SOCIEDAD CHILENA DE INGENIERÍA HIDRÁULICA

XIX CONGRESO CHILENO DE HIDRÁULICA

ANÁLISIS DE FRECUENCIA DE EVENTOS HIDROLÓGICOS MULTIVARIADOS USANDO CÓPULAS

JACKELLINE GONZÁLEZ B.[1]
BONIFACIO FERNÁNDEZ L.[2]

RESUMEN

El análisis de frecuencia es ampliamente usado en el estudio de eventos hidrológicos tales como crecidas, tormentas y sequías. Estos se caracterizan por varias propiedades, como por ejemplo la duración, volumen y peak, por lo tanto se requieren análisis de frecuencia multivariados para representar mejor estos eventos, considerando la dependencia que existe entre dichas propiedades. Las cópulas son una herramienta nueva en el campo de la hidrología que permite construir fácilmente distribuciones multivariadas a partir de distribuciones marginales. El objetivo principal de este trabajo es dar a conocer las cópulas como una herramienta versátil en el análisis de frecuencia de eventos hidrológicos multivariados, presentando las principales características de las cópulas y sus usos en el campo de la hidrología.

[1]Profesor Titular, Carrera de Ingeniería Civil, Pontificia Universidad Católica de Valparaíso - mail: jackelline.gonzalez@ucv.cl
[2]Profesor Titular, Departamento de Ingeniería Hidráulica y Ambiental, Pontificia Universidad Católica de Chile – mail: bfernand@ing.puc.cl

1. INTRODUCCIÓN

En el desarrollo de muchos proyectos es necesario estimar las probabilidades de ocurrencia de eventos extremos que pueden influir de manera decisiva en el éxito o fracaso de las obras involucradas en el sistema. En muchos casos las condiciones de diseño están determinadas por los riesgos aceptables de falla del sistema, de manera que es necesario poder relacionar estos riesgos con las variables que definen los tamaños de las obras involucradas. Por ejemplo un embalse puede diseñarse de un tamaño tal que durante la vida útil del proyecto no se produzcan desabastecimientos para condiciones de disponibilidad predeterminadas, y que permita evacuar adecuadamente los excesos. Esto requiere de análisis de frecuencia para poder estimar las probabilidades de ocurrencia de tales eventos de sequías y crecientes.

Los eventos extremos hidrológicos, tales como tormentas severas, crecidas y sequías, se definen con varias propiedades, como por ejemplo intensidad, duración, magnitud y extensión. Por tanto, la estimación de sus probabilidades de ocurrencia conduce a estimar probabilidades multivariadas. Para abordar el tema en muchos casos se ha recurrido a estimar probabilidades de ocurrencia relacionadas con alguna de las propiedades mencionadas, despreciando la dependencia que existe entre las propiedades y dándole importancia relativa a alguna de las propiedades de las sequías, según el tipo de sistema o los objetivos de cada estudio. Esto se debe a que los análisis de frecuencia de eventos hidrológicos multivariados requieren tratamientos matemáticos sofisticados y el número de modelos disponibles es limitado, adicionalmente requieren registros históricos extensos.

En el campo de la hidrología, la probabilidad de ocurrencia de eventos extremos se asocia al periodo de retorno, siendo éste uno de los criterios más empleados en la planeación, diseño y operación de los diferentes sistemas de aprovechamiento de los recursos hídricos. Para una adecuada estimación del periodo de retorno deben considerarse estos eventos como eventos multivariados, es decir que se debe tener en cuenta la dependencia que existe entre las propiedades que los caracterizan, ya que diferentes combinaciones de las propiedades generan eventos con igual periodo de retorno. Es por esto que los análisis univariados de eventos extremos hidrológicos no son tan apropiados, ya que se sobreestima o subvalora dicho criterio (vea, por ejemplo, Salvadori y De Michele, 2004, y De Michele et al., 2004).

En la mayoría de estudios, los análisis de frecuencia multivariados han sido abordados considerando uno o más de los siguientes supuestos: 1) las variables tienen el mismo tipo de distribución marginal; 2) las variables siguen la distribución normal conjunta; y 3) las variables son independientes. En realidad, las variables son dependientes, en general no distribuyen normal, y no tienen el mismo tipo de distribución marginal.

En los últimos años, las investigaciones se han centrado en el estudio de los eventos hidrológicos extremos como eventos multivariados (e.g. Yue et al., 2001; González y Valdés, 2003; Kim et al., 2003 y 2006; Favre et al., 2004; Zhang y Singh, 2006 y 2007), debido al desarrollo tecnológico, que permite realizar tratamientos matemáticos sofisticados, y a que se ha recurrido a metodologías nuevas en el campo de la hidrología que permiten una mayor flexibilidad en la estimación de las distribuciones multivariadas.

Una de las metodologías nuevas que ha tenido gran acogida en el campo de la hidrología son las cópulas, que son funciones que unen o acoplan distribuciones marginales para formar funciones de distribución multivariadas (Sklar, 1959), donde la distribución marginal de las variables puede ser de cualquier forma. Durante las pasadas décadas han sido usadas en el campo financiero y de seguros, y en los últimos años se han empleado en el campo de la hidrología, especialmente en el estudio de tormentas y crecidas (vea Salvadori y De Michele, 2007).

El objetivo principal de este trabajo es dar a conocer las cópulas como una herramienta versátil en el análisis de frecuencia de eventos hidrológicos multivariados, para lo cual se exponen las principales características de las cópulas, los avances en el análisis de frecuencia de eventos hidrológicos multivariados y se presenta una aplicación a un caso particular de Chile.

2. CÓPULAS

Las cópulas son funciones que unen o acoplan distribuciones univariadas o marginales para formar funciones de distribución multivariadas (Sklar, 1959). Sklar probó que si H es una función de distribución acumulada conjunta multivariada con marginales $F_1, F_2, \ldots, F_n$, entonces existe una cópula C tal que

$$H(x_1, x_2, \ldots\ldots, x_n) = C[F_1(x_1), F_2(x_2), \ldots\ldots\ldots, F_n(x_n)] \tag{1}$$

La gran ventaja de esta herramienta es que la estructura de la cópula es la que reproduce la relación de dependencia entre las variables, sin importar las distribuciones marginales adoptadas. Lo que la hace más flexible y simple de implementar que las distribuciones multivariadas derivadas como una extensión inmediata de distribuciones univariadas. Por tanto, el problema de determinar la función de distribución acumulada conjunta multivariada H se resuelve determinando la cópula C que mejor reproduzca la estructura de dependencia entre las variables.

Existe una gran variedad de cópulas, pero la clase de cópulas denominada cópulas arquimedianas es de las más usadas en un amplio rango de aplicaciones, incluido los análisis hidrológicos. Las razones principales son: 1) se pueden construir fácilmente, 2) existe una gran variedad de familias de cópulas dentro de esta clase, y 3) los miembros de esta clase poseen numerosas propiedades atractivas (Nelsen, 2006), por ejemplo permiten reproducir tanto correlación positiva como negativa entre las variables.

Las cópulas arquimedianas se definen a partir de la siguiente expresión:

$$C[F_1(x_1), F_2(x_2), \ldots\ldots\ldots, F_n(x_n)] = f^{-1}\left(\sum_{i=1}^{n} f[F_i(x_i)]\right) \tag{2}$$

donde f es una función continua estrictamente decreciente y se denomina “generador de cópula”. Las principales propiedades de esta clase de cópulas son: 1) C es simétrica, es decir que $C(u,v) = C(v,u)$ para todo u y v ; y 2) C es asociativa, es decir que $C(C(u,v),w) = C(u,C(v,w))$ para todo u, v y w.

Las familias de cópulas arquimedianas más usadas corresponden a las de un parámetro, debido a que son los modelos más simples. En la Tabla 1 se presentan las familias de cópulas arquimedianas de un parámetro más empleadas en las aplicaciones hidrológicas con su respectiva función generadora de cópula.

Tabla 1. Cópulas arquimedianas de un parámetro y sus generadores

Familia	Generador, *f(t)*	Parámetro, *q*
Clayton	$t^{-q} - 1$	$q > 0$
Gumbel	$(-\ln t)^q$	$q > 1$
Frank	$-\ln\left(\frac{e^{-q\,t} - 1}{e^{-q} - 1}\right)$	$q \neq 0$
Ali-Mikhail-Haq	$\ln\left(\frac{1 - q\,(1 - t)}{t}\right)$	$-1 \leq q \leq 1$

El parámetro de estas cópulas está determinado por la medida de dependencia de las variables aleatorias. Para el caso bivariado, una de las metodologías empleadas para la estimación del parámetro consiste en que la medida de dependencia se representa por coeficientes de correlación por rango, los cuales miden el grado de dependencia monótona entre el par de variables aleatorias. Los coeficientes de correlación que se emplean son la correlación de Kendall (τ) y la de Spearman (ρ). Según Schweizer y Wolff (1981), las expresiones generales que relacionan estos coeficientes de correlación con las cópulas son:

$$r = 12\iint uvdC(u,v) - 3 \tag{3}$$

y

$$t = 4\int CdC - 1 \tag{4}$$

Genest y MacKay (1986) desarrollaron la ecuación 4 para el caso de cópulas arquimedianas, la cual se reduce a:

$$t = 1 + 4\int_0^1 \frac{f(t)}{f'(t)}dt \tag{5}$$

En la Tabla 2 se presentan las expresiones que permiten estimar el parámetro de las cópulas arquimedianas bivariadas más usadas usando el coeficiente de correlación de Kendall (τ).

El procedimiento general para construir una función de distribución acumulada conjunta multivariada mediante cópulas es el siguiente:

1. Estimar la función de distribución marginal de cada variable aleatoria mediante métodos estadísticos tradicionales, por ejemplo los gráficos cuantil-cuantil.
2. Valorar la dependencia entre las variables aleatorias, lo cual se puede realizar mediante herramientas visuales, como lo son el scatter plot, chi-plot y K-plot

Tabla 2. Cópulas arquimedianas bivariadas de un parámetro

Familia	$C(u,v)$	$t = f(q)$
Clayton	$\left(u^{-1/q}+v^{-1/q}-1\right)^{q-1}$	$t=\dfrac{1}{(2q+1)}$
Gumbel	$\exp\left(-\left[(-\ln u)^{q}+(-\ln v)^{q}\right]^{1/q}\right)$	$t=1-q^{-1}$
Frank	$-\dfrac{1}{q}\ln\left[1+\dfrac{\left(e^{-q\,u}-1\right)\left(e^{-q\,v}-1\right)}{e^{-q}-1}\right]$	$t=1-\dfrac{4}{q}\left(D_1(-q)-1\right)*$
Ali-Mikhail-Haq	$\dfrac{u\,v}{1-q(1-u)(1-v)}$	$t=\left(\dfrac{3q-2}{q}\right)-\dfrac{2}{3}\left(1-\dfrac{1}{q}\right)^{2}\ln(1-q)$

* D_1 es la función de Debey de primer orden, definida como: $D_1(q)=\dfrac{1}{q}\int_0^q\dfrac{t^q}{e^{-t}-1}dt$, $q>0$ y $D_1(-q)=D_1(q)+\dfrac{q}{2}$.

3. Seleccionar las familias de cópulas que se desean emplear.
4. Estimar los parámetros de las cópulas que realmente se pueden aplicar a los datos en estudio mediante el método de máxima verosimilitud, para el caso bivariado pueden emplearse las relaciones que existen con el τ de Kendall y el ρ de Spearman.
5. Seleccionar la cópula que mejor se ajusta a los datos, mediante métodos gráficos o tests de bondad de ajuste.

Genest y Favre (2007) presentan una completa explicación de la metodología para emplear cópulas, adicionalmente existen programas estadísticos que contienen paquetes de cópulas, por ejemplo el software estadístico R que es de libre acceso, que facilitan la construcción y selección de estos modelos (vea Yan, 2006 y 2007).

La principal desventaja que presenta esta herramienta es su flexibilidad, lo cual es su virtud para modelar la dependencia entre varias variables y su aplicabilidad en diferentes campos. La desventaja está asociada a que no se cuentan con herramientas o criterios que indiquen las familias de cópulas a emplear, por lo tanto su aplicabilidad debe verificarse en cada caso.

3. PERIODOS DE RETORNO MULTIVARIADOS

El periodo de retorno de eventos hidrológicos extremos, tales como crecidas y sequias, es uno de los criterios comúnmente usados en el diseño de estructuras hidráulicas y sistemas de aprovechamiento de los recursos hídricos. Tradicionalmente se le da importancia relativa a una de las propiedades que caracterizan estos eventos, según los objetivos de cada estudio o el tipo de sistema en estudio, y se estiman periodos de retorno univariados, con lo cual se sobreestima o subvalora dicho criterio al no considerar la correlación que existe entre dichas propiedades. Esto se ha debido esencialmente a que los análisis multivariados de eventos hidrológicos requieren tratamientos matemáticos sofisticados y el número de modelos disponibles es limitado.

En los últimos años se han ido superando las limitantes que han existido para los análisis multivariados, y se han empezado a difundir los análisis bivariados, los cuales incluyen estimaciones de periodos de retorno bivariados, como lo son el periodo de retorno condicional y conjunto. Yue y Rasmussen (2002), Shiau (2003) y Salvadori y De Michele (2004) presentan los conceptos básicos del análisis de frecuencia de eventos bivariados y la relación que existe entre los periodos de retorno univariados y bivariados.

De forma general, el periodo de retorno puede definirse como el promedio del tiempo transcurrido entre dos eventos con propiedades mayores o iguales a unos valores predefinidos, los cuales corresponden a una condición de diseño. Para el caso univariado, el periodo de retorno se expresa como:

$$T_X = \frac{1}{1 - P(X \leq x)} = \frac{1}{1 - F_X(x)} \tag{6}$$

Para el caso bivariado, podemos definir el periodo de retorno conjunto del evento (X,Y) como:

$$T(x, y) = \frac{1}{1 - P(X \leq x, Y \leq y)} = \frac{1}{1 - F(x, y)} \tag{7}$$

La anterior expresión representa los casos en que x o y o ambos son excedidos ($X > x$, o $Y > y$, o $X > x$ y $Y > y$). El periodo de retorno conjunto del evento donde x y y son excedidos ($X > x$ y $Y > y$) puede definirse como:

$$T'(x, y) = \frac{1}{1 + F(x, y) - F_X(x) - F_Y(y)} \tag{8}$$

El periodo de retorno condicional asociado al evento $X > x$ dado que $Y = y$ es:

$$T(x|y) = \frac{1}{1 - P(X \leq x|Y = y)} = \frac{1}{1 - F_{X|Y}(x|y)} \tag{9}$$

Y el periodo de retorno condicional asociado al evento $X > x$ dado que $Y \leq y$ puede definirse como:

$$T'(x|y) = T(x|Y \leq y) = \frac{1}{1 - F(x|Y \leq y)} \tag{10}$$

En Shiau (2003) se pueden observar las modificaciones que sufren estas ecuaciones de periodo de retorno al considerar series de duración parcial.

A partir de estas expresiones se están empezando a realizar aproximaciones a los casos trivariados, es decir a la estimación de periodos de retorno considerando tres variables, aprovechando las facilidades que ofrecen las cópulas (vea, e.g., Zhang y Shing, 2007; y Wong et al., 2009).

Las anteriores ecuaciones del periodo de retorno las podemos reescribir en términos de cópulas, reemplazando en las ecuaciones 7 y 8 la ecuación 1. Para los casos condicionales, las distribuciones condicionales se pueden definir como:

$$P(X \leq x | Y = y) = \frac{\partial}{\partial y} C(x, y) \tag{11}$$

Un análisis en mayor detalle del uso de las cópulas en estos análisis de frecuencia se puede observar en Salvadori y De Michele (2007).

4. APLICACIONES

Las aplicaciones de las cópulas en el campo de la hidrología se han centrado en el cálculo de probabilidades condicionales y su uso en simulaciones bivariadas, el cálculo de curvas de nivel de distribuciones conjuntas, el cálculo de periodos de retorno de eventos bivariados, y el análisis trivariado de eventos, como por ejemplo la modelación trivariada de la estructura temporal de la secuencia de tormentas.

Estas aplicaciones se han hecho principalmente a eventos de crecidas (e.g. Zhang y Singh, 2006 y 2007b; Grimaldi y Serinaldi, 2006c) y de tormentas (e.g. Zhang y Singh, 2007a y 2007b; Grimaldi y Serinaldi, 2006a; Salvadori y De Michele, 2006), y existen algunas pocas en el análisis de sequías (e.g. Shiau, 2006; Serinaldi et al., 2009; Wong, 2009)

En Chile se han realizado aplicaciones de las cópulas en el estudio de las tormentas en la ciudad de Chillán (Zegpi et al., 2008) y al estudio de las sequías en la cuenca del río Maipo (González y Fernández, 2008).

Zegpi et al. (2008) derivaron la función de probabilidad conjunta del volumen y la duración de las tormentas en la ciudad de Chillan usando cópulas arquimedianas. Estos autores señalan que las cópulas son una herramienta importante para lograr buenos resultados en la estimación de la frecuencia de eventos de escorrentía.

González y Fernandéz (2008) modelaron las distribuciones conjuntas bivariadas y trivariadas de las sequías en el río Maipo usando cópulas arquimedianas y la cópula normal. Estos autores señalan que las cópulas se aplican fácilmente en la construcción de la estructura de dependencia entre variables aleatorias correlacionadas que son tan frecuentes en la hidrología; y que las cópulas son una herramienta útil en el estudio de la asociación de las características cuantificables de las sequías. En la Figura 1 se presentan algunos de los resultados obtenidos en este estudio de sequías, los cuales permiten observar las diferencias que existen en las estimaciones para diferentes eventos.

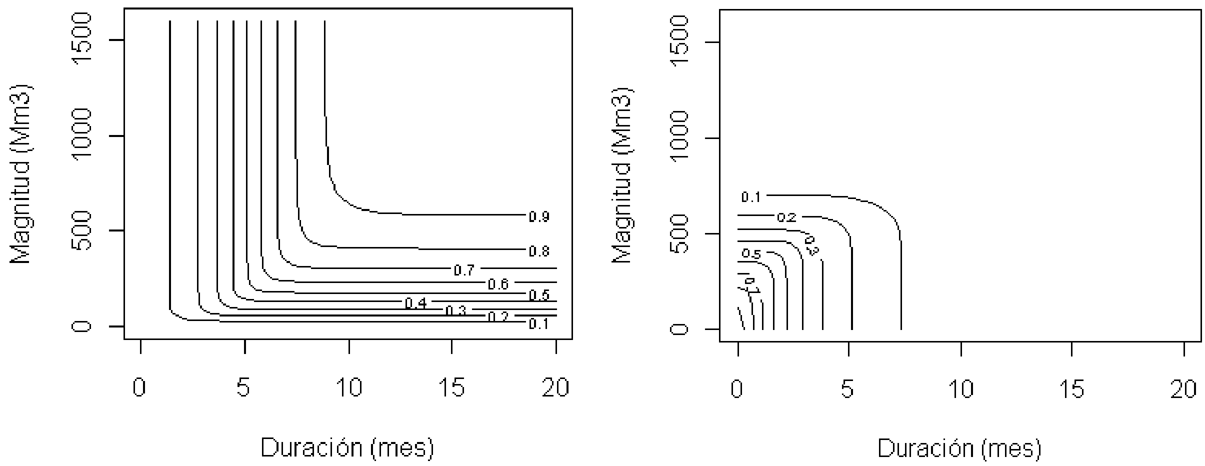

Figura 1. Función de distribución conjunta (izquierda) y función de distribución conjunta de excedencia (derecha) de la duración y la magnitud de los eventos de sequía en el río Maipo
(Fuente: González y Fernández, 2008)

5. CONCLUSIONES

En hidrología es importante el estudio de los eventos extremos, como lo son las tormentas extremas, las crecidas y las sequías, los cuales son eventos multivariados que se caracterizan por variables aleatorias correlacionadas, y para describirlas se emplean las distribuciones multivariadas. Debido al avance tecnológico, han incursionado nuevas tecnologías en el campo de la hidrología que permiten realizar análisis de frecuencia multivariados, como por ejemplo las cópulas. También ha permitido la evolución de los criterios que se emplean en el diseño de los sistemas de aprovechamiento de los recursos hídricos, como es el caso de los periodos de retorno multivariados.

En este artículo se presentaron las principales características de las cópulas, los avances en el análisis de frecuencia de eventos hidrológicos multivariados, y se describieron algunas de las aplicaciones que se han hecho en Chile. A partir de esta revisión podemos concluir y corroborar que las cópulas tienen la ventaja que son más flexibles y simples de implementar que las distribuciones multivariadas derivadas como una extensión inmediata de distribuciones univariadas, lo que la hace una herramienta versátil en el análisis de frecuencia de eventos hidrológicos multivariados.

REFERENCIAS

Beven, K. 1995. Linking parameters across scales: Sub-grid parameterizations and scale dependent hydrological models. Hydrological Processes, 9, 507-525.

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A. y Rosso, R. 2004. Bivariate statistical approach to check adequacy of dam spillway. Journal of Hydrologic Engineering, 10(1), 50-57.

Favre, A., El Adlouni, S., Perreault, L., Thiémonge, N. y Bobée, B. 2004. Multivariate hydrological frequency analysis using copulas. Water Resources Research, 40(1), W01101.

Genest, C. y Favre, A. 2007. Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering, 12(4), 347-368.

Genest, C. y MacKay, J. 1986. The joy of copulas: bivariate distributions with uniform marginals. The American Statistician, 40(4), 280-283.

González, J. y Valdés, J. 2003. Bivariate drought recurrence analysis using tree ring reconstructions. Journal of Hydrologic Engineering, 8(5), 247-258.

González, J. y Fernández, B. 2008. Análisis de frecuencia de eventos multivariados de sequía usando cópulas. XXIII Congreso Latinoamericano de Hidráulica, Cartagena de Indias, Colombia.

Grimaldi, S. y Serinaldi, F. 2006a. Design hyetograph analysis with 3-copula function. Journal Hydrological Sciences, 51(2).

Grimaldi, S. y Serinaldi, F. 2006b. Asymmetric copula in multivariate flood frequency analysis. Water Resources 29, 1155–1167.

Kim, T., Valdés, J. y Yoo, C. 2003. Nonparametric approach for estimating return periods of droughts in arid regions. Journal of Hydrologic Engineering, 8(5), 237-246.

Kim, T., Valdés, J. y Yoo, C. 2006. Nonparametric approach for bivariate drought characterization using Palmer Drought Index. Journal of Hydrologic Engineering, 11(2), 134-143.

Nelsen, R. 2006. An introduction to copulas. Springer, New York.

Salvadori, G. y De Michele, C. 2004. Frequency analysis via copulas: theoretical aspects and applications to hydrological events. Water Resources Research, 40(12), W12511.

Salvadori, G. y De Michele, C. 2006. Statistical characterization of temporal structure of storms. Advances in Water Resources, 29, 827–842.

Salvadori, G. y De Michele, C. 2007. On the use of copulas in hydrology: theory and practice. Journal of Hydrologic Engineering, 12(4), 369-380.

Schweizer, B. y Wolff, E. 1981. On nonparametric measures of dependence for random variables. The annals of statistics, 9, 879-885.

Serinaldi, F., Bonaccorso, B., Cancelliere, A. y Grimaldi, S. 2009. Probabilistic characterization of drought properties trough copulas. Physics and Chemistry of the Earth, 34, 596-605.

Shiau, J. 2003. Return period of bivariate distributed extreme hydrological events. Stochastic Environmental Research and Risk Assessment, 17, 42-57.

Shiau, J. 2006. Fitting drought duration and severity with two-dimensional copulas. Water Resources Management, 20, 795-815.

Sklar, A. 1959. Fonctions de répartition à *n* dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris, 8, 229-231.

Wong, G., Lambert, M., Leonard, M. y Metcalfe, A. 2009. Drought analysis using trivariate copulas conditional on climatic states. Journal of Hydrologic Engineering, *in press*.

Yan, J. 2006. Multivariate modeling with copulas and engineering applications. Capítulo 51 en: Pham, H. (ed.), Springer Handbook of Engineering Statistics, 973-990, Springer, New York.

Yan, J. 2007. Enjoy the joy of copulas: with a package copula. Journal of Statistical Software, 21(4).

Yue, S. Ouarda, T. y Bobée, B. 2001. A review of bivariate gamma distributions for hydrological application. Journal of Hydrology, 246, 1-18.

Zhang, L. y Singh, V. 2006. Bivariate flood frequency analysis using the copula method. Journal of Hydrologic Engineering, 11(2), 150-164.

Zhang, L. y Singh, V. 2007a. Bivariate rainfall frequency distributions using Archimedean copulas. Journal of Hydrology, 332, 93-109.

Zhang, L. y Singh, V. 2007b. Gumbel-Hougarard copula for trivariate rainfall frequency analysis. Journal of Hydrologic Engineering, 12(4), 409-419.

Zhang, L. y Singh, V. 2007c. Trivariate flood frequency analysis using the Gumbel-Hougarard copula. Journal of Hydrologic Engineering, 12(4), 431-439.

Zegpi, M., Gironás, J y Fernández, B. 2008. Modelo analítico para el análisis hidrológico de cuencas urbanas usando cópulas. XXIII Congreso Latinoamericano de Hidráulica, Cartagena de Indias, Colombia.

Yue, S. y Rasmussen, P. 2002. Bivariate frequency analysis: discussion of some useful concepts in hydrological application. Hydrological Processes, 16, 2881-2898.