

Análisis regional de frecuencia de avenidas en la vertiente cantábrica y noratlántica de España

Regional Frequency Analysis of extremes flows in Northern of Spain

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

^a Gas Natural Fenosa, Ingeniería y Desarrollo de Generación, Calle Acanto 11-13 pl. 01 28045 Madrid.
E-mail: ^{a1} jmontesm@gasnatural.com, ^{a2} jmoralo@gasnatural.com, ^{a3} jbaztan@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

Recibido: 20/10/2017

Aceptado: 12/12/2017

Publicado: 30/04/2018

Citar como: Montes, J., Álvarez, M., Pertierra, L., Moralo, J., Baztán, J. 2018. Regional Frequency Analysis of extremes flows in Northern of Spain. *Ingeniería del agua*, 22(2), 93-107. <https://doi.org/10.4995/ia.2018.8782>

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; L-momentos; caudales punta; regresión múltiple

ABSTRACT

This paper presents a study of regional frequency analysis of peak flows in Cantabrian and North Atlantic basins of Spain, based on the index-flood procedure with estimated parameters by L-moments approach (Hosking and Wallis, 1997). Observed annual maximum flow series of 85 gauged catchment areas in natural regime with records from 15 to 72 years were analyzed. The discordancy and heterogeneity measures based on the L-moments suggest 9 statistical homogeneous regions in which the same frequency distribution function is shared except for a scale factor. The main contribution of the study is the distinction of the regions according to their pluvial or snow character and the model capacity to estimate extreme peak flows with greater robustness than other methodologies.

Key words | Statistical techniques; L-moments; peak flows; multiple lineal regression.

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 o la realizada en los Estados Unidos de América respecto a la asimetría regional modelizada bajo la función de distribución log-Pearson tipo III. 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.

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, 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, 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).

En el caso de España, una de las variantes a la anterior metodología aplicada al estudio de los caudales máximos anuales se refiere al procedimiento conocido 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.

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 y Pesca, Alimentación y Medio Ambiente y Centro de Estudios y Experimentación de Obras Públicas. Asimismo, se consultó a todas aquellas entidades privadas con competencia sobre presas ubicadas dentro de la zona en estudio, explorando la posibilidad de que aportaran los registros 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.mapama.es/redes-seguimiento>, <http://ceh-flumen64.cedex.es/general/default.htm>, <http://www.bizkaia.eus>,
<http://www.gipuzkoa.eus>, http://www.navarra.es/home_es, <https://www.chminosil.es/es>, <http://www.meteogalicia.gal>,
<http://augasdegalicia.xunta.gal/>, <https://www.chcantabrico.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, 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 y Robson, 2000) implementadas en el modelo TREND (Chiew y 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 recibo

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 de los 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 resultan aceptables, por 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:

$$k = 1 + 6.1 \cdot 10^{-4} \cdot \frac{UTM_x^{0.453} \cdot P_{max}^{1.271}}{A_c^{1.014} \cdot P_f^{1.709}} \quad (3)$$

donde:

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

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

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

$$UTM_x = \text{Coordenada x asociada al sistema de proyección UTM, huso 30, datum ETRS89 [m]}$$

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

$$A_c = \text{Área de la cuenca [km}^2\text{]}$$

$$P_f = \text{Parámetro de forma de la cuenca (Perímetro [km] / Área [km}^2\text{])}$$

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 UTM_x en la zona en estudio acaba por desplazarla de la ecuación.

Cuenca de tipología nival:

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

donde:

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

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

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

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

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

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

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.

	R ² 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 $k \sim A_c + UTMx + P_f + Pmax$ (sólo cuencas pluviales)	0.82	-0.02%	8.3%

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

	R ² 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%

De este modo, 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).

Resulta importante indicar que sólo un 25% de los caudales máximos instantáneos utilizados en el análisis de frecuencia han sido obtenidos mediante estimación realizada a partir de máximos medios diarios, correspondiendo el 75% restante a registros propios de caudal máximo instantáneo. Esta mayoría de registros aporta mayor fiabilidad a los resultados obtenidos en el análisis de frecuencia, si bien la ampliación de datos realizada ha permitido disponer de series temporales más largas que ayudan a mejorar la estimación de los cuantiles más extremos, al reducir las incertidumbres asociadas a los mismos. Desde un punto de vista estadístico, la ampliación de las series no parece introducir distorsiones significativas de la homogeneidad de partida, teniendo en cuenta que las series ampliadas de las diferentes estaciones han encontrado su pertenencia a regiones aceptablemente homogéneas que resultan consistentes con el análisis *cluster* desarrollado para tal fin. Este análisis *cluster* ha sido realizado mediante el método de Ward (Ward, 1963), tomando como variables de análisis el área de la cuenca (km²), la altura media de la cuenca (m.s.n.m.), la precipitación promedio anual (mm) y las coordenadas de localización (UTM ETRS89). Mediante dicho análisis *cluster* se obtuvo

un total de 6 regiones diferenciadas, cuyas estaciones fueron sometidas a un análisis de heterogeneidad mediante la evaluación de las respectivas medidas D y H .

La identificación de estas 6 regiones proporcionó un punto de partida muy útil, desarrollándose posteriormente 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 sus cuantiles asociados.

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 y Pesca, Alimentación y Medio Ambiente.*

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)	t_5	$D_{\text{máx}}$ (D_{crit})	H_1	H_2	H_3
R1	12	0.3418	0.3539	0.2475	0.1448	1.93 (2.76)	0.52	-0.50	-0.66
R2	14	0.2935	0.1778	0.1449	0.1049	2.72 (2.97)	0.66	-0.49	-0.56
R3	4	0.3579	0.2409	0.1505	0.0663	-	-0.28	1.46	1.88
R4	4	0.2092	0.1170	0.1287	0.0352	-	0.37	-0.55	-0.33
R5	5	0.3398	0.2961	0.2196	0.1093	1.30 (1.33)	-0.16	0.52	0.08
R6	7	0.2604	0.1319	0.1346	0.0394	1.31 (1.92)	-0.28	0.53	0.93
R7	10	0.2599	0.2994	0.2384	0.1135	1.61 (2.49)	0.56	-0.27	-0.30
R8	5	0.2702	0.2581	0.1806	0.0532	1.27 (1.33)	0.40	-0.40	-0.33
R9	10	0.2751	0.2649	0.1708	0.1343	2.37 (2.49)	-0.48	-0.12	1.30
R10	5	0.2424	0.1185	0.0945	0.0316	1.21 (1.33)	-0.91	0.28	0.48

De las tres medidas de heterogeneidad, se puede observar cómo el valor absoluto de la de mayor trascendencia (H_1) se encuentra por debajo de 1 en todas las regiones.

Por otro lado, se puede observar cómo las medidas H_2 y H_3 asociadas a 8 de las 10 regiones propuestas arrojan valores asimismo por debajo de 1, obteniendo únicamente un valor superior a 1 en la región R9 para la medida H_3 , y en la región R3 para las medidas H_2 y H_3 , si bien en cualquier caso por debajo del valor crítico de 2.

Para el caso concreto de la región R3, se ha podido comprobar que la posible heterogeneidad asociada a H_2 y H_3 se debe en gran parte a la inclusión de la estación 1805, cuya cuenca de cierre se corresponde en un gran porcentaje con la laguna de Antela. De hecho, al excluir dicha estación de la región R3 en cuestión, las medidas H_2 y H_3 resultan entonces menores que 1, si bien entonces es la medida H_1 la que asciende levemente por encima de la unidad. En este sentido, puesto que la cuenca de cierre de la estación de Conchas podría verse asimismo influenciada por la presencia de la laguna de Antela, se pasó a analizar la homogeneidad de únicamente las dos estaciones cuyas cuencas de cierre no incluían dicho accidente geográfico, obteniéndose entonces valores de las medidas H_1 , H_2 y H_3 apreciablemente por debajo de 1. De esta manera, todo el análisis realizado sobre la región R3 apunta a que la influencia de la laguna de Antela en 2 de las 4 estaciones que conforman la región parece estar introduciendo una apreciable heterogeneidad, cuyo análisis asociado deberá ser ampliado en futuras investigaciones mediante la inclusión de nuevas estaciones localizadas en la margen izquierda del río Miño, relativas a la zona de Portugal.

Por otro lado, resulta interesante la distinción de las 2 regiones obtenidas en la zona del País Vasco, cuya disposición no responde a una clara separación geográfica sino a las diferentes superficies de cuenca delimitadas por cada una de las estaciones de aforo. Dicha distinción encuentra su interpretación física en la mayor probabilidad de ocurrencia de precipitaciones extremas cuando éstas son acotadas a extensiones de espacio más reducidas. En este sentido, se ha podido comprobar cómo la unión de R9 y R10 en una única región conduce a una alta heterogeneidad asociada a las medidas H_1 , H_2 y H_3 , no siendo recomendable considerar toda la zona cantábrica vasca como una única región estadística.

Adicionalmente al análisis de heterogeneidad anteriormente presentado, se evaluaron las distribuciones candidatas según la bondad de ajuste Z^{Dis} propuesta por Hosking y Wallis, cuyos resultados se muestran en la tabla 4.

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

Región	Umbral orientativo $ Z^{\text{Dis}} \leq 1.64$	Z^{Dis} Mejor
R1	GLO (0.17), GEV (-0.53), GNO (-1.44), PE3 (-3.01), GPA (-2.71)	GLO
R2	GNO (0.03), GEV (0.35), PE3 (-0.75), GLO (2.44), GPA (-4.36)	GNO
R3	PE3 (-0.28), GPA (-0.39), GNO (0.44), GEV (0.86), GLO (1.30)	PE3
R4	PE3 (-0.05), GEV (0.11), GNO (0.16), GLO (1.64), GPA (-3.08)	PE3
R5	GLO (0.20), GEV (-0.41), GNO (-0.89), PE3 (-1.63), GPA (-2.05)	GLO
R6	GNO (0.08), GEV (0.09), PE3 (-0.26), GLO (1.88), GPA (-3.69)	GNO
R7	GLO (-0.37), GEV (-1.16), GNO (-1.77), PE3 (-2.85), GPA (-3.30)	GLO
R8	GEV (0.14), GNO (-0.28), GLO (0.95), PE3 (-1.03), GPA (-1.92)	GEV
R9	GNO (-0.04), GEV (0.54), PE3 (0.85), GLO (1.43), GPA (-1.76)	GNO
R10	PE3 (1.21), GEV (1.41), GNO (1.45), GPA (-1.99), GLO (3.05)	PE3

Desde un punto de vista global, cabe resaltar que la distribución PE3 muestra un ajuste aceptable en 8 de las 10 regiones consideradas, siendo además la distribución que proporciona los menores RMSE en todas y cada una de esas 8 regiones. Es por ello que resulta razonable asumir dicha función de distribución para modelizar la frecuencia de los caudales punta extremos en todas las regiones, excepto en R1 y R7 en donde el mal ajuste obtenido por la función PE3 ha conducido a adoptar la función de distribución WAK como la más adecuada. En este sentido, es importante indicar que pese a la gran flexibilidad aportada por los 5 parámetros de la función WAK para adaptarse a imitar cualquier función de distribución de 2 ó 3 parámetros, dicha función de distribución sólo se presenta como la más apropiada para las regiones R1 y R7, teniendo en cuenta que únicamente es en ellas donde la función WAK llega a proporcionar menores RMSE que el resto de distribuciones.

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 tabla 5 se muestran los resultados obtenidos para toda la zona en estudio, en donde se pueden consultar los principales cuantiles regionales y los valores de RMSE obtenidos.

Tabla 5 | Cuantiles regionales y RMSE relativo obtenidos para cada región.

Región	T (años)	10	100	1000	10000
	F	0.9	0.99	0.999	0.9999
R1	q(F)	1.800	3.899	7.174	12.283
	RMSE rel	0.054	0.115	0.211	0.349
R2	q(F)	1.723	2.657	3.503	4.308
	RMSE rel	0.043	0.099	0.133	0.157
R3	q(F)	1.903	3.235	4.498	5.728
	RMSE rel	0.060	0.188	0.270	0.331
R4	q(F)	1.502	2.067	2.552	2.999
	RMSE rel	0.048	0.101	0.146	0.181
R5	q(F)	1.875	3.313	4.723	6.119
	RMSE rel	0.052	0.142	0.192	0.224
R6	q(F)	1.629	2.362	3.001	3.594
	RMSE rel	0.036	0.102	0.149	0.186
R7	q(F)	1.614	3.038	5.017	7.769
	RMSE rel	0.054	0.125	0.229	0.390
R8	q(F)	1.686	2.733	3.737	4.720
	RMSE rel	0.049	0.122	0.164	0.193
R9	q(F)	1.700	2.784	3.826	4.850
	RMSE rel	0.031	0.068	0.090	0.105
R10	q(F)	1.582	2.239	2.804	3.325
	RMSE rel	0.052	0.124	0.174	0.213

Los anteriores resultados ponen de manifiesto la buena capacidad predictiva del modelo adoptado, cuyos RMSE relativos para periodos de retorno superiores a los 100 años arrojan un valor medio de 0.16. De una forma global, estos resultados presentan una mayor robustez que la de otros trabajos desarrollados para la misma zona en estudio, proporcionando valores de incertidumbre apreciablemente menores. A modo de ejemplo, se muestra a continuación una comparativa de los errores medios obtenidos en este trabajo para la región R6, frente a los alcanzados por el CAUMAX en su región 12 y cuya área de influencia contiene a las 7 estaciones de la región R6. La Figura 2 ilustra la apreciable reducción de los errores medios obtenidos mediante el presente trabajo respecto a los proporcionados por el CAUMAX, alcanzándose reducciones del error cuadrático medio relativo de hasta un 30%.

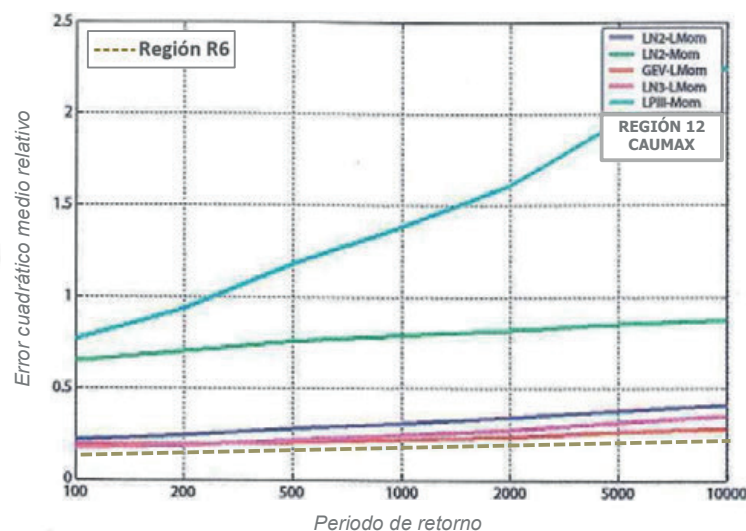


Figura 2 | Comparativa de errores obtenidos por la región R6 del presente estudio frente a la región 12 del CAUMAX. Nota: La fuente original de la figura 2 pertenece a Jiménez *et al.*, 2014, a la cual se han añadido los errores obtenidos por el presente trabajo representándolos mediante una curva discontinua marrón.

Finalmente, como resumen de los resultados obtenidos se presentan a continuación las curvas de cuantil regional asociadas a la función de distribución adoptadas, cuyo análisis cualitativo apunta hacia una menor curvatura de la función cuanto más húmedo resulta el clima de la región en cuestión.

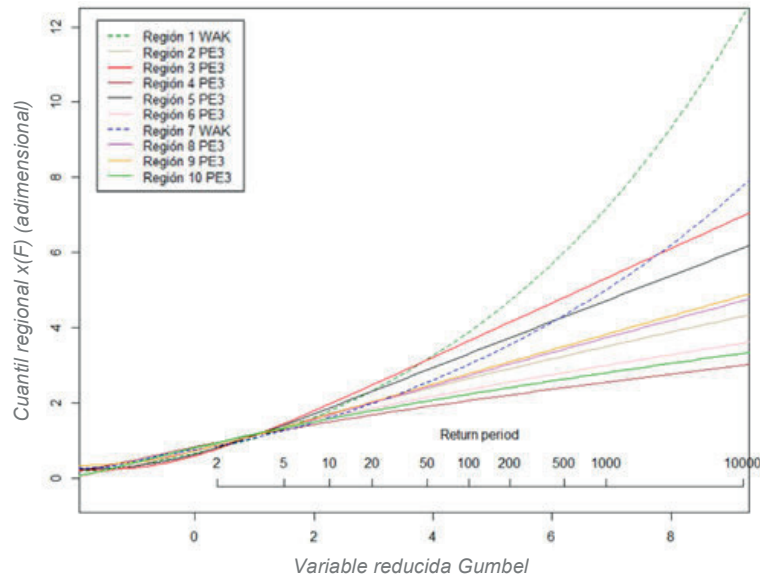


Figura 3 | Curvas de cuantil regional.

El análisis cuantitativo entre la curvatura de las funciones de cuantil regional y la climatología de la zona muestra, nuevamente, cómo la fenomenología asociada a las máximas avenidas anuales que se originan en las cuencas de tipología nival resulta físicamente diferente a las que se originan en las de tipo pluvial, ya que la cantidad de nieve acumulada durante los meses fríos resulta trascendental respecto a la magnitud de las avenidas máximas anuales por el efecto que produce su posterior fusión. En la figura 4 se muestra la buena relación existente entre los cuantiles extremos y la precipitación media anual promediada para cada una de las regiones de tipología nival, correlación que sin embargo se estropea para el caso de las regiones de tipo pluvial.

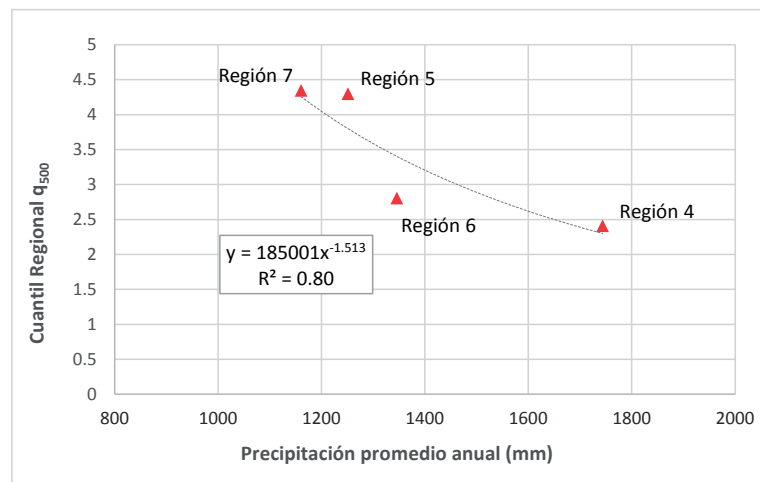


Figura 4 | Relación de cuantiles regionales asociados al periodo de retorno de 500 años respecto a la precipitación promedio anual para cuencas de tipología nival.

Para el caso de las cuencas de tipología pluvial, la figura 5 muestra cómo los cuantiles regionales extremos encuentran en este caso una elevada correlación con la precipitación máxima diaria anual, lo cual resulta consistente con el hecho físico de que

las avenidas máximas anuales en este tipo de cuencas vienen originadas exclusivamente por eventos de precipitación en forma de lluvia.

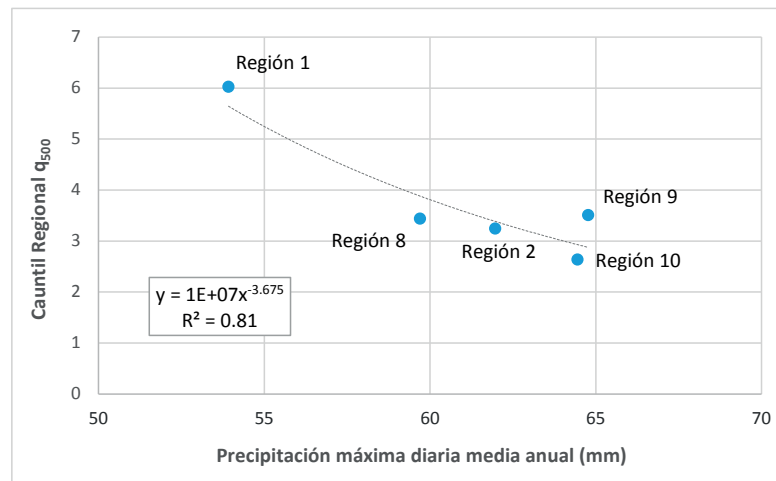


Figura 5 | Relación de cuantiles regionales asociados al periodo de retorno de 500 años respecto a la precipitación promedio anual para cuencas de tipología pluvial.

De la anterior figura 5 se puede observar que la buena correlación obtenida ha sido alcanzada sin contar con la región R3, ya que las dudas existentes sobre su homogeneidad requieren de ser resueltas mediante futuras investigaciones que incluyan nuevas estaciones localizadas en la margen izquierda del río Miño, de la zona de Portugal. De hecho, se ha podido confirmar cómo la inclusión de la región R3 en el análisis de correlación entre los cuantiles y las precipitaciones máximas medias diarias produce una reducción considerable del coeficiente R^2 (de 0.81 a 0.55), lo cual apoya la hipótesis de dudar de la precisión de los cuantiles regionales actualmente obtenidos para la región R3.

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 basando éste en el método del índice de avenida con sus parámetros estimados por los L-momentos. 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 las 7 regiones en las que el ajuste de dicha función resulta aceptable, alzándose la distribución WAK como la más robusta para las dos regiones restantes en las que la PE3 no presenta un ajuste aceptable.

Los resultados de los cuantiles extremos hallados encuentran una buena relación con las principales variables climatológicas de la zona en estudio, habiéndose obtenido para las regiones nivales una ajustada correlación entre el cuantil regional y la precipitación promedio anual, así como una ajustada relación entre el cuantil extremo de las regiones pluviales y la precipitación máxima diaria anual.

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.

REFERENCIAS

- Álvarez, M. 1999. *Análisis regional de frecuencia aplicado a las precipitaciones máximas y avenidas*. Universidad de Santiago de Compostela, Galicia, Spain.
- 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. [https://doi.org/10.1016/0022-1694\(88\)90188-6](https://doi.org/10.1016/0022-1694(88)90188-6)
- Dalrymple, T. 1960. Flood frequency analysis. *USGS Water Supply Paper 1543-A*.
- Fisher, R., Tippett, 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. <https://doi.org/10.1017/S0305004100015681>
- 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. <https://doi.org/10.1017/CBO9780511529443>
- Hosking, J.R.M. 2014a. *Regional frequency analysis using L-moments*. R package: lmomRFA, versión 3.0: July, 2, 2014. Accesible en <https://cran.r-project.org/package=lmomRFA>. Último acceso: abril 2018.
- Hosking, J.R.M. 2014b. *L-moments*. R package: lmom, versión 2.4: July, 2, 2014. Accesible en <https://cran.r-project.org/package=lmomRFA>. Último acceso: abril 2018.
- <http://augasdeg Galicia.xunta.gal>
- <http://www.bizkaia.eus>
- <http://www.chcantabrico.es/index.php/es/>
- <https://www.chminosil.es/es>
- <http://www.gipuzkoa.eus>
- <http://ceh-flumen64.cedex.es/general/default.htm>
- <http://www.meteogalicia.gal>
- http://www.navarra.es/home_es
- <http://sig.mapama.es/redes-seguimiento/>
- 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

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.

Potter, K. W., Lettenmaier, D. P. 1990. A comparison of regional flood frequency estimation methods using a resampling method. *Water Resources Research*, 26(3), 415-424. <https://doi.org/10.1029/WR026i003p00415>

Ward, J. H. 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of American Statistical Association*, 58(301), 236-244. <https://doi.org/10.1080/01621459.1963.10500845>

WMO, 2009. *Guide to Hydrological Practices*. World Meteorological Organization, Geneva, Switzerland.