

TRANSFERENCIA DE INFORMACIÓN DE CRECIENTES MEDIANTE REGRESIÓN LINEAL MÚLTIPLE

• Daniel Francisco Campos-Aranda •
Profesor jubilado de la Universidad Autónoma de San Luis Potosí

Resumen

Los registros de gastos máximos anuales (crecientes) permiten el dimensionamiento hidrológico de las obras hidráulicas de protección y de cruce. Lógicamente, entre mayores sean las series disponibles, más confiables serán sus estimaciones probabilísticas o predicciones. Por tal razón, en este trabajo se describe con detalle la técnica de transferencia de información de crecientes, mediante regresión lineal múltiple, para ampliar registros cortos con base en las series largas cercanas, revisando la conveniencia estadística de dicha transferencia. Se expone la formulación matemática de manera simple, al utilizar la solución matricial y se desarrolla un ejemplo numérico para ampliar la serie de la estación hidrométrica Platón Sánchez del río Tempoal en el estado de Veracruz, México, utilizando diversos registros cercanos. Por último, se formulan las conclusiones, las cuales destacan la sencillez del procedimiento y sugieren su aplicación sistemática.

Palabras clave: media, variancia, coeficiente de correlación múltiple, aleatoriedad, test de Shapiro-Wilk, río Tempoal.

Introducción

La planeación, diseño y manejo de los proyectos de recursos hidráulicos se basa en los registros históricos climáticos e hidrológicos, tales como precipitaciones, evaporaciones, escurremientos líquidos y sólidos, y crecientes. Desafortunadamente, los registros históricos son comúnmente de corta extensión y por ello pueden fallar en proporcionar una estimación confiable de la variabilidad del fenómeno hidrológico estudiado (Salas *et al.*, 2008).

Por lo anterior, resulta importante buscar fuentes de datos adicionales y técnicas de extensión de los registros que permitan obtener series más largas. Las fuentes comunes de información para ampliar registros cortos son otros registros largos cercanos, o bien series amplias de precipitación disponibles dentro de la zona estudiada. Respecto a los procedimientos desarrollados para extender registros

cortos, la mayoría son estadísticos basados en la correlación que puede existir entre las series utilizadas; presentando dos objetivos al realizar la extensión, uno busca mejorar la estimación de los parámetros poblacionales (media y variancia) del registro corto y el otro intenta mantener o preservar sus estimaciones (Salas *et al.*, 2008).

Tomando en cuenta que comúnmente los registros cortos conducen a estimaciones inexactas de los parámetros poblacionales, el propósito de toda extensión de un registro consiste en mejorar la confiabilidad de sus estimaciones, a través de reducir su variabilidad (Gilroy, 1970); ya que una estimación θ'_1 de un parámetro estadístico θ se dice que es más confiable que otra θ'_2 de θ , si $E(\theta'_i) = \theta$, con $i = 1, 2$ y $\text{Var}(\theta'_1) < \text{Var}(\theta'_2)$.

El objetivo de este trabajo consiste en exponer con detalle la técnica de transferencia de información de crecientes (gastos máximos anuales), mediante regresión lineal múltiple,

para ampliar registros cortos con base en las series largas cercanas, revisando la conveniencia estadística de dicha transferencia. Se presenta la formulación matemática de manera simple al utilizar la solución matricial y se desarrolla un ejemplo numérico en la cuenca del río Tempoal, de la Región Hidrológica 26 (Pánuco), de México, para ampliar el registro de la estación hidrométrica Platón Sánchez, empleando diversos registros cercanos de tal cuenca.

Desarrollo

Modelo matemático

La regresión lineal múltiple es aplicada para transferir información de gastos máximos anuales (crecientes) a una localidad con registro corto, cuando existen dos o más sitios cercanos que tienen registros largos. Considerando que la serie corta tiene una extensión de n_1 años, es representada por la variable dependiente y y se tienen p registros de tamaño $n_1 + n_2$ representados por el vector x ; es decir, que se tiene el siguiente arreglo:

$$\begin{matrix} y_1, y_2, \dots, y_{n_1} \\ x_1^1, x_2^1, \dots, x_{n_1}^1, x_{n_1+1}^1, \dots, x_{n_1+n_2}^1 \\ x_1^2, x_2^2, \dots, x_{n_1}^2, x_{n_1+1}^2, \dots, x_{n_1+n_2}^2 \\ \vdots \\ x_1^p, x_2^p, \dots, x_{n_1}^p, x_{n_1+1}^p, \dots, x_{n_1+n_2}^p \end{matrix}$$

en el cual, las observaciones en los diferentes instantes o puntos en el tiempo son independientes. Además, como es lo usual en estadística, se acepta que todos los gastos proceden de una población Normal multivariada, o bien que ellos pueden ser transformados para seguir tal distribución, con parámetros: μ_x^i , μ_y , σ_x^{2i} , σ_y^2 y R , en donde μ_x^i y σ_x^{2i} son los valores poblacionales de la media y la variancia de x^i , respectivamente, para $i = 1, 2, \dots, p$; μ_y y σ_y^2 son las magnitudes poblacionales de la media y variancia de y , y R es el coeficiente

de correlación múltiple poblacional (Gilroy, 1970; Salas, 1993).

Entonces, para transferir información de crecientes desde las p estaciones hidrométricas con registros de tamaño $n_1 + n_2$ al sitio con la serie corta y , y además mejorar las estimaciones de sus parámetros μ_y y σ_y^2 , se usa el siguiente modelo de regresión múltiple que relaciona a y_t con los p registros de x_t (Gilroy, 1970; Salas, 1993):

$$\hat{y}_t = \hat{b}_0 + \sum_{i=1}^p \hat{b}_i \cdot x_t^i + (1 - \hat{R}^2)^{1/2} \alpha \cdot \theta \cdot [s_1(y)]^{1/2} \cdot \varepsilon_t \quad (1)$$

para el cual, la constante \hat{b}_0 y los coeficientes \hat{b}_i se obtienen mediante el arreglo matricial siguiente (Campos-Aranda, 2003):

$$\begin{bmatrix} n_1 & \sum_{t=1}^{n_1} x_t^1 & \sum_{t=1}^{n_1} x_t^2 & \sum_{t=1}^{n_1} x_t^3 & \dots & \sum_{t=1}^{n_1} x_t^p \\ \sum_{t=1}^{n_1} x_t^1 & \sum_{t=1}^{n_1} (x_t^1)^2 & \sum_{t=1}^{n_1} x_t^1 \cdot x_t^2 & \sum_{t=1}^{n_1} x_t^1 \cdot x_t^3 & \dots & \sum_{t=1}^{n_1} x_t^1 \cdot x_t^p \\ \sum_{t=1}^{n_1} x_t^2 & \sum_{t=1}^{n_1} x_t^2 \cdot x_t^1 & \sum_{t=1}^{n_1} (x_t^2)^2 & \sum_{t=1}^{n_1} x_t^2 \cdot x_t^3 & \dots & \sum_{t=1}^{n_1} x_t^2 \cdot x_t^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{t=1}^{n_1} x_t^p & \sum_{t=1}^{n_1} x_t^p \cdot x_t^1 & \sum_{t=1}^{n_1} x_t^p \cdot x_t^2 & \sum_{t=1}^{n_1} x_t^p \cdot x_t^3 & \dots & \sum_{t=1}^{n_1} (x_t^p)^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_p \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{n_1} y_t \\ \sum_{t=1}^{n_1} x_t^1 \cdot y_t \\ \sum_{t=1}^{n_1} x_t^2 \cdot y_t \\ \vdots \\ \sum_{t=1}^{n_1} x_t^p \cdot y_t \end{bmatrix} \quad (2)$$

es decir:

$$X \cdot \underline{b} = Y \quad \therefore \quad \underline{b} = X^{-1} \cdot Y \quad (3)$$

Las otras variables de la ecuación (1) asociadas con el término de ruido son (Salas, 1980):

$$R = \left[\frac{s_1(y) - Se^2}{s_1(y)} \right]^{1/2} \quad (4)$$

ya definido como el coeficiente de correlación múltiple, siendo $s_1(y)$ la dispersión de la variable dependiente y Se^2 es la suma de los cuadrados de los residuos, es decir (Campos-Aranda, 2003):

$$s_1(y) = \frac{1}{n_1} \sum_{i=1}^{n_1} (y_i - \bar{y}_1)^2 \quad (5)$$

con:

$$\bar{y}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} y_i \quad (6)$$

$$Se^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (y_i - \hat{y}_i)^2 \quad (7)$$

El factor α es el requerido para obtener una estimación insesgada de σ_y ; su expresión es (Salas, 1980):

$$\alpha = \left[\frac{n_2(n_1 - 2p - 2)(n_1 - 1)}{(n_2 - 1)(n_1 - p - 2)(n_1 - p - 1)} \right]^{1/2} \quad (8)$$

El factor θ es igual a la unidad si el término de ruido es considerado, de otra manera es cero; caso común cuando se transfiere información de crecientes. Finalmente, ε_i es una variable Normal con media cero y variancia unitaria (Salas, 1980, 1993).

Criterios para mejorar los parámetros

Al aplicar la ecuación (1) para extender el registro corto y , se obtiene la secuencia:

$$y_1, y_2, y_3, \dots, y_{n_1}, \hat{y}_{n_1+1}, \hat{y}_{n_1+2}, \dots, \hat{y}_{n_1+n_2}$$

a partir de la cual se obtienen la nueva media y la variancia $[\bar{y}, s(y)]$. Estos nuevos parámetros

tendrán menor variancia y por lo tanto serán mejores estimaciones muestrales de los valores poblacionales si el coeficiente de correlación múltiple (R) excede ciertos límites. Para la media, su valor crítico está dado por (Gilroy, 1970; Salas, 1980, 1993):

$$R > \left[\frac{p}{n_1 - 2} \right]^{1/2} \quad (9)$$

en las columnas (1) del cuadro 1 se muestran los valores críticos de R para diversas magnitudes de n_1 y de p . Para la variancia, el valor crítico de R tiene una expresión que es función de n_1 , n_2 , θ y p (Gilroy, 1970; Moran, 1974; Salas, 1993). Sin embargo, cuando $\theta = 0$, prácticamente no varía con n_2 , por ello en las columnas (2) del cuadro 1 se presentan sus valores críticos para $n_2 = 60$.

Descripción del sistema de río Tempoal

Al río Tempoal lo forman los ríos Hules y Calabozo, cuyas cuencas de drenaje inician en la frontera del bajo Pánuco (Región Hidrológica 26 Parcial), en los estados de Hidalgo y Veracruz; están aforados por las estaciones hidrométricas Los Hules y Terrerillos, respectivamente. El río Tempoal tiene un recorrido de sur a norte y es uno de los colectores más importantes del río Moctezuma, al que se le une por la margen derecha en el poblado de El Higo, Veracruz. Antes de la estación hidrométrica Tempoal, última del sistema, le llega por margen izquierda el río San Pedro, aforado en la estación El Cardón. Finalmente, cerca del poblado de Platón Sánchez, en Veracruz, está la estación hidrométrica del mismo nombre sobre el río Tempoal. En la figura 1 se muestra la ubicación del sistema del río Tempoal.

Información hidrométrica disponible y su procesamiento

Las áreas de cuenca de las cinco estaciones de aforos citadas: El Cardón, Los Hules,

Cuadro 1. Valores críticos del coeficiente de correlación múltiple (R) para mejorar las estimaciones de media (1) y variancia (2).

n_1	$p = 1$		$p = 2$		$p = 3$		$p = 4$		$p = 5$	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
10	0.354	0.650	0.500	0.763	0.612	0.841	0.707	0.899	0.791	0.945
12	0.316	0.597	0.447	0.707	0.548	0.785	0.632	0.845	0.707	0.892
14	0.289	0.556	0.408	0.661	0.500	0.739	0.577	0.799	0.645	0.847
16	0.267	0.522	0.380	0.624	0.463	0.700	0.535	0.759	0.598	0.808
18	0.250	0.494	0.354	0.592	0.433	0.666	0.500	0.725	0.559	0.774
20	0.236	0.469	0.333	0.565	0.408	0.637	0.471	0.695	0.527	0.744
25	0.209	0.422	0.295	0.510	0.361	0.578	0.417	0.634	0.466	0.681
30	0.189	0.386	0.267	0.469	0.327	0.533	0.378	0.587	0.423	0.632
35	0.174	0.359	0.246	0.436	0.302	0.498	0.348	0.548	0.389	0.592
40	0.162	0.336	0.229	0.410	0.281	0.468	0.324	0.517	0.363	0.559
45	0.152	0.317	0.216	0.387	0.264	0.443	0.305	0.490	0.341	0.531
50	0.144	0.301	0.204	0.368	0.250	0.422	0.289	0.467	0.323	0.506
55	0.137	0.288	0.194	0.352	0.238	0.404	0.275	0.447	0.307	0.485
60	0.131	0.276	0.186	0.338	0.227	0.387	0.263	0.429	0.294	0.466

Terrerillos, Platón Sánchez y Tempoal son, respectivamente: 609, 1 269, 1 493, 4 700 y 5 275 km². El registro de gastos máximos anuales de la estación Tempoal comienza en 1954 y los del resto en 1960, teniendo disponibles tales series hasta 2002 en el sistema BANDAS (IMTA, 2003). En el cuadro 2 se presenta la información que será procesada para completar el registro de la estación hidrométrica Platón Sánchez en el periodo de 1960 a 1977 ($n_2 = 18$ años).

Respecto a los datos faltantes de las estaciones Los Hules (1990 y 1991) y Terrerillos (1981), se intentó deducirlos con base en una regresión entre ambos registros, pero el diagrama de dispersión indica que no existe tal relación. Los datos faltantes de la estación El Cardón (1998 y 2000) tampoco se pudieron ser estimados con cierta garantía. Por lo anterior, el periodo común (n_1) queda definido en veinte años, en el periodo 1978 a 2002, con los cinco años citados excluidos.

Todos los registros disponibles (cuadro 2), tanto de su periodo completo como de 1978 a 2002, se probaron para verificar su aleatoriedad, ausencia de componentes determinísticas y procedencia de una población

Normal. Se aplicaron las pruebas de persistencia, de Kendall, de von Neumann, de Cramer (Campos-Aranda, 1992) y de Shapiro-Wilk (Shapiro y Wilk, 1965; Ruiz-Maya, 1977). Todas las pruebas se realizaron con un nivel de significancia del 5% (95% de confiabilidad).

Se encontró que todos los registros presentan aleatoriedad, es decir, son valores independientes, no presentan componentes determinísticas y no proceden de poblaciones Normales. Debido a esto último, las regresiones que se analizan se realizaron con base en los valores logarítmicos de los datos, que es la situación común en los estudios de transferencia de información de crecientes. Se utilizaron logaritmos naturales.

Regresiones lineales efectuadas

Primeramente se realizaron cuatro regresiones lineales univariadas en el periodo 1978-2002 entre Platón Sánchez y cada una de las otras cuatro estaciones hidrométricas, los coeficientes de correlación (r_{xy}) obtenidos fueron: con El Cardón 0.6761, con Los Hules 0.7881, con Terrerillos 0.8193 y con Tempoal 0.9556. Estos

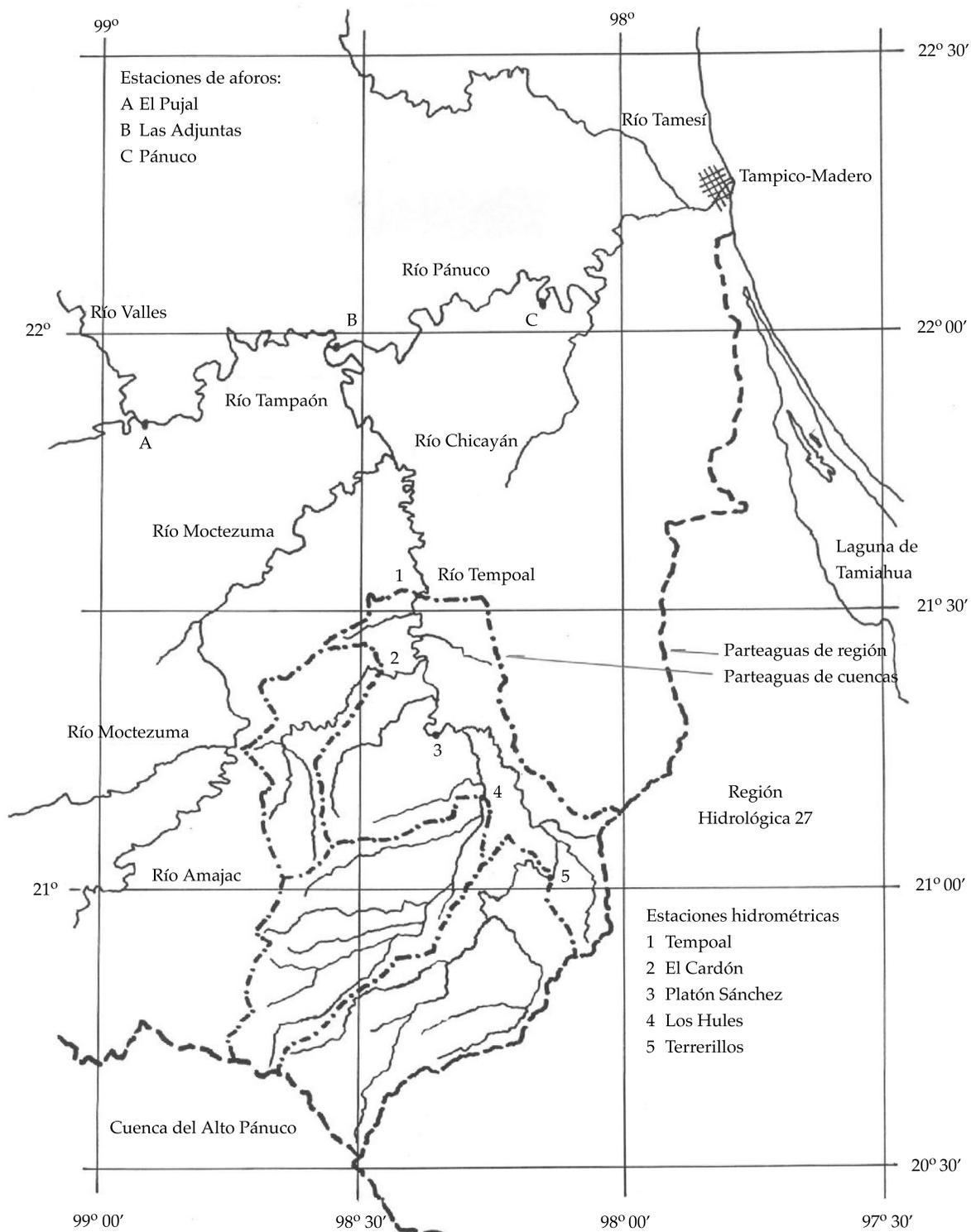


Figura 1. Ubicación del sistema del río Temporal, en la Región Hidrológica 26 Parcial (Bajo Río Pánuco).

Cuadro 2. Gastos máximos anuales (m³/s) en las estaciones hidrométricas de la cuenca del río Tempoal en el periodo común de 1960 a 2002.

Núm.	Año	El Cardón	Los Hules	Terrerillos	P. Sánchez	Tempoal
1	1960	1 080.0	452.6	314.0	–	1 277.0
2	1961	303.5	434.5	525.0	–	852.9
3	1962	262.0	457.5	565.9	–	739.2
4	1963	481.0	947.4	895.9	–	1 800.0
5	1964	188.6	258.0	397.1	–	748.0
6	1965	338.0	414.9	659.4	–	792.7
7	1966	287.0	742.2	1 121.7	–	1 778.0
8	1967	854.2	1 009.4	1 153.0	–	2 245.0
9	1968	476.0	1 096.0	611.2	–	1 145.0
10	1969	555.8	825.0	2 224.2	–	1 948.0
11	1970	560.0	800.0	1 420.0	–	1 418.0
12	1971	720.4	1 064.0	1 488.5	–	1 630.0
13	1972	320.0	1 110.0	529.0	–	989.0
14	1973	392.0	749.0	1 740.0	–	1 668.0
15	1974	1 198.3	1 950.0	3 187.8	–	4 950.0
16	1975	1 204.2	2 470.0	2 085.0	–	4 040.0
17	1976	419.7	937.7	1 000.5	–	1 275.0
18	1977	179.1	559.0	291.2	–	514.0
19	1978	1 390.0	2 874.0	2 152.3	2 898.0	3 725.0
20	1979	667.0	1 082.0	659.1	1 040.0	1 655.9
21	1980	357.0	583.2	994.1	976.0	1 162.0
22	1981	765.2	1 650.3	–	1 940.0	2 020.0
23	1982	182.3	340.0	491.4	589.8	539.6
24	1983	269.8	544.0	768.4	827.3	868.0
25	1984	572.0	2 834.9	2 981.0	4 530.0	4 030.0
26	1985	457.0	938.4	1 487.7	1 608.0	1 882.0
27	1986	192.0	308.0	434.0	462.0	476.0
28	1987	346.8	1 440.0	2 635.0	1 773.0	1 765.0
29	1988	356.0	4 350.0	3 710.0	3 653.0	3 265.0
30	1989	306.0	644.0	2 100.0	653.0	649.0
31	1990	306.0	–	702.0	4 115.0	1 611.0
32	1991	1 248.0	–	2 860.0	1 916.0	3 532.0
33	1992	790.0	762.8	1 607.5	1 494.9	2 291.0
34	1993	865.5	1 684.1	3 422.5	4 380.0	6 120.0
35	1994	412.0	723.8	1 237.9	1 153.8	1 133.0
36	1995	412.2	568.0	531.0	537.0	741.9
37	1996	218.0	804.0	507.6	758.0	683.0
38	1997	348.2	428.4	362.5	1 217.5	905.0
39	1998	–	260.9	1 605.9	1 259.3	1 266.9
40	1999	602.9	630.9	3 328.3	2 776.6	2 693.7
41	2000	–	84.9	753.4	580.4	641.2
42	2001	498.3	278.5	1 512.2	1 201.3	1 847.9
43	2002	134.0	496.7	822.2	774.8	926.4

Nota: los años y datos sombreados no fueron utilizados.

resultados son consistentes, ya que las crecientes que mide la estación El Cardón (*EC*) no tienen relación con las de Platón Sánchez (*PS*); por el contrario, la estación Tempoal (*TP*) está aguas abajo, de cuya área de cuenca la estación con registro corto afora el 89.1%. En el cuadro 3 se presentan las estimaciones de gasto máximo anual en *PS* con esta última estación, cuya ecuación de regresión resultó:

$$PS = 0.4899 + 0.9186 \cdot TP \quad (10)$$

Las siguientes regresiones probadas y sus coeficientes de correlación múltiple fueron: (1) Terrerillos (*TR*) y Tempoal con *PS*, obteniéndose 0.9583, y (2) Los Hules (*LH*) y Tempoal con *PS*, conduciendo a 0.9626, con la expresión siguiente:

$$PS = 0.3171 + 0.1605 \cdot LH + 0.7944 \cdot TP \quad (11)$$

En el cuadro 3 se tienen las estimaciones de gasto máximo anual en *PS* con la ecuación (11).

Finalmente se realizaron otras dos regresiones múltiples, la primera con tres estaciones auxiliares y la segunda con las cuatro. Los resultados fueron:

$$PS = 0.2543 + 0.1472 \cdot LH + 0.0842 \cdot TR + 0.7330 \cdot TP \quad (12)$$

$$PS = 0.6162 - 0.2092 \cdot EC + 0.1325 \cdot LH + 0.0449 \cdot TR + 0.9076 \cdot TP \quad (13)$$

cuyos coeficientes *R* son 0.9640 y 0.9693, respectivamente. Nuevamente en el cuadro 3 se exponen las ampliaciones de registro en *PS* con estos dos modelos de regresión.

Análisis de los resultados

La estación hidrométrica Tempoal, por su ubicación aguas abajo de la estación Platón Sánchez, es una opción muy conveniente para

realizar la transferencia de información, como lo demuestra su coeficiente de correlación (r_{xy}), con un valor de 0.9556, el cual es muy superior a los mínimos requeridos para mejorar la estimación de la media y la variancia, que son 0.236 y 0.469, respectivamente, según el cuadro 1.

Conforme se utiliza más información de las estaciones de aforos cercanas la transferencia mejora, como lo demuestra el valor cada vez mayor de *R*, hasta el máximo alcanzado de 0.9693, obtenido al utilizar las cuatro estaciones auxiliares (ecuación (13)). Para este caso, los valores críticos de *R* son 0.471 y 0.695, mismos que son superados ampliamente, por lo cual es conveniente la transferencia de información.

Por otra parte, con base en los resultados del cuadro 4 mostrados en sus columnas 3 y 5, se concluye que los parámetros estadísticos de la estación de aforos Platón Sánchez se

Cuadro 3. Resultados de la transferencia de información de crecientes en la estación hidrométrica Platón Sánchez en el periodo de 1960 a 1977 con los modelos de regresión indicados.

Año	Modelo de regresión:			
	Ec. (10)	Ec. (11)	Ec. (12)	Ec. (13)
1960	1 164.6	1 075.4	973.8	824.1
1961	803.8	775.3	751.9	758.4
1962	704.8	697.7	686.5	693.9
1963	1 596.3	1 590.3	1 525.1	1 540.6
1964	712.5	642.5	617.8	685.4
1965	751.5	726.1	721.5	696.7
1966	1 578.4	1 514.4	1 485.9	1 660.0
1967	1 955.5	1 914.8	1 848.8	1 702.8
1968	1 053.5	1 136.5	1 083.0	1 026.3
1969	1 716.5	1 656.2	1 709.4	1 642.4
1970	1 282.2	1 280.6	1 298.3	1 199.8
1971	1 457.3	1 497.4	1 505.5	1 344.2
1972	920.9	1 013.7	962.8	971.7
1973	1 488.4	1 441.6	1 473.3	1 498.6
1974	4 042.9	3 988.6	3 961.7	3 713.2
1975	3 354.7	3 525.3	3 410.4	3 122.9
1976	1 162.9	1 207.2	1 193.7	1 163.4
1977	504.8	539.9	512.3	538.5

Cuadro 4. Parámetros estadísticos en la estación hidrométrica Platón Sánchez de su registro original y del ampliado con las cuatro estaciones de aforos cercanas (ecuación (13)).

Parámetro estadístico:	Periodo de registro en años			
	1978-2002	1978-2002	1960-2002	1960-2002
Número de datos	20	25	38	43
Media	1 665.2	1 724.6	1 528.6	1 579.0
Coefficiente de variación	0.762	0.734	0.712	0.703
Coefficiente de asimetría	1.291	1.177	1.539	1.431
Coefficiente de Curtosis	3.843	3.476	4.649	4.245
Coefficiente de correlación serial de orden uno	-0.112	-0.179	0.012	-0.048

conservan con la transferencia, disminuyendo ligeramente la media y el coeficiente de variación, y aumentando sensiblemente los coeficientes de asimetría y curtosis. Además, se conserva la escasa persistencia negativa, es decir, que la serie extendida también muestra una secuencia de valores grandes seguidos por magnitudes reducidas.

Conclusiones

La transferencia de información de gastos máximos anuales mediante regresión lineal múltiple para ampliar un registro corto con base en las series largas cercanas es una técnica sencilla y su conveniencia estadística puede ser verificada fácilmente. El uso de esta técnica permite obtener registros con un periodo común, lo cual es necesario para aplicar ciertos procedimientos de los análisis regionales de estimación de crecientes.

Además, el registro ampliado garantiza una mejor estimación de los parámetros estadísticos (media y variancia) poblacionales, lo cual conducirá a una mayor exactitud de los análisis probabilísticos orientados a obtener las predicciones o avenidas de diseño. Por lo anterior, se recomienda su aplicación sistemática antes de iniciar un análisis de frecuencia de crecientes, ya sea con métodos locales o regionales.

Recibido: 08/07/09
Aprobado: 21/10/10

Referencias

- CAMPOS-ARANDA, D.F. Estudios de homogeneidad en 34 estaciones pluviométricas del altiplano Potosino. *XII Congreso Nacional de Hidráulica (AMH)*. Tema 5: Investigación y Tecnología, Ponencia 5.29, Puerto Vallarta, Jalisco, México, del 7 al 10 de octubre de 1992.
- CAMPOS-ARANDA, D.F. *Introducción a los Métodos Numéricos: Software en Basic y Aplicaciones en Hidrología Superficial*. Capítulo 5: Ajuste de Curvas. San Luis Potosí, México: Librería Universitaria Potosina, 2003, pp. 93-127.
- GILROY, E.J. Reliability of a variance estimate obtained from a sample augmented by multivariate regression. *Water Resources Research*. Vol. 6, No. 6, 1970, pp. 1595-1600.
- IMTA. *Banco Nacional de Datos de Aguas Superficiales (BANDAS)*. Hidrometría y Sedimentos hasta 2002. 8 CD's. Jiutepec, México: CNA-SEMARNAP-IMTA, 2003.
- MORAN, M.A. On estimators obtained from a sample augmented by multiple regression. *Water Resources Research*. Vol. 10, No. 1, 1974, pp. 81-85.
- RUIZ-MAYA, L. *Métodos Estadísticos de Investigación*. Capítulo 9: Condiciones paramétricas del análisis de varianza. Madrid: Instituto Nacional de Estadística, 1977, pp. 233-249.
- SALAS, J.D. Transfer of information to improve estimates of flood frequencies. Chapter 17. *Hydrology for Transportation Engineers*. Thomas G. Sanders (editor). Washington, D.C.: Federal Highway Administration, 1980, pp. 592-653.
- SALAS, J.D. Analysis and modeling of hydrologic time series. Chapter 19. *Handbook of Hydrology*. David R. Maidment (editor in chief). New York: McGraw-Hill, Inc., 1993, pp. 19.1-19.72.

SALAS, J.D., RAYNAL, J.A., TARAWNEH, Z.S., LEE, T.S., FREVERT, D. and FULP, T. Extending short record of hydrologic data. Chapter 20. *Hydrology and Hydraulics*. Vijay P. Singh (editor). Highlands Ranch, USA: Water Resources Publications, 2008, pp. 717-760.

SHAPIRO, S.S. and WILK, M.B. An analysis of variance test for normality (complete samples). *Biométrica*. Vol. 52, No. 3-4, 1965, pp. 591-611.

Abstract

CAMPOS-ARANDA, D.F. *Transfer of flood information through multiple linear regression. Water Technology and Sciences, formerly Hydraulic engineering in Mexico (in Spanish). Vol. II, No. 3, July-September, 2011, pp. 239-247.*

Maximum annual flow records (floods) are used for hydrologic dimensionality of hydraulic structures designed for protection and crossing. Logically, the longer the available series, the more reliable their probabilistic estimates or predictions. Thus, this work describes in detail the procedure for flood information transfer through multiple linear regression. Short records are enlarged based on longer nearby series, examining the statistical advantage of said transfer. The mathematical formulation is presented in a simple way using a matrix solution and a numerical example is developed to enlarge the series at the Platón Sánchez hydrometric station on the Tempoal River in the state of Veracruz, using several nearby records. Lastly, the conclusions are formulated, which point out the simplicity of the procedure and suggest its systematic application.

Keywords: mean, variance, multiple correlation coefficient, randomness, Shapiro-Wilk test, Tempoal river.

Dirección institucional del autor

Dr. Daniel Francisco Campos-Aranda

Profesor jubilado de la Universidad Autónoma de San Luis Potosí
Genaro Codina núm. 240
Col. Jardines del Estadio
78280 San Luis Potosí, S.L.P., MÉXICO
campos_aranda@hotmail.com