

+1 por usar Latex, sin embargo cuidado con el manejo de las figuras y la ubicación de las mismas.. Pensé que habías olvidado hacer muchas cosas porque estaban al final final. Si bien el desarrollo metodológico del test de tendencias parece estar bien. Yo dudo un poco de tus resultado... el z es muy alto teniendo en cuenta la serie, como dice Carlos seamos escépticos y revisemos bien. Buen Trabajo en General

4.0

* Sobre 4.5

Tarea 1 - Análisis de Datos Ambientales

Mónica Andrea Bonilla Rodríguez

*Estudiante Maestría en Ingeniería - Recursos Hidráulicos
Universidad Nacional de Colombia - Sede Medellín*

Resumen

Se presenta el análisis de la serie de caudales medios diarios de la estación MASCOTA LA del IDEAM en un periodo de años desde 1980-01-01 hasta 2015-07-21, mediante las medidas estadísticas robustas y resistentes de localización, dispersión y asimetría. Evaluando además si son estacionarias y tienen alguna tendencia. Para el caso de la variable de caudales medios diarios se encontró que la distribución de probabilidades permite identificar valores extremos de baja probabilidad, y entre aquellos de mayor probabilidad de ocurrencia se identifica variabilidad del ciclo anual y además se identifican variabilidades interanuales que pueden obedecer a procesos físicos hidroclimáticos que se relacionan con la respuesta hidrológica de caudales en Colombia como la oscilación del ENSO, por tanto se descarta la estacionariedad de las medidas estadísticas y se identifica tendencia en los datos con **0% de incertidumbre**, descartando su aleatoriedad.

Recuerda que esto no es posible

1. Serie de Datos

Para el análisis de los datos es recomendable utilizar una serie temporal con mínimo 1000 datos, por lo cual se analizó la serie de caudales medios diarios observados de una de las estaciones del IDEAM en unidades m^3/s , la cual tenía datos diarios en un intervalo de años desde 1980-01-01 hasta 2015-07-21, es decir 12986 datos correspondientes a los casi 36 años de observaciones. A continuación se presenta la información de la estación:

Departamento: Antioquia

Municipio: Yalí

Nombre: MASCOTA LA

Nombre: San Bartolomé

Código: 23107030

Clase: Hidrológica

Categoría: Limnigráfica

Ubicación: (6.6396389°,-74.8754444°)

Para mayor entendimiento de la variabilidad y el comportamiento de los datos observados se grafica la serie temporal en la figura 1.

A partir de la serie de caudales observados puede identificarse que la mayoría de los datos alcanzan valores inferiores a $40m^3/s$, con un comportamiento aparentemente variable interanual, sin embargo es algo que debe evaluarse a lo largo del presente trabajo, además se identificó que el valor máximo de la serie alcanzó un caudal medio diario de $104,1m^3/s$ aunque en la gráfica se observara como un evento de una sola vez.

Caudales Medios Diarios Río San Bartolomé

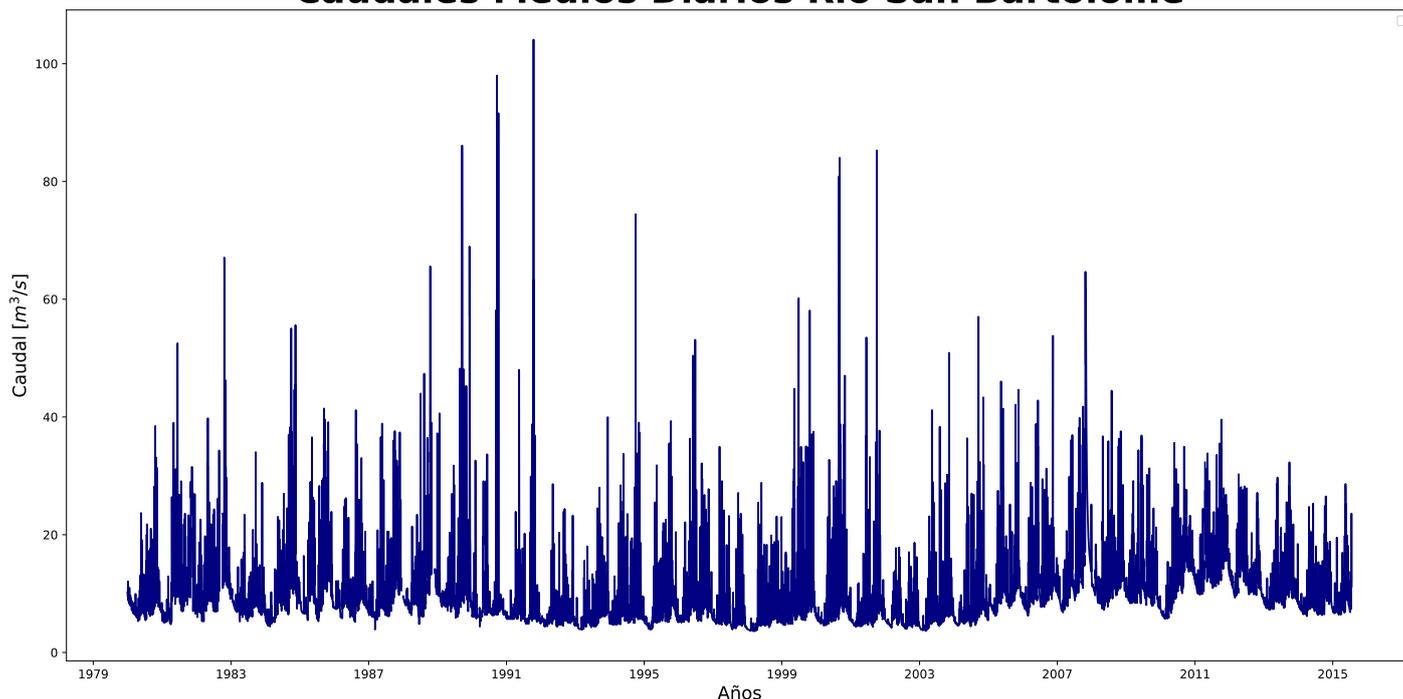


Figura 1: Caudales medios diarios del río San Bartolomé

2. Distribución de Probabilidad

Para empezar a analizar el comportamiento de los datos se obtuvo su histograma de frecuencias con 25 intervalos de clase elegidos arbitrariamente como una buena representación de los datos y posteriormente la función de distribución de probabilidades, que puede observarse en la figura 2, donde puede corroborarse que los datos con mayor probabilidad de ocurrencia son inferiores a $40m^3/s$ y valores de caudal medio diario por encima de este son de mínima probabilidad.

¿Es estacionaria la función de distribución de probabilidades?

Para evaluar estacionariedad se definió una longitud de ventana arbitraria, que para el caso fue de 4 años teniendo en cuenta diferentes aspectos como: los caudales medios diarios tienen una variabilidad anual, cada 4 años se tiene la presencia de años bisiestos y para el caso de Colombia se tiene un fenómeno de variabilidad interanual tan importante como el ENSO que puede oscilar entre los 2 a 7 años y afecta la variabilidad del ciclo anual de los caudales [1]. Dentro de cada una de las ventanas de tiempo se obtuvo la función de distribución de probabilidades en comparación con la correspondiente a la longitud de toda la serie (Figura 2) y para este caso en particular se eligió arbitrariamente mostrar el cambio de las PDF de cada ventana cada 2 años y 10 meses, esto sólo con el fin de abarcar la longitud de toda la serie y comprender las fluctuaciones a lo largo de esta. Como puede observarse en la figura 3 los histogramas y PDF no son estacionarios, sin embargo varían en un rango de los datos muy similar, con variaciones de la probabilidad de ocurrencia de caudales mínimos.

Adicionalmente, la figura 4 en la parte inferior muestra la variación no estacionaria de los PDFs en un recorte hasta un caudal máximo de $20m^3/s$, ya que los valores extremos son de probabilidad baja casi nula por la gran cantidad de datos de la serie y es en estos rangos donde se observa la variabilidad no estacionaria

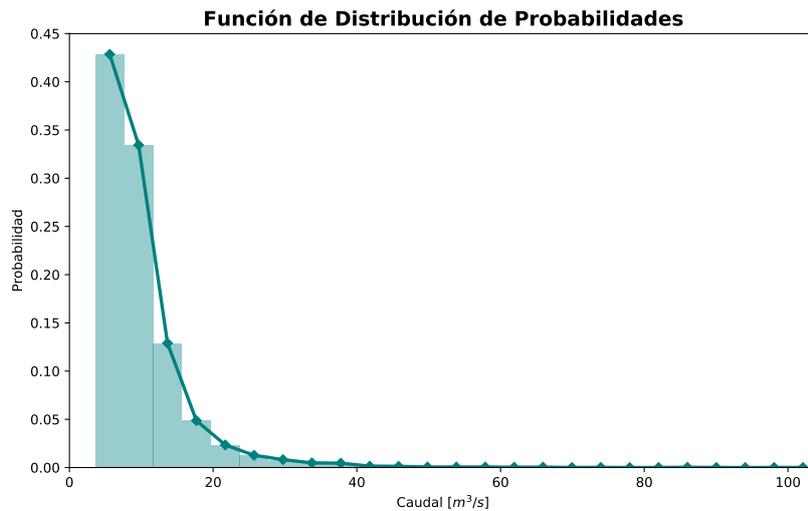


Figura 2: Función de Distribución de Probabilidades de caudales medios diarios del río San Bartolomé

de los índices estadísticos. Como puede verse en la parte superior de la figura 4, se tiene una concentración importante de los datos alrededor de la Media y en percentiles del 90 % y más.

Ambas figuras se apoyan y validan mutuamente ya que fueron calculadas con la misma ventana de tiempo, sin embargo es posible en ellas detectar la no estacionariedad gobernada por un ciclo interanual, que induce a la varaibilidad de medidas estadísticas dentro de un rango de valores.

3. Medidas de Localización, Dispersión y Asimetría

Las diferentes medidas estadísticas describen la serie con el fin de comprender su comportamiento según la localización, dispersión y la asimetría de los datos. Para empezar se identificaron las medidas de posición del conjunto de datos ordenado como los cuartiles que corresponden al 25 %, 50 % y 75 % de los datos, es decir el percentil 25, la mediana y el percentil 75 (Ver figura 5).

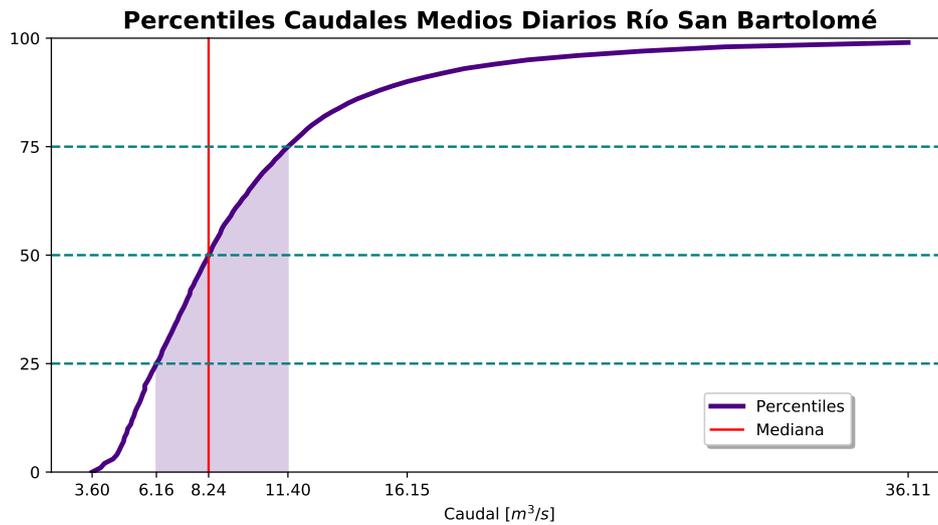


Figura 5: Cuartiles de la serie de caudales medios diarios del río San Bartolomé.

A partir de los análisis de localización realizados hasta el momento, puede detectarse que el 90% de los datos alcanzan valores de caudal medio diario inferiores a $16,15\text{m}^3/\text{s}$, por tanto puede inferirse que haya un comportamiento similar de las medidas de dispersión para toda la serie, así que se procede a analizar la estacionariedad de estas.

¿Son estacionarias las medidas de Localización, Dispersión y Asimetría?

Para estas se parte de las mismas condiciones bajo las cuales se analizaron las diferentes PDFs, teniendo una ventana de ancho igual a 4 años con la cual se calcularon cada una de las medidas. Las de localización: Media, Mediana, Cuartiles y Percentil 90 y 10 pueden observarse en la figura 6. De dispersión: Desviación Estándar y rango intercuartil en la figura 7 y de asimetría se tiene el coeficiente de asimetría Yule - Kendall en la figura 8.

De las gráficas anteriores se resalta que: Se observa que ninguno de los índices descriptivos son estacionarios, sin embargo las medidas de localización presentan variabilidad interanual con indicios de ciclos aproximadamente cada década. Hay mayor amplitud de los datos del percentil 75 al 90 que del 10 al 25. La media móvil se aproxima mucho a los valores del percentil 75, lo cual permite inferir que hay una fuerte influencia de los valores máximos a pesar de ser tan extremos y poco frecuentes, restandole resistencia a la medida de la Media. Se destaca que la desviación estándar sea mayor que el rango intercuartil en la serie de datos anteriores al 2008 y luego se empiece a invertir. Los coeficientes de asimetría Yule-Kendall presentan solamente valores positivos. Todas estas observaciones se discutirán en el análisis de resultados con mayor profundidad.

Índices de Localización - Caudales Medios Diarios

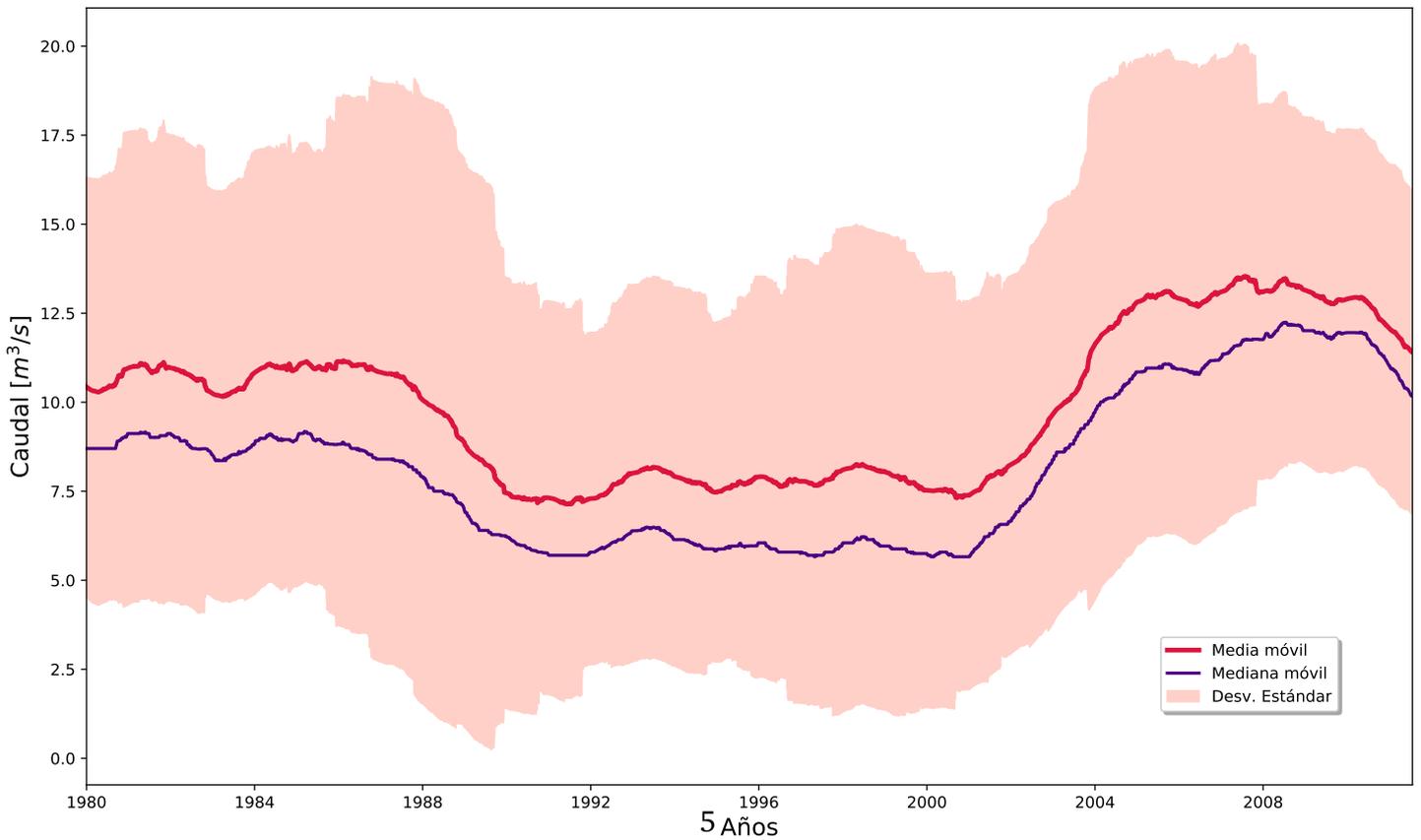
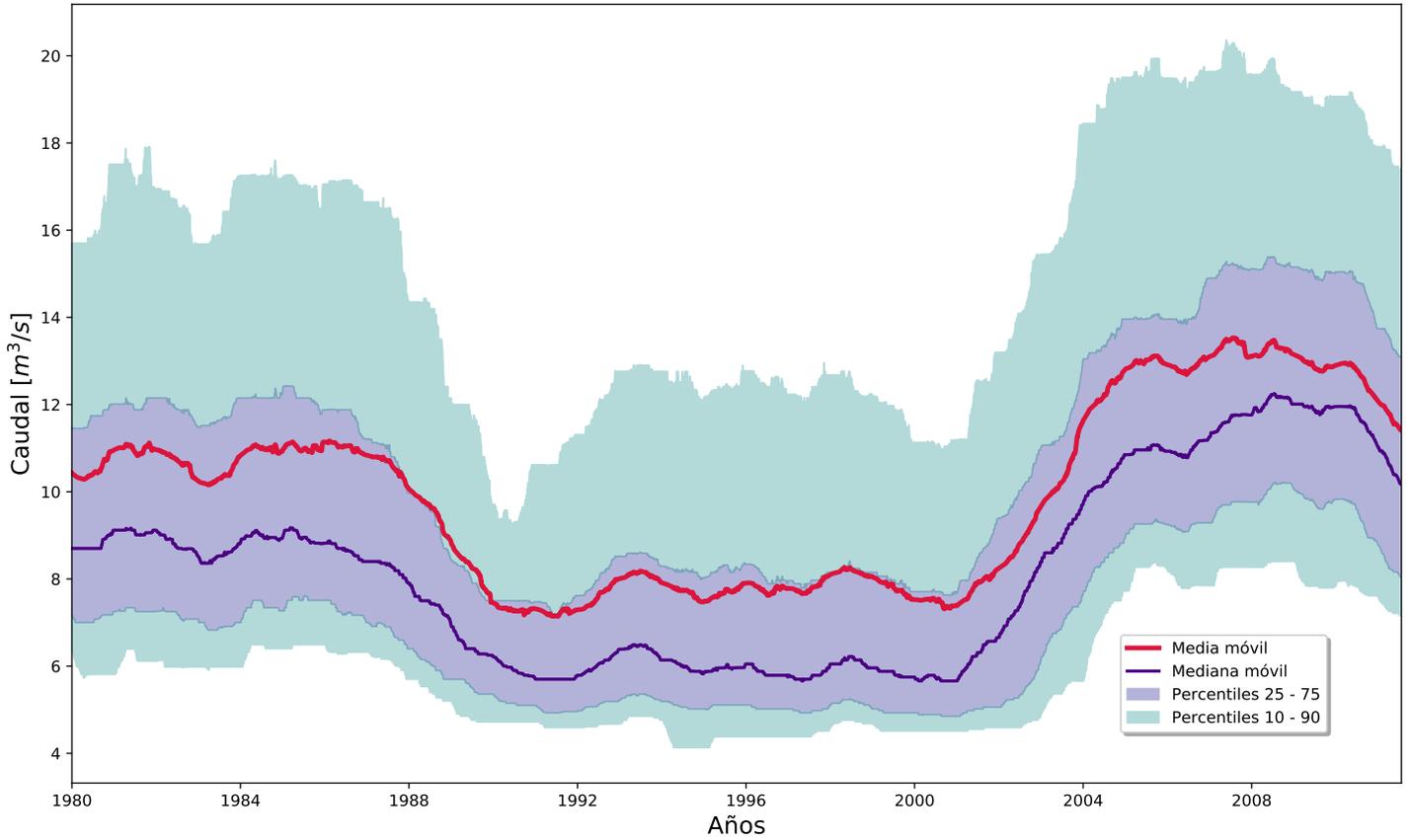


Figura 6: Medidas de localización: Media móvil, mediana móvil, cuartiles móviles, percentiles 90 y 10 móviles y desviación estándar móvil para la serie de caudales medios del río San Bartolomé.

Índices de Dispersión - Caudales Medios Diarios

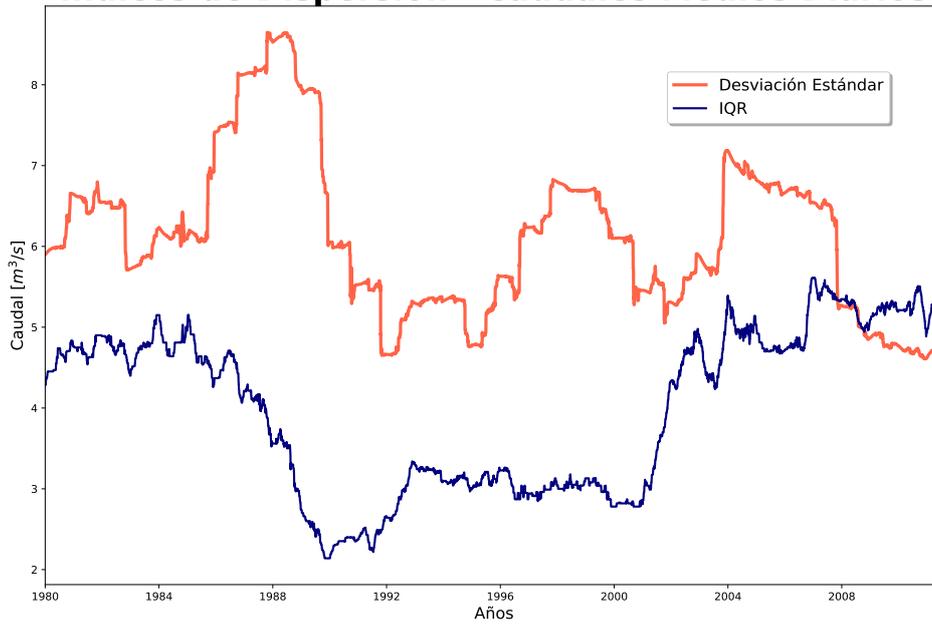


Figura 7: Medidas de dispersión: Desviación estándar móvil y rango intercuartil móvil para la serie de caudales medios del río San Bartolomé.

Índice de Asimetría - Yule Kendall



Figura 8: Medidas de asimetría: Coeficiente de asimetría Yule-Kendall móvil para la serie de caudales medios del río San Bartolomé.

4. Análisis de Tendencias

Partiendo de la variabilidad y no estacionariedad de la serie y los índices estadísticos, se aplica la prueba no paramétrica de Mann-Kendall para detectar tendencias, esta permite evaluar la significancia de tendencias en series hidroclimáticas, de calidad de agua, caudal, temperatura y precipitación a partir de la hipótesis nula de que los datos son independientes y aleatoriamente ordenados, es decir que no hay tendencia o estructura de correlación alrededor de las observaciones [2]. Para ello partimos del hecho de ser un test no paramétrico por tanto se asume que los datos tienen distribución libre y vamos a comparar esta con una distribución normal, que además son datos independientes y se plantea la hipótesis nula de aleatoriedad de los datos, es decir la inexistencia de tendencia.

Para aplicar el test, primero se analizó la variabilidad de la serie de caudales medios alrededor de la media móvil (Ver figura 9) con el fin de entender si la variabilidad era negativa o positiva, en este caso por presentarse ambas se realiza la prueba de dos colas, es decir con $\alpha/2$. Para este test tenemos un estadístico de prueba z , del cual tenemos la hipótesis alternativa de que hay tendencia cuando z es lo suficientemente alta (Tanto negativa como positiva) que su probabilidad de ocurrencia sea baja (menor al 5%, nivel de significancia $\alpha < 0,05$), es decir, nuestra hipótesis alternativa implica que $1 - \alpha > 0,95$ pero por ser de dos colas debe tenerse en cuenta que se rechaza la hipótesis para los valores críticos de $z > 1,95996398$ y $z < -1,95996398$ [3] (Ver figura 10).

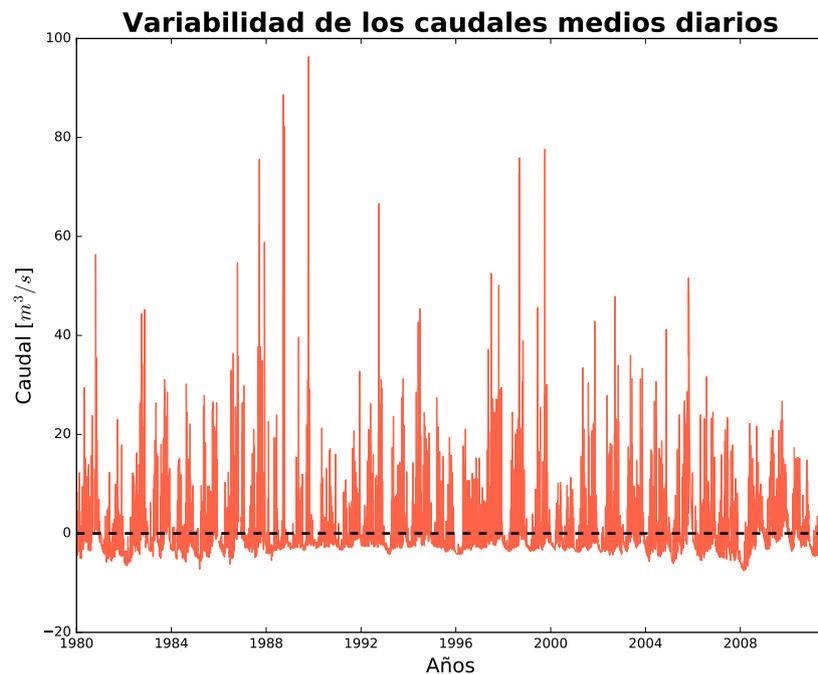


Figura 9: Variabilidad de la serie de caudales.

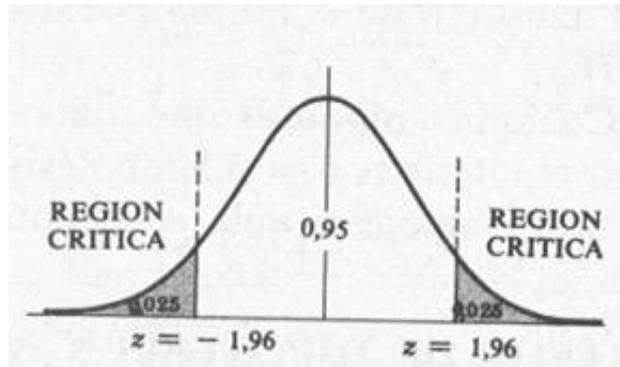


Figura 10: Distribución normal con valores críticos para el 95 % de confianza. [4]

A continuación se presentan los valores z obtenidos para el análisis de tendencia de la serie y los índices:

Serie	z
Caudales medios diarios	14.27319716
Media móvil	27.34065172
Mediana móvil	21.02758703
Percentil 75	27.49692047
Percentil 25	13.71738758

Dado que todos los valores de z para las diferentes series analizadas son mucho mayores que el valor crítico $z > 1,95996398$ podemos rechazar la hipótesis nula, además verificando la probabilidad $1 - \alpha = 0,0000$ por tanto se puede tener certeza de la tendencia de los datos.

5. Análisis de Resultados

La serie de caudales medios diarios contiene unos pocos valores máximos extremos que inducen a revisar eventos importantes de crecidas de caudal en esas fechas, ya que además son de probabilidad de ocurrencia muy baja. Puede considerarse en un próximo análisis filtrar los datos de estos extremos de baja probabilidad, ya que como se observó por ejemplo en la figura 4 y 5, la variabilidad de la serie cobraba importancia en los datos inferiores al percentil 90.

Las figuras 3 y 4 permiten identificar que entre 1990 y 2003 aproximadamente se observa un comportamiento interesante en la serie de caudales, ya que se reduce la media considerablemente, casi a la mitad y además la función de distribución de probabilidades muestra probabilidades mucho más altas de ocurrencia. Para entender este comportamiento que refleja variabilidad interanual de la serie se revisaron los índices del ENSO de la NOAA, encontrando que en este periodo se presentaron 3 periodos cálidos El Niño con apenas un corto y leve periodo La Niña de Julio de 1995 a Abril de 1996, lo cual se puede conjeturar que tiene que ver con esta variabilidad, sin embargo requiere un análisis más profundo que además permita explicar el aumento significativo de la media de los caudales diarios para los años posteriores al 2004.

De la figura 6 llama la atención que para el periodo entre 1990 y 2003 que se mencionó antes, la media se acerca al percentil 75 y en el resto del dominio de los datos se aleja por debajo de este, lo cual da indicios que

para estos periodos de tiempo los eventos máximos fueron representativos a pesar de su poca probabilidad de ocurrencia. Cabe anotar que en todo el dominio de la serie, la media móvil siempre superó el percentil 50 de los datos.

Las medidas de dispersión de los datos indican gran dispersión, arrojando valores de desviación estándar superiores al rango intercuartil, esto debido a los datos máximos tan extremos que se observaron y que además evidencian coeficientes de asimetría Yule-Kendall solamente positivos. También es importante considerar que los valores del rango intercuartil son menores que la desviación estándar por la gran amplitud de los datos que se ubican por encima del percentil 75, si observamos la diferencia de la amplitud entre el percentil 25 y el percentil 10 es muy baja a comparación de la diferencia de amplitud entre el percentil 90 y el 75.

Los análisis permitieron verificar que los índices estadísticos no son estacionarios, sin embargo las medidas de localización como los cuartiles (percentil 25, 50 y 75) mostraban una variabilidad muy similar que parecía inducir en la estacionariedad del rango intercuartil, sin embargo este a su vez tampoco resultó estacionario.

La variabilidad detectada en la serie y en los índices estadísticos permitía conjeturar sobre la tendencia de los datos, sin embargo se analizó la variabilidad de la derivada de la serie (9) alrededor de la media móvil y se encontró variabilidad negativa aunque en menor escala que la positiva, por lo cual se pudo aplicar el test Mann-Kendall de doble cola para así descartar con certeza la aleatoriedad de la serie.

El análisis realizado en este trabajo abre cuestionamientos por resolver, tales como cuál es la variabilidad interanual que puede describir la serie de caudales y los índices estadísticos, analizar a profundidad la oscilación del ENSO, la convergencia de la ZCIT y qué fenómenos físicos hidroclimáticos la afectan y si es posible detectar una variabilidad cíclica o analizar las anomalías de la serie.

6. Conclusiones

Puede evidenciarse que las medidas de localización son robustas y resistentes ya que no dependen de la distribución a la que se ajusten los datos y además no se vieron afectadas por los valores extremos que en la serie fueron de alta dispersión.

Las series de datos de caudales están altamente gobernadas por la variabilidad del ciclo anual, además de variabilidad interanual que puede ser ocasionada por otros procesos físicos globales.

No hay estacionariedad pero si hay tendencias en la serie de caudales medios diarios.

Este análisis de los datos con estadística descriptiva abre la puerta a análisis más profundos y con nuevas herramientas que permitan entender el comportamiento de los datos y obtener información basados en hipótesis.

Referencias

- [1] Chunzai Wang and Paul C. Fiedler. ENSO variability and the eastern tropical Pacific: A review. *Progress in Oceanography*, 69(2-4):239–266, 5 2006.
- [2] Diana Cristina Cantor Gómez and Andrés Ochoa Jaramillo. *EVALUACIÓN Y ANÁLISIS ESPACIOTEMPORAL DE TENDENCIAS DE LARGO PLAZO EN LA HIDROCLIMATOLOGÍA COLOMBIANA*. PhD thesis, Universidad Nacional de Colombia, 2011.
- [3] Sheng Yue and Chun Yuan Wang. The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. *Water Resources Management*, 2004.
- [4] Murray R. Spiegel. *Estadística, teoría y 875 problemas resueltos*. México, 1970.

Variabilidad PDF cada 2 años y 10 meses

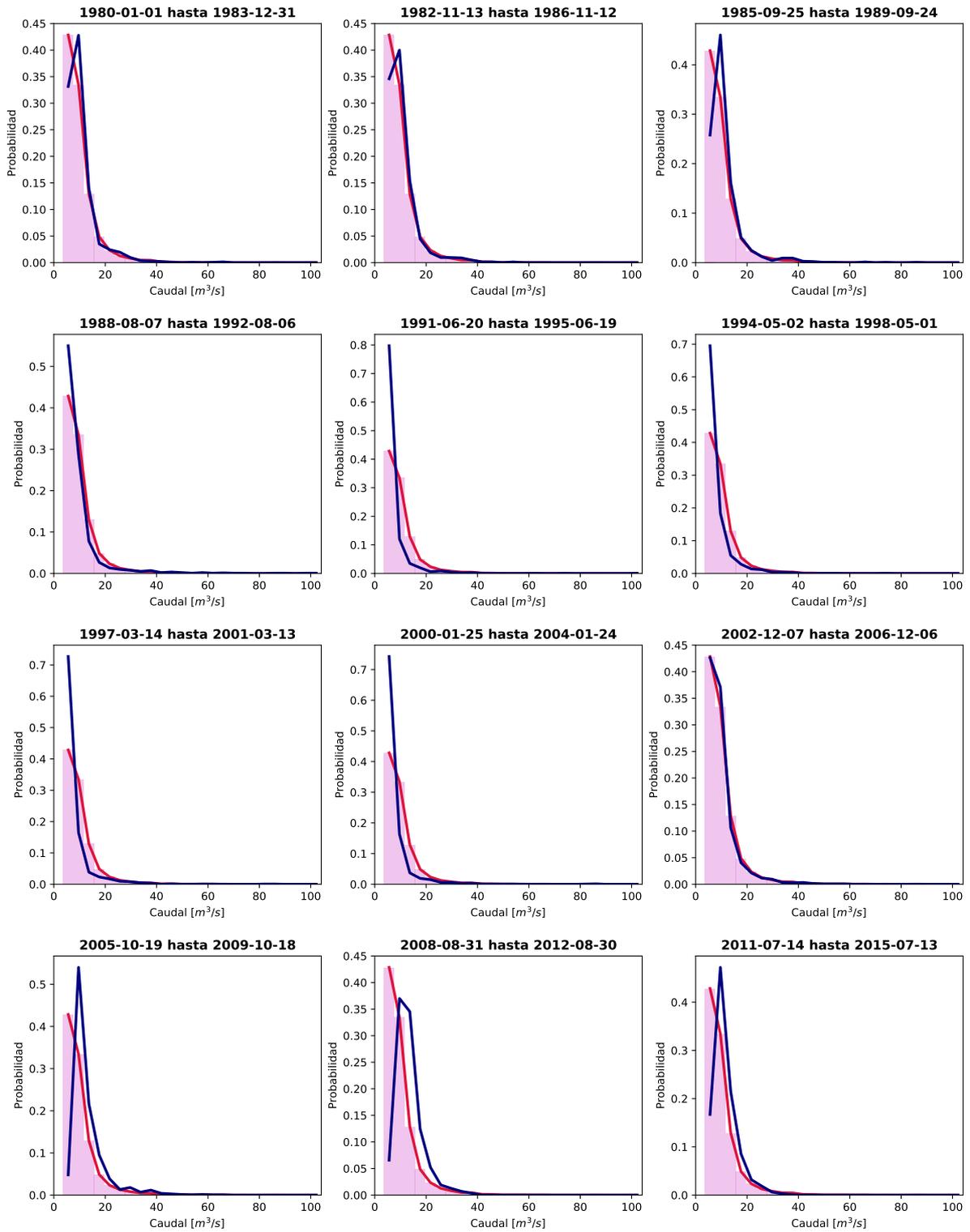


Figura 3: Variación de la PDF de ventana de tiempo a lo largo de la longitud de la serie de caudales

Por qué esto está por acá tan lejos?

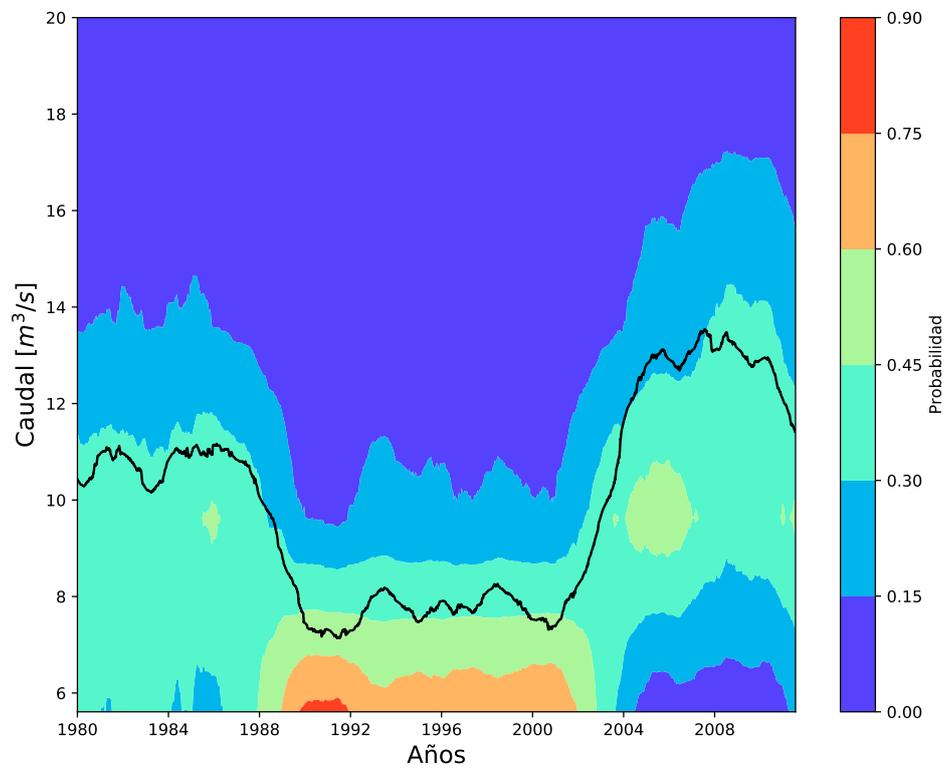
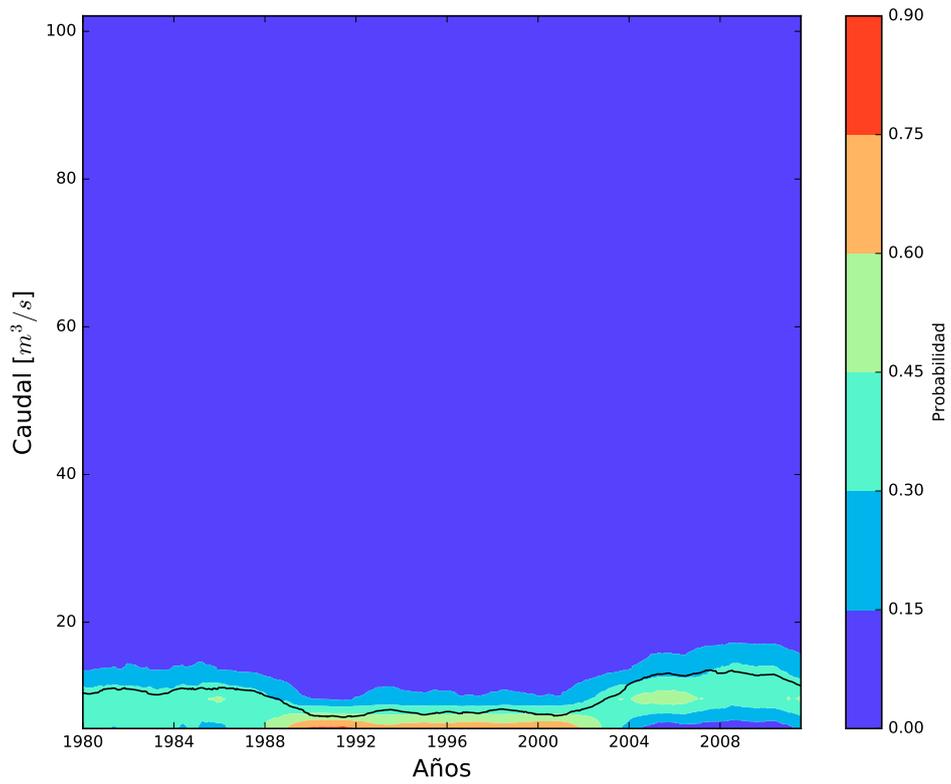


Figura 4: Estacionariedad de las funciones de distribución de probabilidad para cada ventana de tiempo y distribución de la media móvil (línea negra) a lo largo de la serie de tiempo.