

Análisis Regional de Frecuencia de Avenidas en la vertiente cantábrica y noratlántica de España

Montes, J.^{a1}, Álvarez, M.^b, Pertierra, L.^c, Moralo, J.^{a2}, Baztán, J.^{a3}

^{a1} Gas Natural Fenosa, Ingeniería y Desarrollo de Generación, Calle Acanto 11-13 pl. 01 28045 Madrid. E-mail: jmontesm@gasnatural.com

^b Grupo de Ingeniería del Agua y del Medio Ambiente, ETS de Ingenieros de Caminos, Canales y Puertos, Universidade da Coruña, Campus de Elviña s/n. E-mail: ali.alvarez@udc.es

^c Gas Natural Fenosa Explotación, Unidad Hidráulica Generación España, Calle Batundeira 2 pl. 01 32960 Ourense. E-mail: lpertierra@gasnaturalfenosa.com

^{a2} Gas Natural Fenosa, Ingeniería y Desarrollo de Generación, Calle Acanto 11-13 pl. 01 28045 Madrid. E-mail: jmoralo@gasnatural.com

^{a3} Gas Natural Fenosa, Ingeniería y Desarrollo de Generación, Calle Acanto 11-13 pl. 01 28045 Madrid. E-mail: jbaztan@gasnatural.com

Línea temática B | Hidrología, usos y gestión del agua. Riegos. Energía hidroeléctrica.

RESUMEN

En el presente trabajo se ha desarrollado un modelo estadístico para estimar la frecuencia de caudales punta en la vertiente cantábrica y noratlántica de España, basándose en la metodología de Análisis Regional con sus parámetros estimados por los L-momentos (Hosking y Wallis, 1997). Para ello se ha contado con 85 puntos de aforo en régimen natural, con registros de caudal máximo anual entre los 15 y 72 años. Los resultados del estudio revelan la existencia de 9 regiones estadísticamente homogéneas, es decir, regiones que comparten la misma función de distribución de frecuencias excepto por un factor de escala. Como principales conclusiones del estudio cabe subrayar la distinción de regiones según su carácter pluvial o nival, así como la capacidad del modelo propuesto para estimar valores extremos de caudales punta con mayor robustez que otras metodologías. **Palabras clave** | Técnicas estadísticas; análisis regional; caudales punta, regresión múltiple.

INTRODUCCIÓN

El origen de la estadística de valores extremos se remonta al menos al siglo XIX, existiendo una importante cantidad de referencias literarias en las que se plantea el problema de la distribución de máximos y mínimos en una población estadística. Como uno de los mayores logros alcanzados en la primera mitad del siglo XX, resalta la demostración de que sólo tres familias paramétricas de distribuciones de probabilidad límites resultan matemáticamente compatibles con el estudio de valores extremos (Fisher y Tippett, 1928).

Trabajos desarrollados posteriormente por diferentes autores ayudaron a asentar las bases del estudio de la distribución de valores extremos, cuyos principales resultados fueron culminados con la publicación del libro “Statistics of extremes” (Gumbel, 1958), probablemente uno de los hitos más importantes en la historia de la teoría de valores extremos.

Desde entonces, esta teoría ha contribuido de forma muy relevante al avance de una gran variedad de campos de la ciencia y la ingeniería, como es el caso de la hidrología de valores extremos y su concreta aplicación a la frecuencia de las crecidas de los ríos.

Dentro de este marco de aplicación, si bien el uso de información exclusivamente procedente de una determinada estación de aforos se ha visto demasiado extendido durante años, muchos han sido los hidrólogos que desde los propios orígenes de la teoría han mostrado su preocupación acerca de los inconvenientes asociados al uso de dicha metodología local. La escasa longitud temporal de las series de observación de caudales máximos disponibles, unido al hecho de que la propia ocurrencia de los caudales extremos resulta rara por definición, provoca que las estimaciones de los caudales proporcionadas por el análisis local presenten una elevada incertidumbre.

Es por ello que desde hace décadas hayan ido surgiendo diferentes procedimientos enfocados a intentar solventar las limitaciones asociadas a la corta longitud temporal de los registros, aplicando diferentes enfoques regionales mediante los que poder paliar la carencia de datos en el tiempo con la abundancia en el espacio.

De cita casi obligada resulta el método del índice de avenida (Dalrymple, 1960), cuya esencia todavía hoy día sustenta la metodología regional recomendada por diferentes autores y en el cual se asume, excepto por un factor de escala, una misma distribución de frecuencias para las cuencas pertenecientes a una determinada región previamente considerada como homogénea.

Asimismo relevantes resultan las regionalizaciones propuestas por diversos organismos oficiales entre los años 70 y 80 del siglo pasado, tales como la desarrollada para el Reino Unido bajo el método del índice de avenida (Natural Environment Research Council, 1975), o la realizada en los Estados Unidos de América respecto a la asimetría regional modelizada bajo la función de distribución log-Pearson tipo III (Bulletin 17 of the U.D. Water Resources Council, 1981).

Paralelamente a los anteriores procedimientos fueron desarrolladas nuevas técnicas estadísticas mediante las que poder combinar la información regional y local existente, destacando entre ellas la metodología de análisis bayesiano (Wood y Rodríguez-Iturbe, 1975).

Posteriormente, durante los años 80 del siglo pasado surgieron nuevos avances que contribuyeron a mejorar el desarrollo del análisis regional de frecuencia, tales como el procedimiento de regionalización basado en los Momentos de Probabilidad Ponderada (Greis y Wood, 1981), cuya comparativa frente a otros métodos de la época resultó concluyente al señalar a éste como el de mayor robustez y precisión (Cunnane, 1988).

Relacionado con el procedimiento anterior, años más tarde fueron desarrollados los L-momentos como combinación lineal de los Momentos de Probabilidad Ponderada (Hosking, 1990), aportándose una mayor sencillez de aplicación e interpretación a la metodología, así como un incremento de la precisión de resultados.

Ya a finales de los años 80, las dudas que habían sido planteadas sobre la invalidez de la hipótesis de homogeneidad aplicada al análisis regional de frecuencia (Benson, 1962), fueron despejadas por diferentes investigaciones que demostraron cómo el análisis regional es capaz de arrojar estimaciones de valores extremos mucho más precisas que el análisis local, incluso cuando la región resulta moderadamente heterogénea (Potter y Lettenmaier, 1990).

Casi una década más tarde, el desarrollo del análisis regional de frecuencia experimentó un gran avance a través de la formulación planteada por Hosking y Wallis, con una metodología sustentada en el método del índice de avenida y en la incorporación del enfoque de estimación de los parámetros de las funciones de distribución mediante los L-momentos (Hosking y Wallis, 1997). Durante los últimos 20 años, esta metodología ha marcado las pautas de un gran número de estudios de caudales extremos en diferentes países del mundo tales como Sudáfrica (Kjeldsen et al., 2002) ó China (Jingyi y Hall, 2004), así como análisis a escala mundial (Smith et al., 2015).

No obstante, la diversidad de los enfoques de análisis regional empleados durante los últimos años resulta más amplio, habiéndose desarrollado por ello numerosos trabajos con el objetivo de contrastar los diferentes procedimientos existentes (Kochanek et al., 2014).

En el caso de España, una de las variantes aplicadas al estudio de los caudales máximos anuales se refiere a la metodología conocida con el nombre anglosajón de *Regional Shape Estimation* (RSE), en cuyo procedimiento de regionalización se asume un L-coeficiente de asimetría (LCs) idéntico para todas las estaciones de la región, al tiempo que se toma la media y el L-coeficiente de variación (LCv) local de cada una de las estaciones. Este procedimiento se corresponde con la metodología que fue aplicada para la elaboración del mapa de caudales máximos CAUMAX de las cuencas intercomunitarias de España (CEDEX, 2009), cuya precisión de las estimaciones asociadas a las regiones consideradas arrojaron un orden de magnitud aceptable (Jiménez et al., 2014).

En este sentido y en base a investigaciones realizadas para distintas regiones de los Estados Unidos de América con diferente grado de heterogeneidad (Hosking y Wallis, 1997), se sabe que la variante RSE puede llegar a proporcionar estimaciones más precisas que el índice de avenida en casos en los que las regiones resultan homogéneas respecto al LCs , pero claramente heterogéneas respecto al LCv . Dicha metodología, por consiguiente, a priori resulta de aplicación para aquellos casos en los que la dispersión del LCv entre estaciones resulta grande comparado con la variabilidad estocástica asociada al LCv de

cada estación, al tiempo que la dispersión del LCs entre estaciones resulta pequeño comparado con la variabilidad estocástica asociada al LCs de cada estación.

Es por ello que, debido a la heterogeneidad generalizada del LCv asociado a las regiones consideradas por el CAUMAX, la metodología RSE haya encontrado su espacio de aplicación como parte inicial de un proceso jerárquico en el que, como segunda fase de mayor precisión, habrían de intentar identificarse subregiones homogéneas respecto al coeficiente de variación (Fiorentino et al., 1987).

En este sentido, existen estudios de diferentes partes de España en los que se ha mostrado la posibilidad de delimitar regiones aceptablemente homogéneas respecto a los tres principales L-momentos (LCv , LCs y L-coeficiente de curtosis LCK), al evitar que dicha delimitación se vea excesivamente subordinada a factores geográficos como pueden ser los límites de las principales cuencas hidrográficas (Álvarez et al., 1999).

De esta forma, atendiendo a la posibilidad de delimitar regiones completamente homogéneas con las que mejorar la precisión de las estimaciones publicadas hasta la fecha, en el presente trabajo se ha desarrollado un análisis regional de frecuencia de caudales punta basado en la metodología del índice de avenida de Hosking y Wallis, con sus parámetros estimados por los L-momentos, para toda la vertiente cantábrica y noratlántica de España.

ÁREA DE ESTUDIO Y DISPONIBILIDAD DE INFORMACIÓN

La región geográfica analizada en el presente trabajo se ha limitado a la vertiente cantábrica y noratlántica de España, incluyéndose todas aquellas cuencas en las que se ha dispuesto de registros de caudales máximos anuales suficientemente fiables y representativos, en régimen natural.

Con objeto de recopilar la mayor cantidad de información potencialmente aprovechable, se indagó en las bases de datos publicadas por los distintos Organismos Hidrográficos de la zona en estudio, Diputaciones provinciales, Organismos de las Comunidades Autónomas, Ministerio de Agricultura, Alimentación y Medio Ambiente (MAGRAMA) y Centro de Estudios y Experimentación de Obras Públicas (CEDEX). Asimismo, se consultó a todas aquellas entidades privadas con competencia sobre presas y azudes ubicados dentro de la zona en estudio, explorando la posibilidad de que aportaran los registros de caudales máximos anuales asociados a sus infraestructuras.

De este modo se obtuvo información potencialmente aprovechable de 136 estaciones de aforo en régimen natural (24 de ellas localizadas en embalses), con una longitud de registros variable entre los 15 y 88 años comprendidos dentro del periodo hidrológico 1912/1913–2014/2015, y una longitud de serie promedio en torno a los 32 años.

A continuación se muestran los links asociados a los servidores desde los que se puede descargar la información hidrométrica utilizada, a excepción de los datos aportados por las entidades privadas cuyo acceso se encuentra restringido a personal propio de la entidad.

<http://sig.magrama.es/aforos>, <http://hercules.cedex.es/anuarioaforos>, <http://www.bizkaia.eus>, <http://www.gipuzkoa.eus>,
http://www.navarra.es/home_es/, <https://www.chminosil.es/es/>, <http://www.meteogalicia.gal/web/index.action>,
<http://augasdegalicia.xunta.gal/>, <http://www.chcantabrico.es/index.php/es/>.

MÉTODOS DE ANÁLISIS

En el presente apartado se describen brevemente las metodologías seguidas por el estudio para evaluar la calidad y representatividad de los datos de partida, ampliar las series de caudal máximo anual instantáneo y desarrollar el análisis regional de frecuencia.

Calidad y representatividad de los datos de partida

Previo al desarrollo de cualquier análisis de frecuencia, resulta imprescindible realizar un análisis minucioso de la calidad de los datos de partida que resulte coherente con los objetivos perseguidos por el estudio.

En este sentido y atendiendo a las bases estadísticas en las que se fundamenta la teoría de valores extremos, resulta necesario que las series de partida sean aleatorias, independientes y homogéneas. Adicionalmente, las series habrán de ser estacionarias, si lo que se pretende modelizar es una acotada sucesión de décadas para la cual pueda considerarse que la función de distribución de frecuencia resulta invariante respecto al tiempo.

En este sentido, la posible incompatibilidad del análisis regional de frecuencia respecto a las recientes metodologías desarrolladas sobre análisis no estacionarios (Salas y Obeysekera, 2014), ha conducido a plantear el presente trabajo bajo un enfoque estacionario en el que las series de partida habrán de cumplir con los 4 criterios anteriormente mencionados.

Dicha evaluación de las series ha sido realizada a través de análisis exploratorio de datos, así como mediante pruebas de hipótesis estadísticas recomendadas por el Taller de Expertos de la OMM/UNESCO (Kundzewicz and Robson, 2000) implementadas en el modelo TREND (Chiew and Siriwardena, 2005).

Ampliación de las series de caudal máximo anual instantáneo

Con objeto de contar con el mayor número de datos posible de caudal máximo anual instantáneo, resulta recomendable que para aquellos años en los que no se cuenta con observaciones directas de dicha variable, ésta sea estimada a partir de los registros de caudal máximo medio diario. Mediante esta ampliación de información se conseguirá obtener una mayor cobertura espacial de la zona en estudio, así como series temporales más largas que permitan mejorar la precisión de las estimaciones de los caudales punta asociados a los periodos de retorno de interés.

Entre las expresiones utilizadas para transformar el caudal máximo medio diario en caudal máximo instantáneo, probablemente sea la fórmula de Fuller la que mayor uso y aceptación presenta. No obstante, en el presente trabajo se ha comprobado cómo la adaptación de esta fórmula a la zona en estudio arroja bajos coeficientes de correlación R^2 , a priori asociados al hecho de que su expresión matemática únicamente incluye el área de cuenca como predictor.

Por este motivo, en el presente trabajo se ha explorado la posibilidad de mejorar la relación entre caudales máximos mediante el desarrollo de un análisis de regresión múltiple, en el que se ha incluido una gran cantidad de variables hidrológicas adicionales al tamaño de cuenca, así como toda una serie de variables climáticas capaces de proporcionar, matemáticamente, la capacidad de diferenciar entre regiones climáticas distintas. Las variables incorporadas como punto de partida del análisis han sido las siguientes: área y perímetro de cuenca, longitud de cauces, parámetro de forma, densidad de drenaje, rango de altura, altura y pendiente media de la cuenca, precipitación media anual y máxima diaria promediada en la cuenca, aportación total anual, umbral de escorrentía P_0 , temperatura promedio anual mínima, máxima y media en la cuenca, escorrentía anual, evapotranspiración real y potencial media anual, recarga media anual de acuíferos y coordenadas UTM x/y (huso 30, ETRS89).

De este modo, para cada una de las cuencas en estudio, se determinó el valor de las anteriores variables indicadas a partir de la recopilación, evaluación y posterior procesamiento con la herramienta ArcGIS de información publicada por diferentes organismos oficiales, bajo formato asociado a Sistemas de Información Geográfica.

En base a dicha información, se llevó a cabo el análisis de regresión múltiple citado, evaluándose ecuaciones potenciales del tipo:

$$k = 1 + a \cdot A^b \cdot B^c \cdot C^d \cdot D^e \cdot \dots \quad (1)$$

donde:

$$k = Q_{mi}/Q_{md}$$

Q_{mi} = caudal máximo anual instantáneo

Q_{md} = caudal máximo medio diario anual

A, B, C, D, \dots = variables hidrológicas/climáticas

a, b, c, d, e, \dots = coeficientes de la relación potencial

Análisis regional de frecuencia de caudales punta

El análisis regional de frecuencia basado en el método del índice de avenida de Hosking y Wallis, con sus parámetros estimados por los L-momentos, se sustenta en las siguientes cuatro etapas:

- a) Filtrado de las series de observaciones
- b) Identificación de regiones homogéneas
- c) Selección de la función de distribución de frecuencia
- d) Estimación de cuantiles extremos asociados a periodos de retorno de interés.

Los procedimientos de cálculo asociados a estas cuatro etapas consecutivas se encuentran implementados en los paquetes *lmomRFA* (Hosking, 2014a) y *lmom* (Hosking, 2014b), publicados bajo el entorno de programación R.

Filtrado de las series de observaciones: La medida de discordancia D

La medida de discordancia D responde a un estadístico que compara los L-momentos muestrales de cada una de las estaciones que conforman la región en análisis con los correspondientes valores medios de dicha región. De esta forma, una estación i se califica como discordante si su valor D_i supera un valor crítico establecido en función del número de estaciones que contiene la región. Se trata por tanto de una medida que sirve para realizar un filtrado de estaciones previo a la fase de identificación de regiones.

Identificación de regiones homogéneas: La medida de heterogeneidad H

Estadísticamente, se considera que una región resulta homogénea si la función de distribución que describe las series de caudales máximos de las n estaciones de dicha región es la misma, excepto por un factor de escala o índice de avenida. Bajo esta definición, la identificación de regiones homogéneas resulta la etapa de mayor dificultad en el análisis regional, habiéndose

desarrollado por ello diferentes metodologías de apoyo que permiten obtener identificaciones previas de carácter orientativo (Ward, 1963).

Siguiendo la misma estela de trabajos desarrollados a tal efecto, en el presente estudio se ha utilizado el método de Ward para obtener una primera tentativa de la delimitación de regiones potencialmente homogéneas, analizadas y posteriormente refinadas en base a las medidas de heterogeneidad regional H_1 , H_2 y H_3 de Hosking y Wallis. De estas tres medidas, la que mayor peso recibe es la H_1 (relativa al análisis de heterogeneidad del LCv), ya que la variabilidad del LCv entre las estaciones de una región provoca un mayor efecto sobre la varianza de los cuantiles estimados que la variabilidad del LCs ó el LCK (Hosking y Wallis, 1997).

El análisis de estas tres medidas propuestas se fundamenta en la comparativa de los valores observados y simulados de la dispersión de los L-momentos de cada una de las estaciones que conforman la región, considerándose ésta como “aceptablemente homogénea” si $H_i < 1$, “posiblemente heterogénea” si $1 < H_i < 2$ y “definitivamente heterogénea” en el caso que $H_i \geq 2$.

En última instancia, el refinamiento definitivo de las regiones ha sido llevado a cabo mediante un proceso iterativo supeditado a obtener la menor incertidumbre posible en los cuantiles estimados.

Selección de la función de distribución de frecuencia

Una vez definidas las regiones aceptablemente homogéneas, corresponde seleccionar la función de distribución de mejor ajuste para cada una de las mismas. Dentro del paquete informático citado con anterioridad, se pueden evaluar hasta 5 distribuciones (DIST) de tres parámetros: Logística Generalizada (GLO), General de Valores Extremos (GEV), Normal Generalizada (GNO), Pareto Generalizada (GPA) y la Pearson Tipo III (PE3). Existen además las opciones de las funciones Kappa (KAP) y Wakeby (WAK), de cuatro y cinco parámetros respectivamente, siendo la primera recomendada para el desarrollo de los algoritmos de simulación destinados a evaluar la medida de heterogeneidad H .

Para discernir entre las DIST candidatas se cuenta con la medida de bondad de ajuste Z^{DIST} , mediante la cual se evalúa si la región en estudio cumple con el criterio $|Z^{DIST}| \leq 1.64$ (ajuste aceptable), recomendándose para su simulación la función de distribución KAP. De todas aquellas distribuciones que resulten aceptables, la de mejor ajuste será la que tenga el mínimo $|Z^{DIST}|$.

No obstante, se ha de tener en cuenta que el criterio $|Z^{DIST}| \leq 1.64$ resulta un tanto arbitrario, ya que la hipótesis de normalidad que ha de ser satisfecha por el estadístico Z asociado requiere que la región sea exactamente homogénea, así como que no exista correlación espacial entre los datos de las distintas estaciones. Es por ello que, debido a la imposibilidad de cumplir rigurosamente con las anteriores dos condiciones, el criterio $|Z^{DIST}|$ ha de tomarse como un resultado meramente orientativo que no es capaz de identificar con suficiente precisión la función de distribución de mayor capacidad descriptiva.

Estimación de cuantiles extremos asociados a periodos de retorno de interés

Los cuantiles de caudal máximo anual instantáneo, asociados a los distintos periodos de retorno T , se estiman a partir de cada una de las funciones de distribución regionales que orientativamente resultan aceptables, siguiendo el método del índice de avenida.

En dicho procedimiento se asume que la función distribución de frecuencia de todas las estaciones que conforman una región homogénea es idéntica, excepto por un factor local de escala conocido como el índice de avenida. Para cada una de las estaciones comprendidas dentro de cada región homogénea, las estimaciones de los cuantiles se calcula según la siguiente expresión:

$$Q_i(F) = \mu_i \cdot q(F) \quad (2)$$

donde:

$Q_i(F)$ = estimación de los cuantiles asociados a la estación i , para una probabilidad de no excedencia F

μ_i = índice de avenida estimado por el valor medio de la serie anual de caudales máximos instantáneos de la estación i

$q(F)$ = curva regional de frecuencia adimensional o curva de crecimiento, común para todas las estaciones de la región

La precisión de los cuantiles estimados para cada T se evalúa mediante simulaciones de Monte Carlo, obteniéndose el error medio cuadrático relativo (RMSE) y los límites de confianza asociados a un determinado grado de significancia. La DIST con menores RMSE es la más robusta y de mejor capacidad predictiva.

RESULTADOS Y DISCUSIÓN

Como resultado del análisis de calidad de datos realizado, se identificaron un total de 85 estaciones de aforo con información suficientemente fiable y representativa del régimen natural estacionario en estudio, de las cuales 21 de ellas se localizan en embalses. Dentro de dicha información, se incluyen tanto los registros de caudal máximo medio diario de las series anuales identificadas, así como registros parciales de caudal máximo instantáneo asociados a esas mismas series.

Bajo este contexto y con el fin de rellenar las lagunas de información relativas a los registros de caudal máximo instantáneo, se desarrolló un análisis de regresión múltiple cuyos resultados han revelado que la relación entre los caudales máximos medios diarios y los máximos instantáneos se encuentra fuertemente condicionada al hecho de que la cuenca analizada presente un régimen pluvial o predominantemente nival. Estos resultados ponen de manifiesto que el razonamiento asociado a la fórmula de Fuller, por el cual la relación entre caudales máximos disminuye según aumenta el tamaño de cuenca, no explica adecuadamente la fenomenología asociada a las crecidas máximas anuales que se originan en cuencas de tipología nival, marcadamente condicionadas por la fusión de nieve previamente acumulada en su área de captación.

Estos argumentos quedan claramente manifestados por las ecuaciones potenciales derivadas del análisis de regresión múltiple realizado, cuyas expresiones matemáticas se muestran a continuación:

Cuenca de tipología pluvial:

$$k = 1 + 6.1 \cdot 10^{-4} \cdot \frac{UTMx^{0.453} \cdot Pmax^{1.271}}{Ac^{1.014} \cdot Pf^{1.709}} \quad (3)$$

donde:

$$k = Q_{mi}/Q_{md}$$

Q_{mi} = Caudal máximo instantáneo anual [m^3/s]

Q_{md} = Caudal máximo medio diario anual [m^3/s]

$UTMx$ = Coordenada x asociada al sistema de proyección UTM, huso 30, datum ETRS89 [m]

P_{max} = Precipitación máxima diaria promedio en la cuenca asociado al periodo de retorno de 2.33 años [mm]

A_c = Área de la cuenca [km^2]

P_f = Parámetro de forma de la cuenca (Perímetro [km] / Área [km^2])

Como se puede ver, la ecuación obtenida para el modelo pluvial resulta ser una ampliación de la expresión de Fuller, de modo que según aumenta el tamaño de cuenca disminuye la relación k . Las tres variables adicionales presentan una interpretación física razonable, de modo que k aumenta cuanto más circular es la cuenca (menor parámetro de forma), cuanto mayor resultan las precipitaciones máximas en ella, y cuanto más hacia el este se ubica (mayor brusquedad de los episodios de lluvia). Cabe añadir que la pendiente media de la cuenca también resulta un predictor relevante de la relación k , si bien su alta correlación con el área de cuenca y la coordenada $UTMx$ en la zona en estudio acaba por desplazarla de la ecuación.

Cuenca de tipología predominantemente nival:

$$k = 1 + 0.0117 \cdot \frac{T_{media}^{1.593}}{T_{min}^{0.08} \cdot D^{0.202}} \quad (4)$$

donde:

$$k = Q_{mi} / Q_{md}$$

Q_{mi} = Caudal máximo instantáneo anual [m^3/s]

Q_{md} = Caudal máximo medio diario anual [m^3/s]

T_{media} = Temperatura media anual promediada en la cuenca [$^{\circ}C$]

T_{min} = Temperatura mínima media anual en la cuenca [$^{\circ}C$]

D = Densidad de drenaje de la cuenca ($L_{total \text{ de cauces}} [km] / \text{Área}_{cuenca} [km^2]$)

Como se puede observar, la ecuación obtenida para el modelo nival no muestra explícitamente la variable *área de cuenca* entre sus predictores, lo cual parece manifestar que la cantidad de agua que se acumula en este tipo de cuencas en forma de nieve (potencialmente fundible durante eventos de lluvia o episodios propios de deshielo), resulta mucho más significativa respecto a la forma del hidrograma de avenida que la precipitación acaecida en forma de lluvia. Debido a ello, las variables incluidas como predictores se corresponden con la T_{min} y T_{med} , relacionadas respectivamente con la cantidad de nieve acumulada y la capacidad de que ésta se funda con mayor rapidez. La tercera variable incluida en la anterior ecuación se corresponde con la densidad de drenaje, interpretándose de modo que cuanto mayor longitud recorra el flujo difuso de nieve fundida hasta alcanzar un cauce, mayor será el arrastre adicional de nieve que aportará agua al punto de desfogue de la cuenca durante el proceso de fusión. Relacionado con el anterior razonamiento, cabría esperar que la pendiente media de la cuenca

también tuviera su influencia predictiva sobre la relación k , lo cual efectivamente ocurre si bien con un grado de significancia que no llega a cumplir con el umbral máximo del 10% que generalmente se asume en las pruebas de hipótesis estadísticas.

Más allá de su posible interpretación física, cabe resaltar que las anteriores ecuaciones aportan una mayor explicación de la varianza asociada a la relación entre los registros de caudal máximo medio diario y los máximos instantáneos, obteniéndose sesgos y errores cuadráticos medios apreciablemente menores a los obtenidos con la fórmula de Fuller adaptada a la zona en estudio, tal y como se muestra a continuación:

Tabla 1 | Resultados de la bondad de ajuste del modelo pluvial

	R² ajustado	Sesgo	RMSE
Modelo $k \sim A_c$ (con todas las cuencas)	0.38	- 1.59 %	11.5 %
Modelo $k \sim A_c$ (sólo cuencas pluviales)	0.46	-1.09 %	11.3 %
Modelo $k \sim A_c + UTM_x + P_f + P_{max}$ (sólo cuencas pluviales)	0.82	-0.02 %	8.3 %

Tabla 2 | Resultados de la bondad de ajuste del modelo nival

	R² ajustado	Sesgo	RMSE
Modelo $k \sim A_c$ (con todas las cuencas)	0.38	- 1.59 %	11.5 %
Modelo $k \sim A_c$ (sólo cuencas nivales)	0.062	-0.3 %	7.7 %
Modelo $k \sim T_{media}$ (sólo cuencas nivales)	0.44	-0.1 %	5.6 %
Modelo $k \sim T_{media} + T_{min} + D$ (sólo cuencas nivales)	0.81	-0.1 %	2.7 %

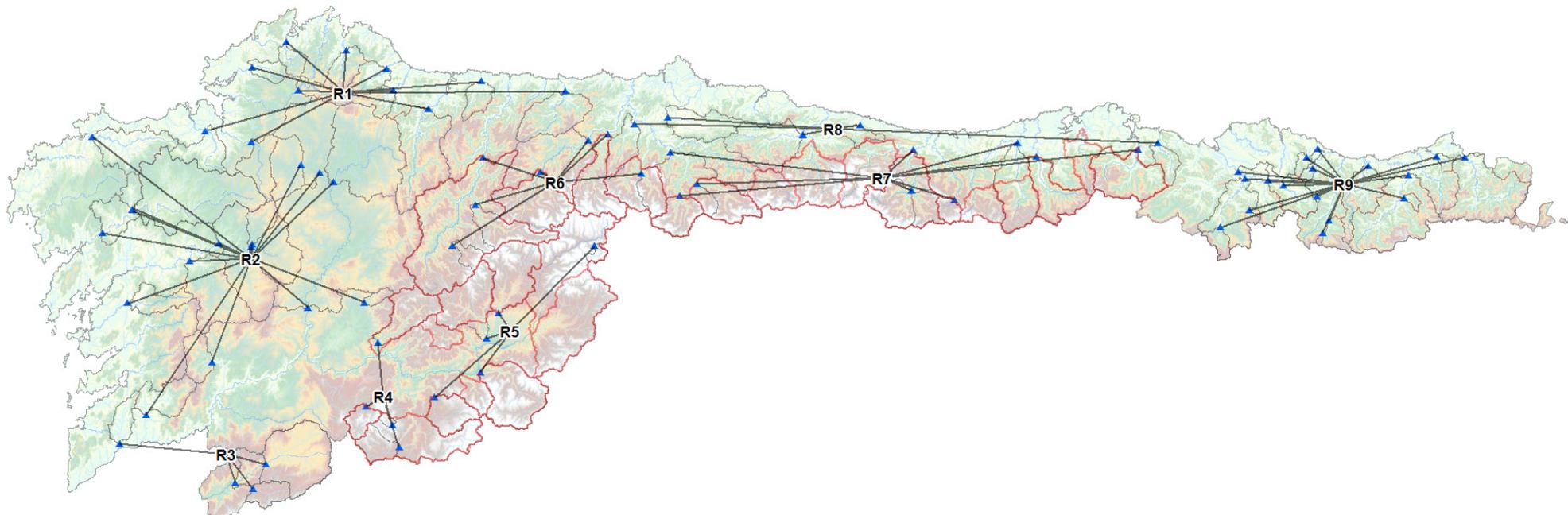
Aplicando las anteriores ecuaciones (3) y (4) a aquellos años y estaciones en las que, existiendo registros de caudal máximo medio diario, no se cuenta con datos de caudal máximo instantáneo, se ha conseguido disponer de un total de 85 series de caudal máximo instantáneo en régimen natural entre los 15 y 72 años de datos. El intervalo de tiempo comprendido por dichas series abarca el periodo hidrológico 1943/1944–2014/2015, con una longitud de serie promedio en torno a los 31 años y una densidad espacial de información de 1.58 estaciones/1000 km², valor superior a los recomendados por la Organización Mundial de Meteorología (WMO-No.168).

Una vez preparadas las series de datos anteriormente indicadas, se realizó un análisis clúster mediante el método de Ward (Ward, 1963) tomando como referencia las siguientes características locales de las estaciones: área de la cuenca (km²), altura media de la cuenca (m) y coordenadas de localización (en grados). De esta manera se obtuvo un total de 5 regiones diferenciadas, cuyas estaciones fueron sometidas a un análisis de heterogeneidad mediante la evaluación de las respectivas medidas D y H , obteniéndose una acusada heterogeneidad generalizada.

No obstante, el análisis cualitativo de la repartición de estaciones entre estas 5 regiones proporcionó un punto de partida útil, de cara a desarrollar un proceso iterativo de refinamiento de las regiones con el objetivo final de obtener una mayor homogeneidad. En esencia, dicho proceso iterativo consistió en la subdivisión y/o combinación de regiones, incluir estaciones en una u otra región, omitir algunas de ellas, etc., comprobando en cada iteración las nuevas medidas de D y H . En última instancia, el refinamiento de las regiones definitivas ha sido llevado a cabo mediante un proceso iterativo adicional, en el que cada una de las estaciones se incluía definitivamente en una u otra región atendiendo a obtener la menor incertidumbre posible de los cuantiles asociados a dicha estación.

Los resultados con la configuración final de las regiones identificadas se pueden observar en la figura 1 a continuación adjunta, junto con una tabla resumen en la que se incluyen los L-momentos promedio de cada región y las medidas D y H correspondientes. En dicha figura se puede observar la diferenciación por regiones entre cuencas pluviales y aquéllas de tipología predominantemente nival, derivada de la mera aplicación del procedimiento de análisis regional adoptado.

Figura 1 | Regiones finales asociadas a las series de caudales máximos instantáneos



NOTA: Se remarca en color rojo el perímetro de las cuencas de tipología nival identificadas por el Ministerio de Agricultura, Alimentación y Medio Ambiente (MAGRAMA)

Tabla 3 | Resumen estadístico de los L-momentos regionales, medidas de discordancia y heterogeneidad

Región	Nº Estaciones	t (LCv)	t ₃ (LCs)	t ₄ (Lck)	D _{máx} (Dcrit)	H1	H2	H3
R1	11	0.3407	0.3719	0.2637	1.79 (2.63)	0.50	-0.73	-0.81
R2	16	0.3036	0.1933	0.1496	2.83 (2.97)	1.41	-0.21	-0.61
R3	4	0.3595	0.2474	0.1514	1.00 (3.00)	-0.28	1.57	1.94
R4	4	0.2143	0.1319	0.1216	1.00 (3.00)	0.38	-0.54	-0.36
R5	5	0.3382	0.2932	0.2135	1.30 (1.33)	-0.16	0.54	0.13
R6	7	0.2566	0.1206	0.1267	1.31 (1.92)	-0.22	0.49	0.94
R7	10	0.2583	0.3155	0.2575	1.61 (2.49)	0.56	-0.28	-0.31
R8	5	0.2759	0.2632	0.1732	1.27 (1.33)	0.39	-0.43	-0.35
R9	16	0.2667	0.2160	0.1474	2.52 (3.00)	-0.42	1.97	2.90

De las tres medidas de heterogeneidad, se puede observar cómo la de mayor trascendencia (H_1) se encuentra por debajo de 1 en todas las regiones, a excepción del valor obtenido para la región R2 en donde dicho estadístico asciende a 1.41.

Estos resultados reflejan la gran homogeneidad de las regiones delimitadas, si bien el caso concreto de la región R2 apunta hacia una posible heterogeneidad, cuya relevancia ha sido analizada comparando los RMSE obtenidos en dicha región respecto a los que se alcanzan al dividir la misma en dos subregiones ya completamente homogéneas. El proceso iterativo desarrollado para llevar a cabo tal subdivisión, sustentado en la misma metodología aplicada a la obtención de las 9 regiones anteriormente presentadas, condujo a incluir en la primera de las dos subregiones las estaciones de Albarellos, 1564 y 1645 (ubicadas en la zona de mayor pluviometría e intensidades máximas de Galicia), circunscribiéndose el resto de estaciones por la segunda de las subregiones planteadas. Como resultado de este análisis se obtuvieron medidas de H_1 , H_2 y H_3 claramente por debajo de 1 para ambas subregiones, si bien los valores del RMSE resultaron en torno a un 20% mayores a los obtenidos con la región R2 completa. Es por ello que se haya optado por no dividir la región R2, teniendo en cuenta además que los cuantiles regionales resultan algo más conservadores bajo una única región.

Por otro lado, se puede observar cómo las medidas H_2 y H_3 asociadas a 7 de las 9 regiones propuestas arrojan valores asimismo por debajo de 1, obteniendo únicamente valores superiores a 1 en las regiones R3 y R9.

Para el caso de la región R3, se ha podido comprobar que la posible heterogeneidad asociada a H_2 y H_3 se debe a la inclusión de la estación 1805, cuya cuenca de cierre se corresponde en gran porcentaje con la Laguna de Antela. De hecho, se ha comprobado que al excluir dicha estación de la región R3 en cuestión, las medidas H_1 , H_2 y H_3 resultan entonces claramente menores que 1, obteniéndose estimaciones de los cuantiles regionales de mayor magnitud. Es por ello que pudiera resultar más acertado excluir dicha estación de la región R3, con objeto de manejar unas estimaciones de caudal punta más conservadoras.

Por otro lado, la heterogeneidad asociada a las medidas H_2 y H_3 de la región R9, conllevó asimismo a analizar la conveniencia de dividir esta región en subregiones de menor heterogeneidad. Los resultados derivados de este análisis mostraron la posible presencia de dos subregiones homogéneas respecto a H_1 , H_2 y H_3 , conteniendo la primera de ellas las 6 estaciones ubicadas más al oeste de la región R9, y la segunda subregión las 10 estaciones restantes. No obstante, la importante reducción de los cuantiles regionales obtenidos para la primera de las subregiones mencionadas, unido a la gran similitud de los RMSE alcanzados para todos los casos analizados, aconseja adoptar la región R9 de una perspectiva única de cara a manejar estimaciones de caudal punta más conservadores. En cualquier caso, puesto que las 6 estaciones que parecen conformar la primera de las subregiones mencionadas se corresponden con puntos de aforo relativamente modernos, esta incertidumbre podrá ser contrastada a corto/medio plazo una vez se cuente con un mayor número de registros.

Adicionalmente al análisis de heterogeneidad anteriormente presentado, se evaluaron las distribuciones candidatas según la bondad de ajuste Z^{Dis} , cuyos resultados se muestran en la siguiente tabla.

Tabla 4 | Resultados de la bondad de ajuste para las distribuciones candidatas

Región	Umbral orientativo $ Z^{Dist} \leq 1.64$	Z^{Dist} Mejor
R1	GLO (0.20), GEV (-0.41), GNO (-1.28), PE3 (-2.78), GPA (-2.35)	GLO
R2	GNO (-0.01), GEV (0.41), PE3 (-0.95), GLO(2.55), GPA(-4.48)	GNO
R3	PE3 (-0.24), GNO (0.37), GEV (0.71), GLO(1.50), GPA(-1.24)	PE3
R4	PE3 (-0.05), GEV (0.11), GNO (0.16), GLO(1.61), GPA(-2.99)	PE3
R5	GLO (0.20), GEV (-0.41), GNO (-0.89), PE3(-1.64), GPA(-2.09)	GLO
R6	GNO (0.08), GEV (0.09), PE3 (-0.26), GLO (1.92), GPA(-3.75)	GNO
R7	GLO (-0.37), GEV (-1.14), GNO (-1.75), PE3 (-2.81), GPA(-3.25)	GLO
R8	GEV (0.14), GNO (-0.28), GLO (0.95), PE3 (-1.03), GPA(-1.91)	GEV
R9	GNO (0.01), GEV (0.52), PE3 (-0.97), GLO (1.96), GPA(-2.97)	GNO

Desde un punto de vista global, cabe resaltar que la distribución PE3 muestra un ajuste aceptable en 7 de las 9 regiones consideradas, siendo ésta la distribución que ha proporcionado los menores RMSE en todas y cada una de las 9 regiones. En este sentido, es importante subrayar que la desviación del ajuste obtenido para las regiones R1 y R7 ha de tomarse con cautela, teniendo en cuenta la imprecisión del estadístico Z derivada de asumir en sus cálculos que las regiones sean exactamente homogéneas, y que acaso no existiera correlación espacial alguna entre los datos de las distintas estaciones que las conforman. De hecho, es razonable que una misma tipología de función de distribución de frecuencias resulte la más adecuada para modelizar una zona en estudio en la que no se presenten climatologías considerablemente dispares (Hosking y Wallis, 1997).

En base a lo anteriormente expuesto, como criterio para la elección definitiva de la mejor distribución de frecuencias se ha optado por escoger aquella que se corresponde con el modelo más robusto para las 9 regiones, es decir, la PE3. La robustez de cada una de las distribuciones candidatas, cuyos resultados aquí se omiten por evitar adjuntar información excesiva, ha sido calculada mediante el empleo de simulaciones de Monte Carlo siguiendo las recomendaciones de Hosking y Wallis (1997).

En la siguiente tabla se muestran los resultados correspondientes a la distribución PE3 finalmente adoptada para toda la zona en estudio, en donde se pueden consultar los principales cuantiles regionales, los parámetros de la función y los valores de RMSE obtenidos.

Tabla 5 | Parámetros, cuantiles regionales y RMSE de la distribución Pearson tipo III para cada región.

Región	Parámetros DIST			T (años)	10	100	1000	10000
	u	α	k	F	.9	.99	.999	0.9999
R1	1.000	0.706	2.197	q(F)	1.906	3.615	5.352	7.103
				RMSE	0.126	0.136	0.144	0.149
R2	1.000	0.542	1.082	q(F)	1.727	2.667	3.521	4.332
				RMSE	0.099	0.108	0.117	0.125
R3	1.000	0.677	1.451	q(F)	1.903	3.235	4.498	5.728
				RMSE	0.137	0.172	0.200	0.220
R4	1.000	0.376	0.714	q(F)	1.502	2.067	2.552	2.999
				RMSE	0.071	0.091	0.112	0.131
R5	1.000	0.663	1.777	q(F)	1.875	3.313	4.723	6.119
				RMSE	0.129	0.153	0.171	0.183
R6	1.000	0.470	0.803	q(F)	1.629	2.362	3.001	3.594
				RMSE	0.008	0.104	0.121	0.136
R7	1.000	0.508	1.797	q(F)	1.669	2.778	3.866	4.945
				RMSE	0.100	0.114	0.125	0.133
R8	1.000	0.515	1.552	q(F)	1.686	2.733	3.737	4.720
				RMSE	0.103	0.126	0.144	0.158
R9	1.000	0.498	1.370	q(F)	1.666	2.620	3.517	4.387
				RMSE	0.111	0.122	0.132	0.140

Los anteriores resultados ponen de manifiesto la buena capacidad predictiva del modelo adoptado, cuyos RMSE para periodos de retorno superiores a los 100 años arrojan un valor medio de 0.14.

CONCLUSIONES

En el presente trabajo se ha podido comprobar la conveniencia de tratar bajo un enfoque regional el análisis de frecuencia de los caudales punta en la vertiente cantábrica y noratlántica de España, concretamente basado éste en el método del índice de avenida con sus parámetros estimados por los L-momentos.

En este sentido, la modelización estadística realizada se ha acotado a un régimen estacionario suficientemente representativo de la hidrología actual de la zona en estudio, con aquellas 85 series de datos que no solamente resultan aleatorias, independientes y homogéneas, sino también estacionarias.

La distinta fenomenología asociada a la hidrología de las cuencas pluviales respecto a las de tipología predominantemente nival, ha desembocado en una clara diferenciación de las expresiones matemáticas inferidas para relacionar los caudales máximos medios diarios con los máximos instantáneos en cada una de las dos tipologías mencionadas. Asimismo, la aplicación del procedimiento de análisis regional adoptado ha conducido a una distribución de estaciones coherente con el punto anterior, identificándose 9 regiones aceptablemente homogéneas en las que no se mezclan las cuencas pluviales con las de tipología predominantemente nival.

El análisis de robustez basado en las simulaciones de Monte Carlo ha reconocido a la PE3 como la distribución más robusta para todas y cada una de las 9 regiones identificadas, obteniéndose los menores RMSE para los cuantiles asociados a periodos de retorno comprendidos entre los 10 y 10000 años. Debido a esto, dicha distribución se alza como la mejor candidata para modelizar la frecuencia de ocurrencia de los caudales punta en la vertiente cantábrica y noratlántica de España, pese a haber resultado la función de mejor ajuste en sólo 2 de las 9 regiones.

Hasta donde conocen los autores, los resultados obtenidos a partir del trabajo desarrollado proporcionan estimaciones de cuantiles más fiables y precisas que otros publicados hasta la fecha, arrojándose valores RMSE de menor orden de magnitud.

AGRADECIMIENTOS

La investigación presentada ha sido parcialmente financiada por el Grupo de Tecnología e Innovación de GAS NATURAL FENOSA, al cual se quiere dar un reconocimiento especial por apostar por el desarrollo de este trabajo. Adicionalmente, los autores agradecen la colaboración mostrada por ENDESA e HIDROCANTÁBRICO, cuya información facilitada ha resultado muy útil para enriquecer el estudio con una mayor cantidad de registros.

REFERENCIAS

- Álvarez, M. 1999. *Análisis regional de frecuencia aplicado a las precipitaciones máximas y avenidas*. Universidad de Santiago de Compostela, Galicia, Spain.
- Benson, M. A. 1962. Evaluation of methods for evaluating the occurrence of floods. *Water Supply Paper 1550-A*. (1983), 589-590. United States Geological Survey, Reston, Va., USA.
- CEDEX, 2009. *Mapa de caudales máximos de avenida para la red fluvial de la España peninsular. Informe Técnico para la Dirección General del Agua del MARM*. Centro de Estudios Hidrográficos del CEDEX, Madrid, Spain.
- Chiew, F., Siriwardena, L. 2005. *Trend/Change detection software. USER GUIDE. CRC for Catchment Hydrology*, Australia.
- Cunnane, C. 1988. Methods and merits of regional flood frequency analysis. *Journal of Hydrology* 100, 269-290.
- Dalrymple, T. 1960. Flood frequency analysis. *USGS Water Supply Paper 1543-A*.
- Florentino, M., Gabriele, S., Rossi, F., Versace, P. 1987. Hierarchical approach for regional flood frequency analysis. *D. Reidel Publishing Company* 35-49.

- Fisher, R., Tippet, L. H. C. 1928. Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* 24, 180-190.
- Greis, N. P., Wood, E. F. 1981. Regional flood frequency estimation and network design. *Water Resources Research* 19 (1983), 589-590.
- Gumbel, E. J. 1958. *Statistics of extremes*. Columbia University Press, New York, USA.
- Hosking, J. R. M., (1990). L-moments: Analysis and Estimation of Distributions using Linear Combinations of order Statistics. *Journal of the Royal Statistical Society, Ser. B*, 52(1), 105-124.
- Hosking, J. R. M., Wallis, J. R. 1997. *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press., UK.
- Hosking, J.R.M. 2014a. *Regional frequency analysis using L-moments*. R package: lmomRFA, versión 3.0: July, 2, 2014. (<http://cran.r-project.org/web/packages/lmomRFA/index.html>).
- Hosking, J.R.M. 2014b. *L-moments*. R package: lmom, versión 2.4: July, 2, 2014. (<http://cran.r-project.org/web/packages/lmom/index.html>).
- Jiménez, A., Mediero, L., García C. 2014. Análisis y selección de modelos estadísticos para el ajuste de la ley de frecuencia de caudales máximos anuales en España. *Ingeniería Civil* 174, 5-31
- Jingyi, Z., Hall, M. J. 2004, Regional flood frequency analysis for the Gan-Ming River basin in China. *Journal of Hydrology* 296, 98-117.
- Kjeldsen, T. R., Smithers, J. C., Schulze R. E. 2002. Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. *Journal of Hydrology* 255, 194-211.
- Kochanek, K., Renard, B., Arnaud, P., Aubert, Y., Lang, M., Cipriani, T., Sauquet, E. 2014. A data-based comparison of flood frequency analysis methods used in France. *Natural Hazards and Earth System Sciences* 14, 295–308.
- Kundzewicz, Z.W., Robson, A. (Editors) 2000. *Detecting Trend and Other Changes in Hydrological Data*. World Climate Program – Water, WMO/UNESCO, WCDMP-45, WMO/TD 1013, Geneva.
- Natural Environment Research Council 1975. *Flood Studies Report, vol I*. Natural Environment Research Council, London, UK.
- Potter, K. W., Lettenmaier, D. P. 1990. A comparison of regional flood frequency estimation methods using a resampling method. *Water Resources Research* 26, 415-424.
- Salas, J., Obeysekera, J. 2014. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *American Society of Civil Engineers*, doi:10.1061/(ASCE)HE.1943-5584.0000820.
- Smith, A., Sampson, C., Bates, P. 2015. Regional flood frequency analysis at the global scale. *Water Resources Research* 51, 539-553. doi:10.1002/2014WR015814.
- U. S. Water Resources Council 1981. Guidelines for determining flood flow frequency. *Bulletin 17B*. Hydrology Committee, Washington D. C., USA.
- Ward, J. H. 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58, 236–244.
- WMO, 2009. *Guide to Hydrological Practices*. World Meteorological Organization, Geneva, Switzerland.
- Wood, E. F., Rodriguez-Iturbe, I. 1975. Bayesian Inference and Decision Making for Extreme Hydrologic Events. *Water Resources Research* vol. 11, No 4, 533-542.