

TAREA 2

* Sobre 4.5

Laura Lorena Barreto

Estudiante Especialización en aprovechamiento de recursos hidráulicos.

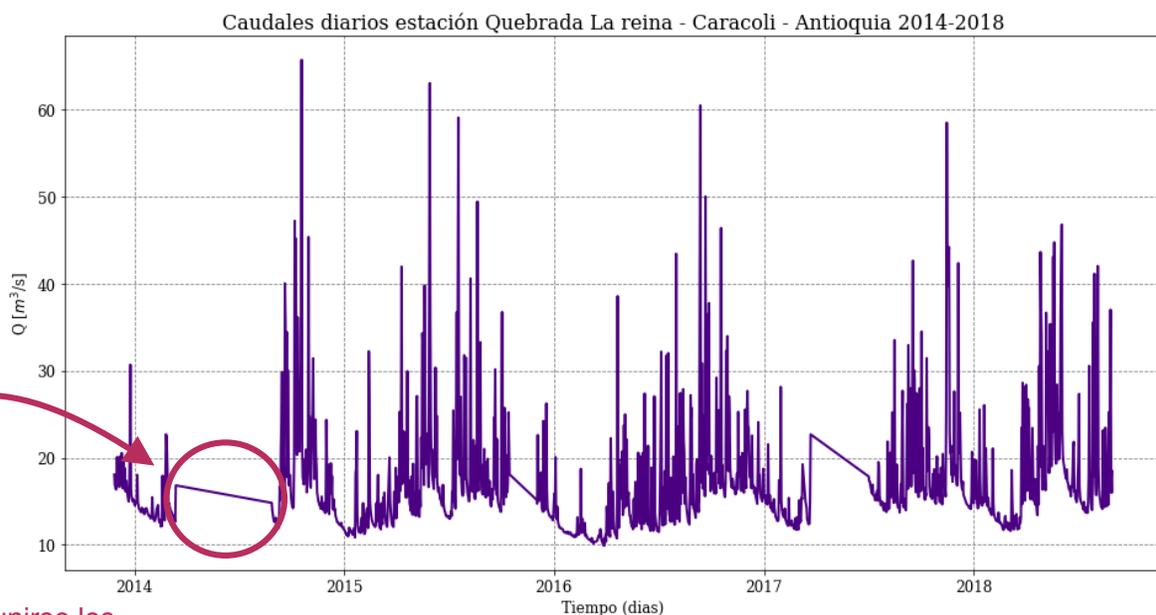
lbarreto@unal.edu.co**1. Selección de la Serie**

El archivo de datos CaudalesCaracoli.csv almacena registros de caudal en una resolución temporal diaria, los datos en esta serie son correspondientes al periodo 2013/11/27 al 2018/08/30 de la estación de la quebrada La Reina ubicada en el municipio de Caracolí – Antioquia.

Este archivo de datos fue obtenido directamente de la página del programa red de agua: Piragua de Corantioquia¹.

2. Lectura de la serie

La serie se compone de 1385 datos de caudal, el vector de datos organizado para realizar los cálculos mediante el software Python tiene 1385 filas y 2 columnas la primera columna corresponde a la fecha y la segunda columna al valor de caudal correspondiente.

3. Grafica

No deberían unirse los
datos faltantes

Grafica 1 Caudales diarios estación Quebrada La Reina – Caracolí – Antioquia 2014 – 2018.

La representación gráfica de la serie evidencia una ausencia de datos en los periodos de marzo de 2014 a agosto de 2014, marzo de 2017 y julio de 2017 entre otros datos aislados.

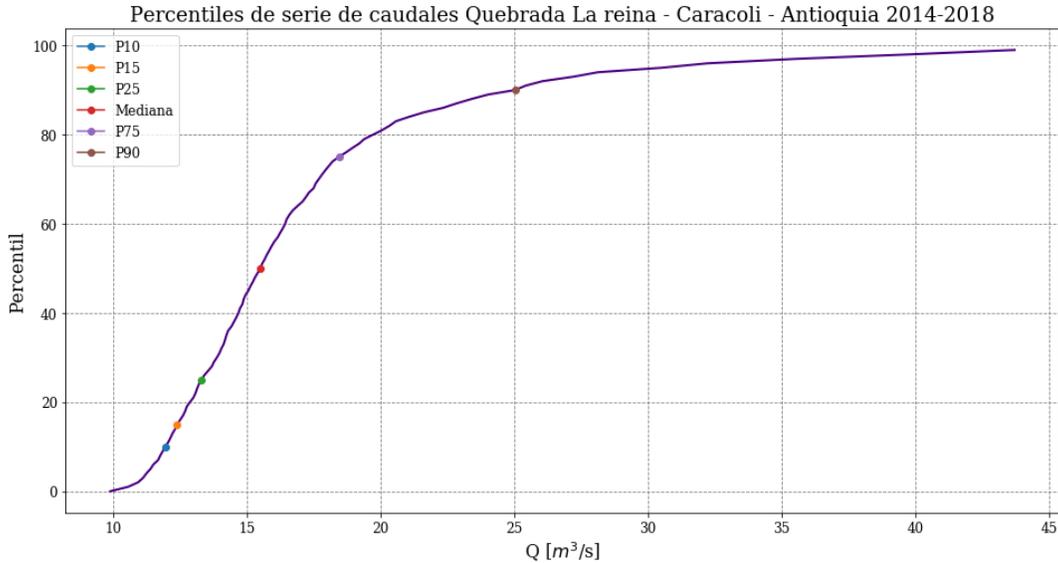
A pesar de la ausencia de datos la gráfica refleja en los caudales representados la distribución del régimen de lluvias bimodal característico del país, también se puede observar que el caudal base de este cauce esta entre los $10 \text{ m}^3/\text{s}$ y los $15 \text{ m}^3/\text{s}$.

¹ <http://www.piraguacorantioquia.com.co/piragua/>

4. Histogramas, Percentiles y probabilidad.

a. Percentiles

Del percentil 90 se concluye que el 90% de los datos de caudal para el intervalo de tiempo analizado son inferiores a 25 m³/s, del percentil 75 que el 75% de los datos es inferior a 18.5 m³/s y del percentil 25 que el 25% de los datos es inferior a 13.2 m³/s. Los percentiles P10, P15, incluso P25 pueden ser muy útiles para establecer los caudales ecológicos en una corriente, en este caso se tienen un P10 y P15 de 11.96 m³/s y 12.4 m³/s correspondientemente.

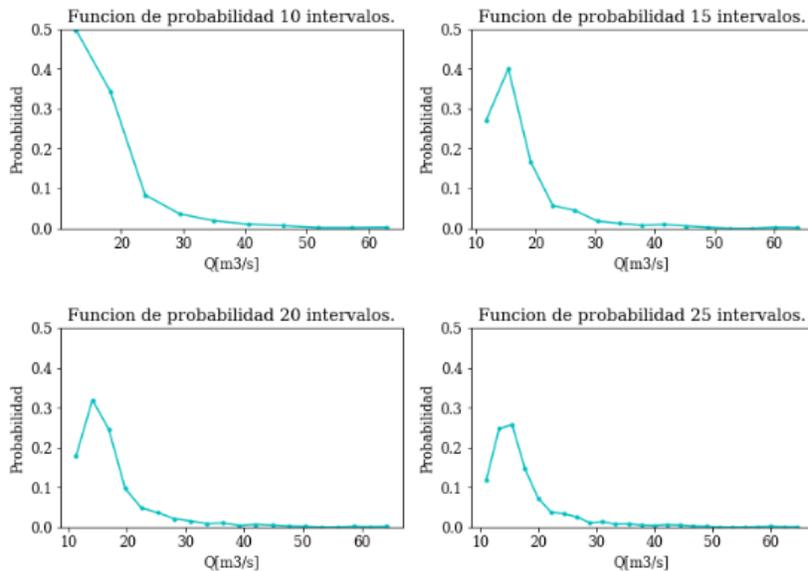


Gráfica 2 Percentiles de la serie de la serie de datos de caudales.

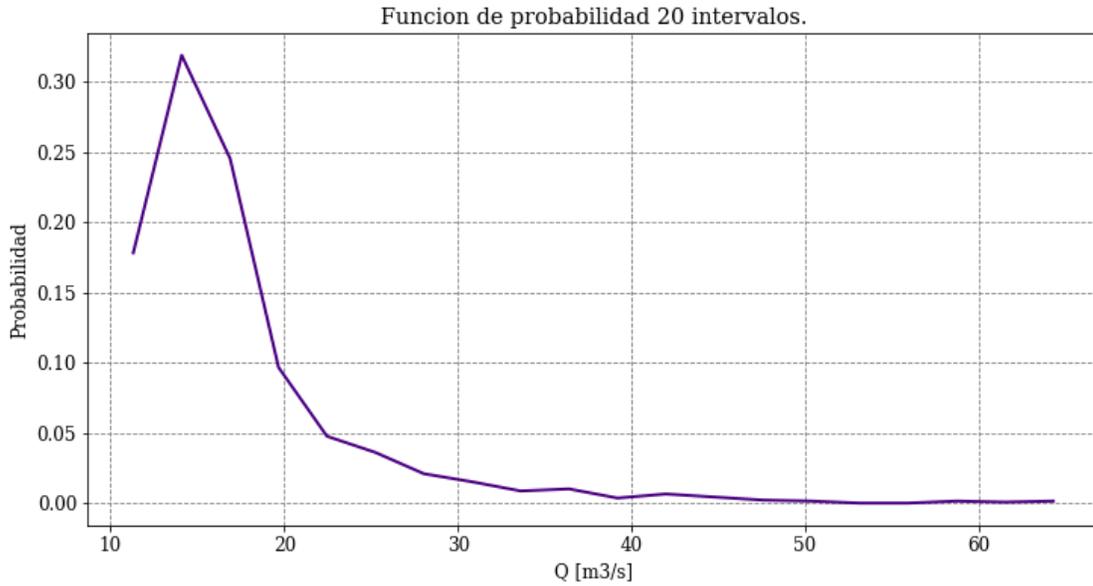
b. Probabilidad

La grafica 3 se muestra la función de probabilidad dependiendo de la cantidad de intervalos de clase escogidos, para este ejercicio tomaremos 20 intervalos de clase, pues para este rango se visualiza con más claridad la tendencia de las probabilidades.

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



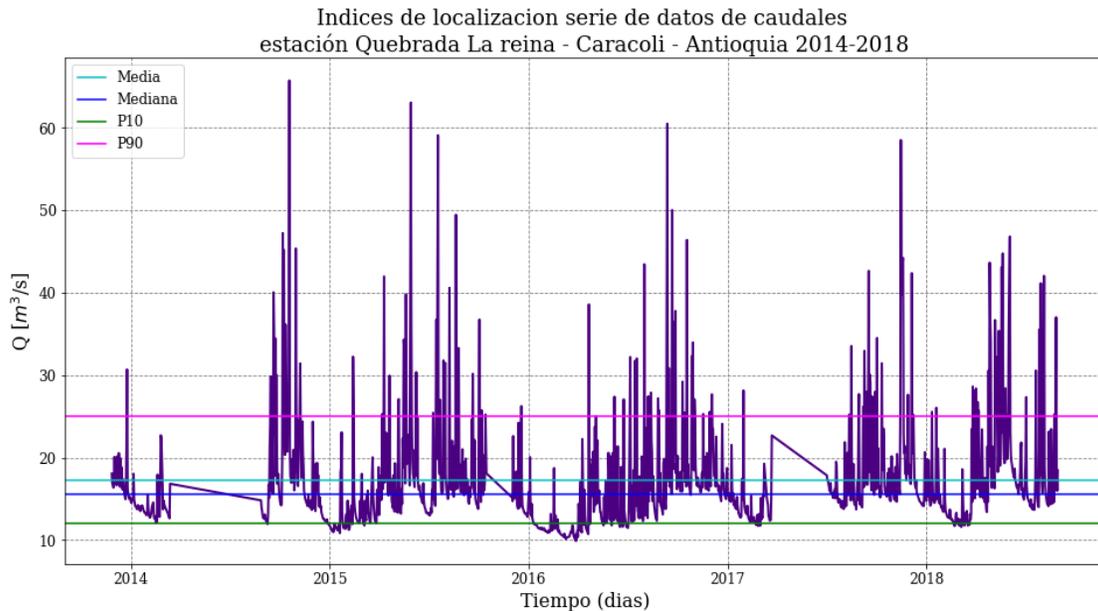
Gráfica 3 Probabilidad de la serie de datos de caudales, para 10, 15, 20 y 25 intervalos.



Gráfica 4 Probabilidad de la serie de datos de caudales, para 20 intervalos de clase.

Al analizar el comportamiento de la función probabilidad se observó en la Gráfica 4 que la probabilidad de que se presenten caudales superiores a 20 m³/s es inferior al 10% y superiores a 30 m³/s es inferior al 5%. A pesar de que la serie de datos corresponde a caudales medios diarios la Gráfica 3 tiene una distribución de campana asimétrica.

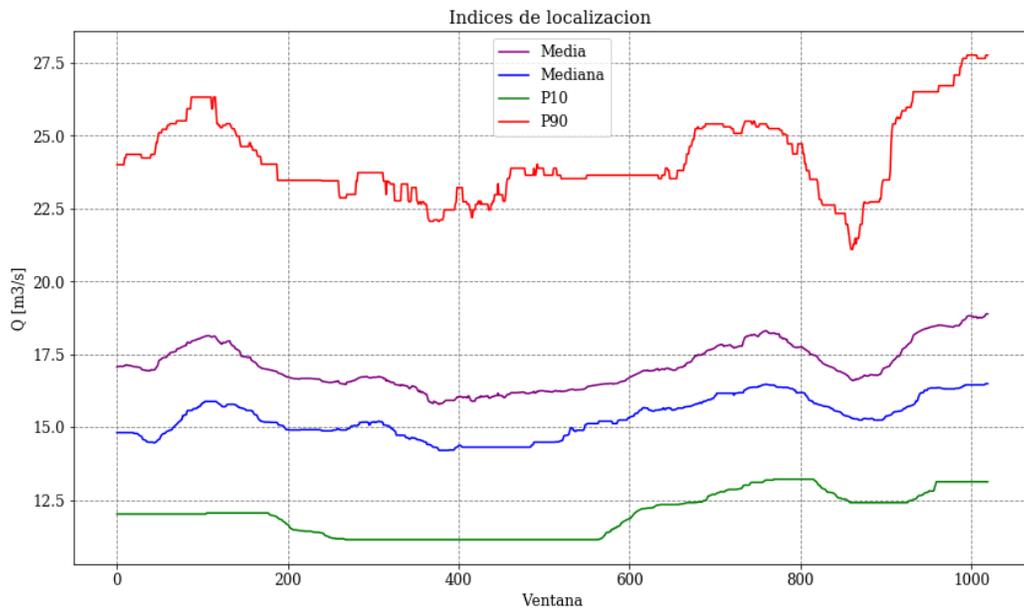
5. Índices de localización dispersión y asimetría de la serie de datos de caudales.



Gráfica 5 Probabilidad de la serie de datos de caudales.

El valor obtenido para la media de esta serie fue: 17.3 m³/s el valor de la mediana: 15.52 m³/s el valor del percentil 10: 11.96 m³/s y el valor del percentil 90 25.03 m³/s.

a. Índices de localización

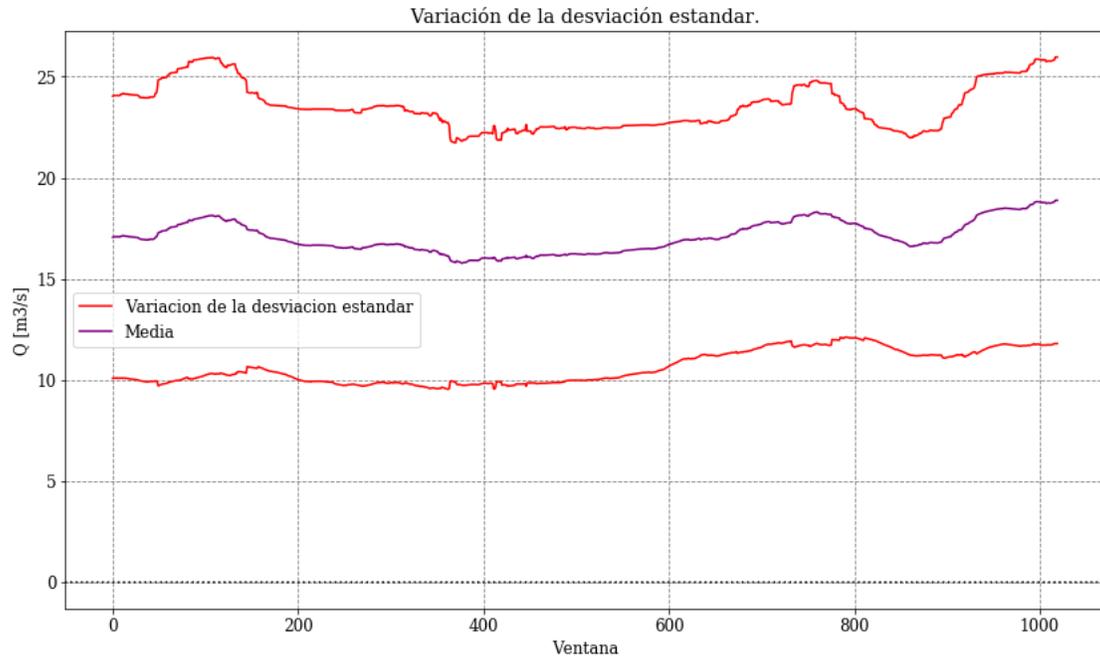


Gráfica 6 Probabilidad de la serie de datos de caudales.

La mediana y la media móvil calculadas en la gráfica 6 muestran que el valor de la media es superior a la mediana, esto indica que la distribución de los datos es asimétrica positiva, y se visualiza con claridad en la gráfica 4 donde la "cola" de la distribución apunta hacia la derecha.

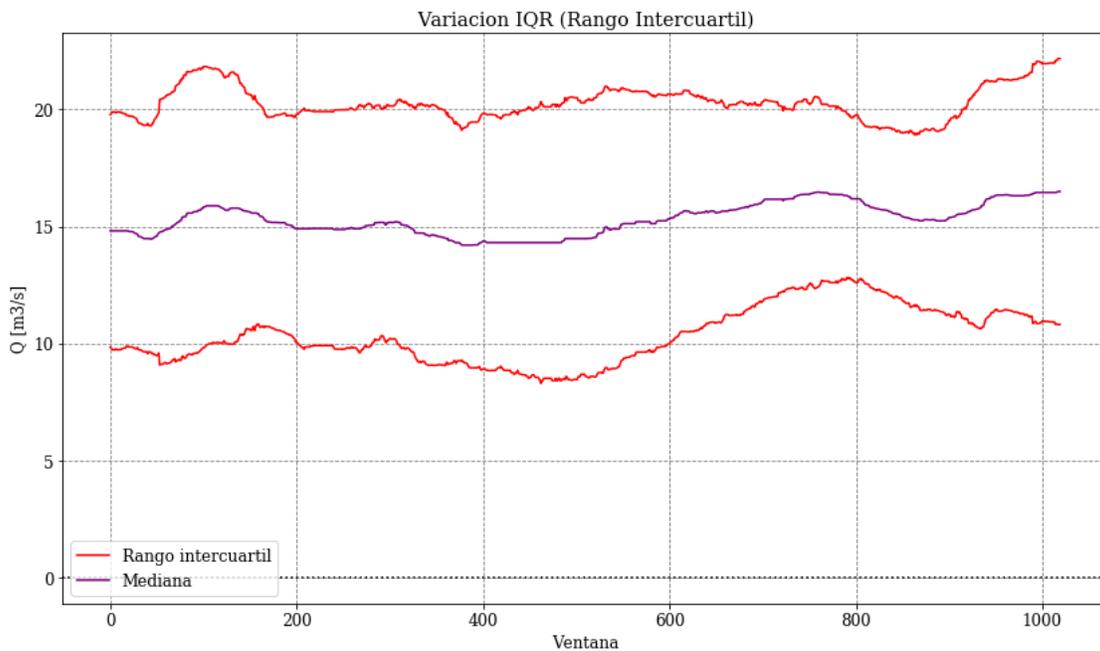
b. Índices de dispersión

En la siguiente grafica al representar la variación de la desviación estándar con respecto a la media, para cada una de las ventanas de datos seleccionadas, vemos como los valores de esta serie de datos se alejan en altas proporciones del caudal promedio, lo cual indica que existe una alta dispersión en los datos de la serie.

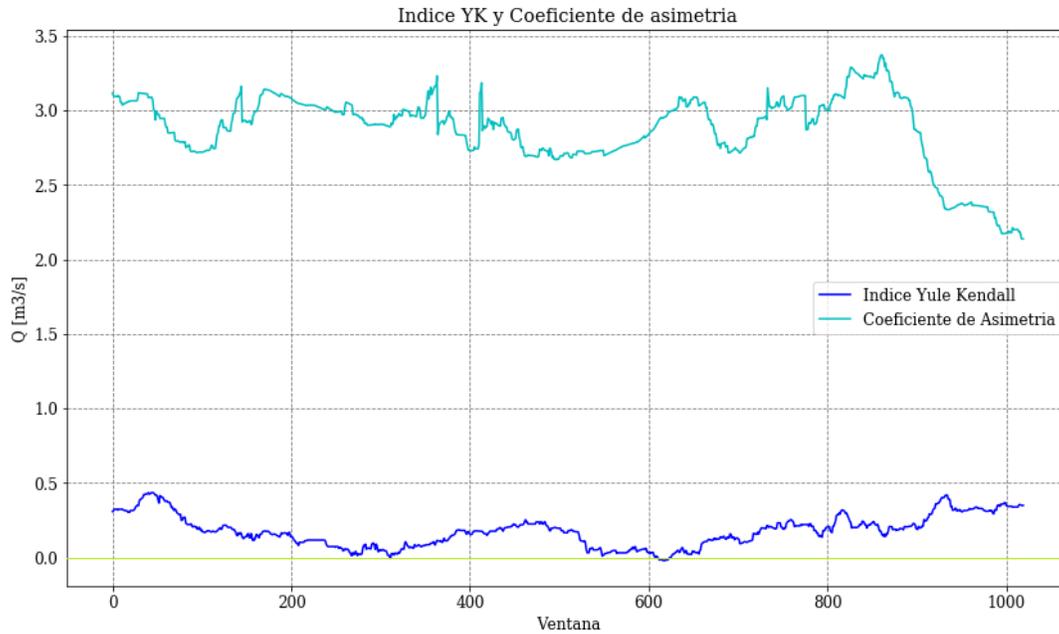


Grafica 7 Probabilidad de la serie de datos de caudales.

A medida que aumenta la dispersión de los datos, el rango intercuartil se hace más grande en la gráfica 8 vemos como este rango se mantiene constante es decir la dispersión de esta serie de datos es constante en el tiempo.



Grafica 8 Probabilidad de la serie de datos de caudales.



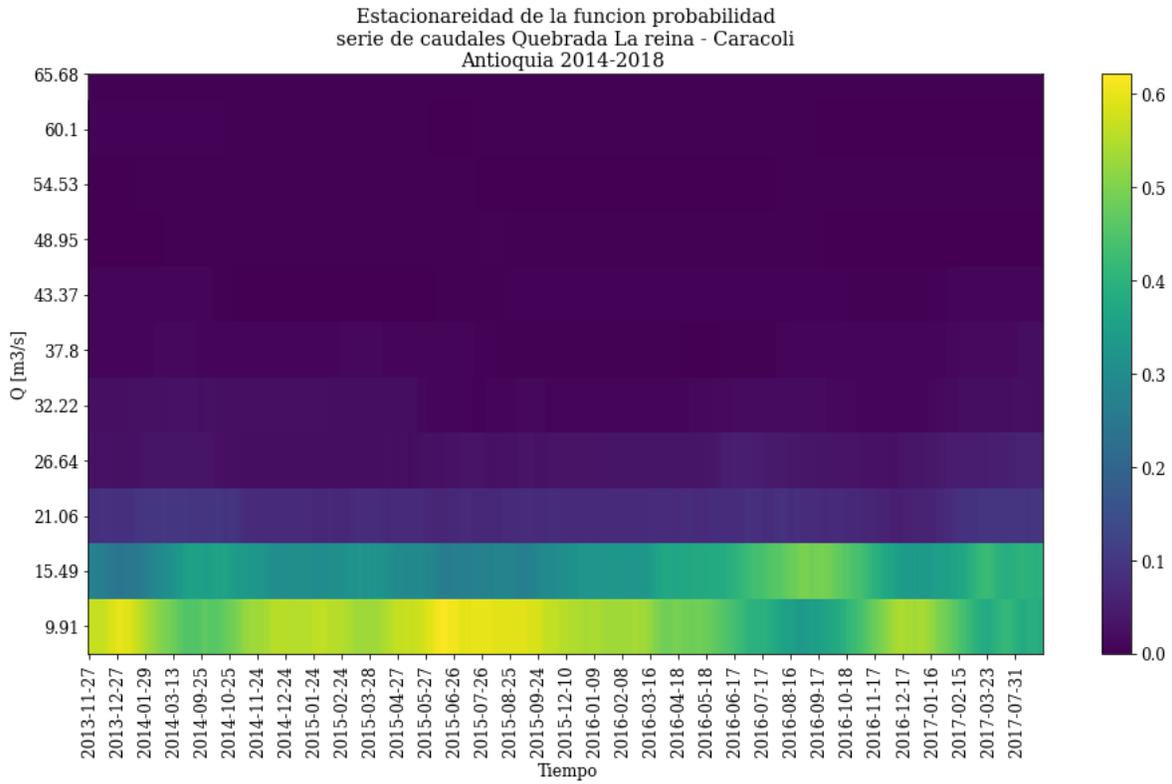
Grafica 9 Probabilidad de la serie de datos de caudales.

Las distribuciones asimétricas positivas o hacia la derecha tienen coeficientes mayores que cero es decir positivos, la grafica 9 muestra como el coeficiente de asimetría para las ventanas de la serie es positivo, conclusión que también se hizo en el análisis de la gráfica 6.

El índice de Yule Kendall es principalmente positivo lo que evidencia una distribución de probabilidad alta de los datos con respecto a la mediana.

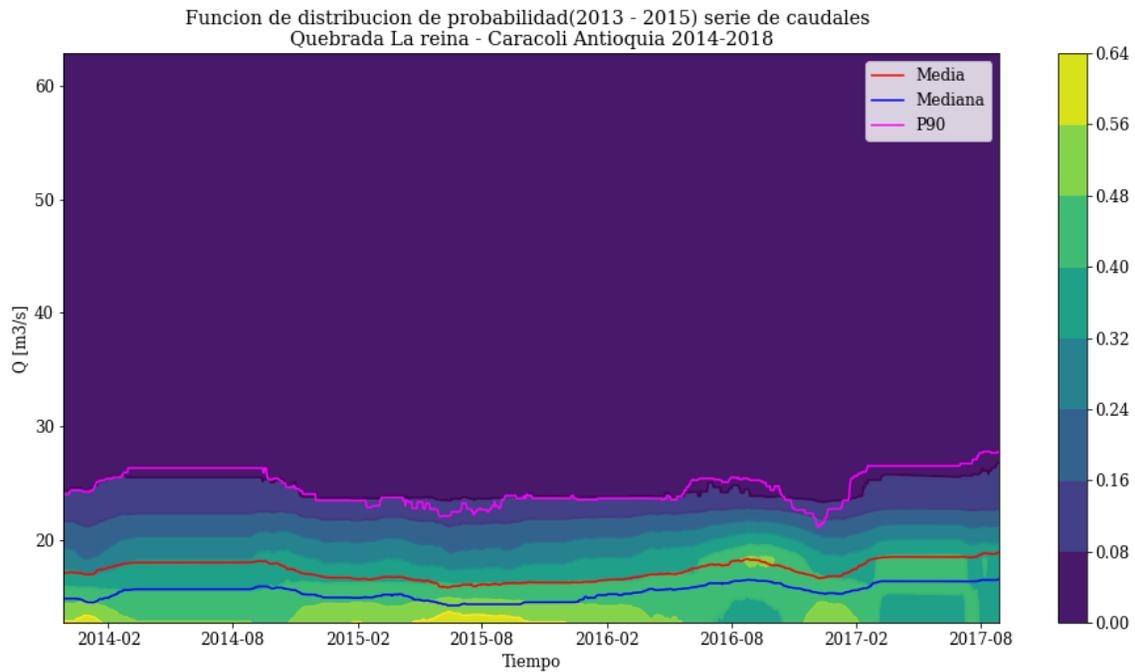
6. Evaluación de la estacionareidad de los histogramas.

La grafica de estacionareidad de los histogramas nos muestra que la función de distribución de probabilidades no cambia con el tiempo. Es decir según vemos en la gráfica 10 los datos son estacionarios, sin embargo debido a que la variable que se analiza es hidroclimatica la hipótesis puede perder validez ante las evidencias de los impactos del cambio y la variabilidad climática, entre otros sobre dinámica de los regímenes de caudales en los ríos de Colombia, a pesar de esta incertidumbre el rango de datos evaluados se comportan de manera estacionaria.



Grafica 10 Estacionareidad de la función de probabilidad de la serie de datos de caudales.

7. Evaluación de la estacionareidad de los índices.



Grafica 11 Distribución de la función Probabilidad de la serie de datos de caudales e índices..

Son o no son estacionarios? No pueden ser ambas cosas a la vez, me confunden un poco las conclusiones con las dos gráficas anteriores

Como se evidencia en el gráfico 10 el comportamiento de los percentiles la media y la mediana no es estacionario pues su representación sobre la función de distribución de probabilidad no es de valor constante, es decir en un instante de tiempo fijo la distribución de probabilidad no es la misma para todas las posiciones o ventanas. En consecuencia, parámetros tales como la media la mediana y los demás percentiles varían a lo largo del tiempo.

8. Evaluación de la tendencia de la serie y de los percentiles en la serie.

Para la evaluación de tendencia se ha utilizado el test no paramétrico de Mann-Kendall. La hipótesis nula de esta prueba es que los datos son independientes y aleatoriamente ordenados, es decir que no hay tendencia o estructura de correlación alrededor de las observaciones.

La prueba de Mann Kendall se basa en el cálculo del estadístico S obtenido mediante la ecuación²:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i)$$

$$\begin{aligned} \text{Si } (x_j - x_i) > 0 & \quad \text{sgn}(x_j - x_i) = 1 \\ \text{Si } (x_j - x_i) = 0 & \quad \text{sgn}(x_j - x_i) = 0 \\ \text{Si } (x_j - x_i) < 0 & \quad \text{sgn}(x_j - x_i) = -1 \end{aligned}$$

*Donde n es Tamaño de la muestra
Donde $x_j - x_i$ Son datos secuenciales*

Si S es positivo se infiere que la tendencia es creciente, cuando S es negativo se infiere que hay tendencia decreciente.

De acuerdo a estos indicadores se estima la varianza para el estadístico de Mann-Kendall, que considera el caso de los empates, el cual ocurre cuando $\text{sgn}(x_j - x_k) = 0$ obtenidos mediante la anterior ecuación.

Expresión matemática:

$$\text{Var}(S) = \frac{1}{18}(n(n-1)(2n+5) - \sum_{i=1}^q t_i(t_i-1)(2t_i+5))$$

Donde:

q es el número de grupos de medidas iguales

t_p es el número de datos en el grupo i-ésimo.

El cálculo del estadístico de comparación (z) se realizó mediante el software Python, cuyo código se adjunta, los valores obtenidos fueron:

² (Van Belle & Hughes, 1984, Yue et al., 2002; Kahya y Kalay, 2004)

	SERIE Q	IQR
z	5.00097	7.71
S	85970.0	4950
var	295511625	411475.83
Sum_T	21738	1110

Al evaluar el valor z obtenido para la serie Q en la ecuación obtenemos como resultado $2.866515719235352e-07$ es decir cero lo que indica que la hipótesis es nula.

La hipótesis siempre es nula, debemos rechazarla o no

Esto nos permite concluir que los datos no son aleatorios y si existe tendencia y esta tendencia es positiva.

9. Conclusiones y Discusión de los resultados.

A pesar de la ausencia de datos las representaciones graficas de los datos reflejan la distribución del régimen climático bimodal característico del país.

La mediana y la media móvil calculadas nos evidenciaron que la distribución de los datos era asimétrica positiva sin embargo la visualización grafica de la serie de datos ya nos había dado índices de este comportamiento pues la "cola" de la distribución apuntaba hacia la derecha, esto muestra la relevancia del estudio de la representación gráfica de los datos.

A pesar de que la serie de datos es estacionaria el comportamiento de los percentiles evaluados en las ventanas de tiempo no es estacionario.

En los resultados del análisis de datos deducimos que estos eran estacionarios, sin embargo debido a que la variable que se analiza es hidroclimatica la hipótesis puede perder validez ante las evidencias de los impactos del cambio y la variabilidad climática, entre otros sobre dinámica de los regímenes de caudales en los ríos de Colombia.