Análisis de series de hidrológicas

Profesor Luis Fernando Carvajal Serna

Facultad de Minas

Departamento de Geociencias y Medio Ambiente

Análisis de homogeneidad. Qué es el análisis de homogeneidad de una series hidrológica?

El análisis de homogeneidad es un procedimiento que usa pruebas gráficas y estadísticas para analizar una serie hidrológicas (caudales, precipitación, temperaturas, etc.). El objetivo es revisar y determinar las características y comportamientos que estas series puedan tener desde el punto de vista de sus estadígrafos (media, desviación estándar, coeficiente de asimetría, etc.), también permite determinar que tipo de tratamiento necesitará la serie para poderla usar en modelos. Así como determinar posibles cambios en el comportamiento de una variable hidrológica debidas a fenómenos locales o macro-climáticos (deforestación, usos del suelo, construcción de embalses,calentamiento global.)

Que es una serie hidrológica?

PROCESO DETERMINISTICO + PROCESO ALEATORIO → SERIE HIDROLÓGICA

La componente determinítica o patrones climáticos:

1. Ciclo anual
2. El fenómeno ENSO (El Niño Oscilación del Sur)
3. Oscilación Decanal del Pacificó (PDO)
4. Fases de la luna.

La componente aleatoria:

1. Turbulencia de la atmosfera.
2. Manchas solares
3. Tormentas solares

Conceptos generales de estadística

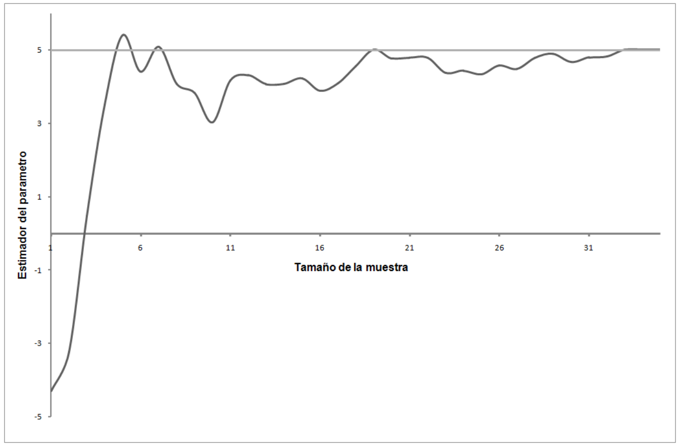
Población

↓

Muestra

↓

Inferir el comportamiento de la población a través de la muestra.



Media	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Varianza	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}({x_i}^2 - \bar{X})^2$
Desviación estándar	$s = \left[\frac{1}{n-1}\sum_{i=1}^{n}({x_i}^2 - \bar{X})^2\right]^{\frac{1}{2}}$
Coeficiente de variación	$CV = \frac{s}{\bar{x}}$
Coeficiente de asimetría	$C_s = \frac{n\sum_{i=1}^{n}({x_i}^3 - \bar{X})^3}{(n-1)(n-2)s^3}$

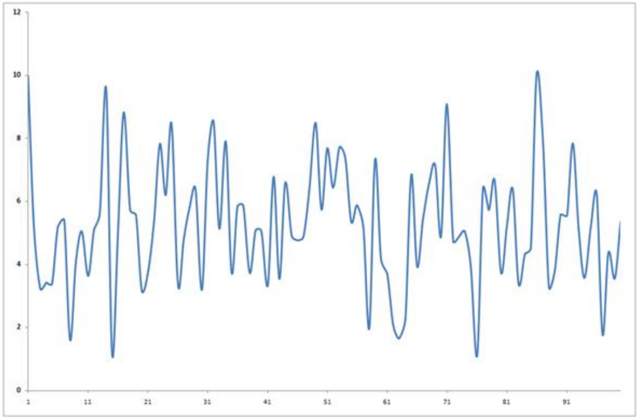
Homogeneidad

- Una serie es estadísticamente homogénea, cuando sus datos provienen de la misma población, es decir que tienen una sola media, una sola varianza, etc. Series estacionarias de segundo orden.

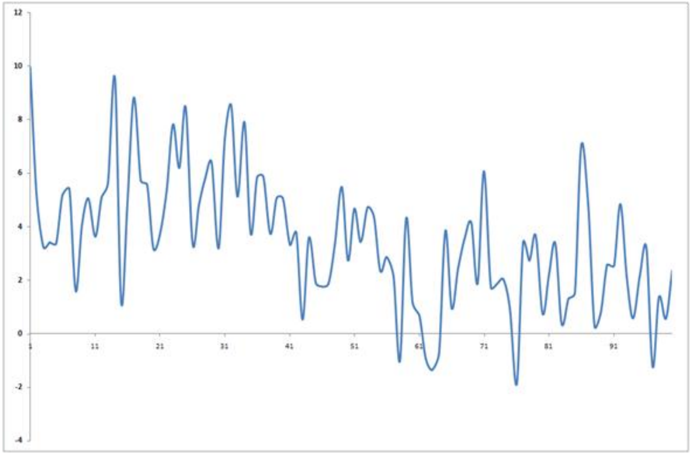
- ¿Qué significa que una serie tenga una sola media y una sola varianza?

Si se tiene una serie estacionaria de segundo orden, esta y cualquier parte de ella deben tener la misma media y la misma varianza, teniendo claro que es necesario que la porción de la serie tenga los suficientes datos como para considerar el calculo de los parámetros significativo.

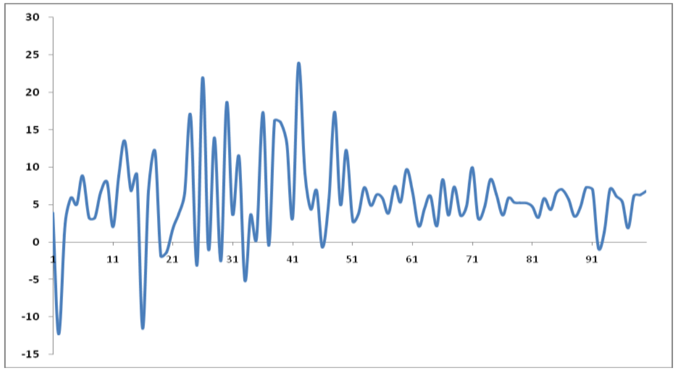
Método gráfico



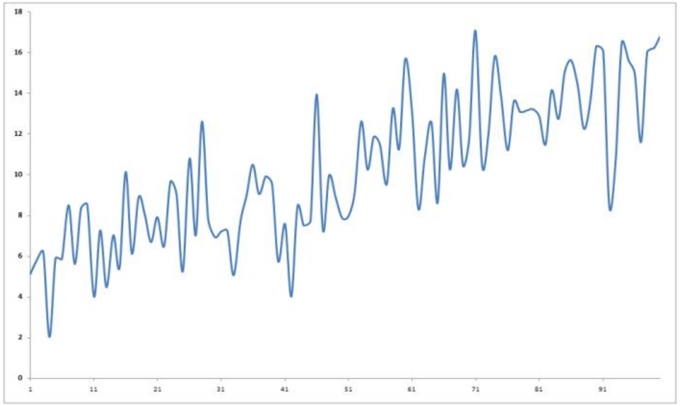
Los datos varían alrededor de un valor y que su variación es aproximadamente constante (serie homogénea)



La variabilidad permanece aproximadamente constante, la media alrededor de la cual varían los datos cambia en un punto, existe un cambio en la media.



La variabilidad cambia alrededor de la misma media, existe un punto de cambio en la varianza



Hay un cambio gradual en la media de los datos, diremos que existe una tendencia en la serie.

Métodos cuantitativos o comprobatorios

Estos métodos se basan en el uso de las pruebas de hipótesis de la estadística muestral:

En una prueba de hipótesis definimos una hipótesis nula (H_o) y una hipótesis alternativa (H_a). Una hipótesis estadística es una afirmación que se hace acerca de la distribución de una muestra. La prueba consiste en comparar un valor analizado asociado a H_o o a H_a contra un valor crítico que define la aceptación o el rechazo de la hipótesis evaluada.

1. Una característica estadística del parámetro analizado.

2. Un nivel de "significancia" de la prueba (α) asociado a la confiabilidad del resultado. En términos estadísticos α corresponde a la probabilidad de rechazar la hipótesis nula cuando era verdadera, o en términos mas sencillos α nos da una indicación de cuanta es la probabilidad de equivocarnos realizando la prueba. Generalmente se usa una α correspondiente a un 5% de probabilidad.

3. Grados de libertad de la prueba. La incertidumbre asociada a la medición de un parámetro depende del tamaño de la muestra, los grados de libertad son una indicación de esta incertidumbre debido al tamaño muestral.

Detección de outliers

Un outlier es un valor de la variable en uno o varios puntos de la serie de tiempo que parece no corresponder a la distribución estadística de los demás valores de la serie.

Tipos de outliers:

1. Outliers que corresponden a realizaciones de eventos de carácter extremo, que son explicables físicamente y que en realidad ocurrieron. Son puntos muy importantes para conocer el verdadero comportamiento estadístico de las series hidrológicas y que tienen un peso físico y muy importante en el momento de usar los datos para análisis hidrológicos mas avanzados.

2. Outliers que corresponden a errores de medición o a manipulación de los datos. Son puntos que no ocurrieron en la realidad y que perturban las características estadísticas de los demás datos. Es necesario eliminar estos puntos de la serie de tiempo, para esto se necesita de la experiencia y conocimiento del comportamiento de la variable y del análisis de las características macro-climaticas en el momento del valor dudoso.

Pruebas para detectar outliers

Prueba de puntos por fuera del rango.

Es una prueba paramétrica basada en el uso de la función de distribución Normal.

Procedimiento consiste en lo siguiente:

1. De la serie analizada se "separa" el valor sospechoso a evaluar.
2. Para la serie restante se calcula la media y la desviación estándar.
3. Si el valor sospechoso se encuentra en el intervalo $\bar{x} \pm N * S_x$, no es outlier. N puede tomar el valor de 2 o 3 dependiendo de que tan estricta quiera hacerse la prueba.
4. Si el valor sospechoso se considera outlier se saca temporalmente de la serie y se procede a analizar el próximo valor sospechoso.

Una vez determinados todos los outliers deben clasificarse en: Explicables o No explicables. Todos los outliers que se consideren no explicables deben ser eliminados definitivamente de la serie.

Detección de outliers

Prueba del autocorrelograma de Anderson.

Consiste en estimar los coeficientes de auto correlación de una serie de tiempo para diferentes rezagos temporales.

$$\frac{-1 - Z_{1-\frac{\alpha}{2}}\sqrt{N-K-1}}{N-K} < R(K) < \frac{-1 + Z_{1-\frac{\alpha}{2}}\sqrt{N-K-1}}{N-K}$$

Una serie de tiempo puede aceptarse como serialmente independiente, si su autocorrelograma se encuentra (totalmente) contenido en el intervalo:

Para $K>1$, donde K es el rezago temporal , R(K) es el coeficiente de auto correlación de rezago K y $Z_{1-\alpha/2}$ es el quantil $1-\alpha/2$ de la distribución normal estándar.

Métodos cuantitativos

Se investiga con estás pruebas posibles cambios en la media, la varianza o tendencia.

1. Cambio en la media

Se dice que una serie de tiempo presenta cambio en la media cuando se observa un cambio abrupto en el nivel o la magnitud de determinada variable. En general estas pruebas evalúan la hipótesis nula de igualdad de las medias entre las dos subseries resultantes de dividir la serie original por un punto de cambio (PC).

Las pruebas para detectar cambio en la media son: Prueba T simple, prueba T modificada, prueba de Mann-Whitney y la prueba del signo.

Métodos cuantitativos

Prueba de Signo.

Se asume que la serie Y_t, $t=1,2,\ldots, N$, es una serie hidrológica que puede ser dividida en dos partes con el mismo numero de observaciones, $Y_1, Y_2, \ldots, Y_m$ y Y_{m+1}, $Y_{m+2}, \ldots, Y_{2m}$, *con* $N=2m$. Se define una nueva serie para $i=1, 2, \ldots, m$ cómo:

$W_1=1$ *si* $Y_i<Y_{m+i}$

$W_1=0$ *si* $Y_i>Y_{m+i}$

Si $Y_i=Y_{m+1}$, se eliminan de la prueba.

El estadístico de la prueba Signo puede ser calculado como:

$$U_c = \left|\frac{2\sum_{i=1}^{m} W_i - m}{\sqrt{m}}\right|$$

La hipótesis de igualdad de medias para los dos grupos de observaciones no puede ser rechazada si: $|U_c| \leq Z_{1-\frac{\alpha}{2}}$, donde $Z_{1-\frac{\alpha}{2}}$ representa el quantil 1-α/2 de la distribución normal estándar.

La prueba de Signo es recomendada para el caso m>20.

Métodos cuantitativos

2. Cambio en la varianza

Se dice que una serie de tiempo presenta salto en la varianza cuando se observa un cambio abrupto en el nivel o la magnitud de la varianza de determinada variable. En general estas pruebas evalúan la hipótesis nula de igualdad de varianzas entre las dos subseries resultantes de dividir la serie original por un punto de cambio (PC).

Las pruebas para detectar cambio en la varianza son: Prueba F simple, prueba F modificada, prueba de Ansari-Bradley, prueba de Bartleff, prueba de Levene.

Métodos cuantitativos

Prueba *F* simple

Se tiene una serie hidrológica, Y_t, $t=1,2,\ldots,N$, que se divide en dos subseries (antes y después del PC, $t=N_1$). La primera subserie, definida como Y_t, $t=1,2,\ldots,N_1$, se asume normalmente distribuida con media μ_1 y varianza σ_1^2. La segunda subserie, definida como, Y_t, $t=N_1+1,\ldots,N$, se asume normalmente distribuida con media μ_2 y varianza σ_2^2. La prueba *F* simple puede ser usada para probar la igualdad de las varianzas de las dos subseries. Se considera la siguiente prueba: H_o: $\sigma_1^2=\sigma_2^2$ *vs.* H_a: $\sigma_1^2\neq\sigma_2^2$

Donde H_o y H_a son la hipótesis nula y alternativa, respectivamente. El estadístico de la prueba F simple está definido como:

$$F = \frac{\widehat{\sigma_M}^2}{\widehat{\sigma_m}^2}$$

En donde $\widehat{\sigma_M}^2$ y $\widehat{\sigma_m}^2$ representan las varianzas estimadas del mayor y el menor entre σ_1^2 y σ_2^2 respectivamente. La hipótesis de igualdad de varianzas no puede ser rechazada si:

$F_{\alpha/2}(N_1-1, N-N_1-1) \leq F \leq F_{1-\alpha/2}(N_1-1, N-N_1-1)$. Donde $F_\beta(N_1-1, N-N_1-1)$ es el quantil β de la distribución F con N_1-1 *y* $N-N_1-1$ grados de libertad.

Métodos cuantitativos

Tendencia en la serie

Se dice que una serie presenta tendencia en la media, cuando se presenta un cambio progresivo y gradual en la magnitud o el nivel de determinada variable.

Las pruebas aquí presentadas para detectar cambio en la media son: Prueba T, prueba de Hotelling- Pabst, prueba de Mann-Kendall y la prueba de sen.

Prueba T para la detección de tendencias lineales

Se supone que Y_t, $t=1,2,\ldots, N$ es una serie hidrológica anual, en la cual se pretende determinar si se presenta una tendencia significativa. La prueba *T* puede ser usada para determinar si una serie dada tiene una tendencia lineal significativa.

Si la serie analizada Y_t tiene una tendencia significativa, una ecuación de regresión puede ser escrita de la siguiente forma: $Y_t = a + b*t$, $t=1,\ldots, N$.

Donde N representa el tamaño muestral y, a y b representan los parámetros o coeficientes de la ecuación de regresión.

Métodos cuantitativos

Si la serie Y_t tiene una tendencia, significa que el parámetro b es diferente de cero.

Luego, probar la hipótesis de que la serie tiene una tendencia significativa es equivalente a probar l a hipótesis de que el parámetro b es diferente de cero. Se calcula el estadístico T_c a partir de la expresión:

$$T_c = \hat{\rho}Y,t\sqrt{\frac{N-2}{1-\hat{\rho}Y,t^2}}$$

Donde: $\hat{\rho}Y,t$ es el coeficiente de correlación simple entre Y_t y t. La hipótesis de que el parámetro b es igual a cero es rechazada si $|T_c| > T_{1-\alpha/2}(N-2)$, $T_{1-\alpha/2}(N-2)$ es el quantil $1-\alpha/2$ de la distribución t de Student con $N-2$ grados de libertad.

Red de monitoreo cantidad y calidad de agua del río Medellín.

Tabla 3.4 Pruebas utilizadas para el análisis de homogeneidad.

Cambios en la media	Cambio en la varianza	Tendencia en la media
* Prueba T simple	* Prueba F simple	* Prueba T
* Prueba T modificada	* Prueba F modificada	* Hotelling-Pabst
* Prueba del signo	* Ansari-Bradley	* Mann-Kendall
* Prueba de Mann-Whitney	* Bartleff	* Sen
	* Levene	

Tabla 3.5 Tabla de convenciones

Media	PRUEBA		
1 -	Mann-Whitney	A-	Acepta media estacionaria
2 -	T Simple	R-	Rechaza media estacionaria
3 -	T Modificada		
4 -	Signo		
Varianza	PRUEBA		
1 -	F Simple	A-	Acepta varianza estacionaria
2 -	F Modificada	R-	Rechaza varianza estacionaria
3 -	Ansari-Bradley		
4 -	Barlett		
5 -	Levene		
Tendencia	PRUEBA		
1 -	T	A-	Acepta serie sin tendencia en la media
2 -	Hotelling-Pabs	R-	Rechaza serie sin tendencia en la media
3 -	Mann-Kendall		
4 -	Sen		
# -	El tamaño de la serie no es el suficiente para aplicar la prueba		

Red de monitoreo cantidad y calidad de agua del río Medellín

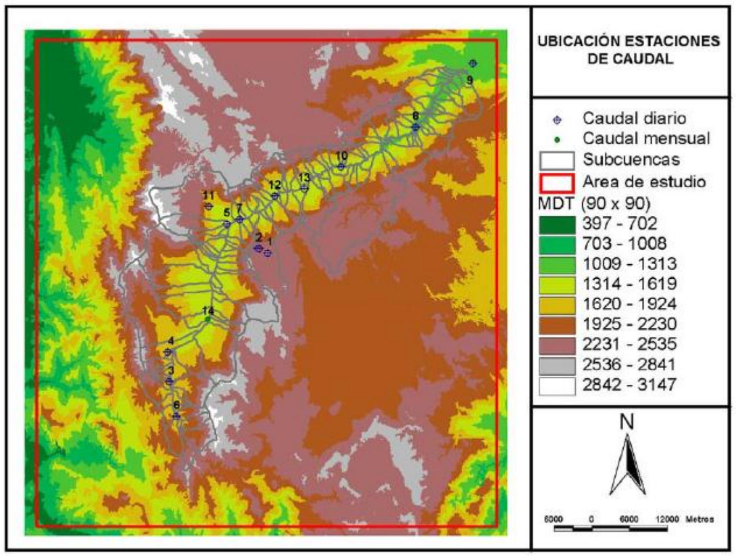


Figura 3.4 Localización de estaciones de caudal

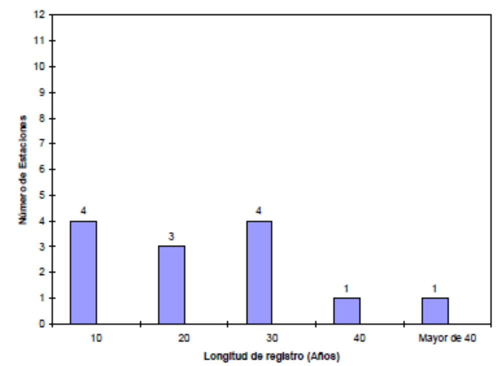


Figura 3.5 Longitud de registros de las estaciones de caudal diario en la zona de estudio.

Series de caudal

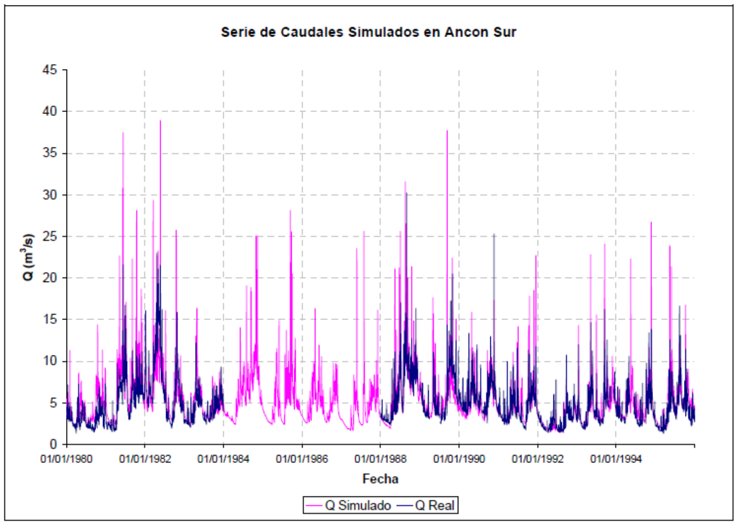


Figura 5.6 Serie de caudal simulada y observada para la subcuenca sobre el río Medellín con salida en la estación Ancón Sur.

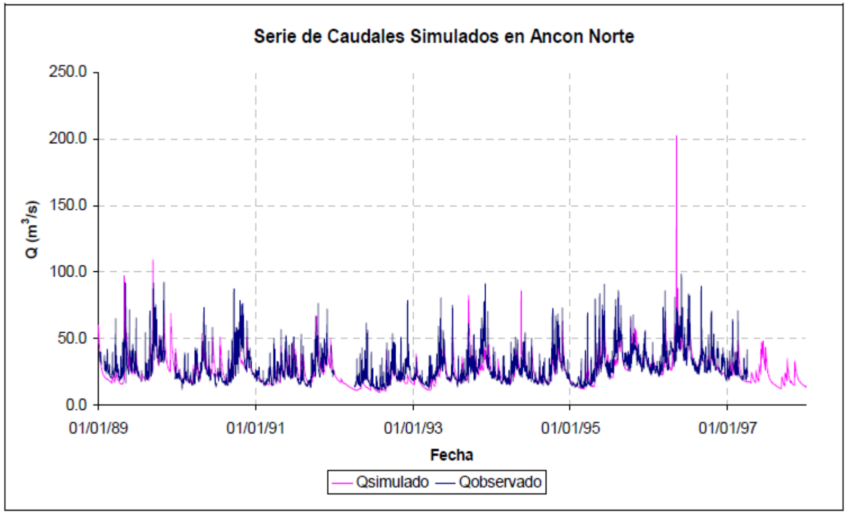


Figura 5.10 Serie de caudal simulada y observada para la subcuenca sobre el río Medellín con salida en la estación Ancón Norte.

Red de monitoreo cantidad y calidad de agua del río Medellín.

Tabla 3.8 Pruebas de homogeneidad a nivel mensual de las estaciones de caudal

Estación		Media					Varianza					
Nombre	Código	Cambio	1	2	3	4	Cambio	1	2	3	4	5
RM-7 Piedras Blancas	2701709	01/1992	R	R	R	R	10/1986	R	R	R	R	R
RM-8 Chorrillos	2701710	10/1970	A	R	R	R	01/1976	R	R	R	R	R
MEDELLIN @ CALDAS RM	2701716	05/1984	A	R	A	R	06/1986	R	A	R	R	R
MEDELLIN @ ANCON SUR	2701727	01/1991	R	R	R	R	12/1990	R	R	R	R	R
Q. EL HATO @ RM-9	2701731	09/1969	A	A	A	R	12/1976	R	R	R	R	R
MEDELLIN @ SALADA LA	2701733	01/1976	R	R	R	R	01/1976	R	R	R	R	R
MEDELLIN @ MACHADO R	2701734	04/1981	A	R	R	R	06/1982	R	A	R	R	R
MEDELLIN @ YARUMITO	2701735	06/1995	A	R	R	R	08/1988	R	R	R	R	R
RMS-15 Gabino	2701736	07/1995	A	R	A	R	03/1996	A	A	R	A	R
MEDELLIN @ HATILLO EL	2701738	04/1994	A	R	R	R	09/1986	R	R	R	R	R
Q. LA GARCIA @ EMBAL	2701765	05/1995	A	R	A	R	05/1995	A	A	R	A	R
MEDELLIN @ ANCON NOR	2701781	10/1994	A	R	R	R	07/1994	A	A	R	A	R
RMS-22 GIRARDOTA	2701803	12/2000	R	R	R	R	11/2000	A	A	R	A	R

Tabla 3.9 Pruebas de homogeneidad a nivel anual de las estaciones de caudal

Estación		Media					Varianza						Tendencia			
Nombre	Código	Cambio	1	2	3	4	Cambio	1	2	3	4	5	1	2	3	4
RM-7 Piedras Blancas	2701709	1992	R	R	R	R	1988	R	R	R	R	R	R	R	R	R
RM-8 Chorrillos	2701710	1970	A	R	R	R	1976	R	A	R	R	R	A	A	A	A
MEDELLIN @ CALDAS RM	2701716	-	A	#	#	A	1985	A	#	R	A	R	A	A	A	A
MEDELLIN @ ANCON SUR	2701727	1991	R	R	R	A	1988	R	A	R	A	R	R	R	R	R
Q. EL HATO @ RM-9	2701731	-	A	#	#	A	1976	R	#	R	R	R	A	A	A	A
MEDELLIN @ SALADA LA	2701733	1976	R	#	#	A	1978	A	#	R	A	R	A	A	A	A
MEDELLIN @ MACHADO R	2701734	-	A	#	#	A	1982	R	#	A	A	R	A	A	A	A
MEDELLIN @ YARUMITO	2701735	1995	A	R	R	A	1991	R	R	R	R	R	R	R	R	R
RMS-15 Gabino	2701736	-	A	A	A	A	1992	A	A	R	A	R	A	A	A	A
MEDELLIN @ HATILLO EL	2701738	1994	A	R	R	A	1986	R	R	R	R	R	R	R	R	R
Q. LA GARCIA @ EMBAL	2701765	-	A	#	#	A	1991	A	#	R	A	R	A	A	A	A
MEDELLIN @ ANCON NOR	2701781	-	A	#	#	A	1992	A	#	R	A	R	A	A	A	A
RMS-22 GIRARDOTA	2701803	-	A	#	#	A	1999	R	#	R	A	R	A	A	A	A

Gabino (2701736): En general las pruebas mas concluyentes acerca de cambios en la media, varianza y tendencias en la media no detectaron brincos o irregularidades de la serie.

Red de monitoreo cantidad y calidad de agua del río Medellín.

Medellín_Caldas RM (2701716): Es una serie de muy pocos datos, por lo que los resultados que arrojan las pruebas son poco confiables. En general los cambios que algunas pruebas afirman son debidos posiblemente a datos atípicos que generan ruido. En general los resultados de las pruebas más relevantes no detectaron cambios ni tendencias.

Medellín _Ancon_Sur (2701727): Las pruebas detectan un cambio en la media en el año de 1991, cambio que levemente se puede apreciar gráficamente. Por su parte la varianza también registra cambios en el año de 1990. La serie registra tendencias en la media.

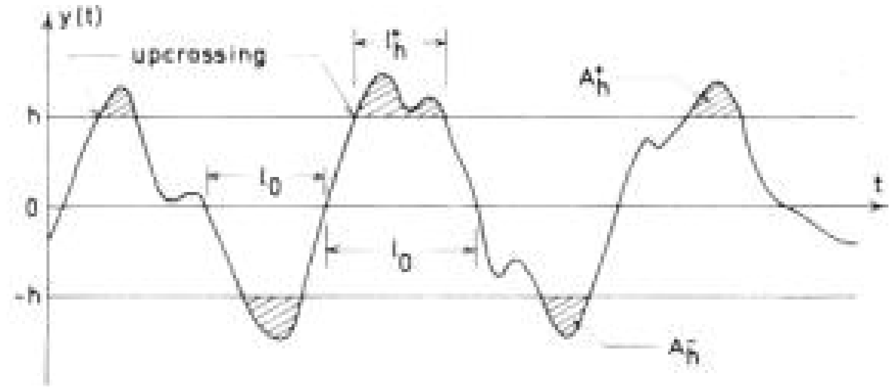
RM-9 Q. El Hato (2701731): Las pruebas no detectaron cambios en la media ni tendencias de la misma. Alrededor de 1976 se detectó un cambio en la varianza, pero en realidad este es más atribuible a los datos atípicos que se presentaron en ese período el cual genera ruido en las pruebas. No se considera por tanto que tal cambio exista.

La Salada (2701733): En general la serie tiene muy pocos registros. Tanto para la media como para la varianza se detectan cambios en 1976 lo cual es verificable de la gráfica. No hay tendencias en la media.

Machado (2701734): Solo se tienen registros de 1980 a 1984 lo que es una muestra muy corta para el análisis homogéneo.

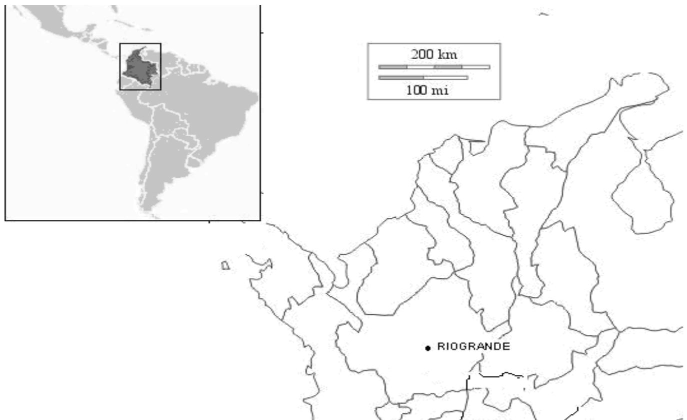
Yarumito (2701735): Según las pruebas el cambio en la media se presenta en el año de 1995; esto es explicable ya que en ese período de tiempo comenzó a descargar sobre el río la central de tasajera lo que indudablemente aumento los caudales como se aprecia en la serie. En la varianza también se presenta cambio pero alrededor del ***año de 1991.*** Además de lo anterior la serie registra tendencias en la media.

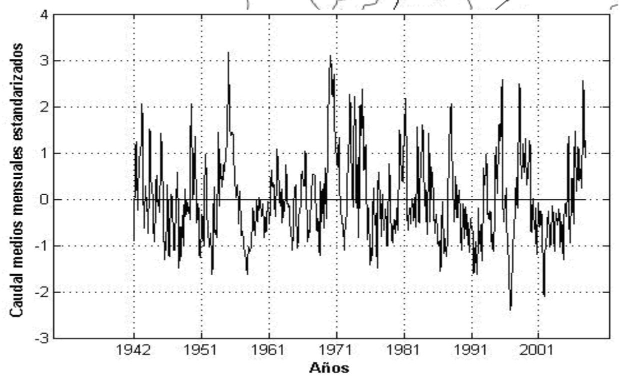
Teoría de cruces (rachas)



En el estudio de series de lluvias y caudales es de especial interés el análisis de las características relacionadas con las sequías y los excesos, para tal efecto se utiliza el concepto de excursiones del proceso aleatorio por encima y por debajo de un valor límite. Los parámetros se conocen como excursiones o cruces positivos y excursiones o cruces negativos. Se determinan el número y la longitud de las excursiones por encima (N+, L+) y por debajo (N-, L-) de un valor de referencia además del área por encima(A+) y por debajo (A-). Para series estandarizadas con media cero y varianza 1 el valor de referencia o umbral para medir los cruces es u=1.

Ejemplo: estación RG8





	HISTÓRICA	ONDITAS	AR(2)	ARMA(2,1)
Criterios estadísticos				
Media	0	0	-0.02	-0.05
Desviación estándar	0.99	0.99	1.02	0.98
Asimetría	0.76	0.69	0.58	0.53
Kurtosis	3.67	3.48	3.02	3.30
Criterios relacionados con eventos extremos				
Núm. de excursiones positivas (NL+)	58	56.2	54	53
Longitud de excursiones Positivas (L+)	2.0	2.20	2.33	2.06
Área de excursiones positivas (A+)	3.53	3.51	3.56	3.13
Número de excursiones Negativas (NL-)	60	65.0	53	71
Longitud de excursiones Negativas (L-)	1.72	1.49	2.28	1.92
Área de excursiones Negativas (A-)	-2.28	-2.06	-3.54	-2.9

Análisis de frecuencia
La Transformada rápida de Fourier

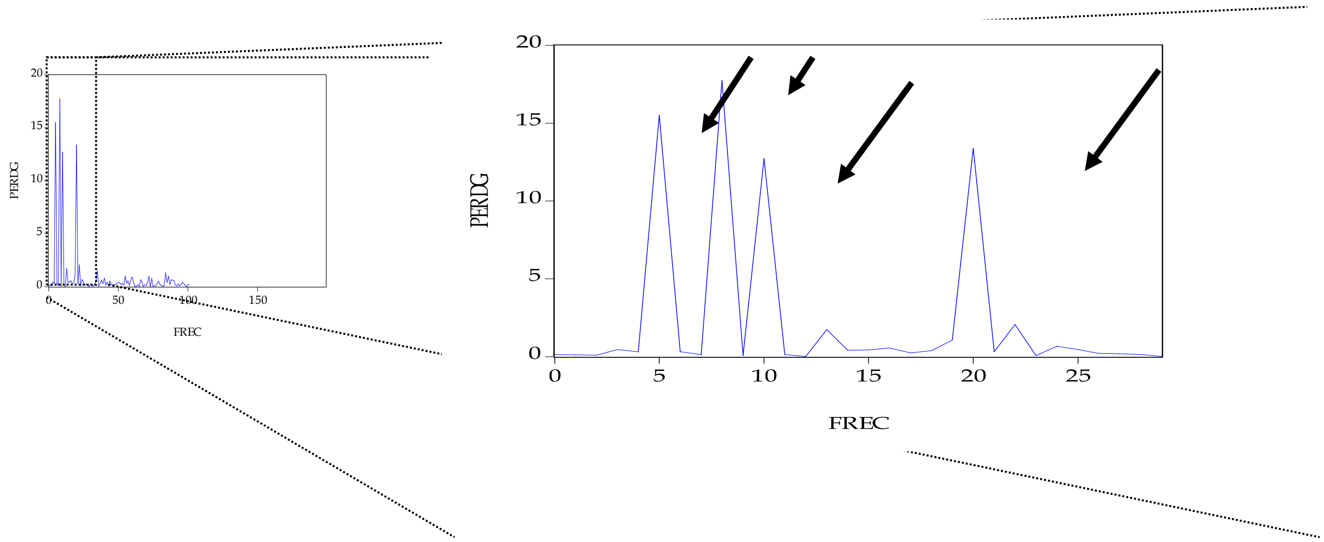
Una señal puede estar compuesta por tantos armónicos como el número de datos de la serie lo permita (*hasta n/2*) y cada armónico tiene asociado su respectiva frecuencia. Es la representación de una serie de tiempo en series de senos y cosenos. Una serie que está en el dominio del tiempo se lleva al dominio de la frecuencia. Es muy usado para descomponer una serie de tiempo en sus modos de vibración (frecuencias). Se aplica el principio de superposición para obtener nuevamente la serie original. Matemáticamente la representación de una señal en series de *Fourier.*

$$Y_t = \sum_{i=1}^{k} (a_p \cos \omega_i t + b_i \operatorname{sen} \omega_i t) + \varepsilon_t \qquad \omega_i = \frac{2\pi p_i}{N} \qquad p_i = 1,\ldots,k$$

$$\hat{a}_p = \frac{2}{N} \sum_{t=1}^{N} Y_t \cos p\omega_o t \qquad \hat{b}_p = \frac{2}{N} \sum_{t=1}^{N} Y_t \operatorname{sen} p\omega_o t$$

$$\hat{a}_0 = \sum_{t=1}^{N} \frac{Y_t}{N} \qquad \hat{a}_{N/2} = \frac{1}{N} \sum_{t=1}^{N} Y_t \cos \pi t \qquad I(\omega_p) = \frac{(a_p^2 + b_p^2)}{2\omega_0}$$

Periodograma



El periodograma está basado en una herramienta matemática denominada Transformada de Fourier, según la cual una serie, que cumpla determinados requisitos, puede descomponerse como suma de un número finito o infinito de frecuencias.

Ejemplo: TRF, estación Santa Rita