

caso. Sin embargo, dado que el coeficiente de correlación de rezago 1 es 0,1388, se puede estimar el factor K (ver Ecuación 5.139) para la corrección por sesgo de la desviación estándar; así, $K = 1,313$ y la desviación estándar corregida es $695 \text{ m}^3/\text{s}$, calculada con la Ecuación 5.138. La desviación estándar de la serie mensual global es $1394 \text{ m}^3/\text{s}$. En la Figura 5.9 está el correlograma de esta serie, donde se observa su estacionalidad. El valor del coeficiente de correlación de rezago 1 es 0,6827, y dado que son 480 datos mensuales, el factor de corrección de la desviación estándar K es 5,2746 y la desviación estándar corregida es $1400 \text{ m}^3/\text{s}$.

El coeficiente de asimetría de la serie anual es 0,14874, mientras que el de la serie mensual global es 0,5195. En el primer caso, la muy baja correlación para diferentes rezagos permiten considerar la serie anual como compuesta por datos independientes; además, las pruebas Chi-cuadrado y Kolmogorov-Smirnov indican que no se puede rechazar que esta serie provenga de una distribución Normal o Lognormal o Weibull con un nivel de confianza de al menos 90 %. Esto significa que no procede hacer ninguna corrección por sesgo al valor del coeficiente de asimetría. Para la serie mensual global, los coeficientes de correlación para varios rezagos son significativos por su carácter estacional (rezagos 1, 2, 5, 6, 9, 11, 12, por ejemplo); adicionalmente, las pruebas de hipótesis mencionadas antes establecen que no se puede rechazar que los datos de esta serie procedan de una distribución Gamma con al menos 90 % de nivel de confianza. Por consiguiente, la corrección propuesta por Fernandez y Salas (1990) puede aplicarse, resultando un valor corregido del coeficiente de asimetría de 0,53276.

Cuando se consideran las series individuales de caudales mensuales, las Figuras 5.14, 5.15, 5.16 y 5.17 presentan sus correspondientes valores multianuales de media y desviación estándar, los coeficientes mensuales de asimetría, al igual que los correlogramas para los meses de Febrero, Agosto, Septiembre y Diciembre, junto con los límites de confianza de 95 %. La gráfica de caudales mensuales multianuales muestra el régimen bimodal del río Magdalena en esta estación hidrométrica, siendo Febrero el mes más seco y Noviembre el más húmedo. Las desviaciones estándar mensuales presentan un comportamiento que en general indica que sus valores son aproximadamente mayores con caudales mensuales mayores. Los doce coeficientes de asimetría son todos mayores que cero, indicando que las distribuciones de caudales mensuales tienen colas más largas hacia la derecha, y que esto es más acentuado en los meses de Junio a Septiembre. Para todos los meses del año el coeficiente de correlación de rezago 1 es alto (varían entre 0,6763 y 0,9 correspondientes a los meses de Abril y Noviembre, respectivamente), y en general la correlación es significativa para los cuatro a seis primeros rezagos, como se observa en la Figura 5.17 con los límites de confianza. La corrección de los valores mensuales de desviación estándar se pueden justificar a partir del coeficiente de correlación de rezago 12, pues éste indica para un mes cualquiera la correlación con rezago de un año. En el caso de la estación de El Banco, los valores de este coeficiente varían entre $-0,1635$ y $0,1919$ (meses de Diciembre y Junio, respectivamente), y estando todos dentro de la franja definida por los límites de confianza. Esto indica que la corrección no debería proceder pues es marginal como se muestra en la Figura 5.15.

5.5.1.5. El Fenómeno de Hurst

Harold E. Hurst, durante su larga estadía en Egipto (1906-1968), fue responsable de estudiar la hidrología del río Nilo y de buscar su aprovechamiento con un enfoque integral, desde su nacimiento en los grandes lagos africanos y planicies etíopes hasta el gran delta en su desembocadura en el mar Mediterráneo [Graves et al. (2017)]. En sus análisis de caudales para identificar proyectos de embalses, Hurst (1951) usó un método similar al de Rippl [éste se resume en una curva de masa acumulada de caudales con superposición de líneas con pendiente de masa acumulada de caudal regulado constante de salida, para identificar el volumen requerido de almacenamiento; McMahon (1993)], analizando una estadística de caudales acumulados a lo largo del tiempo que denominó “rango”.

Para una serie de tiempo discreta y_t , $t = 1, 2, \dots, N$ (por ejemplo de caudales anuales) defínanse la suma acumulada parcial w_l como

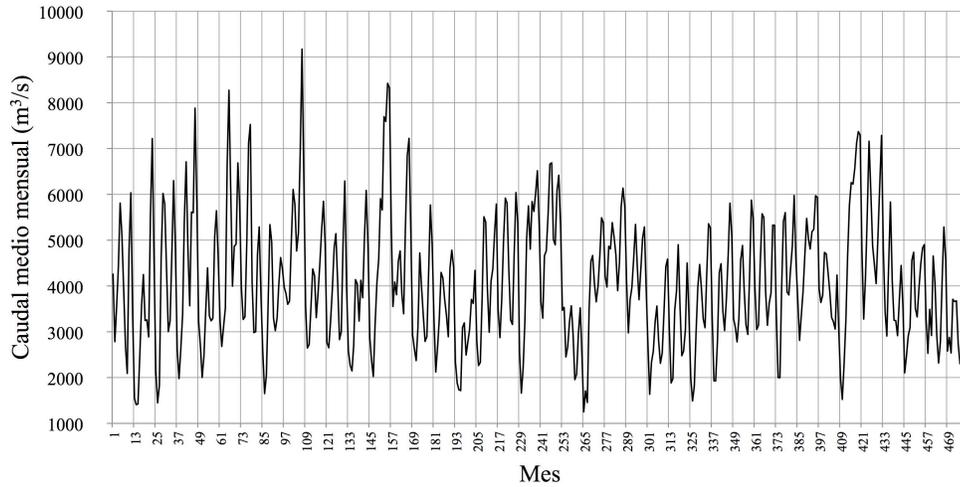


Figura 5.11: Serie de caudales medios mensuales - El Banco

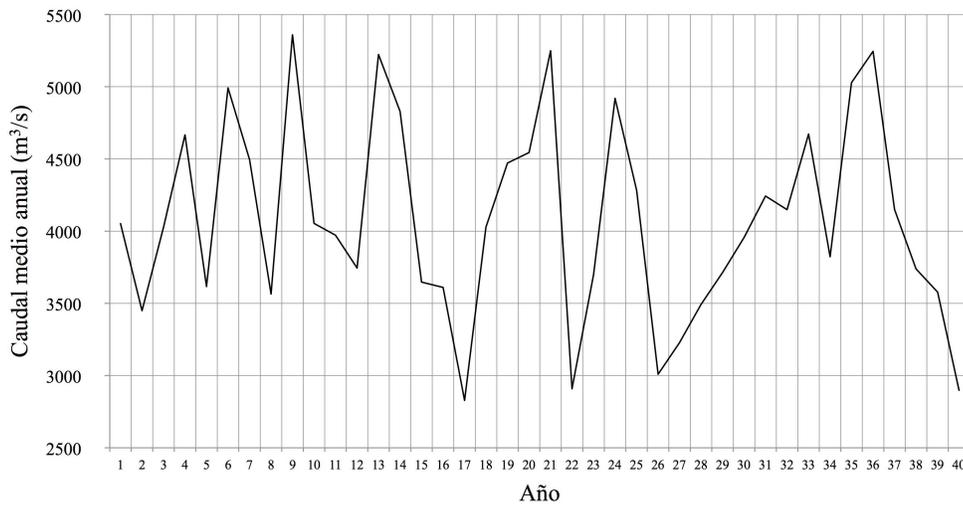


Figura 5.12: Serie de caudales medios anuales - El Banco

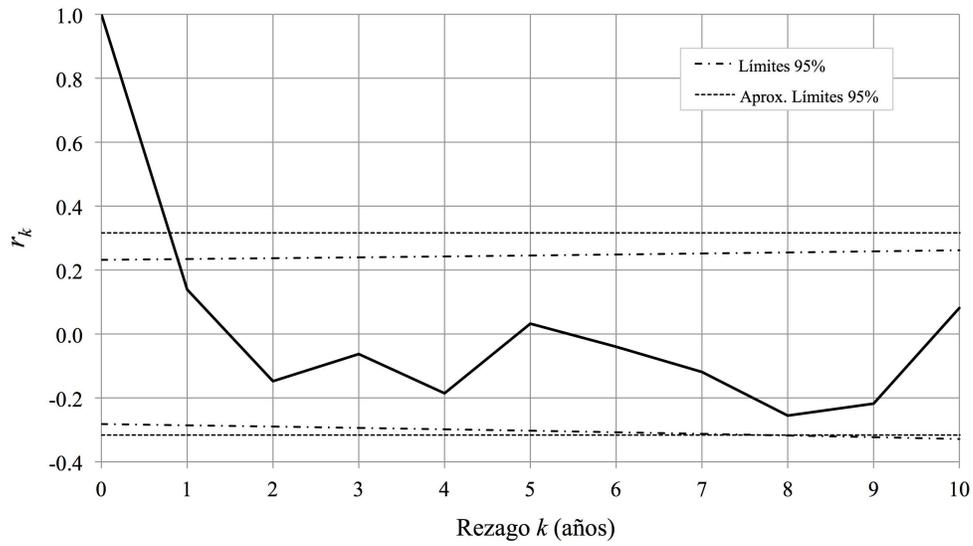


Figura 5.13: Correlograma anual - El Banco

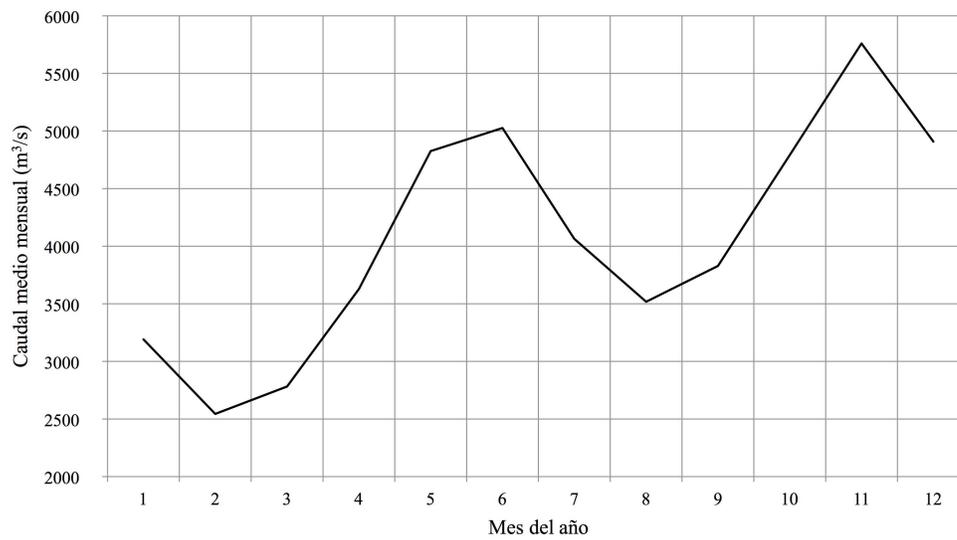


Figura 5.14: Caudales mensuales multianuales - El Banco

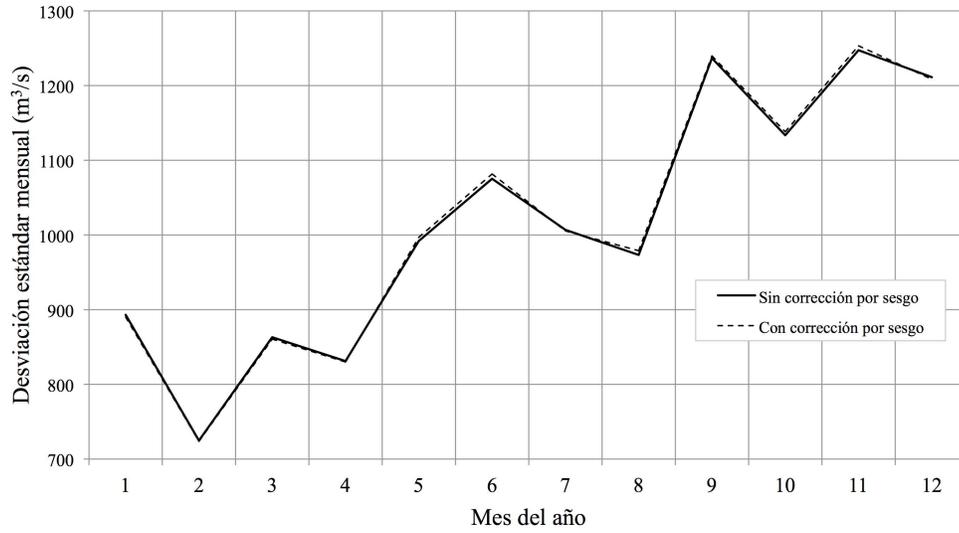


Figura 5.15: Desviaciones estándar mensuales multianuales - El Banco

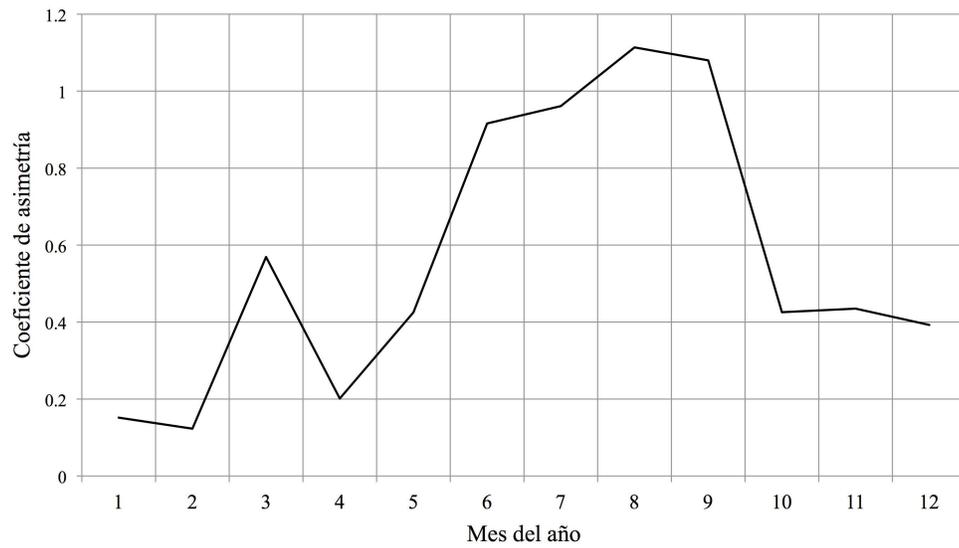


Figura 5.16: Coeficientes mensuales de asimetría - El Banco

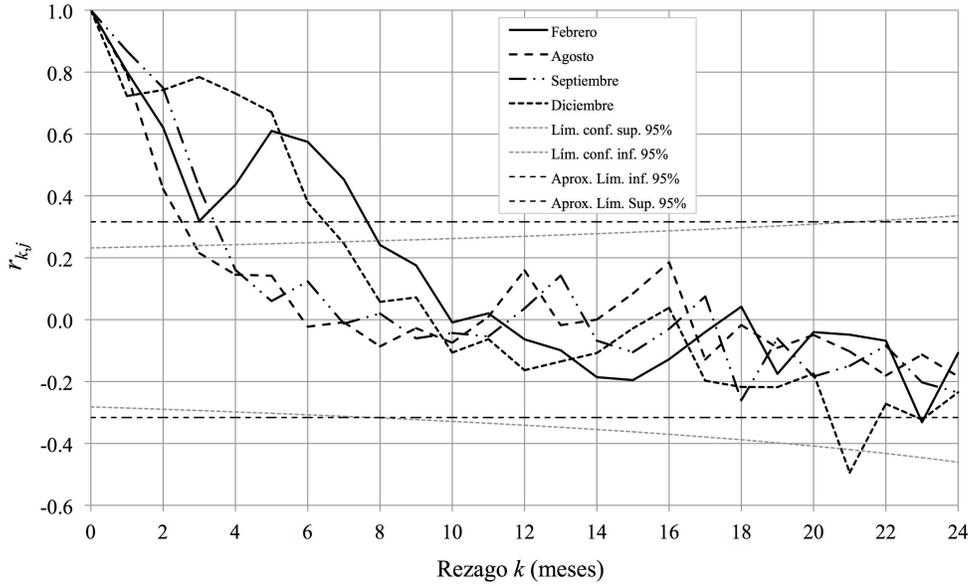


Figura 5.17: Correlogramas mensuales - El Banco

$$w_l = y_1 + y_2 + \dots + y_l \quad (5.159)$$

con $1 \leq l \leq N$, el promedio parcial \bar{y}_l

$$\bar{y}_l = \frac{w_l}{l} \quad (5.160)$$

con $1 \leq l \leq N$, y la desviación estándar parcial [Anis y Lloyd (1976)] s_l

$$s_l = \left[\frac{1}{l} \sum_{i=1}^l (y_i - \bar{y}_l)^2 \right]^{1/2} \quad (5.161)$$

también con $1 \leq l \leq N$ (algunos autores dividen por $l - 1$). Además, sea S_l^* igual a

$$S_l^* = w_l - l\bar{y}_l \quad (5.162)$$

El “rango ajustado” [Bras y Rodriguez-Iturbe (1985), Graves et al. (2017)] para $l = N$ es

$$R^*(N) = \max_{1 \leq i \leq N} \{S_i^*\} - \min_{1 \leq j \leq N} \{S_j^*\} \quad (5.163)$$

Hurst (1951) normalizó $R^*(N)$ con la desviación estándar s_N para obtener el “rango ajustado reescalado”, es decir

$$R^{**}(N) = \frac{R^*(N)}{s_N} \quad (5.164)$$

cuyo atractivo es que para un período de tiempo de N años por ejemplo, $R^*(N)/s_N$ es un parámetro adimensional subrogado del volumen ideal de almacenamiento de un embalse durante ese período de tiempo N (el volumen ideal es el menor almacenamiento requerido para un período de tiempo dado,

durante el cual la salida del embalse es constante, el embalse inicia lleno y termina lleno, y no hay vertimientos de excesos). En el caso del río Nilo, esto significaba que el tamaño de un embalse, que en promedio fallase una vez cada N años en suplir el caudal medio multianual, era aproximadamente proporcional a $N^{0.72}$ [Eltahir (1996)], en contradicción con sus propios análisis teóricos que indicaban que asintóticamente debía ser $N^{0.5}$.

Dado lo anterior, Hurst aplicó el rango ajustado re-escalado a 690 series de tiempo de 75 fenómenos geofísicos como el espesor de anillos de árboles, lluvia, caudal, temperatura y presión atmosférica [Mesa y Poveda (1993), Sutcliffe et al. (2016), Graves et al. (2017)], y estableció que

$$\frac{R^*(N)}{s_N} = KN^H \quad (5.165)$$

con un valor promedio de H de 0,73, desviación estándar de 0,08 y distribución aproximadamente normal. La tendencia de las series de tiempo de fenómenos geofísicos de tener valores de $H > 0,5$ se conoce hoy en día como el “fenómeno de Hurst”, y esto contrasta con el comportamiento de series largas de números aleatorios normales estándar independientes, para las cuales el valor esperado del rango ajustado re-escalado, obtenido también por Hurst, es $E[R^*(N)] = (\pi N/2)^{0.5}$, es decir $H = 0,5$. Esta ecuación es la misma obtenida por Feller usando la teoría de movimiento browniano sin suponer normalidad [O’Connell et al. (2016)].

Por ese entonces los modelos de generación de series sintéticas de caudales, útiles para establecer capacidades requeridas de embalses, eran incipientes y poco representativos [como por ejemplo el de Hazen y Sudler de caudales Normalmente distribuidos e independientes, Graves et al. (2017)], claramente estaban en desacuerdo con la evidencia empírica de Hurst pues tenían $H = 0,5$, y por lo tanto su uso en el diseño de embalses para horizontes largos de planeamiento podría resultar en subdimensionamientos con potenciales consecuencias de inundaciones (es decir, llevarían a embalses subdimensionados cuya capacidad de almacenamiento realmente debería ser $N^{0.22}$ veces mayor).

Así pues, el rango ajustado re-escalado se constituyó en una estadística que podría dar la base de una medida de variabilidad a largo plazo en series de tiempo de variables geofísicas [O’Connell et al. (2016)]. Los resultados obtenidos por Hurst generaron múltiples investigaciones, no solo dentro de las geociencias sino también en las matemáticas y la estadística.

Desde principios de los años 50’s hasta mediados de los 60’s se plantearon diferentes posibles explicaciones del fenómeno de Hurst [Bras y Rodriguez-Iturbe (1985), Mesa y Poveda (1993), O’Connell et al. (2016), Graves et al. (2017)]:

1. Procesos con fdps inusuales con asimetrías fuertes. Sin embargo, estos tipos de fdps no representan los datos analizados por Hurst, y en general a las series geofísicas.
2. Presencia de no estacionaridad en la media del proceso. Esto efectivamente puede llevar al fenómeno, como lo mostraron Bhattacharya et al. (1983) con un proceso estacionario con estructura de dependencia temporal débil y una tendencia suave monótona. Este tipo de dependencia significa que la función de correlación ρ (ver Ecuación 5.16) decrece suficientemente rápido para garantizar convergencia de la escala de fluctuación dada por la Ecuación 5.18 [Mesa y Poveda (1993)].
3. Transitoriedad o comportamiento pre-asintótico debido a series no suficientemente largas, donde la convergencia a $H = 0,5$ es lenta e inalcanzable por la longitud de éstas.
4. Efectos de autocorrelación a corto plazo (o falta de independencia en el proceso) que pueden llevar a comportamientos pre-asintóticos. Sin embargo, Mandelbrot y sus colaboradores [Graves et al. (2017)] mostraron más tarde que cualquier forma de autocorrelación conocida en ese entonces asintóticamente convergía a un comportamiento con $H = 0,5$.

La explicación relacionada con una fuerte dependencia temporal o memoria infinita (cuando $\theta = \infty$) se concretó a finales de los 60’s con Mandelbrot, Wallis y Van Ness [Mandelbrot y Van Ness (1968),

Mandelbrot y Wallis (1968)], quienes con base en el Movimiento Browniano Fraccionado (MBF, definido como una derivada fraccionada del movimiento browniano estándar), propusieron el denominado Ruido Gaussiano Fraccionado (RGF), proceso estacionario que preserva el coeficiente de Hurst y cuya función de correlación se presentó previamente (ver Ecuación 5.33) [Bras y Rodríguez-Iturbe (1985), Mesa y Poveda (1993)]. Mandelbrot introdujo el concepto de auto-similaridad, de tal forma que si $Y(t)$ es un proceso estocástico continuo, éste es auto similar, con parámetro H de auto-similaridad, si para $c > 0$, $Y(ct) \stackrel{d}{=} c^H Y(t)$, donde el símbolo $\stackrel{d}{=}$ indica identidad de distribución. Esto significa que un re-escalamiento temporal del proceso luce estadísticamente idéntico al original [Mesa et al. (2012), Graves et al. (2017)]. Cuando $H > 0,5$ los incrementos de MBF son de memoria infinita. El coeficiente H está relacionado con la denominada dimensión fractal D , que es una medida de auto-similaridad. En el caso de una serie de tiempo, esta relación es $D = 2 - H$, con $1 \leq D \leq 2$, siendo D una medida de la lisura de la serie; si una serie de tiempo sigue un movimiento browniano estándar, $D = 1,5$, pero si $1 < D < 1,5$ la serie muestra persistencia y mayor lisura o suavidad.

Posteriormente, Hosking, Granger y Joyeaux [Granger y Joyeux (1980), Granger (1980), Hosking (1981)] introdujeron una función de correlación asintótica de ley de potencia $\rho(\tau) \propto \tau^{2d-1}$ en el modelo FARIMA (ver Sección 5.6.1.7), donde d es un parámetro de diferencia fraccionado tal que $-0,5 \leq d \leq 0,5$; si $0 < d < 0,5$ el proceso es estacionario con fuerte dependencia temporal, pero con capacidad de inclusión de propiedades de corta y larga memoria. Por comparación con la Ecuación 5.33, $d = H - 0,5$.

Es de anotar que no existe explicación física para la ocurrencia de memoria infinita en series geofísicas, por lo cual son plausibles aproximaciones como la de Boes y Salas (1978) en la cual, en un proceso estacionario de memoria corta con media fluctuante aleatoriamente a lo largo del tiempo, el proceso resultante en si es estacionario, aunque una realización puede parecer no estacionaria y mostrar características pre-asintóticas, o como la antes mencionada de Bhattacharya et al. (1983), por ejemplo. Mesa y Poveda (1993) indican que el fenómeno de Hurst en series de tiempo geofísicas probablemente sea el resultado de una mezcla de escalas más que consecuencia de memoria infinita. La discusión sigue siendo válida aún hoy en día. Inclusive, se han presentado metodologías como la propuesta por Poveda y Mesa (1991) y Mesa y Poveda (1993) en la cual, con base en el teorema funcional del límite central y la escala de fluctuación, y en diagramas denominados “GEOS” (N vs. $R^{**}(N)/N^{0,5}$) y “GEOS-h” (N vs. $R^{**}(N)/N^h$, $h > 0,5$), se muestra que la mayoría de las 23 series usadas por Hipel y McLeod (1978) no exhibe el fenómeno de Hurst. En la Sección 5.6 se presentan algunos modelos estocásticos para la generación sintética de series hidrológicas de tiempo que ilustran la variedad de aproximaciones y sus características de persistencia.

5.5.1.6. Estimación del Coeficiente de Hurst

Existen diferentes procedimientos para estimar el coeficiente H . El estimador original de rango ajustado re-escalado R/S fue el propuesto por Hurst (1951), el cual fue posteriormente modificado [Mesa y Poveda (1993), Graves et al. (2017)]. Otros métodos temporales son el análisis de fluctuación con eliminación de tendencias [DFA, por “Detrended Fluctuation Analysis”, Weron (2002), Benmehdi et al. (2011), Mesa et al. (2012)], las varianzas agregadas [Koutsoyiannis (2003), Chronopoulou y Viens (2010), Benmehdi et al. (2011), Ceballos y Largo (2017)] y la escala de fluctuación [Mesa y Poveda (1993)]. Dentro de los métodos espectrales están procedimientos basados en el periodograma [Weron (2002), Benmehdi et al. (2011), Wu (2020)], el variograma [Wu (2020)], el estimador basado en onditas [“wavelets”, Chronopoulou y Viens (2010), Wu (2020)], entre otros. En este libro se describen los métodos R/S , de varianzas agregadas y de periodograma.

Método R/S : el procedimiento comprende los pasos que se describen a continuación [Weron (2002), Chronopoulou y Viens (2010), Kale y Butar (2011), O’Connell et al. (2016), Ceballos y Largo (2017)].

1. Dividir la serie de tiempo de longitud N en k subseries de longitud $l_k = \lfloor N/k \rfloor$, donde $\lfloor \cdot \rfloor$ es la función piso que determina el mayor entero menor que N/k ;

2. Para cada subserie calcular el rango ajustado re-escalado $R^*(l_k)/s_{l_k}$ de acuerdo con las Ecuaciones 5.159 a 5.164;
3. Calcular el promedio $\overline{[R^*(l_k)/s_{l_k}]}$ de los k valores de $R^*(l_k)/s_{l_k}$ y asociarlo con la longitud de serie l_k ;
4. Repetir los pasos (1) a (3) para valores de $k = 1, \dots, m$ donde $m = \lfloor N/d \rfloor$, siendo d la longitud mínima de subserie considerada adecuada para calcular el rango ajustado re-escalado;
5. En escala log-log graficar l_k vs. $\overline{[R^*(l_k)/s_{l_k}]}$ con $k = 1, \dots, m$ [esta gráfica fue denominada diagrama “pox” por Mandelbrot y Wallis (1969)] y verificar si existe comportamiento lineal;
6. Hacer la regresión potencial y determinar el valor de H como exponente resultante.

De acuerdo con Anis y Lloyd (1976) y Weron (2002), para valores pequeños de k existe sesgo en el estimador $R^*(l_k)/s_{l_k}$, y se han propuesto correcciones, como por ejemplo [Ceballos y Largo (2017)]

$$\overline{[R^*(l_k)/s_{l_k}]_*} = \overline{[R^*(l_k)/s_{l_k}]} - E[(R/S)_k] + E[(R/S)_k]_* \tag{5.166}$$

donde $\overline{[R^*(l_k)/s_{l_k}]_*}$ es el estimador corregido, y

$$E[(R/S)_k] = \begin{cases} \left(\frac{k-0,5}{k}\right) \left(\frac{\Gamma[k/2-0,5]}{\sqrt{\pi}\Gamma[k/2]}\right) \sum_{i=1}^{k-1} \left(\frac{k-i}{i}\right)^{0,5} & k \leq 340 \\ \left(\frac{k-0,5}{k}\right) \left(\frac{2}{\pi k}\right)^{0,5} \sum_{i=1}^{k-1} \left(\frac{k-i}{i}\right)^{0,5} & k > 340 \end{cases} \tag{5.167}$$

y

$$E[(R/S)_k]_* = \sqrt{0,5\pi k} \tag{5.168}$$

donde $\Gamma[\cdot]$ es la función Gamma. El valor condicionante en la Ecuación 5.167 de 340 es el valor de k que hace que los valores de $E[(R/S)_k]$ sean iguales con ambas fórmulas.

Varianzas Agregadas: el análisis de las varianzas agregadas se basa en la relación asintótica de la varianza muestral de valores promedio del proceso agregado en bloques sin traslapo obtenidos de la serie original para diferentes niveles de agregación [Montanari et al. (1997), Ceballos y Largo (2017)], y que queda sintetizado en el climacograma [Koutsoyiannis (2003), ver Secciones 5.2.4 y 5.2.5]

La función de varianza $\gamma(T)$ define la reducción de la varianza puntual σ^2 del proceso continuo $Y(t)$, por promediar éste sobre un intervalo T (ver Ecuación 5.21), y que genera el proceso continuo $Y_T(t)$ y el discreto $Y_i^{(T)}$ (ver Secciones 5.2.4 y 5.2.5). Una forma de expresar esta reducción es mediante el climacograma que es la gráfica log-log del intervalo T vs. $\gamma_T(T)$, donde $\gamma_T(T) = \sigma^2\gamma(T)$ (ver Ecuación 5.30). Así, el climacograma permite analizar la variabilidad del proceso a través de diferentes escalas temporales, donde la variabilidad se cuantifica en términos de varianza.

Sea y_t una serie de tiempo con $t = 1, 2, \dots, N$, donde la escala de tiempo es $k = 1$ y T es el nivel original de agregación temporal (por ejemplo, esta serie puede ser horaria, diaria, mensual, anual, etc.). Para esta escala de tiempo ($k = 1$), la media \bar{y}_1 y la varianza s_1^2 corresponden a (ver Ecuaciones 1.55 y 1.61)

$$\bar{y}_1 = \frac{1}{N-1} (y_1 + y_2 + \dots + y_N) \tag{5.169}$$

y

$$s_1^2 \equiv \hat{\gamma}_{(1)} = \frac{1}{N} \sum_{t=1}^N (y_t - \bar{y}_1)^2 \quad (5.170)$$

donde el subíndice 1 representa la escala original.

Para otras escalas $k = 2, 3, \dots, m$ de agregación progresiva, y donde $m = \lfloor N/10 \rfloor$, siendo $\lfloor \cdot \rfloor$ la función piso que determina el mayor entero menor que $N/10$ (se considera que hasta con m datos se puede calcular adecuadamente la varianza del proceso agregado), se definen las correspondientes series de tiempo asociadas, cuyos valores se calculan como

$$y_i^{(k)} = \frac{1}{k} \sum_{t=(i-1)k+1}^{ik} y_t \quad (5.171)$$

con $i = 1, \dots, \check{N}_k$ con $\check{N}_k = \lfloor N/k \rfloor$.

La varianza para la serie de orden k es

$$s_k^2 \equiv \hat{\gamma}_{(k)} = \frac{1}{\check{N}_k - 1} \sum_{i=1}^{\check{N}_k} (y_i^{(k)} - \bar{y}_1)^2 \quad (5.172)$$

El climacograma empírico es la gráfica log-log de k vs. $\hat{\gamma}_{(k)}$ con $k = 1, 2, \dots, m$. En el caso de un proceso de muestreo discreto $\check{Y}_i^{(T)}$ (ver Sección 5.2.6), el climacograma se calcula de forma similar.

Si el climacograma empírico tiene un comportamiento que decae linealmente con k (ver Figura 5.3) un ajuste de ley de potencia (Ecuación 5.35) con propiedades invariantes de escala puede ser plausible [Koutsoyiannis (2003)], es decir

$$\gamma_{(k)} = c_{(k)}^2(H) k^{2-2H} \sigma^2 \quad (5.173)$$

donde σ^2 es la varianza del proceso discreto y $c_k(H)$ es un factor corrección por sesgo para la varianza (que depende de H , descriptor de la dependencia temporal del proceso, y del tamaño N de la serie) dado por

$$c_{(k)}(H) = \left[\frac{N/k - (N/k)^{2H-1}}{N/k - 1/2} \right]^{1/2} \quad (5.174)$$

Para estimar \hat{H} y $\hat{\sigma}$ a partir de la serie de tiempo, Koutsoyiannis [Koutsoyiannis (2003), Tyralis y Koutsoyiannis (2011)] propone un procedimiento que consiste en minimizar el error de ajuste entre el climacograma empírico (Ecuación 5.172) y el modelo de ley de potencia (Ecuación 5.173), es decir

$$e^2(\hat{\sigma}, \hat{H}) = \sum_{k=1}^m \frac{\left[\ln \hat{\sigma} + (\hat{H} - 1) \ln k + \ln c_k(\hat{H}) - \ln \hat{\gamma}_{(k)}^{1/2} \right]^2}{k^p} \quad (5.175)$$

siendo $m = \lfloor N/10 \rfloor$ y donde el peso $1/k^p$ disminuye con k para valores de $p > 0$. Para $p = 0$ el peso es igual para todo k , y para un valor fijo de k , entre mayor sea el valor de p , menor es el correspondiente peso. El uso del peso se justifica debido a que entre mayor sea el valor de k el tamaño de muestra es menor para calcular $\gamma_{(k)}$, y por lo tanto la incertidumbre es mayor.

Cuando se compara el climacograma teórico dado por la Ecuación 5.173 (con \hat{H} y $\hat{\sigma}$, estimados con el procedimiento descrito) y el climacograma empírico, éstos no necesariamente coinciden pues hay un sesgo presente (debido al tamaño N de la serie de tiempo) que debe ser restado al primero [Koutsoyiannis (2021)]. Este sesgo se puede determinar analíticamente partiendo de la expresión del valor esperado de $\gamma_T(T)$, es decir

$$E[\gamma_T(T)] = \frac{1}{N} E \left[\sum_{t=1}^N \{(Y_t^T - \mu) - (\hat{\mu} - \mu)\}^2 \right] \quad (5.176)$$

donde $\gamma_T(T)$ es el estimador de la varianza del proceso promedio Y_t^T en intervalos disyuntos T (ver Sección 5.2.5) durante un tiempo total NT , μ es el valor esperado del proceso estocástico y $\hat{\mu}$ es su estimador obtenido al considerar solo una duración NT . Este último está dado por

$$\hat{\mu} = \frac{Y(NT)}{NT} \quad (5.177)$$

donde $Y(t)$ es el proceso continuo $Y(\xi)$ acumulado entre 0 y t (ver Ecuación 5.40).

Al desarrollar la Ecuación 5.176 se llega a [Koutsoyiannis (2021)]

$$\begin{aligned} E[\gamma_T(T)] &= \gamma_T(T) - \gamma_T(NT) \\ &= \left[1 - \frac{\gamma_T(NT)}{\gamma_T(T)} \right] \gamma_T(T) = \left(1 - \frac{1}{N'} \right) \gamma_T(T) \end{aligned} \quad (5.178)$$

donde

$$\gamma_T(NT) = \text{Var} \left[\frac{Y(NT)}{NT} \right] \quad (5.179)$$

y $N' = \gamma_T(T)/\gamma_T(NT)$, y representa el tamaño equivalente o efectivo de la serie de tiempo si ésta fuera independiente, teniendo en cuenta el sesgo generado por la dependencia temporal que realmente tiene la serie.

La Ecuación 5.178 escrita en términos de la escala k es

$$E[\gamma(k)] = \gamma(k) - \gamma(N) = \left[1 - \frac{\gamma(N)}{\gamma(k)} \right] \gamma(k) = \left(1 - \frac{1}{N'} \right) \gamma(k) \quad (5.180)$$

Diferencias de Varianzas Agregadas: como ya se ha mencionado, series sin estructura de dependencia a largo plazo pero con cambios en la media o tendencias decrecientes suaves pueden mostrar valores de $\hat{H} > 0,5$, calculados con procedimientos de estimación como el Método R/S o el de varianzas agregadas [Bhattacharya et al. (1983)]. Teverovsky y Taquq (1997) propusieron un método para detectar dependencia a largo plazo en series de tiempo aún con la presencia de los anteriores tipos de cambios y tendencias, mediante un estimador basado en la diferencia de varianzas agregadas, es decir

$$\Delta s_k^2 = s_k^2 - s_{k-1}^2 \quad (5.181)$$

cuyo logaritmo se relaciona con H como

$$\ln[\Delta s_k^2] = (2H - 2) \ln k + c \quad (5.182)$$

En consecuencia, una gráfica log-log de k vs. Δs_k^2 debería esperarse con un comportamiento lineal con pendiente igual a $2H - 2$. Según Montanari et al. (1997) y Montanari et al. (2000), debido a que la diferencia de varianzas agregadas incrementa la dispersión de esta gráfica, su pendiente se debe estimar con la porción central mediante mínimos cuadrados. De acuerdo con Montanari et al. (1999), este método es recomendable usarlo junto con el de varianzas agregadas.

Método del Periodograma: usando los valores de la serie de tiempo y_t se puede calcular el periodograma según [Montanari et al. (2000)]

$$I(\lambda_l) = \frac{1}{2\pi N} \left| \sum_{t=1}^N y_t e^{jt\lambda_l} \right|^2 \quad (5.183)$$

donde N es el tamaño de la serie y $\lambda_l = 2\pi l/N$ es la frecuencia, con $l = 1, \dots, N^*$, siendo $N^* = \lfloor N/2 - 1/2 \rfloor$. El periodograma es un estimador de la densidad espectral, y por lo tanto una serie con dependencia de largo plazo debería exhibir un periodograma proporcional a $|\lambda_l|^{1-2H}$ en las cercanías al origen. En consecuencia, la pendiente de una regresión lineal entre $\ln(\lambda_l)$ y $\ln[I(\lambda_l)]$ debería ser $1 - 2H$, de donde se puede estimar el valor de H . Montanari et al. (2000) sugieren usar el 5% menor de las N^* frecuencias para hacer esta regresión.

Ejemplo 5.2 El Ejemplo 2.1 se basa en 662 datos de la serie anual de niveles mínimos en el río Nilo durante el período comprendido entre los años 622 y 1284 A.C. Una serie más larga (años 622 a 1470 A.C., es decir 849 datos), también publicada por Toussoun, fue complementada con siete años faltantes dispersos por Koutsoyiannis (2013) y está disponible en ITIA (2013). En la Figura 5.18 se grafica esta serie anual.

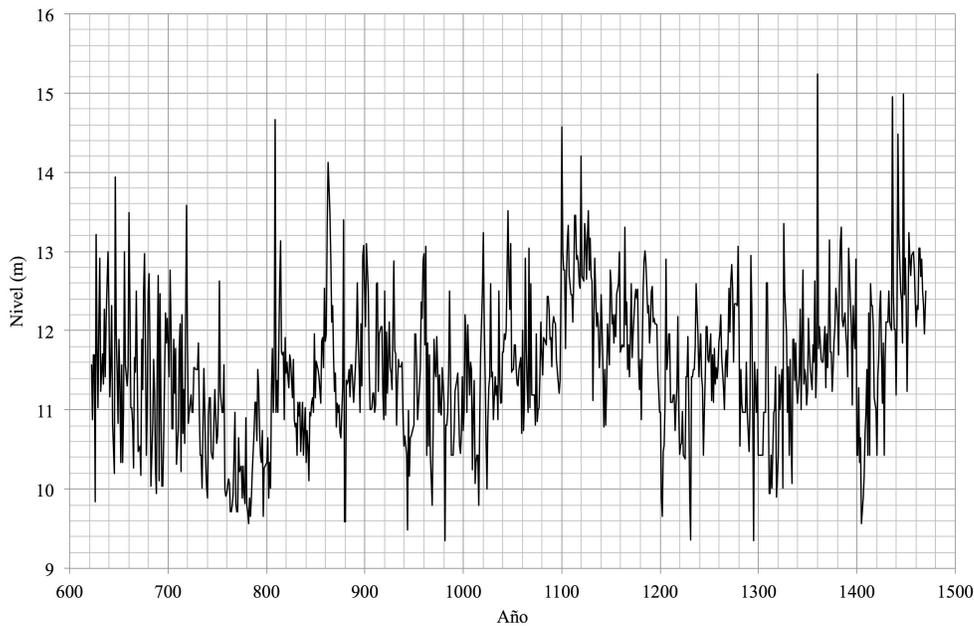


Figura 5.18: Niveles mínimos anuales - Río Nilo

El método R/S se aplicó a la serie anual de niveles mínimos considerando que $d = 42$, de tal manera que $m = \lfloor N/d \rfloor = \lfloor 849/42 \rfloor = 20$. Usando el procedimiento descrito en la Sección 5.5.1.6, en la Tabla 5.1 se resumen los cálculos respectivos. La gráfica de l_k vs. $[R^*(l_k)/s_{l_k}]$ se presenta en la Figura 5.19, donde se puede apreciar el comportamiento lineal en escalas log-log. Esto permite ajustar una regresión potencial cuyo resultado es $[R^*(l_k)/s_{l_k}] = 0,4213l_k^{0,86916}$, con $R^2 = 0,9958$. Con base en esta regresión, el coeficiente de Hurst es $H = 0,86916$.

En la Tabla 5.1 también están los valores de $E[(R/S)_k]$ y $E[(R/S)_k]_*$, requeridos para hacer la corrección por sesgo debido a valores bajos de k definida con la Ecuación 5.162, y calculados con las Ecuaciones 5.163 y 5.165. Así, los valores corregidos $[R^*(l_k)/s_{l_k}]_*$ están en la última columna de esta tabla. Su respectiva gráfica, similar a la Figura 5.19, muestra un comportamiento igualmente lineal en