

Universidad Nacional de Colombia-Sede Medellín

Facultad de Minas-Departamento de Geociencias y Medio Ambiente

Análisis de Datos Ambientales- Primer Taller

Manuel Alejandro Salazar Gómez C.C. 1037658852

Juan Simón Rodas C.C. 1152455346

Serie de datos

La Serie de datos a analizar corresponde a los caudales medios diarios históricos del Río Apartadó, comprendidos entre el mes de enero de 1984 y diciembre del año 2012. Los datos están en un formato inicial .CSV (Comma Separated Values) y fueron obtenidos de la página web del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). Esta base de datos contiene la fecha de medición en formato YYYY-MM-DD HH:MM, seguida por una coma y seguida por la magnitud del caudal medido.

Lectura

Para la lectura se usó la librería Panda y la función `read_CSV`.

Gráfica de la Serie de Datos

En la Figura 1 se muestra el comportamiento de los pares de datos durante el tiempo de mediciones. En esta grafica se nota un pico significativo en la segunda mitad de la década de los años noventa; estas magnitudes de caudal podrían estar relacionadas con el fenómeno del ENSO (específicamente, La Niña) durante los periodos comprendidos entre los años 1995-1996 y mayormente entre finales del año 1998 y principios del 2001 (National Weather Service , 2018). No se ve con claridad una tendencia definida; sin embargo, en la figura se observan transiciones de caudales, en los que se presentan caudales pico seguidos inmediatamente de caudales muy bajos o casi nulos.

Estimación de P.D.F, Histograma y Percentiles

En la Figura 2 y en la Figura 3 se muestra el histograma de frecuencias y la función de distribución acumulada respectivamente. Ambas gráficas representan la cantidad de datos de caudal que hay dentro de ciertos rangos; estos rangos fueron escogidos por defecto usando la función “bins” de Phyton, por lo que todo el rango de caudales se dividió en diez sub rangos de igual longitud. El histograma de probabilidades tiene en las abscisas los rangos de caudales determinados por los “bins”, mientras que en las ordenadas se encuentra el número de datos que entran dentro de cada uno de estos rangos (Frecuencia), dividido el número total de datos de la serie (probabilidad). De la gráfica del histograma se observa una fuerte concentración de los datos en torno a los valores inferiores a los 4 m³/s. Del otro lado del espectro el número de datos es mucho

menor; sin embargo, su magnitud es del orden de 14 veces la media de la serie. Este comportamiento también se observa en la Figura 1, en la que son notables los cambios abruptos de caudal.

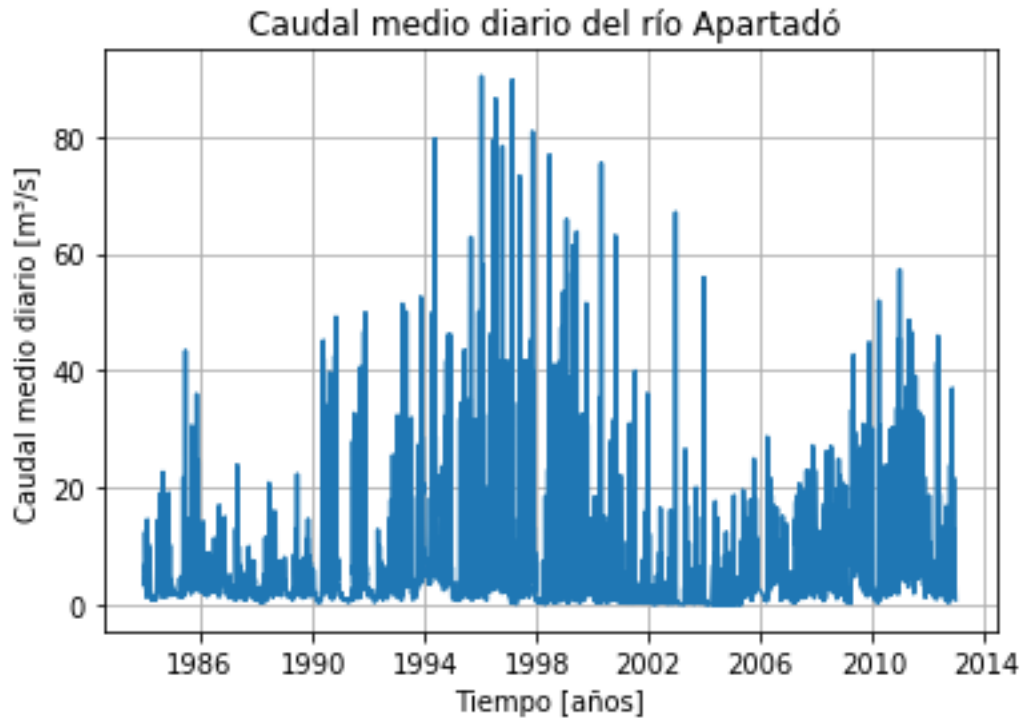


Figura 1. Serie de caudales en el periodo de medición

Por otro lado, la figura de la Función de Distribución Acumulada muestra los valores de caudal de la serie en las abscisas y en las ordenadas el porcentaje del total de los datos que representa los datos menores a su correspondiente en las abscisas; es decir, para un caudal determinado, un punto en la gráfica muestra el porcentaje de datos de la serie menores a este caudal. En esta figura también se pueden observar los percentiles, que son representados por los porcentajes de las ordenadas; cada lectura en las abscisas de la intersección del porcentaje con la gráfica, es el caudal que representa dicho percentil en la serie. Como era de esperarse, la función de distribución acumulada crece de manera acelerada, con el 50% de los datos por debajo de 2.8 m³/s y el 75% por debajo de 6 m³/s.

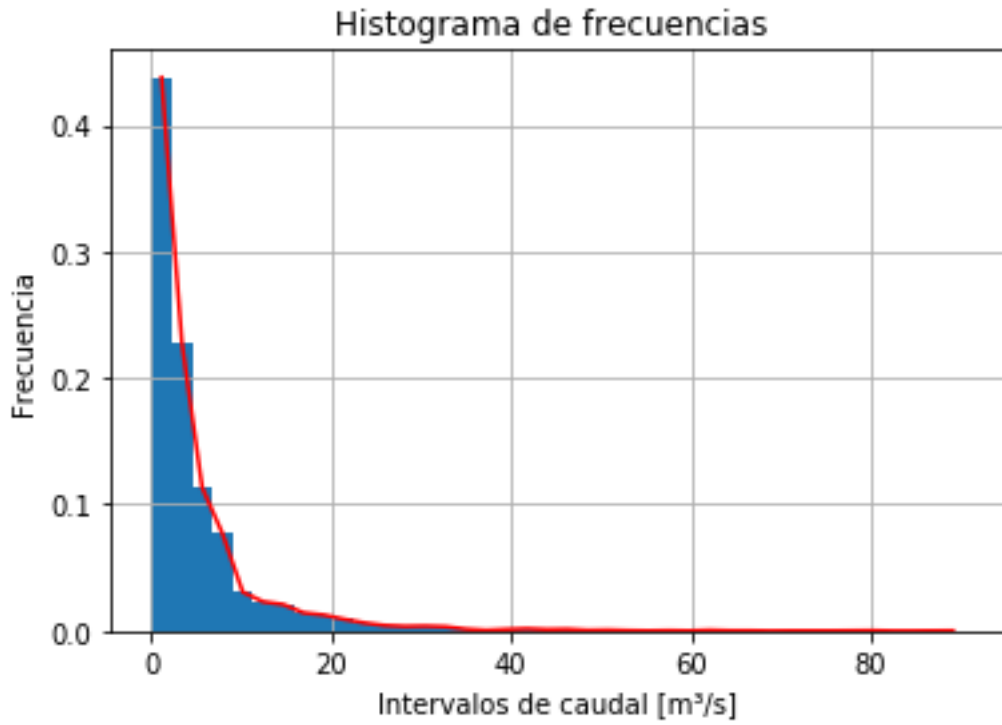


Figura 2. Histograma de probabilidades

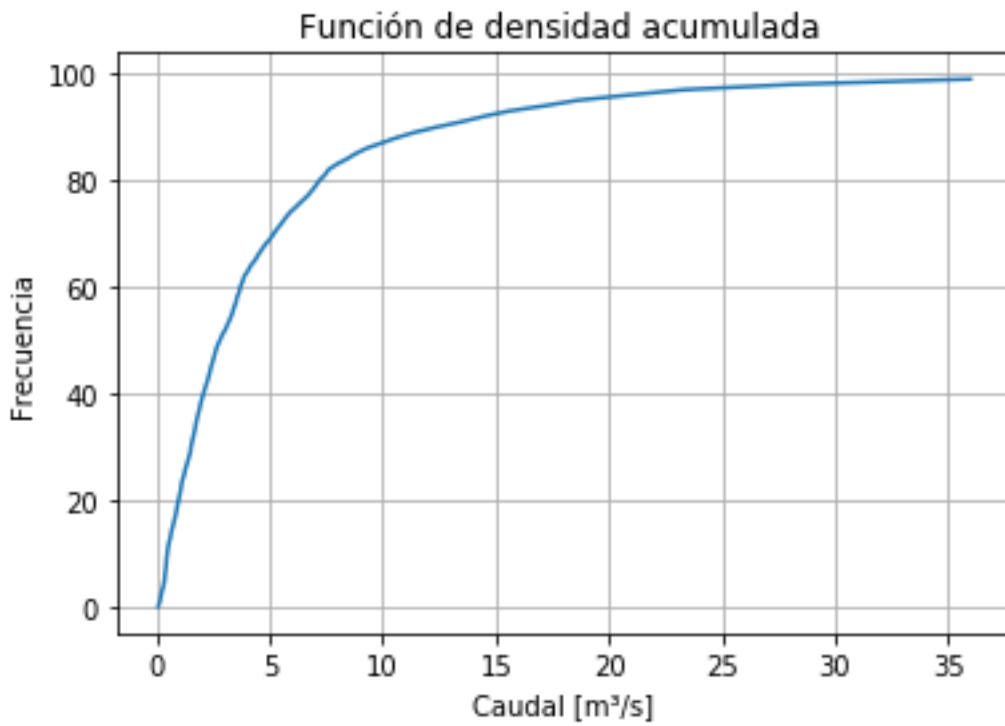


Figura 3. Función de Distribución Acumulada (Percentiles)

Estimación de índices

Los índices de localización, dispersión y simetría fueron calculados teniendo en cuenta su variación intertemporal, mediante la construcción de unas ventanas con tamaño fijo de un año, en las que se evaluó cada uno de estos índices. El tamaño de la ventana se considera representativo de los fenómenos globales en el comportamiento de un río y además se facilita el cálculo de los diferentes índices y matrices. Las ventanas son construidas empezando en cada uno de los datos de la serie y por ventana se incluyen 365 datos a partir de este. Todos estos índices presentan variaciones temporales entre las ventanas determinadas; por esta razón se podría determinar que son no estacionarias.

- **Localización**

- Media: la media para cada ventana se encuentra sumando todos los números dentro de la ventana y dividiendo dicha sumatoria por el número total de datos.
- Mediana: es el valor que se encuentra localizado en el medio de la serie de datos dentro de la ventana ordenados de menor a mayor.

A continuación se adjunta la Figura 4, en la cual se muestran las dos medidas de localización anteriormente descritas y los percentiles 10 y 90. En la gráfica se puede apreciar cierta coherencia entre la mediana y la media, especialmente en los primeros y los últimos años de la serie. En la década de los noventa estas tienen una diferencia mayor en las magnitudes, pero conservan de manera significativa su relación de forma. De esta manera, se puede ver que la media muestra poca resistencia en esta serie de datos, por lo que su mediana ayuda más a la descripción de esta serie.

Por qué?

Por otro lado, el percentil 90 presenta fuertes fluctuaciones y caudales significativamente mayores, aunque no superan los 30 m³/s, lo cual genera dudas con respecto a los outliers presentes en la serie. A pesar de esto, también parece mantener la tendencia en la forma de las otras tres medidas. En las fechas del 2000 al 2004, muestra una aproximación a las otras medidas de localización, alcanzando valores de caudal cercanos a los 2,5 m³/s; esto quiere decir que el 90% de los valores en este momento son menores a 2,5 m³/s, lo cual se podría mostrar un período de estiaje (esta hipótesis la asevera el hecho de que esta tendencia se presenta por varios años, al menos cuatro)

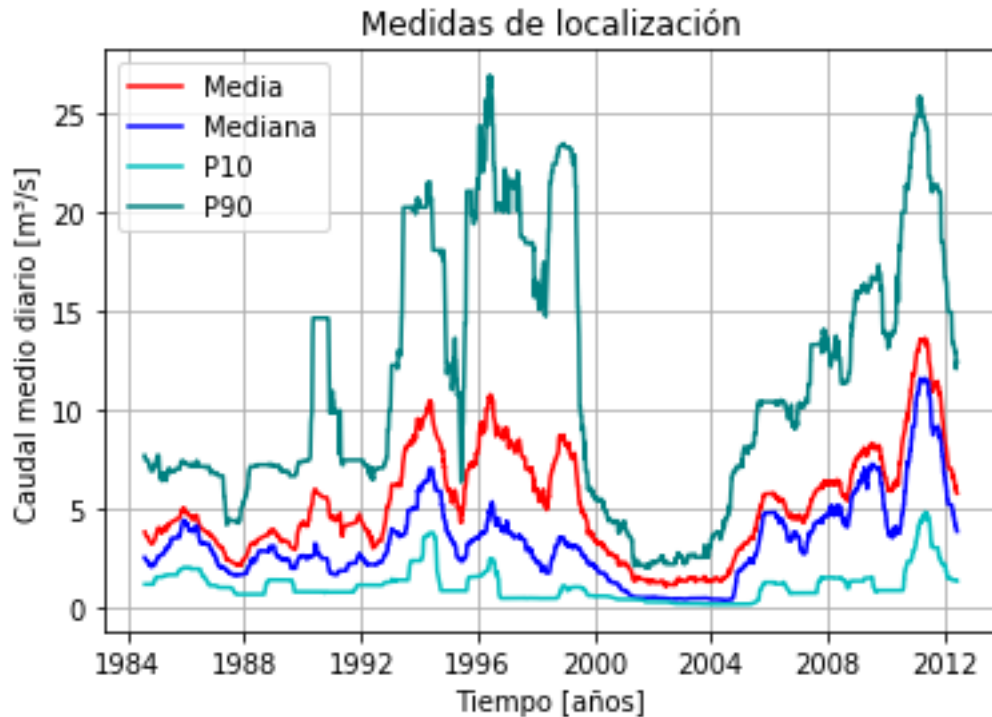


Figura 4. Índices de localización

- **Dispersión**

- Desviación estándar: este estadístico mide cuánto difiere en promedio cada dato del promedio general de los datos (media) y se calcula como la raíz cuadrada de la varianza.
- Rango intercuartil(IQR): Consiste en la diferencia entre el tercer y el primer cuartil ($Q_{75}-Q_{25}$); su utilidad está en que elimina valores extremadamente alejados y su resultado puede ser comparado con la mediana al ser medidas basadas en la posición.

A continuación se adjunta la Figura 5, en la cual se muestran las dos medidas de dispersión. En esta se puede observar que las dispersiones de los datos de la ventana móvil son más relevantes en los periodos donde se presentaron los caudales de mayor magnitud. Es importante señalar que en el año 95 y desde finales de los 90 hasta principios del 2005 la desviación estándar muestra vulnerabilidad a los datos extremos.

Ojo con la redacción

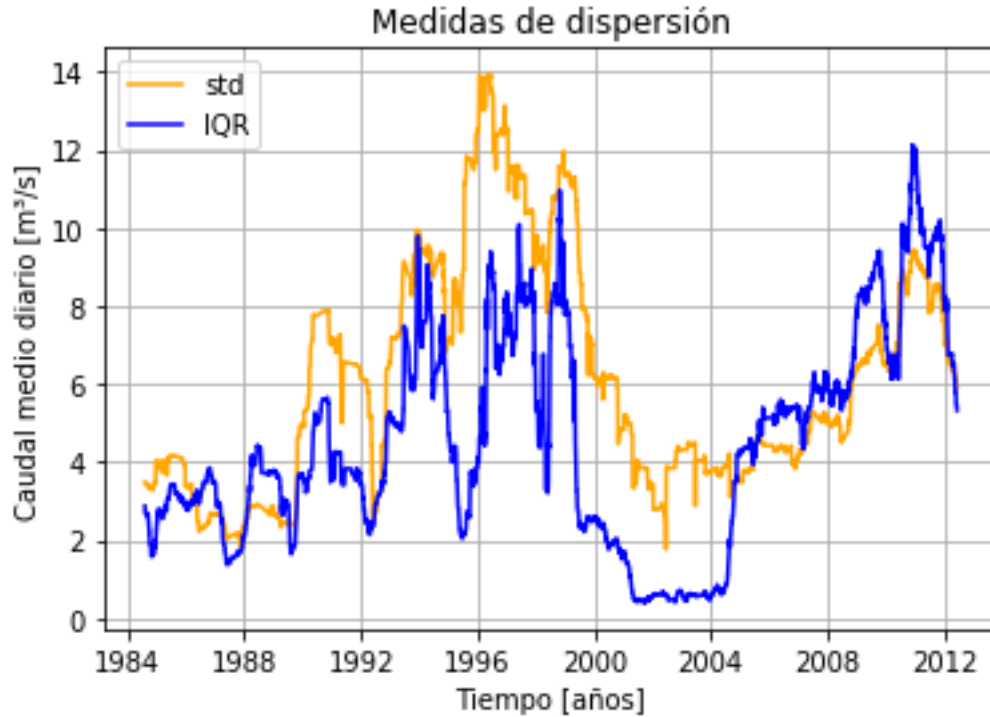


Figura 5. Medidas de dispersión

- **Simetría**

- Simetría: este índice se calcula utilizando la siguiente ecuación

$$\gamma = \frac{\frac{1}{n-1} * \sum_{i=1}^n (X_i - \bar{X})^3}{S^3}$$

- Yule-Kendall: este índice se calcula utilizando la siguiente ecuación

$$\gamma_y = \frac{(q_{75} - q_{50}) - (q_{50} - q_{25})}{IQR(\text{Rango Intercuartil})}$$

A continuación se adjunta la Figura 6, en la cual se muestran las dos medidas de simetría. En esta se puede observar que el índice de simetría se ve afectado por los valores extremos (especialmente en esta serie que tiene magnitudes diametralmente diferentes), lo cual era de esperarse por tener parámetros elevados a la tercera potencia.

Por otro lado, casi la totalidad de los índices se encuentran por encima de la línea de 0, de tal manera que, al igual que en la serie completa de caudales, las series de periodo de 1 año (Figura 8) muestran inclinación hacia la izquierda. Esta información refuerza la hipótesis de que la serie es estacionaria.

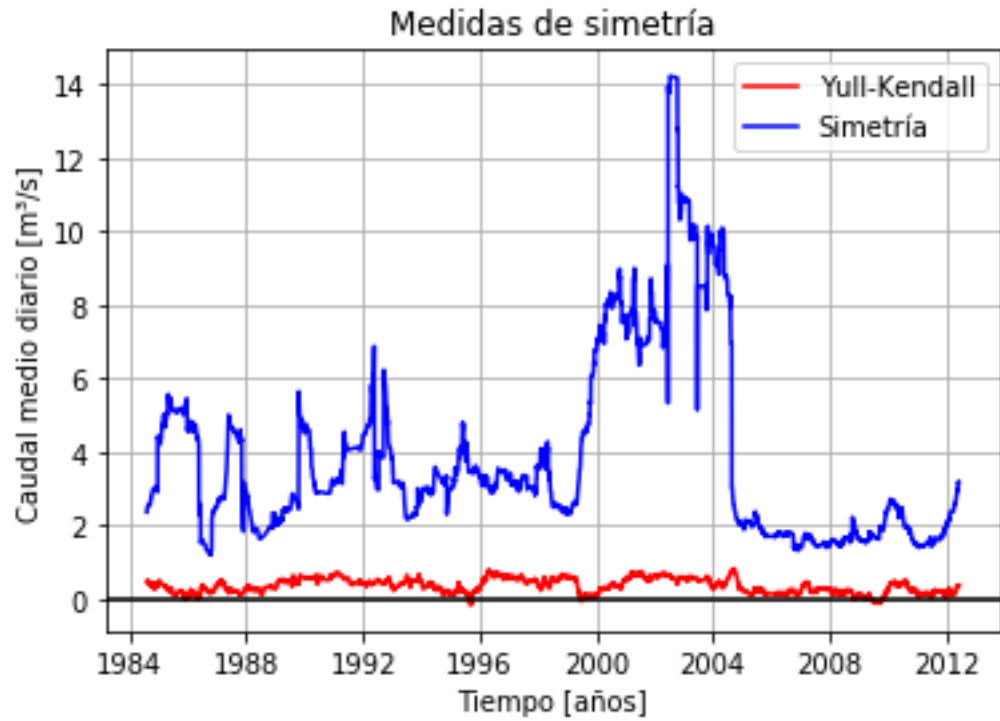


Figura 6. Medidas de simetría

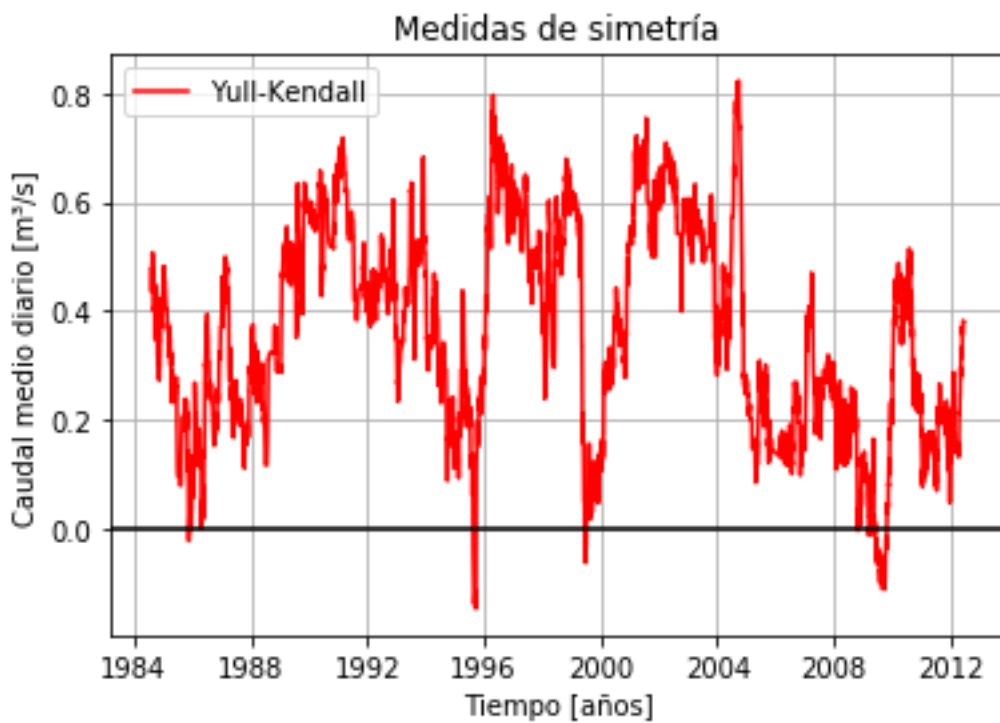


Figura 7. Yule-Kendall para la serie de datos

Estacionariedad Ventanas móviles es diferente a periodos de tiempo

La estacionariedad en una serie de datos temporales se da cuando la distribución de probabilidad es igual o similar para sus diferentes periodos de tiempo. En la Figura 8 se observa en el sentido de las ordenadas, una matriz conformada por la distribución de probabilidad de cada una de las ventanas estudiadas, siendo el color amarillo donde se dan las mayores probabilidades correspondientes a los caudales del margen izquierdo.

Ciertamente, el comportamiento de los picos de los histogramas de las ventanas móviles muestra pocos desvíos de los valores inferiores. Únicamente en periodos de anomalías relevantes estos pierden esta tendencia. Aunque su estacionariedad no es definitiva, sí existe una fuerte inclinación, en periodos de tiempos considerables, a presentar una mayor probabilidad en los primeros rangos.

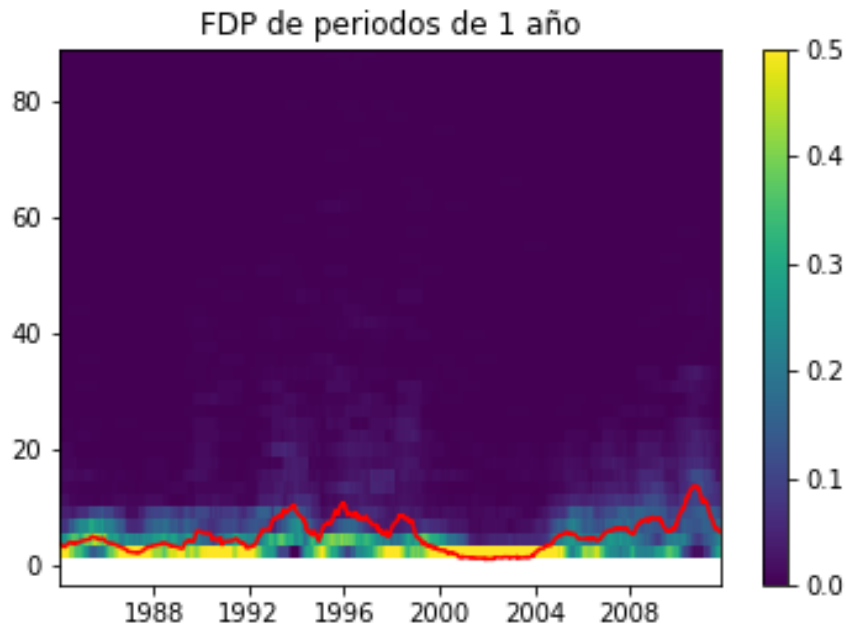


Figura 8. FDP de períodos de un año

Tendencia

En el test de Mann - Kendall calculamos el estadístico que se distribuye de manera normal para poder rechazar o no la hipótesis nula, si este está lo suficientemente apartado del 0 (con un nivel de confianza de $1 - \alpha = 0.95$), se puede rechazar o no que la serie es aleatoria (hipótesis nula), y si el nivel de confianza es mayor ($1 - \alpha/2$) se puede aceptar que la serie presenta una tendencia creciente cuando el estadístico es positivo y una tendencia decreciente cuando el estadístico es negativo (Arun Mondal, 2012)

Los valores de los estadísticos de Mann - Kendall para las series de caudal, P10, P90 y media móvil son respetivamente:

Tabla 1. Estadísticos Mann- Kendall

| | |
|---------|--------|
| z_Q | 2.75 |
| z_P10 | -20.05 |
| z_P90 | 27.89 |
| z_media | 25.53 |

En una distribución normal, el 95% ($\alpha = 0.05$) del área bajo la curva se encuentra en un valor del estadístico $Z = 1.65$. Todos los valores de los estadísticos del Mann - Kendall son mayores en valor absoluto al Z anterior, por lo que se acepta que la serie de caudales muestra una tendencia.

Para poder aceptar con suficiente confianza si la tendencia de estas series es creciente o decreciente, se compara con un $Z = 1.95$, que corresponde a un nivel de confianza de 97.5% ($\alpha/2$), como los valores absolutos de todos los estadísticos calculados anteriormente son mayores a este Z , se acepta que la serie tiene una tendencia creciente para los valores positivos del estadístico (Serie de caudales, percentil 90 y media móvil) y tendencia decreciente para los negativos (Percentil 10)

Consideraciones finales y conclusiones

A partir de los análisis hechos en este informe se puede considerar que:

- Debido a la existencia de datos **significativamente diferentes** de caudal, la mediana sería el mejor índice estadístico para definir un valor medio de la serie (además, las medidas que tienen en cuenta la posición y no el valor de los datos, son más resistentes)
- Estas **diferencias abruptas** de datos consecutivos se presentan en repetidas ocasiones a lo largo de la serie, por lo que sería poco probable que se debiera a errores en las mediciones y se podría considerar una característica inherente a las dinámicas del río.
- El río Apartadó presenta una variabilidad importante de caudales en periodos cortos de tiempo, por lo que con posteriores estudios de otras variables como la precipitación, el tamaño y la forma de la cuenca, se podría llegar a hacer conclusiones acerca del funcionamiento hidrológico del río.

- Por la distribución que se muestra en la Figura 8 y por la información obtenida en la función de distribución acumulada de la Figura 3, se puede identificar que el caudal base de este río controla el hidrograma y que este es menor a $5\text{m}^3/\text{s}$; esto, ya que este rango del caudal se mantiene cuasi-constante por periodos prolongados de la serie, lo cual no sería posible si el caudal de la escorrentía manejara el hidrograma.

Bibliografía

- Arun Mondal, A. M. (2012). *Rainfall trend analysis by Mann- Kendall test: a case study of North-Eastern part of Cuttack district, Orissa*. Retrieved from Researchgate:
https://www.researchgate.net/profile/Anirban_Mukhopadhyay6/publication/268438767_Rainfall_trend_analysis_by_Mann-Kendall_test_a_case_study_of_North-Eastern_part_of_Cuttack_district_Orissa/links/54e44fd50cf2b2314f600105/Rainfall-trend-analysis-by-Mann-Kend
- National Weather Service . (2018). *Climate Prediction Center*. Retrieved from National Oceanic and Atmospheric Administration:
http://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php