

## Análisis del flujo del Río Lebrija

A partir de los datos de caudal diario del río Lebrija, se construyeron otras 3 series de tiempo. La primera serie de tiempo corresponde a caudales mensuales, para esto se tomaron grupos de 30 datos a los cuales se les calculó la media y se tomó este valor como el caudal mensual representativo. Luego, se construyó una serie anual, tomando grupos de 12 datos de la serie mensual, a partir de estos grupos, se halló la media y se tomó esta como el valor representativo de cada año. Finalmente, se construyó la serie mensual multianual, para esta el valor representativo de cada mes se obtuvo promediando los valores de la serie mensual que correspondían con el mes de interés, asumiendo que el primer dato correspondía a enero.

En primer lugar, analizaremos los estadísticos de cada una de las 4 series construidas:

Estadístico	Caudales Diarios	Caudales Mensuales	Caudales Anuales	Caudales Mensuales Multianuales
Media	93.785	93.7577	93.6276	93.9658
Desviación Estandar	72.9917	51.7733	29.4067	19.6525
Varianza	5327.7854	2680.4766	864.7569	386.221
Coefficiente de Variación	0.7783	0.5522	0.3141	0.2091
Coefficiente de Asimetría	2.0168	1.431	0.6218	0.2184
Coefficiente de Curtosis	7.653	5.5285	3.1278	2.1728
Covarianza	351.1322	224.0586	138.4721	386.221
Coefficiente de Correlación	0.1368	0.3475	0.2399	1

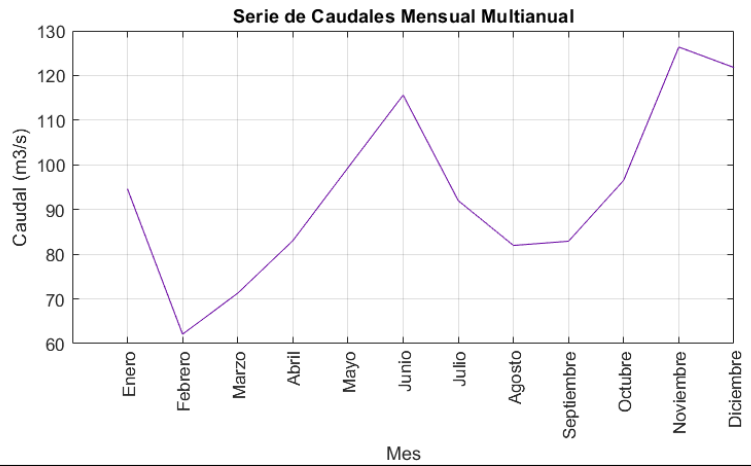
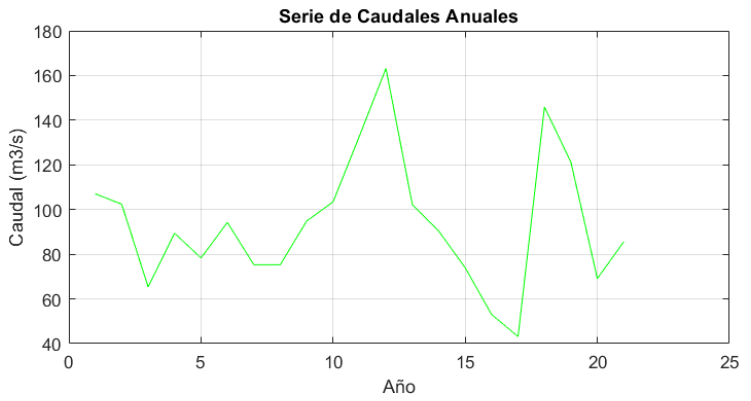
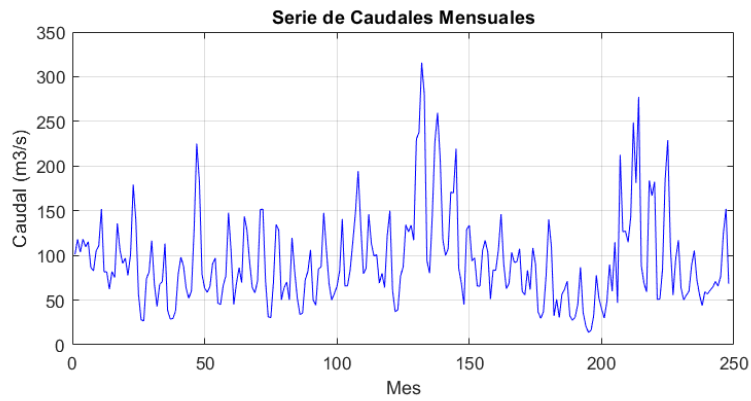
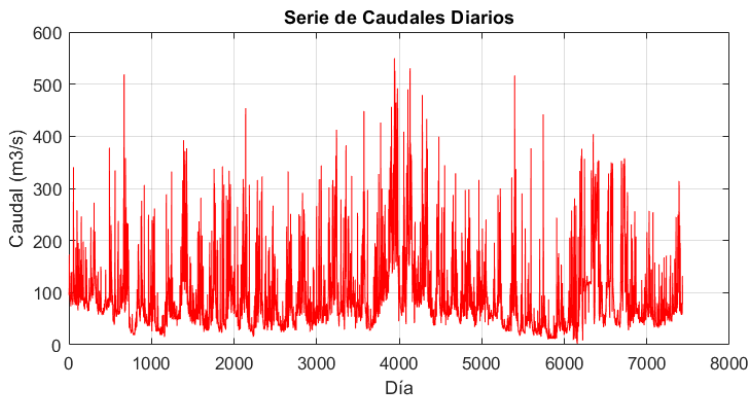
Cabe aclarar que, para la covarianza en los datos diarios, se evaluó la relación existente entre los datos del primer año y del segundo año de datos. En el caso de los caudales mensuales, se evaluaron nuevamente el primer año y el segundo. Para los caudales anuales, se compararon los primeros 10 años con los 10 siguientes. Finalmente, a los caudales mensuales multianuales, se aplicó la covarianza con los mismos datos, por lo cual este valor es igual a la varianza y el coeficiente de correlación es igual a 1.

De los estadísticos es posible destacar como la varianza va disminuyendo a medida que se va agregando la serie diaria a periodos de tiempo más largos. Esto implica que la serie es más homogénea al compararla anual o mensualmente, mientras entre día y día la variación en los valores es mayor, lo cual hace que se presenten picos grandes en la serie de tiempo diaria, los cuales son difíciles de modelar.

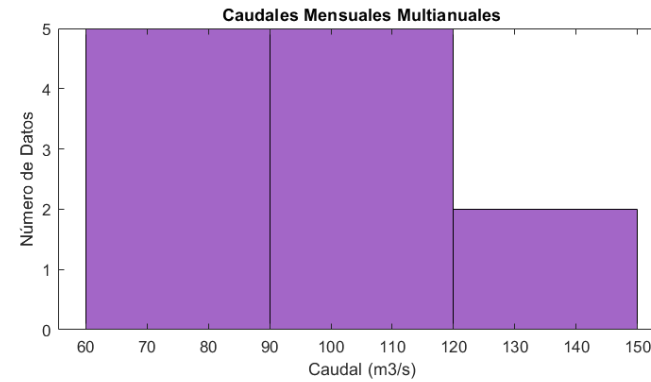
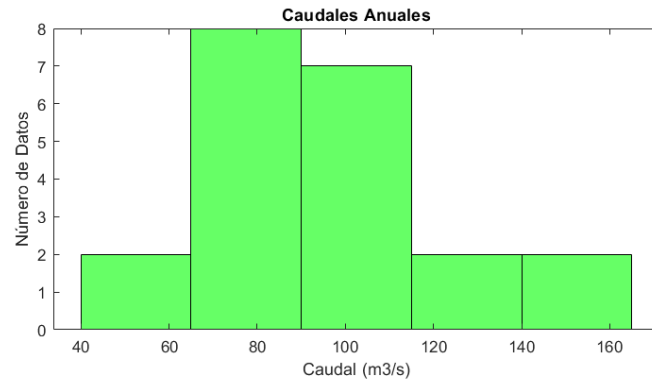
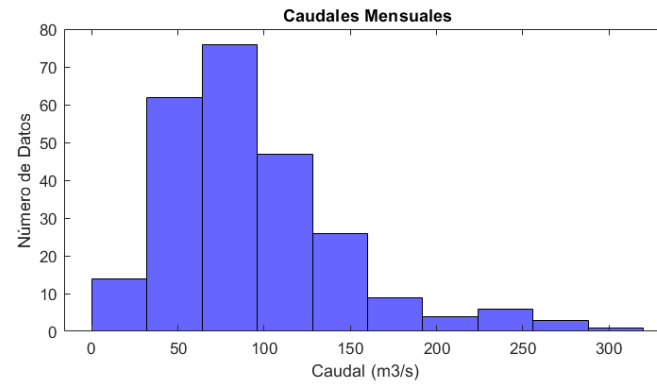
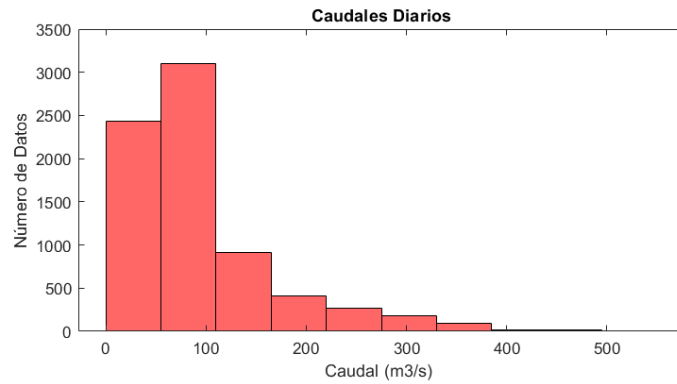
Otro aspecto destacable es ver como a medida que se agrega la serie, la asimetría va disminuyendo, aunque siempre se tiene un sesgo a la izquierda en las distribuciones. Adicionalmente, se observa una alta concentración de datos, principalmente en la serie diaria, ya que, el coeficiente de curtosis es mucho mayor a 1.

Luego se realizaron las correspondientes gráficas que caracterizan cada una de las series de tiempo:

De las series de tiempo es interesante observar el rango de valores y como este se va acortando a medida que se agrega la serie, pasando de diferencias de más de 500 m<sup>3</sup>/s entre el mayor y el menor valor diario hasta apenas un poco más de 10 entre el menor y el mayor valor anual. A pesar de esto, tener un promedio anual que duplique el promedio de otro año en un mismo río indica una gran variabilidad en su caudal.

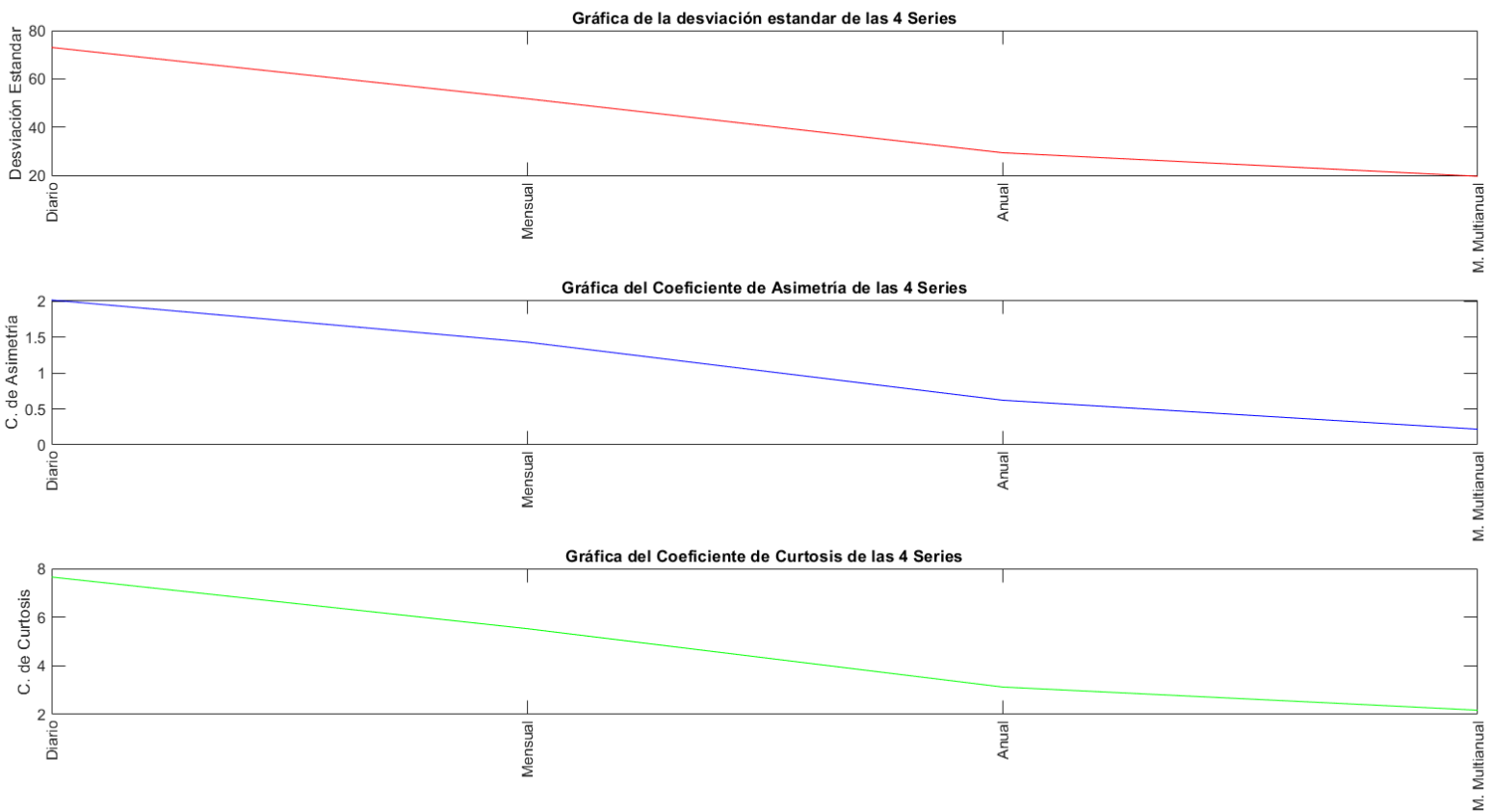


- Histograma



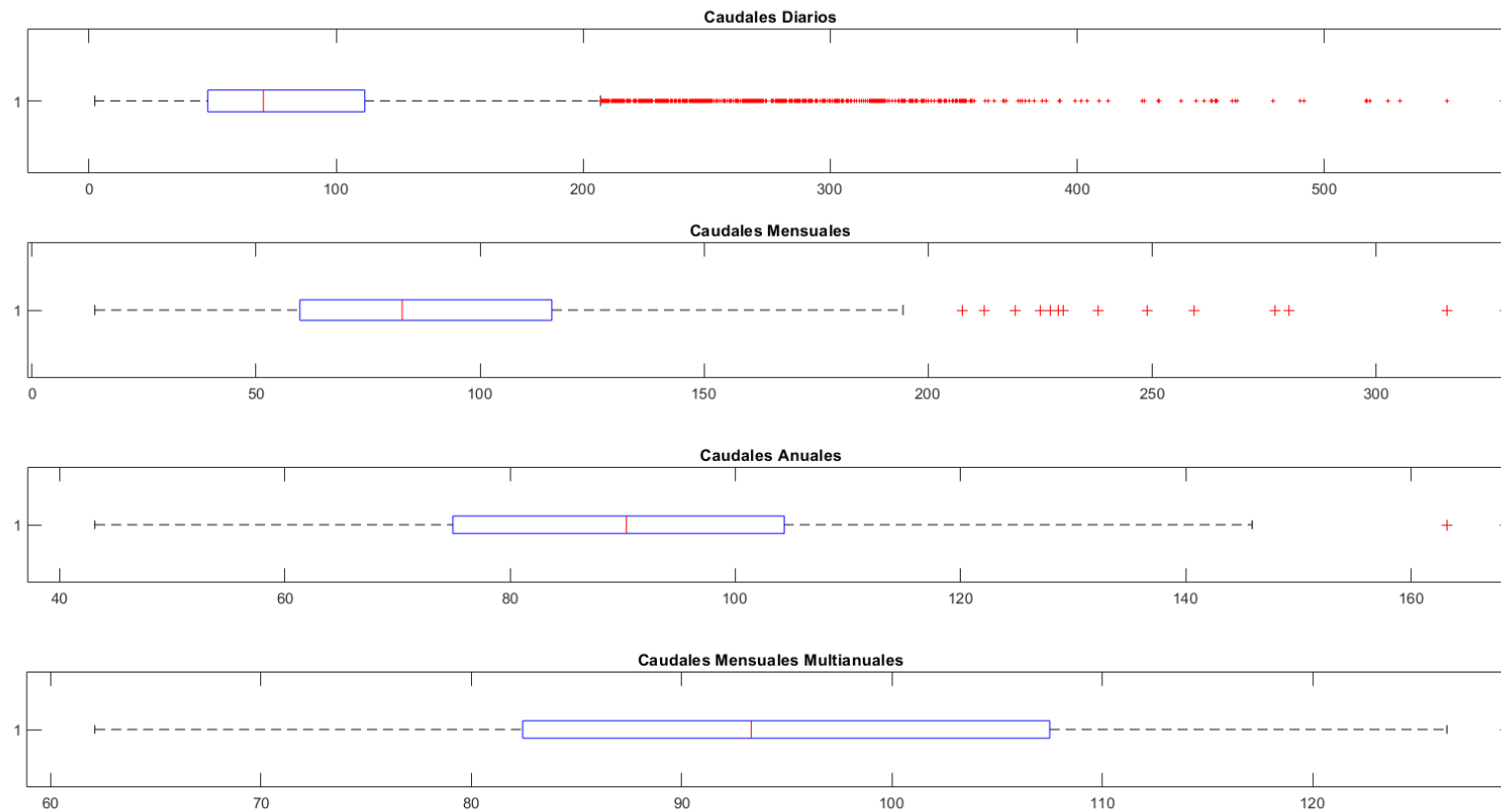
En los histogramas es posible observar el sesgo hacia la izquierda que tienen los datos y como este va disminuyendo a medida que se agrega la serie a periodos más largos, tal y como lo indica el coeficiente de asimetría. Esto se puede deber principalmente a la reducción del rango de valores, ya que, es en los caudales diarios en donde se tienen los valores más extremos.

- Comparación de los estadísticos



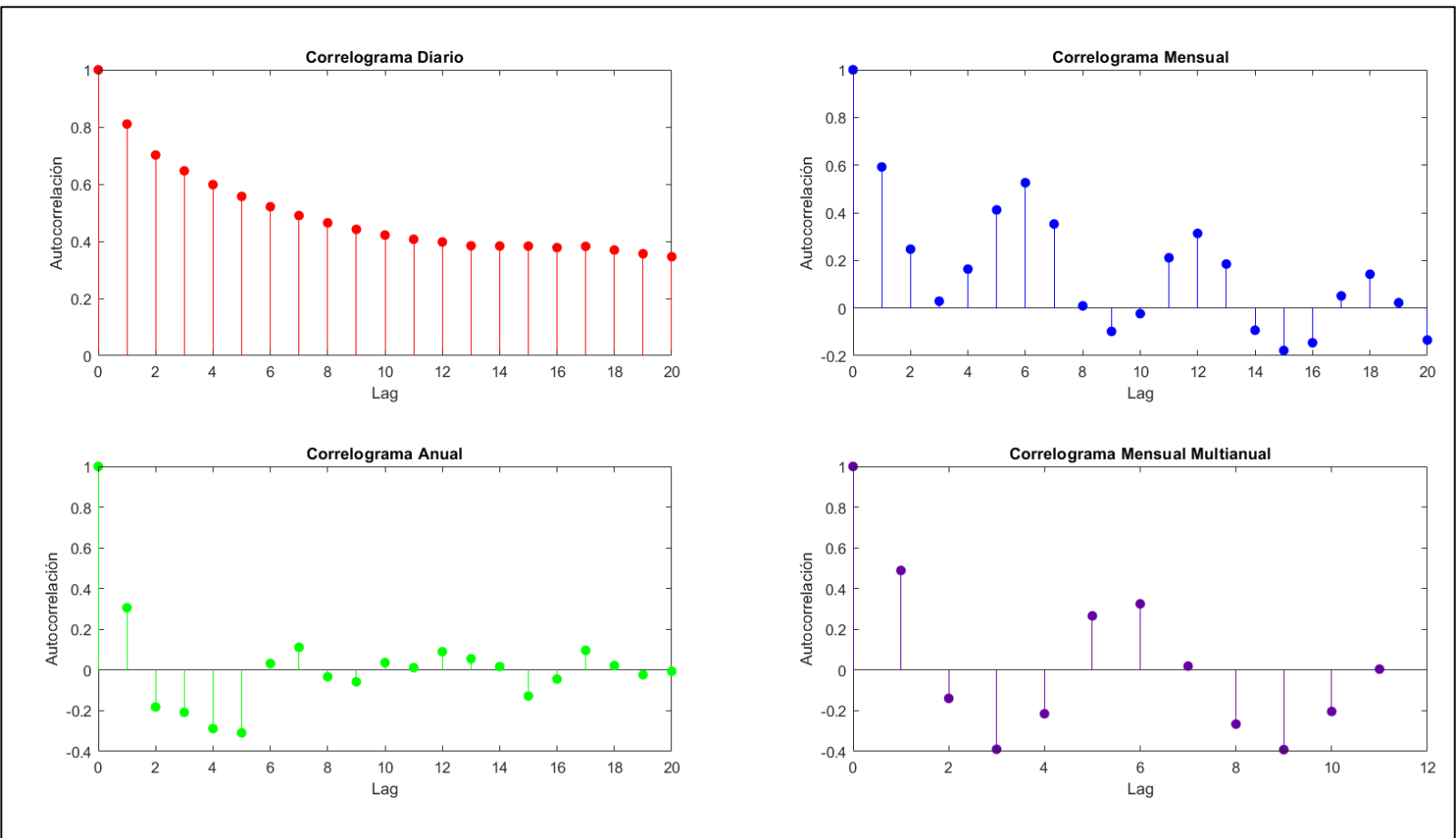
Es interesante observar como los estadísticos que indican la variabilidad y el sesgo de, los datos, van disminuyendo en la misma proporción a medida que se agrega la serie de tiempo.

- Diagrama de cajas y bigotes

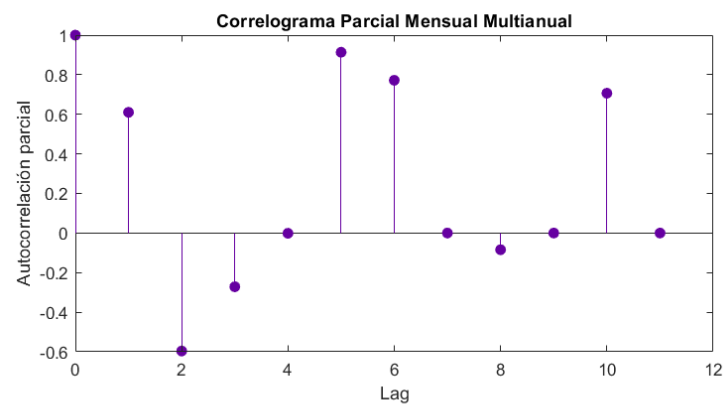
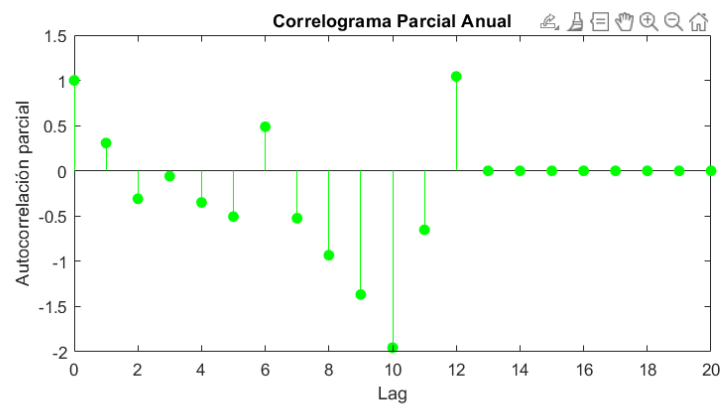
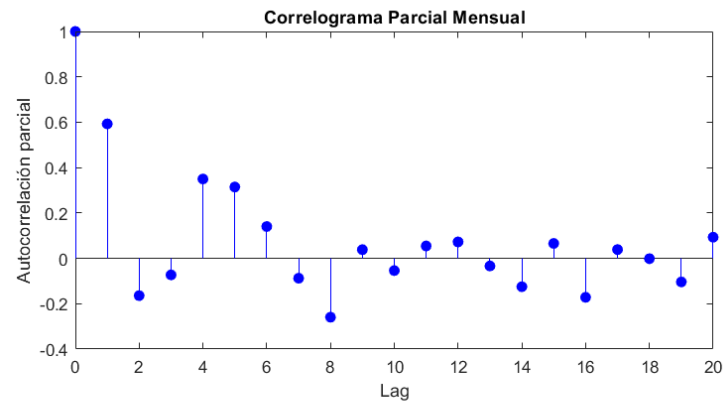
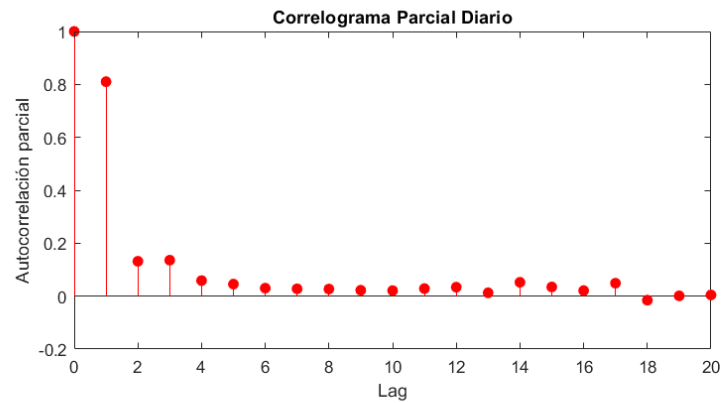


Este gráfico nuevamente muestra como en los caudales diarios se presentan muchos valores atípicos, representados por picos en la serie de tiempo, que son difíciles de modelar adecuadamente. Mientras que para las series que consideran rangos de tiempo más largo casi no se presentan valores extremos. También cabe resaltar que los valores atípicos son valores muy altos de caudal, mostrando picos hasta 6 veces mayores que la media.

- Correlogramas



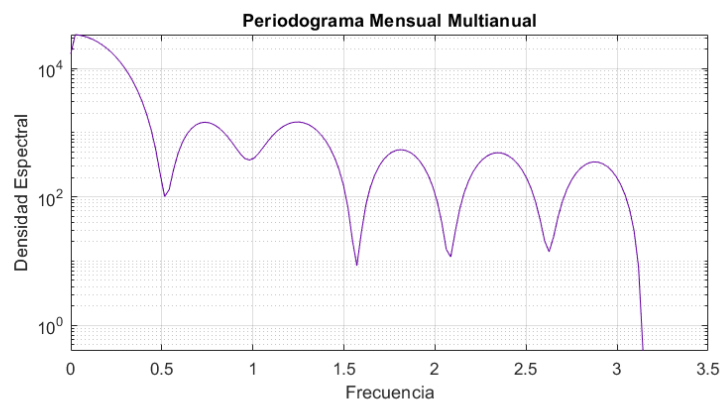
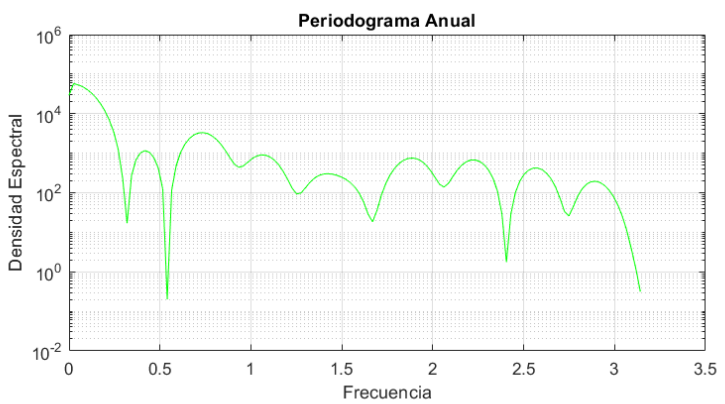
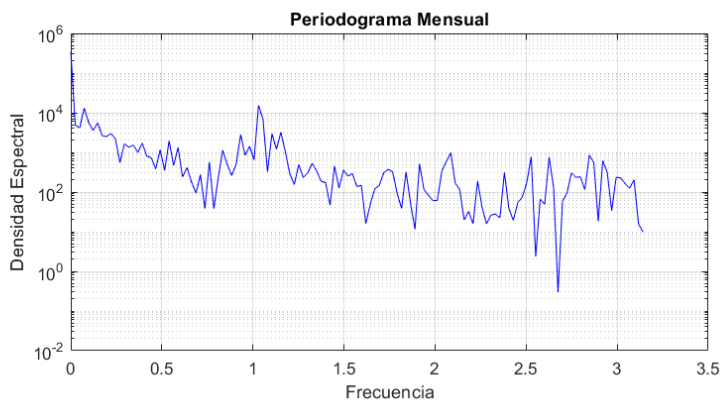
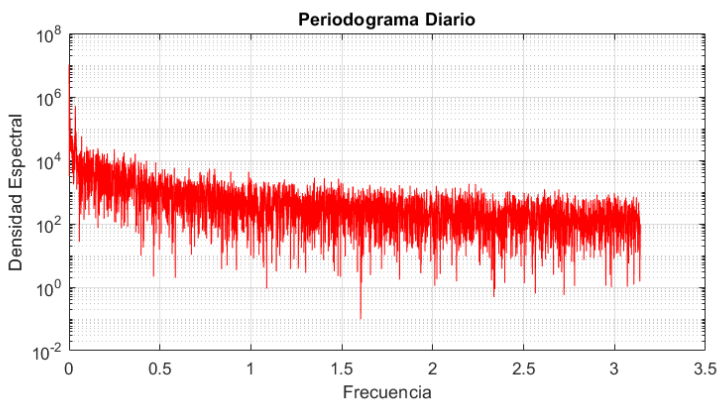
- Correlogramas parciales



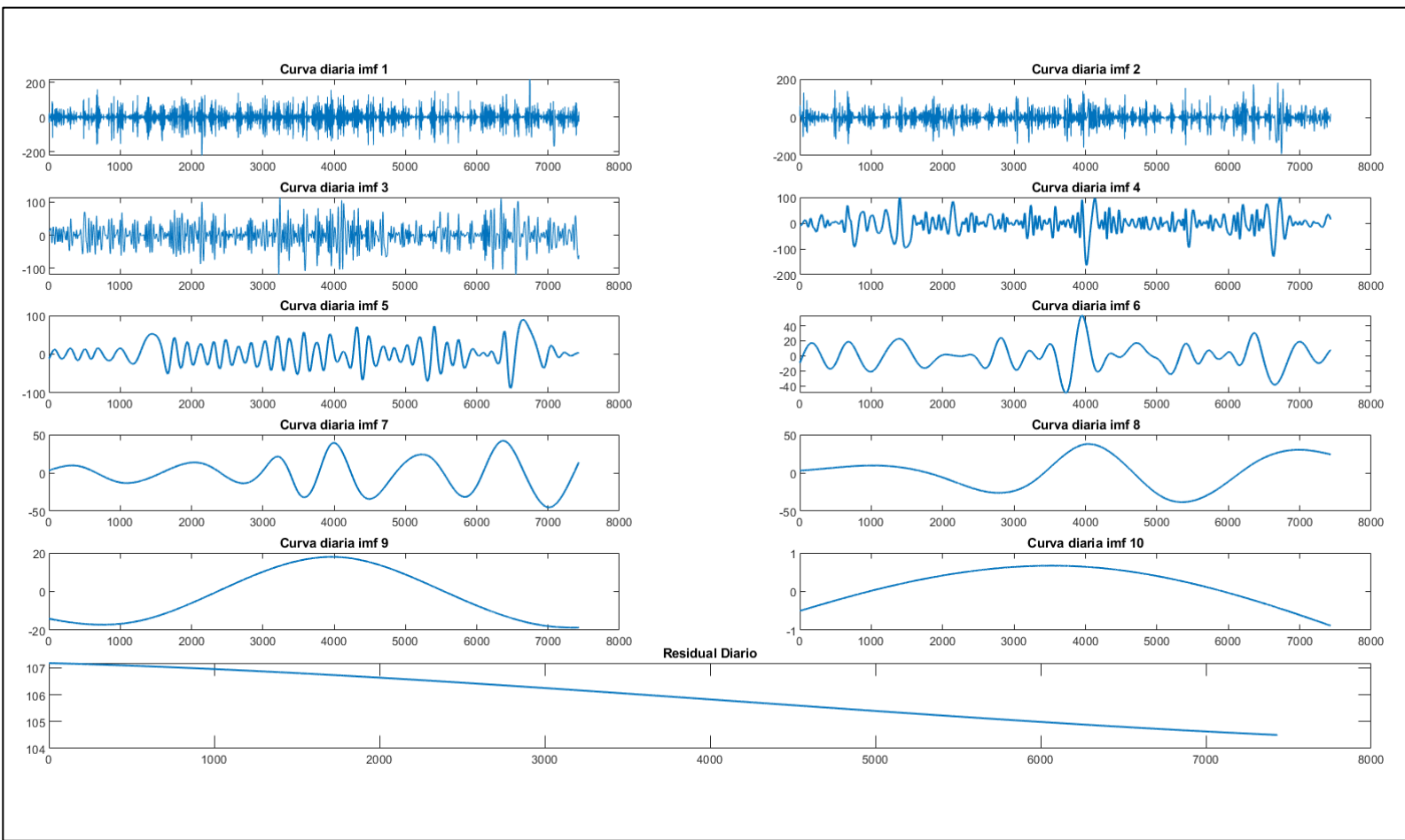
Algo muy interesante al analizar los correlogramas mensual, anual y mensual multianual es la oscilación del valor de autocorrelación, esto es un indicativo de que el caudal en el río Lebrija sigue un proceso cíclico que se repite cada cierto tiempo que en algún momento los valores posteriores dependan en mayor medida de sus predecesores.

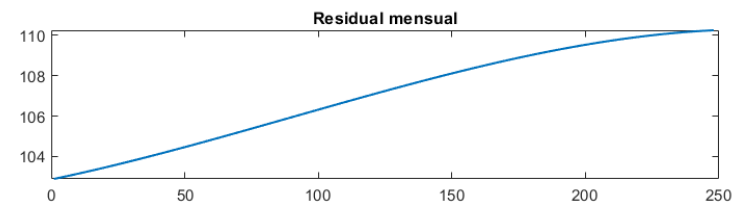
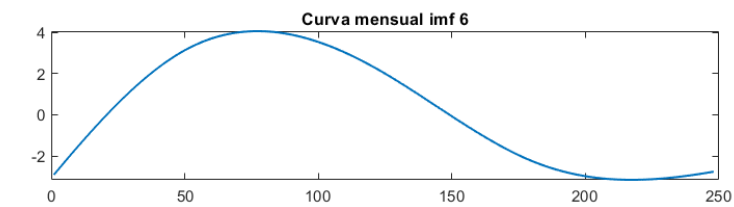
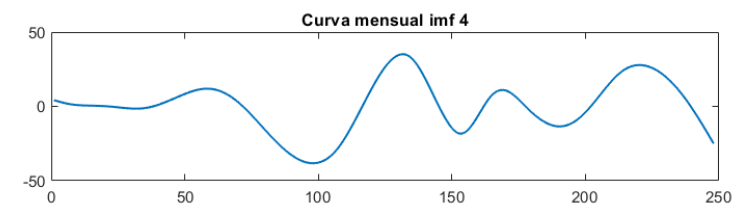
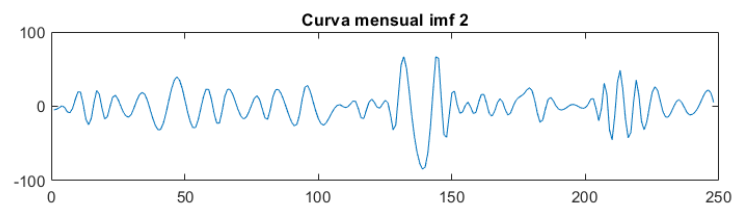
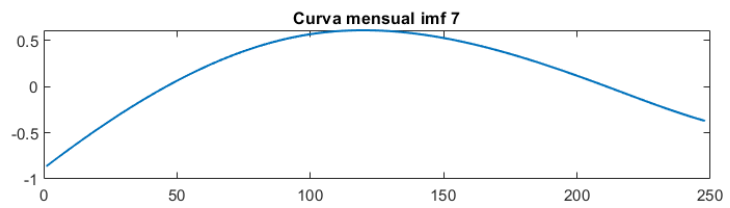
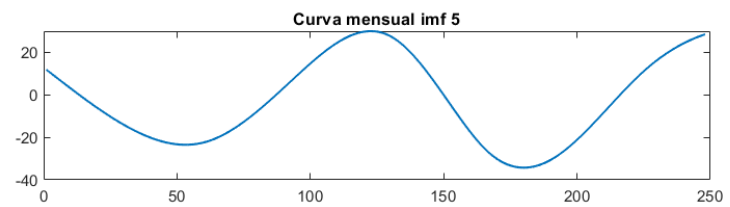
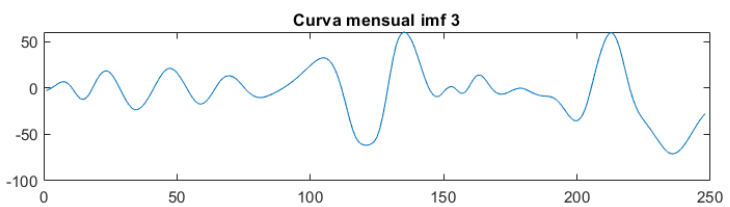
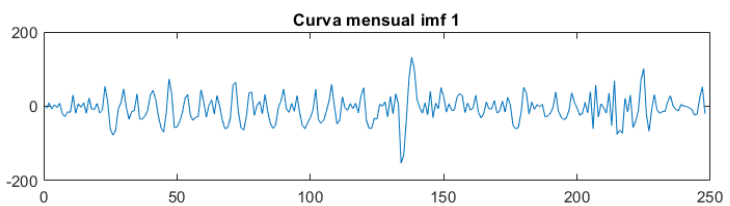


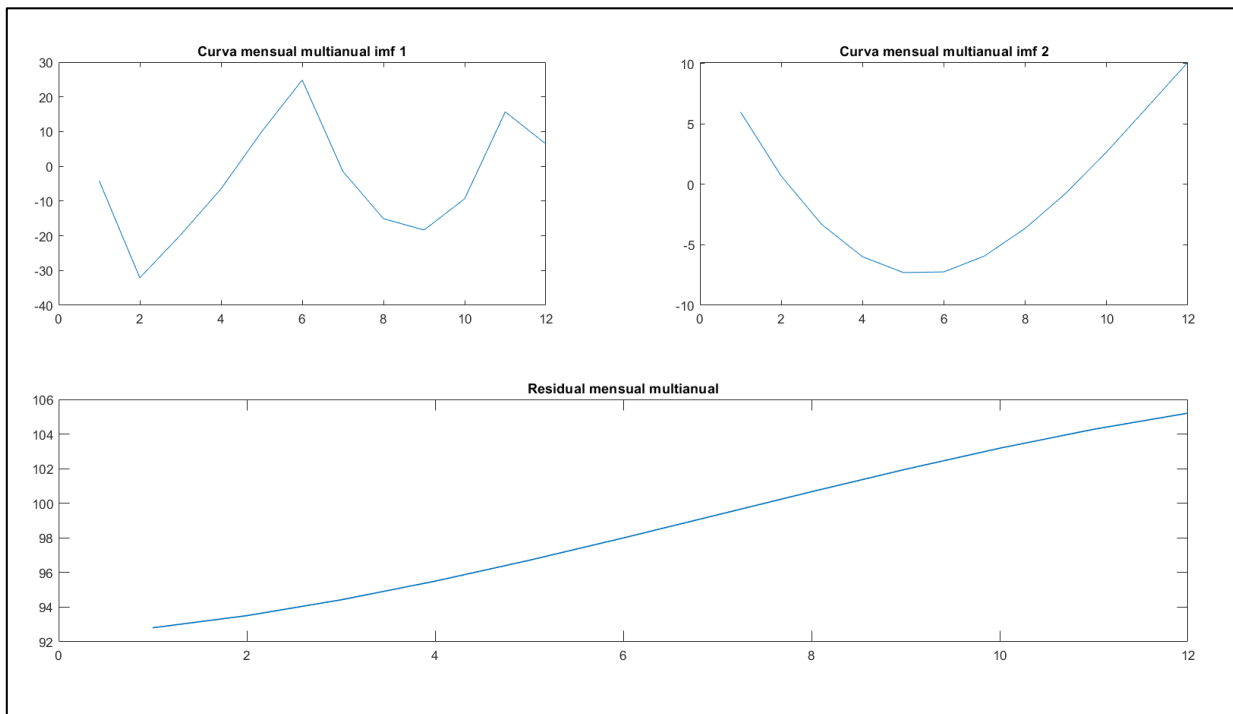
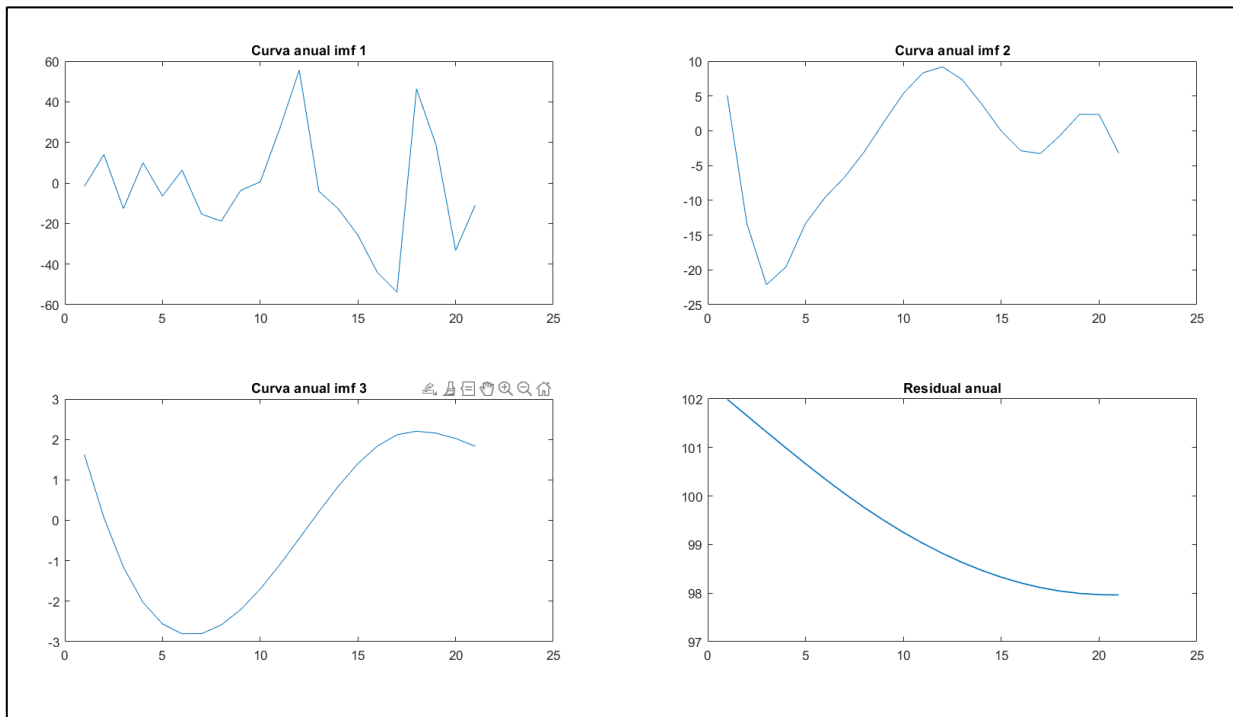
- Periodogramas



• Curvas IMD







## Pruebas Estadísticas

- **Análisis de saltos**

En primer lugar, tenemos la prueba de Pettitt, la cual arroja los siguientes resultados:

Prueba de Pettit	
Serie	Valor p
Diaria	1.27E-43
Mensual	0.025889
Anual	1.0067
M. Multianual	0.36701

A partir de esto es posible concluir que la serie de valores diarios no es homogénea y presenta saltos, pero, al agregar la serie mensual o anualmente, en promedio los valores se vuelven más parecidos y no se presentan saltos en la serie de acuerdo con la prueba.

La siguiente prueba realizada es la prueba de suma de rangos, de la cual se obtuvieron los siguientes resultados:

Prueba de suma de rangos		
Comparación	H	P
Diaria-Mensual	1	0.001269
Diaria-Anual	1	0.049304
Diaria-M. Multianual	0	0.069314
Mensual-Anual	0	0.383207
Mensual-M. Multianual	0	0.341506
Anual-M. Multianual	0	0.836922

Las comparaciones en las que se rechaza la hipótesis nula (Es decir, cuando  $H=1$ ) muestran que los 2 grupos de datos no tienen la misma mediana para un nivel de significancia del 5%, mostrando que la serie de datos diaria tiene una mayor variación en sus estadísticos con respecto a las otras series, nuevamente resaltando como se homogenizan los datos a agregar mensual o anualmente.

La siguiente prueba aplicada fue la de Kruskal-Wallis, obteniendo los siguientes resultados:

Prueba Kruskal-Wallis	
Serie	P
Diaria	0
Mensual	0.2482
Anual	0.7624
M. Multianual	0.4233

En esta prueba se compararon los datos diarios del primer año con los del segundo, los datos mensuales del primer año y el segundo, los datos anuales de los primeros 10 años y los siguientes 10 y los datos mensuales multianuales de los primeros 6 meses y los siguientes 6. Los resultados nuevamente confirman la poca homogeneidad de los datos diarios, ya que, la prueba lo que muestra es que estos datos del primer año y los del segundo año en la serie diaria no provienen de la misma distribución de probabilidad. En los demás casos el p valor nos indica que la hipótesis nula no puede ser rechazada, indicando que los datos comparados pueden venir de la misma distribución de probabilidad.

El siguiente test realizado fue el de CUSUM, que arrojó los siguientes resultados:

Prueba CUSUM	
Serie	H
Diaria	1
Mensual	1
Anual	0
M. multianual	0

Prueba t de student:

Prueba t student	
Serie	H
Diaria	0
Mensual	0
Anual	0
M. multianual	0

Aunque la prueba t de student se aplica a muestras con pocos datos, es posible usarla para determinar si los datos provienen de una distribución normal con una media determinada. En este caso se usó el valor de media calculado para cada una de las series, obteniendo en todos los casos un resultado que impide rechazar la hipótesis nula, por lo cual los datos del Río Lebrija podrían seguir una distribución normal de acuerdo con esta prueba.

- **Pruebas de hipótesis de tendencias**

Prueba de coeficiente de correlación de Spearman

Coeficiente de Spearman	
Serie	H
Diaria	0.084318
Mensual	-0.097902.
Anual	0.260606.
M. multianual	0.885714

Para la prueba del coeficiente de Spearman, se comparó en primer lugar, los valores diarios para el primer y segundo año de datos, para la serie mensual se compararon los primeros 12 mese contra los siguientes 12, en al caudal anual se compararon los primeros 10 años contra los siguientes 10 y finalmente, en la serie mensual multianual se compararon los primeros 6 meses del año contra los siguientes 6. A partir de los resultados del test, es posible evidenciar la baja correlación de los datos, excepto en el caudal mensual multianual, lo cual puede hablar de 2 periodos de comportamiento similar a lo largo de cada año, es decir, un comportamiento bimodal que puede ser observado de mejor manera con el paso de varios años.

#### Prueba de Mann-Kendall

Prueba de Mann-Kendall		
Serie	H	P
Diaria	1	0
Mensual	0	0.401691
Anual	0	0.694644
M. multianual	0	0.064104

#### Prueba de Mann-Kendall modificada

Prueba de Mann-Kendall Modificada		
Serie	H	P
Diaria	0	0.344024
Mensual	0	0.394068
Anual	0	0.69464
M. multianual	0	0.064104

Las pruebas de Mann-Kendall para un grado de significancia de 0.05 arrojan resultados muy similares, exceptuando en la serie diaria, en donde la primera nos indica que los datos de esta serie no presentan ninguna tendencia mientras que la segunda prueba nos indica que si tienen tendencia y podrían estar asociados a una distribución de probabilidad. En este caso no podemos rechazar la hipótesis nula y los datos pueden tener tendencia.

## Prueba de regresión lineal

Prueba de regresión lineal	
Serie	R <sup>2</sup>
Diaria	0.07121
Mensual	0.238141
Anual	0.018553
M. multianual	0.736551

Para realizar esta prueba, en el caso de las series mensual y diaria, se compararon los datos correspondientes al 3 y 4 año, nuevamente observando una baja correlación entre estos datos, lo cual indicaría, que el caudal del río Lebrija no presenta ciclos marcados a lo largo de periodos anuales. Por otro lado, nuevamente es posible apreciar una alta correlación entre el primer y segundo semestre en la serie mensual multianual.

- **Prueba de hipótesis de cambio de distribución**

### Prueba de Wald-Wolfowitz

Prueba de Wald-Wolfowitz		
Serie	H	P
Diaria	1	0
Mensual	1	0
Anual	0	0.184091
M. multianual	0	0.350649

- **Pruebas de hipótesis de normalidad**

### Prueba de Kolmogorov-Smirnov

Prueba de Kolmogorov-Smirnov		
Serie	H	P
Diaria	1	0
Mensual	1	0.003175
Anual	0	0.800237
M. multianual	0	0.942268

### Prueba Chi cuadrado



Prueba Chi Cuadrado	
Serie	H
Diaria	1
Mensual	1
Anual	0
M. multianual	0

#### Prueba de Anderson Darling

Prueba de Anderson-Darling		
Serie	H	P
Diaria	1	>0.0005
Mensual	1	>0.0005
Anual	0	0.46427
M. multianual	0	0.681819

Los resultados de las pruebas de hipótesis de normalidad son muy concordantes entre sí, indicando de forma rotunda que las series mensual y diaria no siguen una distribución normal, mientras que, al agregar anualmente de manera mensual multianual, si parece seguirse una distribución normal, con un P valor alto que parece confirmar la hipótesis, aún así hay que tener en cuenta que estas series tienen muy pocos datos y estas pruebas pueden no ser las más indicadas para definir su normalidad.

- **Pruebas de aleatoriedad e independencia**

Al aplicar la prueba de puntos de cambio en la serie de datos diarios, teniendo en cuenta un umbral igual a 2 veces la desviación estándar, se encontraron un total de 72 puntos por fuera de este rango, lo cual habla de la variabilidad de los datos diarios.

#### Prueba de diferencia de rangos

Prueba de Diferencia de Rangos		
Serie	H	P
Diaria	1	0
Mensual	1	0
Anual	1	0.00006
M. multianual	1	0.000488

El resultado de la prueba de diferencia de rangos (O prueba de Wilcoxon) muestra que los datos de las diferentes series no parecen provenir de una misma distribución de probabilidad.

Al aplicar la prueba de Von Newman, se obtuvieron los siguientes resultados:

Prueba de Von Newman	
Serie	Q'
Diaria	-0.00000126
Mensual	-0.0001523
Anual	-0.00339521
M. multianual	-0.01351389

Los bajos valores en el parámetro  $Q'$  nos indica que en ninguna de las 4 series de tiempo se tiene dependencia con los valores anteriores.

En conclusión, el flujo del río Lebrija parece ser muy cambiante con el paso del tiempo, ya que, incluso la serie de datos anuales presenta una gran variabilidad. Asimismo, se puede decir, que es más complejo modelar una serie diaria que una serie mensual, esto debido a que en la serie diaria hay picos más altos de caudal, mientras que, a medida que se agregan los datos, los valores de la serie tienden a acercarse a la media mostrando mayor uniformidad.