

Recibido: 22.06.2022 • Aceptado: 10.10.2023

Palabras clave: Momentos L, distribución GVE, covariables, error estándar de ajuste, distribuciones estacionarias GVE_1 y GVE_2



Análisis de frecuencias con datos hidrológicos extremos no estacionarios

DANIEL FRANCISCO CAMPOS ARANDA
campos_aranda@hotmail.com
PROFESOR JUBILADO DE LA UASLP

Las estaciones hidrométricas de los ríos estiman el gasto (m^3/s) máximo instantáneo diario y con el mayor de cada año se integra el registro de crecientes o avenidas máximas. Tal serie o muestra estadística se procesa de manera probabilística a través del llamado análisis de frecuencias, para obtener las crecientes de diseño (CD) o gastos máximos anuales asociados a bajas probabilidades de ser excedidos. Con las CD se dimensiona y se brinda seguridad hidrológica a los diques y muros de defensa de las planicies de inundación, a los puentes carreteros y del ferrocarril, así como a todas las obras del drenaje urbano. Cuando en la cuenca o área de terreno que drena a la estación hidrométrica, ocurren urbanizaciones o deforestaciones, los gastos observados tienden a ser mayores y el registro de crecientes mostrará una tendencia ascendente. Ahora tal serie o muestra es no estacionaria.

El análisis de frecuencias permite estimar eventos de diseño de variables aleatorias como crecientes, niveles de ríos, sequías, vientos y lluvias máximas. Con tales eventos se dimensionan hidrológicamente las obras de infraestructura hidráulica, consta de cinco etapas: 1) recopilación de datos y verificación de su calidad estadística; 2) selección de un modelo probabilístico o función de distribución de probabilidades acumuladas (FDP); 3) aplicación de uno o varios métodos de estimación de los parámetros de ajuste de la FDP; 4) adopción de la FDP que mejor representa al registro disponible y 5) con base en la FDP seleccionada, se estiman las predicciones buscadas o eventos de diseño (Khaliq *et al.*, 2006).

La validez y exactitud de los resultados del análisis de frecuencias depende, de manera preponderante, de las dos suposiciones clásicas del registro: que sus datos son independientes y que sus propiedades estadísticas no cambian en el tiempo o condición de estacionario. En la práctica, la independencia se pierde cuando el muestreo de los datos es próximo o cercano y entonces un valor puede estar ligado con el anterior. En registros anuales, tal problema prácticamente no ocurre. Lo no estacionario se origina por el cambio climático global o regional o por los efectos de las actividades humanas en la cuenca, principalmente los aprovechamientos hidráulicos, o bien, los cambios de uso del suelo, sobre todo la deforestación y la urbanización (Khaliq *et al.*, 2006).

Por lo general y de acuerdo con la zona geográfica, el cambio climático puede generar tendencias sutiles ascendentes o descendentes en los registros de datos hidrológicos extremos. Las tendencias fuertes ascendentes de los registros de crecientes anuales y de precipitación máxima diaria (PMD) anual, se originan por la urbanización ocurrida en la cuenca o en las inmediaciones de la estación pluviométrica y la tendencia severa descendente se asocia en los registros de crecientes con los aprovechamientos hidráulicos y en las series de PMD con las deforestaciones o la desecación de lagos, ambas cercanas.

Solución del problema

El ajuste de una FDP no estacionaria implica dos supuestos básicos: 1) se acepta que la no estacionariedad de las series de datos hidrológicos anuales es causada por cambios graduales del entorno geográfico o por el cambio climático global, de esta forma se genera una alteración ligera de sus parámetros estadísticos y 2) se acepta que la FDP es independiente del tiempo, entonces una distribución con parámetros de ajuste variables con el tiempo o alguna otra covariable es aceptable para modelar datos extremos no estacionarios (Katz, 2013).

Planteamiento con la distribución GVE

La extensión de la teoría estadística de valores extremos, al caso de registros hidrológicos no estacionarios, ha seguido varios enfoques descritos por Khaliq *et al.* (2006). Uno de esos enfoques, quizás el más simple, aplica la FDP clásica de esta teoría, la distribución general de valores extremos (GVE) con tres parámetros (u , a , k), permitiendo un ajuste o traslado gradual al introducir el tiempo t como una

covariable en su parámetro de ubicación u , conservando constantes los otros dos.

Varios autores han establecido una nomenclatura para estas FDP no estacionarias, por ejemplo, Adlouni y Ouarda (2008) definen la función GVE estacionaria como GVE_0 , la que tiene su parámetro de ubicación variable linealmente con el tiempo ($u=\delta_1+\delta_2 \cdot t$) como GVE_1 y cuando la variación es curva, es decir, cuadrática ($u=\delta_1+\delta_2 \cdot t+\delta_3 \cdot t^2$) es la GVE_2 . En el modelo GVE_{11} varían linealmente con el tiempo los parámetros de ubicación y de escala.

En los modelos citados, otra covariable que se ha utilizado es algún indicador de la variabilidad climática global o regional, como el índice de la oscilación del sur (SOI, por sus siglas en inglés), que cuantifica la diferencia de la presión del aire en superficie entre Darwin, Australia y Tahití en la Polinesia francesa.

Ajuste de la distribución GVE_0

El método de los momentos L (λ_i), que son combinaciones lineales de los momentos de probabilidad ponderada (β_i), se ha establecido como un procedimiento consistente y exacto de estimación de los parámetros de ajuste de las FDP utilizadas en hidrología. Su estimación comienza con el cálculo de los β_i de la muestra (b_i) insesgados que son:

$$1) \quad b_0 = \sum_{i=1}^n \frac{X_i}{n} = \bar{x}$$

$$2) \quad b_1 = \sum_{i=1}^{n-1} \frac{(n-i) \cdot X_i}{n \cdot (n-1)}$$

$$3) \quad b_2 = \sum_{i=1}^{n-2} \frac{(n-i)(n-i-1) \cdot X_i}{n \cdot (n-1)(n-2)}$$

donde X_i es el dato hidrológico extremo anual ordenado de mayor a menor y n es su número. Los momentos λ_i de la muestra son:

$$4) \quad \lambda_1 = b_0$$

$$5) \quad \lambda_2 = 2 \cdot b_1 - b_0$$

$$6) \quad \lambda_3 = 6 \cdot b_2 - 6 \cdot b_1 + b_0$$

$$7) \quad t_3 = \frac{\lambda_3}{\lambda_2}$$

El cociente de momentos L de la muestra con similitud con el de asimetría es:

$$8) \quad x(p) = u + \frac{a}{k} [1 - (-\ln p)^k]; \quad k \neq 0$$

cuyos parámetros de forma (k), escala (a) y ubicación (u) son estimados con las ecuaciones 9 a 13:

$$9) \quad c = \frac{2}{3 + t_3} - 0.63093$$

$$10) \quad k \cong 7.8590 \cdot c + 2.9554 \cdot c^2$$

$$11) \quad a = \frac{l_2 \cdot k}{(1 - 2^{-k}) \cdot \Gamma(1+k)}$$

$$12) \quad u = l_1 - \frac{a}{k} [1 - \Gamma(1+k)]$$

Fórmula de Stirling que se utilizó para estimar el valor de la función Gamma:

$$13) \quad \Gamma(\varepsilon) \cong e^{-\varepsilon} \cdot \varepsilon^{\varepsilon-1/2} \cdot \sqrt{2\pi} \cdot \left(1 + \frac{1}{12 \cdot \varepsilon} + \frac{1}{288 \cdot \varepsilon^2} - \frac{139}{51840 \cdot \varepsilon^3} - \frac{571}{2488320 \cdot \varepsilon^4} + \dots\right)$$

Ajuste de la distribución GVE_1

La generalización del método de momentos L propuesta por Adlouni y Ouarda (2008) para el ajuste de la FDP no estacionaria tipo GVE_1 , considera variable en el tiempo t la media (μ), cuya ecuación es:

$$14) E[X] = \mu_t + u \frac{a}{k} [1 - r(1+k)]$$

La inspección, como serie cronológica, del registro de datos hidrológicos extremos anuales permite definir si se adopta una tendencia de la media (μ) lineal o curva. Cuando es lineal y se introduce en la ecuación 14, se puede despejar el parámetro de ubicación (u) que ahora será variable con respecto al tiempo t , que fluctúa de 1 a n ; su expresión es:

$$15) u_t = \delta_1 + \delta_2 \cdot t - \frac{a}{k} [1 - r(1+k)]$$

De esta ecuación se intuye que un estimador $\hat{\delta}_2$ de δ_2 se puede obtener mediante una regresión lineal simple entre la variable X y la covariable t . Para ello se define o forma una nueva variable S_1 sin tendencia, con la expresión (Khaliq *et al.*, 2006):

$$16) S_1 = X - \hat{\delta}_2 \cdot t$$

Se deduce de la ecuación 16, que la nueva variable S_1 se distribuye según una FDP tipo GVE_0 con parámetros u , a y k que se estiman con las ecuaciones 9 a 13 del método de momentos L, se destaca que u (ecuación 12) es igual a δ_1 , con lo cual quedan estimados los cuatro parámetros de ajuste del modelo GVE_1 , definido por la ecuación 15, que se aplica en la 8 para estimar las predicciones buscadas.

Ajuste de la distribución GVE_2

El mismo enfoque se utiliza para introducir una dependencia curva o cuadrática en el parámetro de ubicación u , con la ecuación:

$$17) u_t = \delta_1 + \delta_2 \cdot t + \delta_3 \cdot t^2 - \frac{a}{k} [1 - r(1+k)]$$

Ahora, las estimaciones $\hat{\delta}_2$ y $\hat{\delta}_3$ de δ_2 y δ_3 se obtienen mediante una regresión lineal múltiple de la variable X contra t y t^2 . La nueva variable S_2 sin tendencia cuadrática será:

$$18) S_2 = X - \hat{\delta}_2 \cdot t - \hat{\delta}_3 \cdot t^2$$

Con las ecuaciones 9 a 13 aplicadas a la muestra de datos corregidos S^2 , se definen los tres parámetros restantes k ,

a y $\delta_1 = u$ del modelo GVE_2 , definido por la ecuación 17, que se aplica en la 8 para estimar las predicciones buscadas.

Otra distribución GVE_2

Varios autores consideran que usar una variación lineal en el parámetro de ubicación es demasiado simple para representar una muestra no estacionaria de gastos máximos anuales, ya que algunos registros exhiben mayor variabilidad en ciertos años, indicando con ello alguna influencia de factores climáticos. La distribución no estacionaria GVE_2 también se puede aplicar con dos covariables (w y h), quedando el parámetro de ubicación variable (u), según la ecuación 15, igual a:

$$19) u_t = \delta_1 + \delta_2 \cdot w + \delta_3 \cdot h - \frac{a}{k} [1 - r(1+k)]$$

Las estimaciones $\hat{\delta}_2$ y $\hat{\delta}_3$ de δ_2 y δ_3 se obtienen mediante una regresión lineal múltiple de la variable X contra w y h . La nueva variable S_3 sin tendencia o serie estacionaria es:

$$20) S_3 = X_i + \delta_2 \cdot w - \delta_3 \cdot h$$

Con las ecuaciones 9 a 13 aplicadas a la muestra de datos corregidos S_3 , se definen los tres parámetros restantes k , a y $\delta_1 = u$ del modelo GVE_2 , definido por la ecuación 19, que se aplica en la 8 para estimar las predicciones buscadas.

Error estándar de ajuste

Definido por la ecuación 21 evalúa la desviación estándar de las diferencias entre los valores observados y los estimados con la FDP que se prueba; en este caso los modelos: GVE_1 y GVE_2 .

$$21) EEA = \left[\frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{(n \cdot np)} \right]^{1/2}$$

En tal ecuación, n y np son el número de datos de la muestra y de parámetros de ajuste, ahora cuatro y cinco. X_i son los datos ordenados de menor a mayor y \hat{X}_i son los valores estimados con la solución inversa $x(p)$ o función de cuantiles (ecuación 8) que utiliza el parámetro de ubicación variable (u), para una probabilidad de no excedencia estimada con la fórmula de Weibull o ecuación 22.

$$22) P = (X \leq x) = p = \frac{m}{n+1}$$

en la cual, m es el número de orden del dato, con 1 para el menor y n para el mayor.

Planteamiento de los análisis probabilísticos

En los análisis no estacionarios de una o dos covariables, con base en la solución inversa (ecuación 8) se calculan predicciones con periodos de retorno (Tr) de 2, 25, 50 y 100 años, a través del periodo de registro, con el parámetro de ubicación u variable (ecuaciones 15, 17 o 19). La primera predicción corresponde a la mediana, ya que su probabilidad de no

excedencia (p) es del 50 por ciento y las tres siguientes se calculan para probabilidades complementarias, para definir su valor superior e inferior, es decir, para los valores siguientes: $p = 0.96$ y $p = 0.04$ para el Tr de 25 años; $p = 0.98$ y $p = 0.02$ para el Tr de 50 años y $p = 0.99$ y $p = 0.01$ para el Tr de 100 años. Además, en estos análisis se pueden hacer predicciones a futuro, con $t > n$.

Primer ejemplo numérico

La figura 1 muestra los datos y predicciones durante el periodo de registro de la PMD anual en milímetros en la estación Andong de Corea del Sur, con tendencia lineal ascendente. El registro abarca de 1973 a 2007 con cuatro años faltantes ($n = 31$). Su EEA de la GVE₁ fue de 9.8 mm. Otros resultados se pueden consultar en Campos (2018).

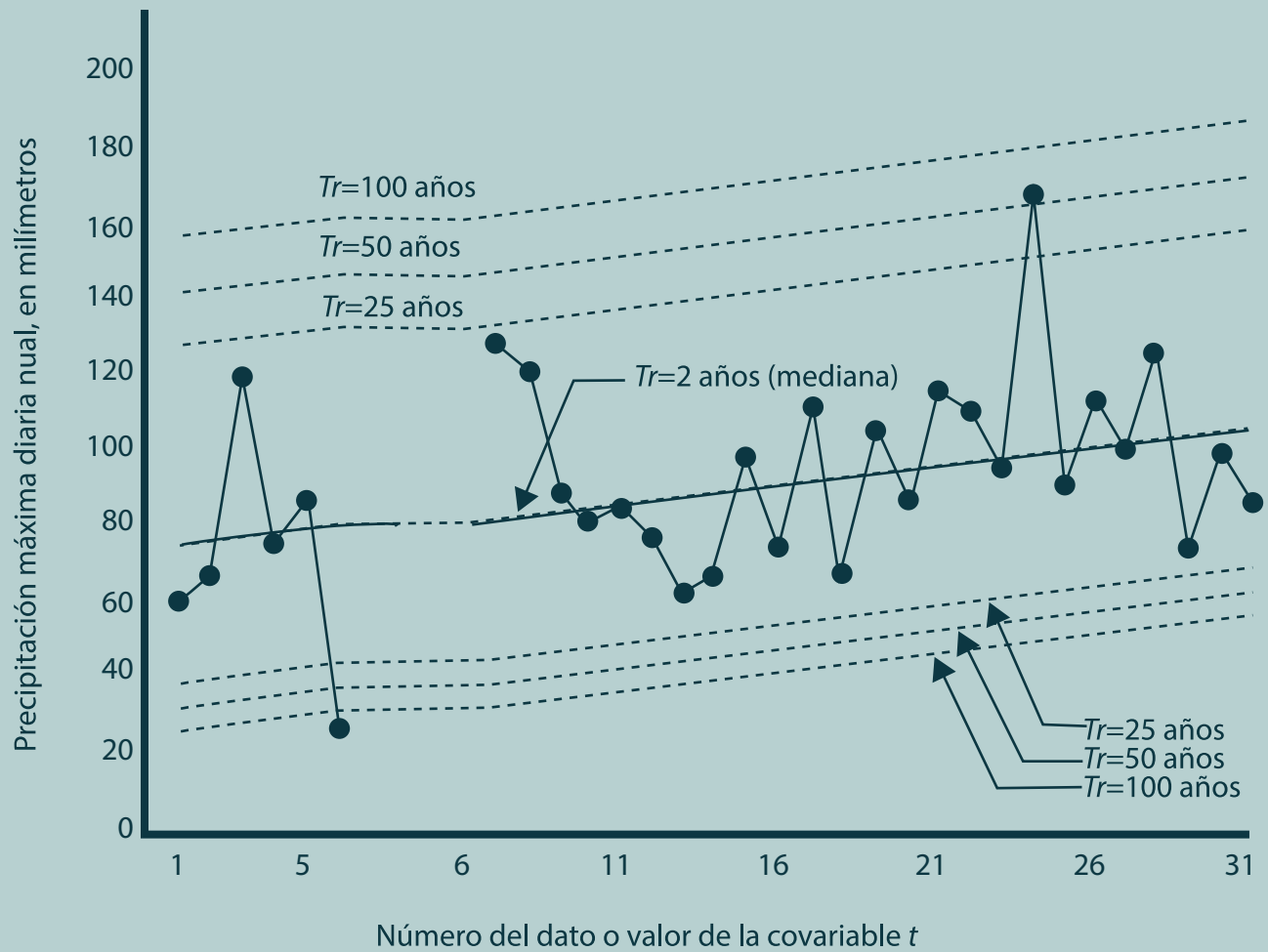


Figura 1. Diagrama de datos y de predicciones estimadas con la distribución GVE₁ en la estación pluviométrica Andong, Corea del Sur.

Segundo ejemplo numérico

En la figura 2 se muestra el registro de la estación pluviométrica Tehachapi del sur de California, EUA, de PMD anual las

ordenadas contra sus valores correspondientes del SOI en las abscisas, que fungió como covariable; se observa o deduce una tendencia descendente curva. Estos datos proceden de El Adlouni y Ouarda (2008) y otros resultados se pueden consultar en Campos (2018).

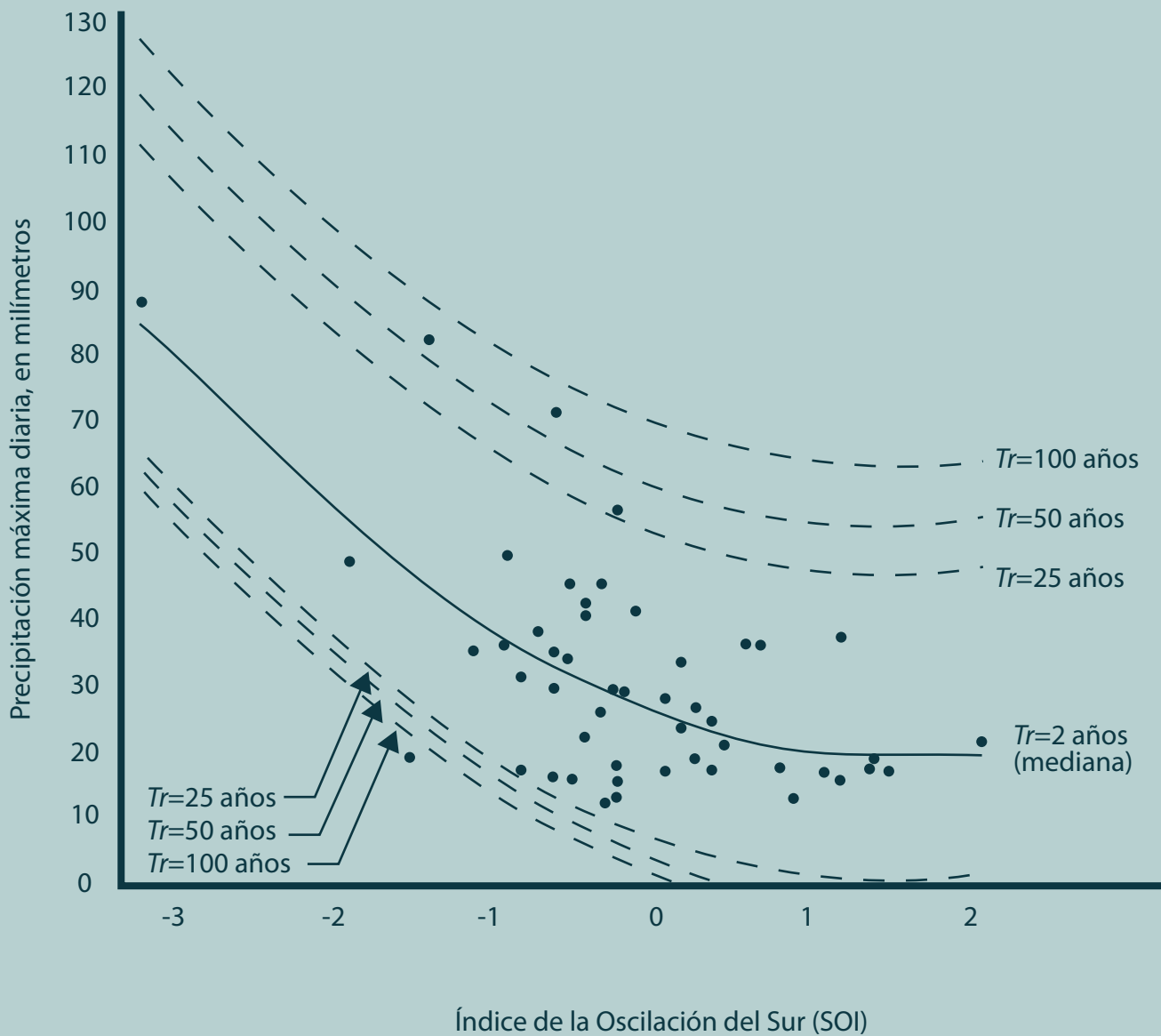


Figura 2. Nube de puntos y curvas de predicciones estimadas con la distribución GVE_2 en la estación pluviométrica Tehachapi del sur de California, EUA.

Tercer ejemplo numérico

Una cuenca urbanizada en el río Lostock de Inglaterra, tuvo un registro de 33 crecientes anuales durante los años 1974 a 2006; las cuales muestran tendencia ascendente y variabilidad mayor en su inicio. Se conoce que su extensión del área urbana (EAU) pasó de un 6.3 por ciento en 1970, a los siguientes porcentajes: 10.2 por ciento, 11.3 por ciento, 12.2 por ciento y 16.4 por ciento en los años 1980, 1990, 2000 y 2010. Además, se tienen los valores anuales de la PMD de $Tr = 100$ años en su cuenca. Estos datos disponibles en Campos (2020) se procesaron con una GVE de una y dos covariables: tiempo, PMD_{100} y EAU. Los resultados se muestran en las tabulaciones siguientes.

En la columna cuatro se observa que el indicador estadístico de la pendiente (DS) sólo es significativo con la covariable PMD_{100} . Para los ajustes con dos covariables, en la columna seis, se deduce que el último es el mejor. Al comparar las predicciones (columnas 13 a 15) de ambos métodos, se selecciona el primero por aportar magnitudes mayores.

Realizó el Doctorado en Ingeniería con especialidad en Aprovechamientos Hidráulicos en la Facultad de Ingeniería de la UNAM. Obtuvo la medalla Gabino Barreda que otorga la UNAM y el Premio Nacional Francisco Torres H. de la Asociación Mexicana de la Hidráulica. Es profesor jubilado de la UASLP.



| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------------------|------------------|------------|---------|------------|------------------|--------------|
| FDP | Covariable: | δ_2 | DS | δ_3 | r_{xy} o R^2 | $\delta_1=u$ |
| GVE ₁ | t | 0.1446 | 1.323 < | – | 0.2312 | 18.722 |
| GVE ₁ | PMD_{100} | 0.8404 | 3.059 > | – | 0.4815 | 3.147 |
| GVE ₁ | EAU | 1.0516 | 1.631 < | – | 0.2811 | 9.420 |
| GVE ₂ | t, PMD_{100} | 0.2129 | – | 0.9588 | 0.3432 | –3.109 |
| GVE ₂ | t, EAU | –0.3694 | – | 3.1851 | 0.1028 | –8.524 |
| GVE ₂ | PMD_{100}, EAU | 0.9791 | – | 1.4744 | 0.3809 | –16.158 |

| 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|-------|--------|---------------------|---------|------|--|---------|----------|
| a | k | EEA | Mediana | | Predicciones máximas (m ³ /s) | | |
| | | (m ³ /s) | mín | máx | $Tr=25$ | $Tr=50$ | $Tr=100$ |
| 4.567 | –0.015 | 1.16 | 20.5 | 25.2 | 38.5 | 41.8 | 45.2 |
| 4.525 | 0.063 | 3.45 | 18.1 | 31.5 | 43.0 | 45.5 | 47.9 |
| 4.506 | –0.013 | 1.17 | 19.3 | 26.6 | 39.6 | 42.9 | 46.2 |
| 3.896 | –0.008 | 3.55 | 17.8 | 30.9 | 41.8 | 44.5 | 47.1 |
| 4.619 | 0.019 | 1.33 | 17.8 | 27.9 | 40.5 | 43.6 | 46.5 |
| 3.991 | 0.063 | 3.49 | 17.6 | 31.9 | 42.1 | 44.3 | 46.4 |

A manera de conclusión

El primer ejemplo muestra cómo procesar un registro de *PMD* anual con pendiente lineal, usando como covariable explicativa el tiempo; en cambio, el segundo ejemplo expone un registro de *PMD* anual con tendencia curva respecto a una covariable climática. Por último, el tercer ejemplo presenta el uso de una covariable y de las combinaciones de dos covariables, para explicar la variabilidad de un registro de gastos máximos anuales.

Así que se han expuesto los fundamentos de los análisis de frecuencias no

estacionarios, que inician con la distribución GVE, con una y dos covariables. Tal método se ha extendido a otras distribuciones y ha evolucionado hacia la GVE₁₁, con parámetros de ubicación y escala con variación lineal. [UR](#)

Referencias bibliográficas:

Campos Aranda, D. F. (2018). Ajuste con momentos L de las distribuciones GVE, LOG y PAG no estacionarias en su parámetro de ubicación, aplicado a datos hidrológicos extremos. *Agrociencia*, 52(2), pp. 169-189.

Campos Aranda, D. F. (2020). Análisis de Frecuencias de Crecientes No Estacionario con una y dos covariables. *Aqua-LAC*, 12(2), pp. 47-61.

El Adlouni, S. y Ouarda, T. B. M. J. (2008). Comparaison des méthodes d'estimation des paramètres du modèle GEV non stationnaire. *Revue des Sciences de l'Eau*, 21(1), pp. 35-50.

Katz, R. W. (2013). Statistical Methods for Nonstationary Extremes, pp. 15-37. En: AghaKouchak, A, Easterling D., Hsu, K, Schubert S., Sorooshian S. (Eds.). *Extremes in a Changing Climate*. Springer. Dordrecht, The Netherlands.

Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J. C., Gachon, P. y Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, 329(3-4), pp. 534-552.