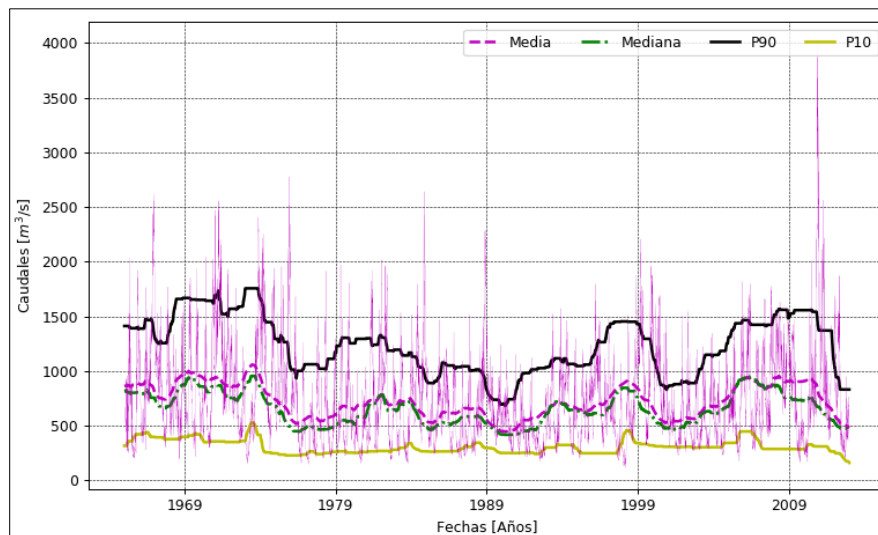
7. ESTIMACIÓN DE ÍNDICES DE PDF

Para determinar los índices de localización se emplean diferentes operadores matemáticos u estadísticos tales como: localización dispersión y asimetría de los datos evaluados.

7.1. ÍNDICES DE LOCALIZACIÓN

Para el análisis de los índices se utilizó una ventana de 365×3 , ya que esta nos ayuda a entender de forma más detallada el comportamiento de la serie evaluada, donde se analizaron variables estadísticas tales como: Media, Mediana y los percentiles P_{10} y P_{90} . De lo que se infiere que los datos evaluados no muestran tendencias definidas ni positivas o negativas ya que siempre están en fluctuaciones que dependen del comportamiento de variables hidrometeorológicas como lo pueden ser precipitación, evapotranspiración y entre otras que siempre van a incidir en el comportamiento de los caudales que se puedan presentar en una cuenca hidrográfica (ver Gráfica 6).

Se observa que las variables estadísticas analizadas muestran el mismo comportamiento a lo largo de la serie, donde una de ellas disminuye la otra tiene hacer lo mismo, por lo que no es evidente una tendencia marcada a ser creciente o decreciente en los datos (ver Gráfica 6).



Gráfica 6. Comportamiento de los índices de Localización

La serie atrás genera un poco de ruido, considera quitarla para la próxima :)

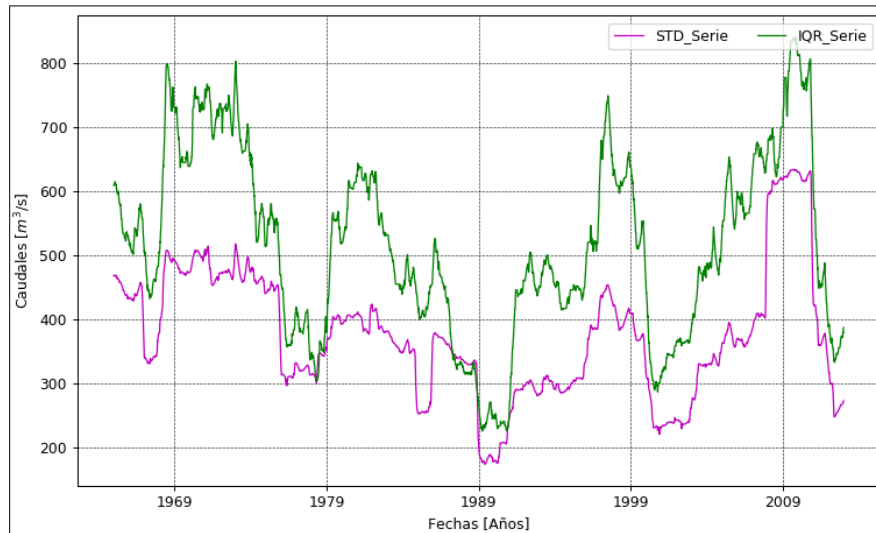
Según con la relación existente entre la media y la mediana se puede concluir que la serie de datos presenta un comportamiento asimétrico, muy sesgada hacia la izquierda, ya que la media es mayor que los valores presentados de mediana. Si bien sabemos que la media se ve afectada por los valores extremos encontrados en la serie analizada, mientras que la mediana es un estadístico más robusto y los valores extremos no se ven alterados.

7.2. ÍNDICES DE DISPERSIÓN

En la Gráfica 7 se presenta el comportamiento de la serie de datos analizados en los índices de dispersión (desviación estándar y Rango Intercuartílico (IQR)), donde se ve una variación semejante en el transcurso del tiempo, siendo mayor los datos obtenidos en el IQR, con respecto a la desviación estándar.

Se observa que la variación entre los valores obtenidos en el IQR, están entre 609 y 387,4 m^3/s , donde el valor mínimo es inferior al encontrado en la media móvil de la ventana evaluada, donde es evidente que el pico más alto presentado es en la temporada invernal presentada durante el año 2010. Mientras que los valores

obtenidos en la desviación estándar oscilan entre 468,02 y 273,14 m³/s, la cual se ve muy afectada por los valores extremos presentados (ver Gráfica 7).



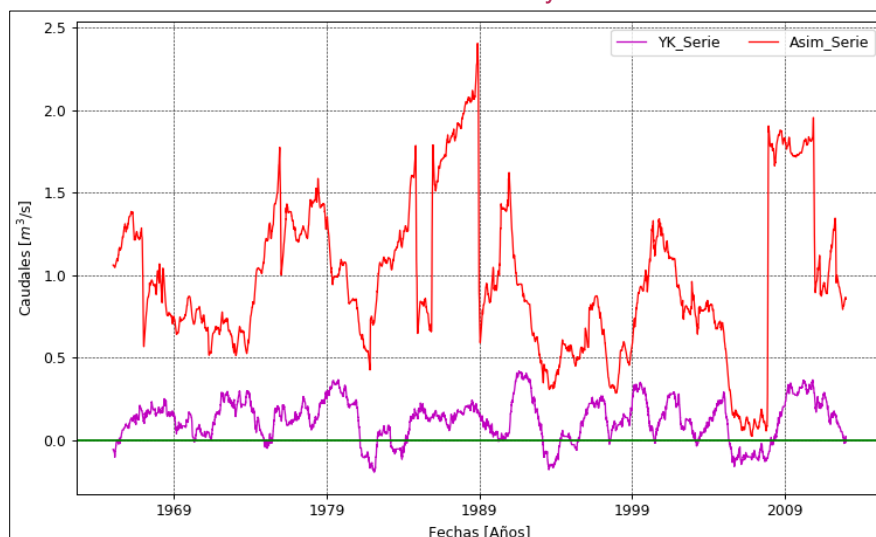
Gráfica 7. Comportamiento de los índices de Dispersión

7.3. ÍNDICES DE ASIMETRÍA

Los índices de asimetría nos muestran la distribución de los datos de la serie respecto al valor central de los mismo, donde en la metodología de Yule-Kendall, nos da una representatividad con valores más alejados a de 0, indicando que dicha distribución no es simétrica en algunos años de evaluación de la serie, ya que en otro se muestra como simétrica, los valores de Yule – Kendall varían desde -0,057 a 0,017, (ver Gráfica 8).

Mientras que la metodología del cálculo de asimetría empleado nos dice que los datos obtenidos en la ventana de la serie analizada no son simétricos ya que en su totalidad se alejan de 0, con picos muy pronunciados durante los años 1.989, 2.008 y 2.010, posiblemente asociados a diferentes fenómenos macroclimáticos ocurridos en estas épocas, los valores obtenidos por el método de asimetría varían desde 1,061 a 0,853, (ver Gráfica 8).

Este índice habla de la distribución y no de la serie como tal

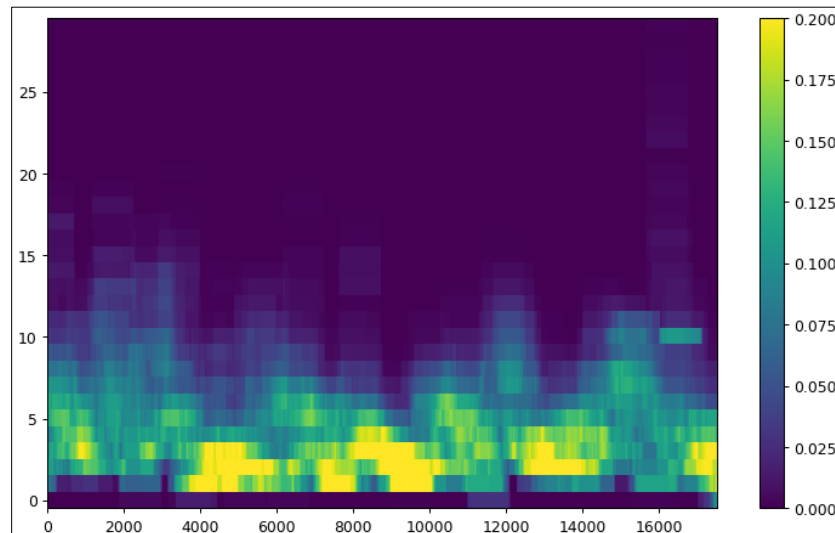


Gráfica 8. Comportamiento de los índices de Asimetría

8. SON ESTACIONARIOS LOS ÍNDICES

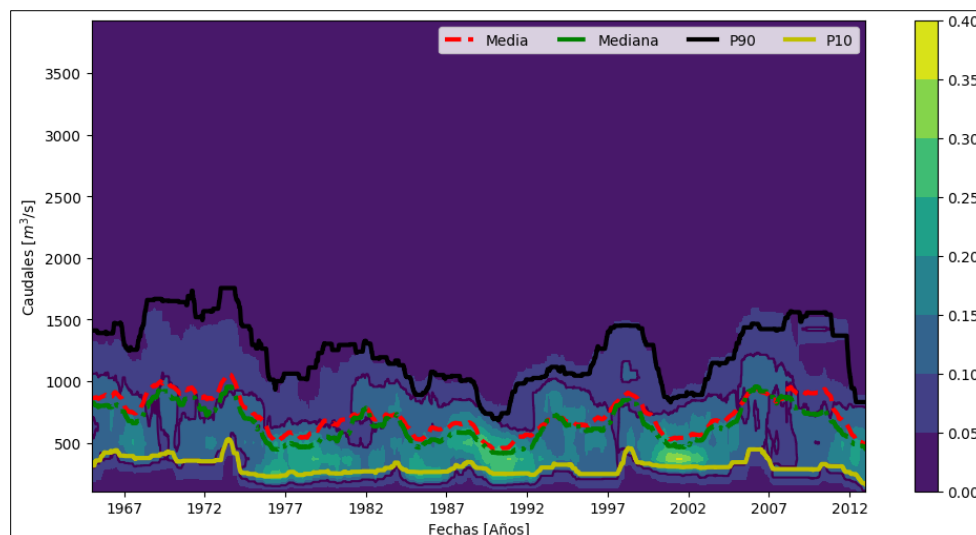
Según lo observado en el comportamiento entre la Gráfica 9 y Gráfica 10, se evidencia que los datos evaluado no presentan una estacionariedad definida ya que la variabilidad de los datos son muy fluctuantes a lo largo del tiempo.

Si bien observamos que en la Gráfica 10, donde se grafican algunos índices de localización, lo que nos ayuda a mostrar que nuestros datos son aleatorios en el tiempo, lo que refleja que durante este periodo analizado sucedieron algunos fenómenos macroclimáticos, estando asociado a sucesos irregulares o no periódicos.



Gráfica 9. Evaluación de estacionariedad de los índices

Adicionalmente se puede decir que en la Gráfica 10, se muestra como es el comportamiento de los índices de localización donde la media de la serie no se mantiene en el tiempo, viéndose muy afectada por los sucesos de ocurrencia de procesos no paramétricos, mostrando así que los datos no son estacionarios en el tiempo.

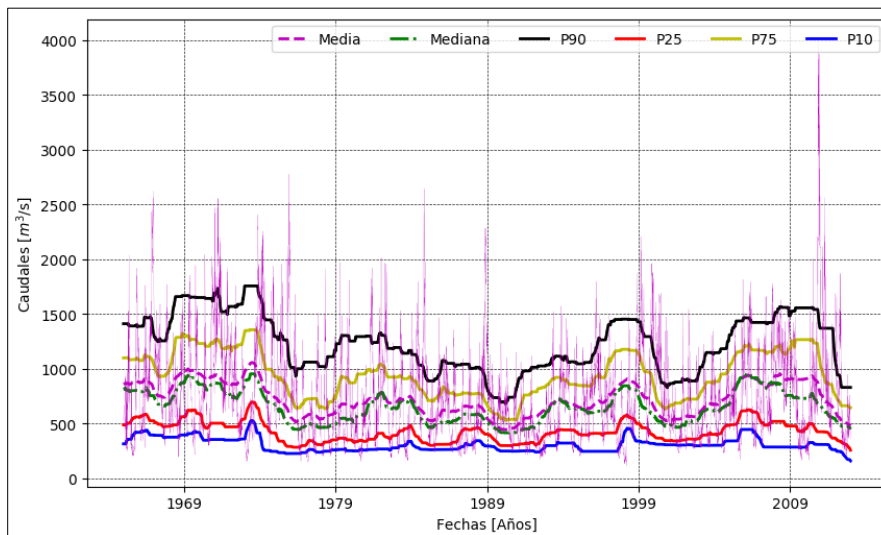


Gráfica 10. Contornos de la estacionariedad de los índices

9. TENDENCIA DE LA SERIE

A partir de la prueba de hipótesis de Mann – Kendall, se evaluó la tendencia en la media de la serie de datos de caudal en el punto de interés para este análisis se cuenta con una significancia del 99,999% se concluye que existen tendencias negativas en la media al considerar la serie de datos, lo cual es congruente con el análisis de tendencia con la serie completa de los datos analizados (ver Gráfica 11).

Considerando los resultados de las pruebas de hipótesis realizadas, además de la inspección visual de las series de caudales correspondientes al sitio de interés para este análisis, es posible concluir que, aunque existen tendencias negativas en la serie que producen periodos no homogéneos, estos cambios hidrológicos están influenciados por los principales eventos macroclimáticos que rigen los aspectos meteorológicos del trópico, especialmente la climatología de montañas.



Gráfica 11. Tendencias de la serie de los datos de caudales analizados

Esta conclusión no tiene mucho sentido, teniendo en cuenta los estadísticos... No observo una tendencia decreciente con esa probabilidad (99.999%). Me falta mucho detalle en este punto, donde están los estadísticos? S, z, la varianza?

10. DISCUSIÓN RESULTADOS ENCONTRADOS

En la Gráfica 1 muestra la serie original, en la cual se puede identificar que tiene una tendencia con ciertos picos indicadores de estacionalidad; es decir, la serie es no estacionaria, como también que los picos están asociados a diferentes eventos macroclimáticos.

En la Gráfica 11 se muestra que P_{75} de los datos analizados están por encima de la media y la mediana, donde encada uno de los índices de localización graficados se observan tendencias hacia abajo.

Este tipo de análisis nos ayuda a conocer alguna ocurrencia de fenómenos macroclimáticos que existen en la tendencia de la serie de datos bajo estudio.

Según el análisis de tendencia por la prueba de tendencia de Mann-K, se concluye que aun que se rechaza la hipótesis H_0 , no se debe hacer totalmente debido a que siempre existe un porcentaje muy pequeño de ocurrencia de esta.

11. CONCLUSIONES

De los datos analizados se concluye que la serie presenta tendencia negativa, aun que se presentan algunos picos en diferentes épocas, siendo más marcada la ocurrida durante el periodo de 2010 y 2011, que se presentaron grandes crecientes en los ríos colombianos, debido a la temporada de humedad relacionada con el fenómeno de la niña.

El análisis de la información hidrometeorológica es de gran importancia ya que nos brinda herramientas para el entendimiento de los diferentes fenómenos o comportamientos de estos a lo largo de una serie de tiempo determinada.

Finalmente podemos decir que la integración de los softwares de programación con la ingeniería es algo muy valioso debido a que nos agiliza en el procesamiento de información extensa adicional nos ayuda a minimizar los errores que están asociados a la manipulación directa de los datos.