

# Análisis regional de frecuencias de las precipitaciones diarias extremas en Navarra. Elaboración de los mapas de cuantiles

## *Regional frequency analysis of annual maximum daily rainfall in Navarra. Quantiles mapping*

López, J.J.<sup>a1</sup>, Goñi, M.<sup>a2</sup>, San Martín, I.<sup>a3</sup>, Erro, J.<sup>a4</sup>

<sup>a</sup>Departamento de Proyectos e Ingeniería Rural; Universidad Pública de Navarra; Campus de Arrosadía; Pamplona; 31006-Navarra.  
E-mail: <sup>a1</sup> [jjlr@unavarra.es](mailto:jjlr@unavarra.es), <sup>a2</sup> [mikel.goni@unavarra.es](mailto:mikel.goni@unavarra.es), <sup>a3</sup> [ibansanmartin@gmail.com](mailto:ibansanmartin@gmail.com), <sup>a4</sup> [juanerroeslava@hotmail.com](mailto:juanerroeslava@hotmail.com).

Recibido:30/04/2018

Aceptado: 20/12/2018

Publicado: 31/01/2019

**Citar como:** J.J. López, Goñi, M., San Martín, I., Erro, J. 2019. Regional frequency analysis of annual maximum daily rainfall in Navarra. Quantiles mapping. *Ingeniería del agua*, 23(1), 33-51. <https://doi.org/10.4995/la.2019.10058>

## RESUMEN

La determinación de la ley de frecuencias de precipitaciones resulta imprescindible para el diseño de diferentes infraestructuras hidráulicas así como para el análisis y determinación de zonas inundables. El objetivo de este artículo es presentar los cuantiles de las precipitaciones diarias extremas en el territorio de la Comunidad Foral de Navarra, obtenidos mediante un análisis regional de frecuencia (ARF), y su representación espacial. Se ha partido de las 142 estaciones pluviométricas manuales, localizadas en la Comunidad Foral de Navarra y en el entorno de la misma, con duración de registro superior a 20 años. El Análisis Regional de Frecuencias (ARF) se ha realizado según la metodología de Hosking y Wallis, obteniendo seis regiones homogéneas y sus funciones de distribución más adecuadas. Para la elaboración de los mapas se han aplicado dos métodos de interpolación: el de la distancia inversa ponderada; y el geoestadístico Kriging ordinario. Después del análisis comparativo se ha elegido el Kriging.

**Palabras clave** | precipitaciones máximas en Navarra; análisis regional de frecuencias de precipitación; representación espacial de precipitaciones; interpolación distancia inversa ponderada; Kriging.

## ABSTRACT

*The determination of the law of frequency of precipitations is essential for the design of different hydraulic infrastructures as well as for the analysis and determination of flood areas. The objective of this paper is to present the quantiles of the extreme daily rainfall in Navarre, obtained through a regional frequency analysis, and its spatial representation. 142 manual rainfall stations, located in the Autonomous Community of Navarra and in the surrounding area, have been considered for the study. The duration of registration of the series is greater than 20 years. The Regional Frequency Analysis has been carried out according to the Hosking and Wallis methodology obtaining six homogeneous regions and their most appropriate distribution functions. The mapping has been made applying two interpolation methods: the inverse distance weighted; and the ordinary geostatistical Kriging. The selected method has been the Kriging.*

**Key words** | maximum rainfall in Navarra; regional frequency analysis; spatial representation of rainfall; inverse distance weighted interpolation; Kriging.

## INTRODUCCIÓN

La determinación de la ley de frecuencias de precipitaciones y caudales máximos resulta imprescindible para el diseño de diferentes infraestructuras hidráulicas así como para el análisis y determinación de zonas inundables. Tradicionalmente la estimación de los cuantiles, precipitaciones asociadas a periodos de retorno, se ha realizado mediante el análisis de frecuencia de la serie de precipitaciones máximas anuales registradas en una estación pluviométrica. Sin embargo, con frecuencia, la escasa longitud de los registros y la presencia de valores atípicos (*outliers*) hace que esta forma de proceder carezca del rigor suficiente y, por lo tanto, no sea la más adecuada. Este problema puede resolverse “paliando la carencia de datos en el tiempo con su abundancia en el espacio” cuando se dispone de observaciones en diferentes estaciones más o menos próximas entre sí. Con esta finalidad surge el Análisis Regional de Frecuencias (ARF). Dentro de las metodologías existentes, en este trabajo se ha aplicado la metodología de Hosking y Wallis (1997) que utiliza el método del índice de avenida (Dalrymple, 1960) conjuntamente con momentos lineales (momentos-L). Se trata de una metodología robusta, aun en presencia de *outliers*, y su correcta aplicación puede conducir a estimaciones suficientemente precisas en diversos problemas reales.

La literatura disponible sobre el Análisis Regional de Frecuencias (ARF) de precipitaciones extremas es numerosa. Algún trabajo como el de Bilham se remonta al año 1936 para el análisis regional de las precipitaciones extremas en Inglaterra y Gales. Más recientemente, Ayoade (1976) realizó el análisis regional de la precipitación diaria en Nigeria. Schaefer (1990), siguiendo la metodología del índice de avenida de Dalrymple (1960), encontró que la distribución de valor extremo II era la mejor distribución para los datos de precipitación anual en el estado de Washington. Pearson *et al.* (1991) aplicaron el ARF en Australia, Vogel *et al.* (1993) en Estados Unidos de América. Cannarozzo *et al.* (1995) realizaron un ARF para la región siciliana usando la distribución de valores extremos de dos componentes para la precipitación máxima anual de varias duraciones. Adamowski *et al.* (1996) aplicaron el análisis regional de la precipitación en Canadá, concluyendo que la distribución de Valores Extremos Generalizada (GEV) utilizando momentos-L proporcionó las estimaciones más fiables. Parida (1998) obtuvo las estimaciones de cuantiles para las lluvias de los monzones de verano utilizando la distribución kappa de cuatro parámetros. Park *et al.* (2001) utilizaron la distribución Wakeby con el método de estimaciones de momentos-L para obtener los cuantiles de las precipitaciones extremas de verano en 61 estaciones de Corea del Sur. Smithers y Schulze (2001) aplicaron el ARF basado en momentos-L para estimar la lluvia de diseño de corta duración en Sudáfrica, también en este país lo aplicaron Kjeldsen *et al.* (2002). Sveinsson *et al.* (2002) también lo aplicaron en el noreste de Colorado. Lee y Maeng (2003) encontraron que las distribuciones GEV y Logística Generalizada, GLO, fueron las más adecuadas para las precipitaciones diarias extremas en Corea. Otros ejemplos de ARF de la lluvia son los trabajos de Trefry *et al.* (2005) para las precipitaciones sobre el estado de Michigan; Caporali *et al.* (2006) para lluvias extremas en Toscana; Lin y Chen (2006) para los datos de precipitaciones en Taiwán; Weaver (2006) para las precipitaciones en la ciudad de Charlotte y Mecklenburg (Carolina del Norte); Kysely y Picek (2007) y Kysely *et al.* (2007) para precipitación extrema en República Checa; y Norbiato *et al.* (2007) para lluvias extremas en los Alpes orientales italianos. Recientemente hay muchos ejemplos más como Yang *et al.* (2010); Ngongondo *et al.* (2011); Shahzadi *et al.* (2013); Malekinezhad y Zare-Garizi (2014); García-Marín *et al.* (2015); Liu *et al.* (2015); Blanchet *et al.* (2016); Halbert *et al.* (2016); Abdi *et al.* (2017); Hailegeorgis y Alfreðsen (2017). En España también se han realizado diversos trabajos como el realizado por el Centro de Estudios Hidrográficos del CEDEX para caracterizar los ríos de la península (Álvarez y Orduña, 2014), o los llevado a cabo para caracterizar las precipitaciones en Galicia (Álvarez *et al.*, 1999), en Andalucía (García-Marín, 2007) y en la vertiente Cantábrica (Montes *et al.*, 2018).

El trabajo, cuyos resultados se presentan en esta comunicación, ha tenido como objetivo la determinación de los cuantiles de las precipitaciones diarias extremas en el territorio de la Comunidad Foral de Navarra y la elaboración de los mapas correspondientes mediante interpolación espacial. Para ello se han dispuesto de todas las series de precipitación diaria registradas en todas las estaciones pluviométricas manuales localizadas en la Comunidad Foral de Navarra y en las comunidades vecinas y Francia próximas al límite de Navarra. En total se han utilizado un total de 142 estaciones. Estas estaciones fueron elegidas a priori por tener una duración de registro superior a 20 años.

En el análisis regional de frecuencias (ARF) propuesto por Hosking y Wallis (1997) una vez determinados los momentos-L de las series, se realiza un filtrado primario de las observaciones, obteniendo la discordancia de las series, y antes de proceder a ajustar las funciones de distribución, es necesario proceder a identificar las regiones homogéneas que se encuentran en la zona de estudio, cuyo objetivo es formar grupos de estaciones que satisfagan aproximadamente la condición de homogeneidad al objeto

de compartir la misma distribución de frecuencias. Definidas las regiones, se procede a determinar la función de distribución más apropiada para cada una de ellas. Estos autores recomiendan la elección entre las siguientes cinco distribuciones: Logística Generalizada (*Generalized Logistic*, GLO), Valores Extremos Generalizada (*Generalized Extreme Value*, GEV), Pareto Generalizada (*Generalized Pareto*, GPA), Normal Generalizada (*Generalized Normal*, GNO), y Pearson tipo III (PE3). Estos autores justifican la elección entre estas cinco distribuciones por el hecho de poseer tres parámetros, lo cual facilita el ajuste en un análisis regional al disponerse de observaciones de varias estaciones, y de incluir entre sus casos particulares las distribuciones más empleadas en el estudio de variables ambientales. Obtenidas las funciones de distribución más adecuadas para cada región se procede a determinar la curva de cuantiles aplicando el método del índice de avenida (Dalrymple, 1960).

Existen diferentes técnicas o métodos de interpolación (Hengl, 2009) que se pueden clasificar en función de diferentes criterios (Akkala *et al.*, 2010). En este estudio, para la elaboración de los mapas de cuantiles en Navarra se han analizado dos métodos de interpolación espacial: el de la distancia inversa ponderada (*Inverse distance weighted-IDW*); y el geostatístico Kriging ordinario (Delhomme, 1978; Oliver and Webster, 1990). El método IDW se clasifica de manera general como un método exacto, local y determinístico, y el método de Kriging como un método aproximado, local y estocástico, aunque en función de la metodología al aplicar cada uno podría cambiar la clasificación.

## MATERIAL Y MÉTODOS

### Descripción de las series de datos

Para el presente estudio se ha partido de todas las estaciones pluviométricas manuales localizadas en la Comunidad Foral de Navarra y en las comunidades vecinas y Francia próximas al límite de Navarra. Las series de datos registradas en estas estaciones han sido aportadas por la Sección de Suelos y Climatología del Servicio de Estructuras Agrarias del Gobierno de Navarra del estudio Agroclimático de Navarra actualizado (G.N., 2001).



Figura 1 | Localización y agrupación por regiones de todas las estaciones seleccionadas.

Conviene señalar que, para la realización del estudio con un mayor rigor estadístico, en principio se prescindió de aquellas estaciones cuya duración de registro fuera inferior a 20 años por considerarse demasiado cortas y no representativas en el tiempo, posteriormente se incluyeron las estaciones de Tarazona y SOS del Rey Católico, con 19 años, por situarse en zonas con poca densidad de estaciones. En la Figura 1 se muestra la localización de las estaciones seleccionadas después del filtrado.

En la Tabla 1 se presenta el número de años de registro de la serie de cada estación. En total se han utilizado un total de 142 estaciones: 86 de Navarra, 16 de Gipuzkoa, 6 de Araba, 7 de La Rioja, 9 de Zaragoza, 6 de Huesca y 12 de Francia.

Una vez seleccionadas las estaciones y hecho un primer chequeo de fallos de datos, se seleccionaron las series de precipitaciones diarias máximas anuales de cada estación.

**Tabla 1 |** Estaciones seleccionadas con la longitud de la serie en años, y el valor de la discordancia,  $D_i$ , considerando una única región.

Estación	Nº años	$D_i$	Estación	Nº años	$D_i$	Estación	Nº años	$D_i$
<b>Navarra (Total Estaciones 86)</b>								
Abaurrea A	60	0.58	Cáseda	41	1.02	Lezáun	27	0.28
Aibar	41	0.66	C. Arrambide	28	0.41	Lodosa	36	0.89
Alloz	58	0.14	Corella	47	<b>3.06</b>	L.Arcos	32	1.06
Altsasu	102	0.32	Doneztebe	85	0.37	Luzaide	41	1.85
Amañur	39	1.27	Epároz	41	0.28	Miranda	32	0.59
Amillano	40	0.44	Erro	40	0.18	Monreal	35	0.32
Andosilla	24	0.62	E.Salazar	41	1.07	Monteagudo	85	0.69
Aoiz	22	1.46	Estella	23	0.88	Navascués	35	0.89
Areso	27	0.81	Etxalar	27	0.97	Noáin	38	0.38
Aribe	44	0.15	Eugi	47	0.34	Olagüe	40	0.42
Arizkun	38	1.4	Falces	46	0.77	Olite	83	0.57
Arróniz	48	0.24	Fitero	82	0.74	Olóriz	33	0.94
Artieda	48	0.57	Galbarra	40	0.73	Oroz B	22	0.76
Artikutza	33	0.9	Genevilla	25	0.12	Otazu	39	0.37
Aurizberri	43	1.21	Goizueta	47	0.28	Pamplona	129	0.47
Azanza	27	1.26	Goñi	40	0.75	P.Reina	43	0.38
Azpirotz	27	1.3	Igúzquiza	32	0.25	Sartaguda	91	1.01
Barásoain	36	0.26	Ilundáin	36	1.27	Sesma	27	0.43
Belate	28	0.62	Iraizotz	44	1.27	Sunbilla	36	0.58
Belzunce	27	0.03	Irotz	36	0.5	Tudela	29	2.26
Bera	35	1.1	Irurtzun	40	0.51	Urbasa	28	2.45
Bertiz	23	0.14	Javier	64	1.09	Urzainqui	34	1.16
Betelu	38	2.5	Larraona	40	0.71	Viana	33	0.27
Buñuel	85	2.04	Leire	37	0.92	Yesa	81	0.12
Cabanillas	33	<b>4.49</b>	Leitza	29	0.45	Zalba	22	2.94
Cábrega	20	1.54	Lerga	40	1.9	Zuazu	28	0.13
Cadreita	48	0.98	Lerín	42	1.01	Zubiri	47	0.35
Caparoso	84	1.26	Lesaka	34	<b>4.14</b>	Zugarramurdi	25	1.19
Carcastillo	82	0.85	Lesaka San Antón	24	2.77			
<b>Gipuzkoa (Total Estaciones 16)</b>								
Andoain	40	0.45	Hernani CS	40	1.38	Mutiloa	40	0.21
Añarbe e	53	0.23	Hondarribia A	50	0.31	Ordizia CA	45	0.97
Arditurri Oiartzun	54	0.89	Irún	25	1.31	Rentería Presa	50	0.31
Beasain	35	0.37	Lareo	50	1.91	Urrieta V	45	0.8
Elduayen	65	0.43	Legazpia B	50	1.08			
Fuenterrabia Z	40	0.35	Legorreta	40	0.39			

Tabla 1, continúa en la página siguiente

Tabla 1, continúa de la página anterior

Estación	Nº años	$D_i$	Estación	Nº años	$D_i$	Estación	Nº años	$D_i$
<b>Álava (Total Estaciones 6)</b>								
Arriola	44	2.09	El Ciego	25	1.77	Opacua	44	0.36
Bernedo	31	0.2	Laguardia	30	<b>3.05</b>	Santa Cruz Campezo DFA	33	0.4
<b>La Rioja (Total Estaciones 7)</b>								
Alfaro e	48	1.66	Cornago	28	2.68	Varea V	43	0.17
Arnedo	58	1.11	Logroño Agoncillo	60	0.31			
Cenicero i	62	0.25	Rincón del Soto	40	1.84			
<b>Zaragoza (Total Estaciones 9)</b>								
Biel	63	0.78	Ejea Caballeros	22	1.57	Sádaba	59	0.52
Borja Ayto	27	<b>4.94</b>	El Bayo	43	0.72	Sos del Rey Católico	19	<b>4.02</b>
Catilsar	22	0.78	Los Pintanos	34	1.31	Tarazona	19	1.37
<b>Huesca (Total Estaciones 6)</b>								
Aisa de Jaca	43	0.58	Bailo PR	37	0.67	Castiello de Jaca	76	0.91
Aratores	36	0.17	Canfrac LA	101	0.54	Jaca ECA	28	0.87
<b>Francia (Total Estaciones 12)</b>								
Accous UED	52	1.99	Espelette GB	37	1.08	Larrau	55	0.47
Arette V	31	0.5	Hosta V	47	1.12	Licq therey	55	0.9
Banca MJ	50	2.32	Irouleguy FM	50	0.19	ST Jean le Vx	32	0.76
Cambo Les Bains E	44	1.05	Irrisarry	36	1.26	Trois Villes	55	1.69

## Descripción del análisis de frecuencias regional

El Análisis Regional de Frecuencias (ARF) establecido por Hosking y Wallis (1997), cuya metodología se ha aplicado en este trabajo, consta de cuatro etapas:

- a. *Filtrado primario de las observaciones.* Es el primer paso en cualquier tipo de análisis estadístico. Está encaminado a detectar posibles errores o inconsistencias en unos datos de partida que han de ser homogéneos en el tiempo.
- b. *Identificación de regiones homogéneas.* Consiste en repartir las diferentes estaciones pluviométricas en regiones que se consideran homogéneas. La región es la unidad fundamental del análisis regional.
- c. *Elección de una distribución de frecuencia para la región.* Consiste en seleccionar la distribución regional de frecuencia,  $q(F)$ , más apropiada para la región, mediante pruebas de bondad de ajuste a las diversas funciones posibles.
- d. *Estimación de los cuantiles regionales de frecuencia.* Escogida la función de distribución adecuada para la región, se estiman los cuantiles asociados a un cierto periodo de retorno. Para ello se emplea el algoritmo regional de los momentos-L, a través del cual los parámetros que definen la distribución se estiman separadamente en cada estación, combinándose después para obtener una media regional. Con el método del índice de avenida, los valores regionales se particularizan en cada estación local. Además, para que los resultados obtenidos alcancen la mayor utilidad práctica posible y para poder calibrar su fiabilidad, se evalúa la magnitud de la incertidumbre de las estimaciones mediante simulación de tipo Montecarlo.

### Filtrado primario de las observaciones

Dos tipos de errores son habituales en las muestras: por un lado, posibles fallos en el grabado o la transcripción de los datos, que conducen a valores incorrectos; por otro, posibles cambios en las circunstancias bajo las cuales la variable fue medida, como puede ser el reemplazo de los aparatos de medición o su traslado de un lugar a otro. Hosking y Wallis (1997) establecieron que los cambios de tendencia, *outliers*, valores incorrectos y demás errores de una muestra quedan reflejados en sus momentos-L. Definen así la medida de la discordancia como un estadístico que compara los L-Coeficientes (L-CV,  $\tau$ , L-asimetría,  $\tau_3$ , L-curtosis,  $\tau_4$ ), que

son relaciones entre los momentos-L y se expresan en la Ecuación 1, de cada estación en una región con los del resto de estaciones y los valores medios de dicha región.

$$L - CV = \tau = \frac{\lambda_2}{\lambda_1} \quad L - \text{asimetría} = \tau_3 = \frac{\lambda_3}{\lambda_2} \quad L - \text{kurtosis} = \tau_4 = \frac{\lambda_4}{\lambda_2} \quad (1)$$

Así, se puede decir que, dado un grupo de estaciones, el objetivo de la discordancia es identificar las que presentan una anomalía o resultan discordantes con respecto al grupo. Definiendo el vector de L-Coeficientes ( $\tau$ ,  $\tau_3$ ,  $\tau_4$ ) de cada estación y considerándolo como un punto en el espacio tridimensional, el grupo de estaciones da lugar a una nube de puntos. Se considera discordante cualquier punto que quede “lejos” del centro de la nube.

Numéricamente, supónganse  $N$  estaciones en el grupo. Sea  $u_i = [t^{(i)} \ t_3^{(i)} \ t_4^{(i)}]^T$  el vector que contiene los valores de los ratios de momentos-L  $t$ ,  $t_3$  y  $t_4$  de la estación  $i$ , con el superíndice T indicando transposición del mismo. La discordancia,  $D_i$ , se define mediante la ecuación 2.

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \quad (2)$$

Donde  $\bar{u}$  es la media de  $u_i$  y  $A$  es la matriz de suma de cuadrados y productos cruzados.

La estación  $i$  se considera discordante si el valor de  $D_i$  es elevado, y ello está condicionado por el número de estaciones en el grupo. Hosking y Wallis (1997) establecieron que una estación ha de ser catalogada como discordante si su  $D_i$  supera un valor crítico establecido en función del número de sitios en la región.

A nivel práctico, se recomienda el empleo de la medida de discordancia en dos ocasiones. En primer lugar, al comienzo del análisis, procurando abarcar la totalidad de las estaciones o regiones suficientemente amplias como para que su número sea elevado. Con esta primera aproximación se trata de detectar aquellos sitios que difieren significativamente de la mayoría, sea por la presencia de valores atípicos o por la existencia de errores de bulto. En segundo lugar, una vez identificadas las diferentes regiones homogéneas, es aconsejable recalcular la medida de la discordancia de cada estación en su región asignada. Aquellas estaciones que resulten discordantes habrá que considerar la posibilidad de trasladarlas a otra región, si bien esta opción no prevalecerá sobre las similitudes físicas con el resto de sitios de su región asignada.

### Identificación de regiones homogéneas

De todas las etapas del análisis regional de frecuencias, la identificación de regiones homogéneas es a menudo la más complicada, al requerir de la toma de decisiones subjetivas. Bajo la hipótesis del método del Índice de Avenida, básicamente el objetivo es formar grupos de estaciones que satisfagan aproximadamente la condición de homogeneidad, esto es, que compartan la misma distribución de frecuencias excepto por un factor de escala.

De forma habitual, los datos disponibles a la hora de regionalizar se reducen a las observaciones registradas en las estaciones y a una serie de descriptores que tienen en cuenta sus características físicas (localización geográfica, altitud...). Es importante esta distinción entre observaciones y descriptores, pues Hosking y Wallis (1997) propusieron basar la definición de las diferentes regiones en los descriptores físicos, reservando las observaciones para testear la homogeneidad de la regionalización propuesta. Ellos consideraron que el análisis clúster es la alternativa más adecuada para agrupar estaciones. No existe un número de agrupaciones “correcto”, sino que es el área estudiada la que marca la necesidad de agrupar en más o menos regiones. En este sentido, es evidente que grupos con pocas estaciones corren el riesgo de mejorar muy poco la precisión en las estimaciones con respecto a los métodos de análisis local, mientras que regiones de gran tamaño pueden vulnerar el criterio de homogeneidad. Por ello, la elección de algoritmos que tienden a formar clústers de tamaño y variabilidad parecida, como el método de Ward o el enlazado promedio, parece la opción más razonable. Por otra parte, el resultado que arroja un análisis clúster no tiene por qué ser definitivo. Son muchos los ajustes de tipo subjetivo que pueden ser empleados para mejorar la coherencia física del reparto e incluso reducir la heterogeneidad de ciertas regiones.

Establecidas y ajustadas las diferentes regiones, llega el momento de emplear las observaciones registradas en las estaciones para comprobar la hipótesis de homogeneidad. Hosking y Wallis (1997) recomiendan el empleo de tests basados en los momentos-L. Proponen una medida de la heterogeneidad -validada en una serie de experimentos de simulación tipo Montecarlo- para evaluar hasta qué punto un grupo de estaciones puede ser tratado como una región homogénea, comparando las variaciones entre los momentos-L muestrales con las que cabría esperar en caso de que lo fuera.

Se puede decir que en una región homogénea todas las estaciones tienen los mismos momentos-L poblacionales. No ocurre lo mismo con los momentos-L muestrales, que son diferentes debido precisamente a la variabilidad que surge del muestreo. Por tanto, la cuestión que se plantea es averiguar si la dispersión entre los momentos-L muestrales de las estaciones que constituyen la región considerada (datos observados) es mayor que la que se obtendría en caso de homogeneidad (datos simulados).

Por otro lado, para definir lo que cabría esperar en caso de homogeneidad se emplea la simulación repetida de regiones con el mismo número de estaciones y las mismas longitudes muestrales que la real. Gracias a estas simulaciones se obtiene la media y la desviación estándar de la medida de dispersión elegida. Con respecto a la función de distribución a partir de la cual se generan los datos simulados, Hosking y Wallis (1997) subrayaron las ventajas que proporciona la función tetraparamétrica Kappa. Siguiendo el procedimiento establecido por estos autores, la medida de heterogeneidad se calcula mediante el estadístico  $H$ , cuya expresión se representa en la Ecuación 3.

$$H = \frac{(V - \mu_V)}{\sigma_V} \quad (3)$$

Donde  $V$  representa la desviación estándar ponderada de los  $L$ -CV locales,  $\mu_V$  su media y  $\sigma_V$  la desviación estándar. La región se declara heterogénea si el valor de  $H$  es suficientemente elevado. Hosking y Wallis (1997) establecieron que para considerar una región aceptablemente homogénea el valor de  $H$  debía ser inferior a 1, entre 1 y 2 se podía considerar posiblemente heterogénea, y para valores de  $H$  superiores a 2 había que considerarla definitivamente heterogénea.

### Elección de una distribución de frecuencia para la región

Definidas las regiones, se procede a determinar la función de distribución más apropiada para cada una de ellas. Es importante remarcar que la distribución elegida no tiene por qué ser la que más se aproxime a las observaciones. El hecho de existir una función con un buen ajuste a los datos no garantiza que, en ella, los valores futuros de la variable vayan a ser coherentes con los del pasado, pues a menudo derivan de procesos físicos propensos a originar *outliers* alejados del resto. En este sentido, es preferible anteponer la elección de un modelo robusto en estimaciones a la selección de la distribución que mejor ajusta.

Como se ha indicado en la Introducción, Hosking y Wallis (1997) recomiendan evaluar las distribuciones: Logística Generalizada (GLO); Valores Extremos Generalizada (GEV); Pareto Generalizada (GPA); Normal Generalizada (GNO); y Pearson tipo III (PE3); por poseer tres parámetros -lo cual, en un análisis regional, facilita el ajuste al disponerse de observaciones de varias estaciones- y por incluir entre sus casos particulares las distribuciones más empleadas en el estudio de variables ambientales.

Son muchas las alternativas disponibles para evaluar la bondad de ajuste de una determinada distribución a una muestra de observaciones, en el análisis regional de frecuencias, Hosking y Wallis (1997) propusieron trabajar directamente sobre los momentos-L regionales. Definen además una medida de bondad de ajuste -validada en una serie de experimentos de simulación tipo Montecarlo- para evaluar hasta qué punto una distribución concreta se aproxima correctamente a los datos, y seleccionar de entre las candidatas aquella que proporciona el mejor ajuste. En este caso se propuso el estadístico  $Z^{DIST}$ , expresado en la Ecuación 4, como medida de bondad de ajuste.

$$Z^{DIST} = (\tau_4^{DIST} - t_4^R + B_4) / \sigma_4 \quad (4)$$

Donde  $t_4^R$  es el valor medio regional de  $L$ -curtosis,  $\tau_4^{DIST}$  el  $L$ -curtosis de la distribución en cuestión,  $B_4$  es el sesgo de  $t_4^R$  y  $\sigma_4$  la desviación estándar de  $t_4^R$ . El ajuste se declara adecuado si  $Z^{DIST}$  es suficientemente cercano a cero; un criterio adecuado, según Hosking y Wallis (1997), es establecer el límite  $|Z^{DIST}| \leq 1.64$ . Conviene aclarar que el valor de  $\sigma_4$  se determina mediante simulaciones

de Montecarlo para ver si el valor de la desviación de  $t_4^R$  respecto a la teórica de la distribución corresponde razonablemente a lo esperable en una muestra que siga esa distribución.

### Estimación de los cuantiles regionales de frecuencia

Hosking y Wallis (1997) consideraron el método del índice de avenida como la opción más conveniente y eficiente para estimar los parámetros de la distribución y los cuantiles de interés. Sin olvidar que asume la independencia de las observaciones tanto dentro de las series como entre estaciones, siempre que exista dependencia y no sea elevada la precisión de las estimaciones no tiene por qué verse afectada. Para una variable genérica,  $Q$ , supónganse que la región tiene  $N$  estaciones, cada estación  $i$  con un tamaño muestral  $n_i$  y datos observados  $Q_{ij}$ ,  $j = 1, \dots, n$ . La función cuantil de la distribución de frecuencia en la estación  $i$ ,  $Q_i(F)$ , para una región homogénea, se define por la ecuación 5.

$$Q_i(F) = \mu_i q(F), \text{ para } i = 1, 2, \dots, N \quad (5)$$

siendo  $0 < F < 1$ ,  $\mu_i$  es el factor de escala local de la estación  $i$ , esto es, el índice de avenida de dicha estación siendo  $\hat{\mu}_i$  su estimador, que se trata de una medida de centralidad de la población, normalmente la media de las observaciones como es el caso que nos ocupa; y  $q(F)$  la curva regional de frecuencia. La estimación de los valores de esta función se realiza combinando los valores de las relaciones de los momentos-L de cada estación de forma ponderada, es decir, aplicando el procedimiento denominado “algoritmo regional de momentos-L”. Los datos reescalados y adimensionalizados son  $q_{ij} = Q_i / \hat{\mu}_i$ ,  $j = 1, \dots, n_i$ ,  $i = 1, \dots, N$ .

Todo el procedimiento de análisis regional descrito se ha codificado y ejecutado en lenguaje de programación R con la ayuda de los paquetes Lmom (Hosking, 2015a) y LmomRFA (Hosking, 2015b).

### Métodos de interpolación espacial

Como ya se ha comentado, se han comparado por una parte el método de interpolación de distancia inversa ponderada (IDW) y el método de interpolación geoestadístico Kriging ordinario. El trabajo se ha realizado con la extensión “Geostatistical Analyst” de ArcGIS 10.3.1 for Desktop de Esri. Este análisis se ha realizado para los periodos de retorno de  $T=2, 5, 10, 20, 50, 100, 200, 500$  y  $1000$  años.

Tanto IDW como Kriging son métodos ponderados (6) que obtienen el valor en un punto donde no existen datos a partir de los datos de estaciones cercanas.

$$z_j = \sum_{i=1}^n w_i z_i \quad (6)$$

Siendo:  $z_j$  el valor de la variable del punto  $j$  que se quiere estimar;  $z_i$  el valor de la variable medida en el punto  $i$ ;  $w_i$  el peso de la variable del punto  $i$  en la variable del punto  $j$ ,  $\sum_{i=1}^n w_i = 1$ ; y  $n$  el número de puntos que afectan al punto  $j$ .

La diferencia entre los métodos consiste en el modo de obtener el valor de los pesos. En el método IDW la ponderación es una función de la distancia inversa (7).

$$w_i = \frac{\frac{1}{d_{ij}^k}}{\sum_{l=1}^n \frac{1}{d_{lj}^k}} \quad (7)$$

Siendo:  $d_{ij}$  la distancia entre el punto  $i$  y  $j$ ;  $k$  el parámetro de potencia, indica como decae el valor del peso en función de la distancia, en este análisis se ha utilizado el valor de 2.

En el método de Kriging los pesos de la ponderación los obtiene a partir del semivariograma del conjunto de la muestra. El semivariograma se construye a partir de los datos conocidos y muestra la semivarianza frente a la distancia entre puntos que es

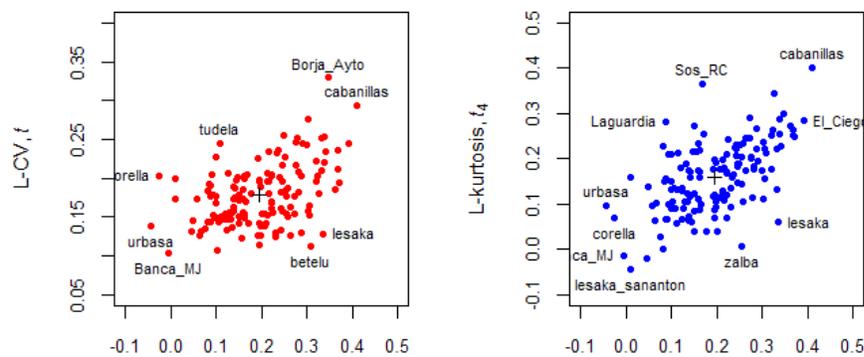
un indicativo de la correlación espacial existente entre puntos en función de la distancia. Para poder aplicar el método es necesario ajustar el semivariograma empírico a una función. La extensión “Geostatistical Analyst” permite utilizar ajustes a diferentes funciones, en este estudio se ha utilizado la función circular.

## RESULTADOS Y DISCUSIÓN

### Análisis Regional de Frecuencias (ARF)

#### Filtrado primario de las observaciones

Antes de comenzar con el análisis propiamente dicho, se procedió al ordenamiento y limpieza de las series de datos, eliminando los errores encontrados y los años con excesivas ausencias. Una vez depuradas las series se procedió a determinar los momentos-L y los L-Coefficientes de cada serie. Representando gráficamente los valores de L-CV y L-curtosis frente a L-asimetría, de esta forma se obtuvo una primera aproximación de las estaciones que podían presentar Discordancia (Figura 2).



**Figura 2** | Representación gráfica de los ratios de los L-Coefficientes: L-CV ( $t$ ) vs. L-asimetría ( $t_3$ ) y L-curtosis ( $t_4$ ) vs. L-asimetría ( $t_3$ ).

A primera vista, las series aparecieron bastante agrupadas excepto algunas excepciones como Corella, Urbasa, Banca\_MJ, Betelu, Lesaka, Cabanillas, Borja\_Ayto, Sos\_RC, Laguardia, Lesaka San Antón, Zalba y El\_Ciego, cuyos valores de los ratios se encontraron alejados del valor medio (representada por el símbolo +). Estas estaciones podrían ser discordantes con respecto al resto de estaciones. Los valores de la discordancia,  $D_i$ , obtenidos mediante la Ecuación 2 se presentan en la Tabla 1. El valor crítico de la Discordancia,  $D_i$ , en un conjunto de 15 o más estaciones es de 3, según Hosking y Wallis (1997). Según esto, las estaciones que salieron discordantes (en negrita Tabla 1) son Cabanillas, Corella, Lesaka, Laguardia, Borja Ayto y Sos del Rey Católico. En cada una de la estaciones discordantes se procedió a analizar de manera más exhaustiva los datos de dichas estaciones con el fin detectar posibles anomalías. Hay que destacar que su revisión individualizada no reflejó evidencias de posibles anomalías salvo la presencia de determinados *outliers* cuya coherencia fue comprobada con estaciones próximas. En cualquier caso, la decisión de incluir, o no, las estaciones discordantes en el análisis, quedó a expensas de los resultados del test de homogeneidad y del estudio de regionalización que se desarrolló a continuación

#### Identificación de regiones homogéneas

Realizado el filtrado primario de las observaciones y antes de proceder a ajustar las funciones de distribución, es necesario proceder a identificar las regiones homogéneas que se encuentran en la zona de estudio. De todas las etapas del análisis regional de

frecuencias, la identificación de regiones homogéneas es a menudo la más complicada, al requerir la toma de decisiones discrecionales. El objetivo que se plantea es formar grupos de estaciones que satisfagan aproximadamente la condición de homogeneidad al objeto de compartir la misma distribución de frecuencias.

En primer lugar, se comprobó si el territorio definido por todas las estaciones podía ser establecido como una región homogénea en el comportamiento de sus precipitaciones diarias máximas anuales. Se calculó entonces, para el total de las 142 estaciones consideradas, la medida de heterogeneidad  $H$  planteada en la ecuación 3. El resultado obtenido para una única región, fue  $H=14.1$ , el valor se situó muy por encima del límite, 2, a partir del cual la región ha de ser considerada “definitivamente heterogénea”. Eliminando las estaciones que han salido discordantes en el apartado anterior, el valor de  $H$  siguió siendo muy elevado, 13.1 por lo que se desechó la opción de 1 región. Fue preciso, por lo tanto, identificar las regiones homogéneas.

Para determinar las regiones homogéneas el procedimiento que se considera como la alternativa más apropiada por muchos autores como Hosking y Wallis (1997), es el análisis clúster, aunque éste puede venir apoyado por otro tipo de información física o climatológica. En este trabajo, se ha optado por un análisis clúster en el que se parte de cinco variables físicas que se consideran relacionadas con la precipitación diaria máxima anual: latitud (1.0), longitud (0.5), altitud (1.5), temperatura media anual (0.5) y precipitación media anual (1.5), a las que se aplica el coeficiente de ponderación indicado. El algoritmo empleado en este caso para realizar el análisis clúster es el método Ward, procedimiento jerárquico aglomerativo que agrupa las estaciones de manera que la varianza intragrupal de la estructura formada sea mínima (Hosking y Wallis, 1997).

El procedimiento es algo complejo en el que primero se fija un número de regiones a considerar, en este caso se ha partido de cuatro regiones que son las establecidas en el Estudio Agroclimático de Navarra (G.N., 2001), a partir de ese momento, mediante el análisis clúster, se asignan las estaciones a cada región, y posteriormente se comprueban los criterios de discordancia y de homogeneidad de cada región. En función de los valores de discordancia y de homogeneidad obtenidos se van tomando decisiones algo discrecionales como subdividir regiones, eliminar estaciones o cambiar estaciones de región teniendo en cuenta criterios de localización, etc.. Es, por lo tanto, un proceso de aproximaciones sucesivas donde se van probando el número de regiones y las estaciones asignadas a las mismas hasta llegar a conseguir que los valores de los índices de heterogeneidad y la discordancia sean aceptables.

En este caso, tras probar el análisis clúster para distintos números de regiones, después de hacer pruebas cambiando estaciones de región, subdividir regiones y/o eliminar estaciones, evidentemente siempre en coherencia con la regionalización climática de Navarra (G.N., 2001), se llegó a conseguir 6 regiones aceptablemente homogéneas, que fue el mejor resultado obtenido. Para llegar a esta regionalización se tuvieron que eliminar 6 estaciones: Lesaka San Antón, Fuenterrabía, Laguardia, Borja Ayto, Sádaba y Sos del Rey Católico; por sus elevados valores de discordancias y por empeorar la homogeneidad de las regiones que las contenían.

En la Figura 1 se presentan las estaciones agrupadas en las regiones resultantes y en la Tabla 2 se muestran los valores del índice de heterogeneidad,  $H$ , de cada región así como los valores de la discordancia,  $D$ , en cada estación para la regionalización definitiva. El número de estaciones es bastante parecido en las regiones centrales 2, 3, 4 y 5, siendo respectivamente de 30, 29, 35 y 21, la región 6, más al sur, tiene 14, y, por último, en el norte se localiza una región centrada en la zona divisoria entre Gipuzkoa, Francia, y Navarra que solo tiene 7 estaciones. Los valores de discordancia,  $D$ , en todas las estaciones están por debajo del umbral crítico en cada caso: región 1 (7 estaciones) por debajo de 1.917; región 6 (14 estaciones) por debajo de 2.971; y el resto de estaciones (>15 estaciones) por debajo de 3; únicamente en Banca MJ el valor es de 3. Hay que indicar que la regionalización obtenida es bastante coherente con las características de la precipitación de Navarra.

Los valores del índice de heterogeneidad,  $H$ , en cuatro regiones, 2, 3, 4, y 6, están por debajo de 1,  $H < 1$ , por lo que se pueden considerar, según Hosking y Wallis (1997), “aceptablemente homogéneas”. En las otras dos regiones, 1 y 5, los valores de  $H$  se encuentran en el intervalo  $1 \leq H < 2$  por lo que se pueden considerar como “posiblemente heterogéneas”. La decisión de contar con dos regiones calificadas como “posiblemente heterogéneas” se fundamenta en varios aspectos. En el caso de la región 1, es una región con pocas estaciones pero que al introducirlas en otras regiones empeoran la heterogeneidad de las mismas y la discordancia de las otras estaciones. En el caso de la región 6 con 14 estaciones, el valor de  $H$  es 1.19 muy próximo al límite para ser considerada como “aceptablemente homogéneas”, después de realizar diversos cambios este es el mejor resultado obtenido.

**Tabla 2** | Valores de la heterogeneidad,  $H$ , de las distintas regiones y de la discordancia,  $D_i$ , en todas las estaciones para esa regionalización.

Estación	$D_i$	Estación	$D_i$	Estación	$D_i$	Estación	$D_i$
REGIÓN 1 (7 estaciones)				Heterogeneidad ( $H=1.60$ )			
Lesaka	1.52	Arditurri Oiartzun	0.32	Bera	1.03	Hondarribia A	0.99
Irún	1.11	Etxalar	0.13	Zugarramurdi	1.89		
REGIÓN 2 (30 estaciones)				Heterogeneidad ( $H=0.42$ )			
Amaiur	1.13	Central Arrambide	0.07	Elduayen	1.10	Irrisarry	2.12
Areso	0.87	Doneztebe	0.44	Andoain CI	0.23	Irouleguy_Fm	1.68
Arizkun	1.41	Goizueta	0.18	Urnieta V	0.67	Licq Therey	0.47
Artikutza	1.48	Leitza	0.29	Anarbe E	0.49	St Jean Le Vieux	0.59
Azpirotz	1.10	Luzaide Valcarlos	1.52	Arette V	0.20	Trois Villes	1.24
Belate	0.41	Sunbilla	0.68	Cambo les Bains E	1.54	<b>Banca MJ</b>	<b>3.00</b>
Bértiz	1.53	Renteria Presa	0.19	Espelette Gb	0.73		
Betelu	2.87	Hernani Cs	1.31	Hosta V	1.44		
REGIÓN 3 (29 estaciones)				Heterogeneidad ( $H=0.47$ )			
Abaurrea Alta	0.36	Larraona	0.40	Aisa de Jaca	0.86	Oroz Betelu	2.47
Altsasu	0.12	Olagüe	0.69	Canfranc La	0.36	Legorreta	0.38
Aribe	0.18	Urbasa	2.30	Accous Ued	1.80	Beasain	0.63
Erro	0.16	Urzainqui	1.30	Larrau	0.54	Ordizia Ca	0.93
Esparza Salazar	0.85	Zubiri	0.25	Arriola	2.08	Irotz	0.83
Eugi	0.55	Mutiloa	0.26	Goñi	1.39		
Iraizotz	2.23	Lareo	1.66	Aurizberri	2.53		
Irurtzun	0.79	Legazpia B	0.88	Opacua	1.21		
REGIÓN 4 (35 estaciones)				Heterogeneidad ( $H=0.31$ )			
Aibar	0.82	Cábrega	1.84	Lezaun	0.17	Zuazu	0.34
Alloz	0.39	Epároz	0.27	Monreal	0.50	Bernedo	0.24
Amillano	1.14	Estella	0.64	Navascués	1.90	Castiello de Jaca	1.67
Aoiz	1.46	Galbarra	1.27	Noain	0.96	Jaca Eca	1.49
Arróniz	0.29	Genevilla	0.19	Olóriz	2.46	Sta Cruz Campezo Dfa	0.29
Artieda	0.54	Iguzquiza	0.20	Otazu	0.70	Puente la Reina	0.63
Azanza	2.13	Ilundain	1.30	Pamplona	0.90	Aratores	0.71
Barasoain	0.70	Javier	2.22	Yesa	0.29	Los Arcos	0.88
Belzunce	0.16	Leire	2.39	Zalba	2.88		
REGIÓN 5 (21 estaciones)				Heterogeneidad ( $H=1.19$ )			
Andosilla	0.42	Lerga	1.52	Viana	0.97	Catiliscar	2.21
Caparroso	1.32	Lodosa	2.15	Logroño Agoncillo	0.59	Biel	0.74
Carcastillo La Oliva	0.80	Miranda	0.54	Varea V	0.74	Bailo Pr	1.07
Cáseda	0.95	Olite	0.46	El Ciego	1.86		
Falces	1.31	Sartaguda	0.49	Cenicero I	0.34		
Lerín	0.68	Sesma	1.03	Los_Pintanos	0.79		
REGIÓN 6 (14 estaciones)				Heterogeneidad ( $H=0.73$ )			
Buñuel	0.70	Fitero	0.13	El Bayo	1.72	Ejea Caballeros	0.44
Cabanillas	2.39	Monteagudo	0.14	Rincón del Soto	1.85	Tarazona	2.22
Cadreita	0.15	Tudela	0.86	Cornago	1.30		
Corella	1.13	Alfaro E	0.53	Arnedo E	0.44		

### Elección de una distribución de frecuencia para la región

La elección de la distribución de frecuencias se realiza por medio del estadístico de bondad de ajuste  $Z^{DIST}$ . En la Tabla 3 se presentan los valores para cada una de las distribuciones propuestas por Hosking y Wallis (1997): Logística Generalizada (GLO), Valores Extremos Generalizada (GEV), Normal Generalizada (GNO), Pearson tipo III (PE3) y Pareto Generalizada (GPA).

El ajuste se declara adecuado, según Hosking y Wallis (1997), cuando  $|Z^{DIST}| \leq 1.64$ , siendo mejor cuanto más cercano esté a cero. Por lo tanto, a tenor de los valores indicados en la Tabla 3, se puede decir que las funciones de distribución que proporcionan los valores de  $Z$  más adecuados (valores en rojo) son: la GNO para las regiones 1, 2, 4 y 5; la GLO para la región 3; y la GEV

para la región 6. En azul se han representado los valores de  $Z$  para las funciones de distribución que también cumplen. También se comprueba que la GEV es la función que obtiene valores aceptables en las 6 regiones, la GNO en 5, la PE3 en 4, la GLO en 3 y la GPA no obtiene en ninguna región un valor aceptable de  $Z$ .

**Tabla 3** | Valores del estadístico  $Z$  para las funciones de distribución.

	$Z^{GLO}$	$Z^{GEV}$	$Z^{GNO}$	$Z^{PE3}$	$Z^{GPA}$
Región 1	1.24	0.43	-0.27	-1.48	-1.85
Región 2	4.14	0.36	0.09	-0.92	-7.84
Región 3	1.52	-1.60	-2.19	3.50	-8.71
Región 4	4.94	1.61	0.78	-0.91	-6.15
Región 5	3.35	0.68	-0.02	-1.42	-5.56
Región 6	1.23	-0.75	-1.40	-2.64	-5.49

Después de analizar las curvas de frecuencias regionales de las funciones de distribución, cuyo ajuste es aceptable según el estadístico  $Z$ , se observa que la diferencia entre las mismas es muy acusada en la Región 1, sobre todo para periodos de retorno elevados, en la Región 3 también la diferencia es apreciable aunque menor, y, sin embargo en el resto de regiones las diferencias entre son pequeñas. También se aprecia también como la función de distribución GLO tiende a estimar los cuantiles por encima del resto de funciones, así como la PE3 lo hace por debajo. En la Región 3, en la que solo son aceptadas la GLO y la GEV, se ve como a partir del  $T=50$  la primera estima los valores de los cuantiles por encima de la GEV, aumentando las diferencias conforme aumenta el periodo de retorno, al igual que ocurre con la Región 1.

A efectos prácticos, la elección de una u otra distribución, dada la similitud de los resultados en las regiones 2, 4, 5 y 6 puede resultar indiferente, sin embargo en las regiones 1 y 3 puede ser más importante dadas las diferencias entre las distintas distribuciones. Es, por lo tanto, recomendable optar por un modelo de selección más robusto, que no implica necesariamente el mejor ajuste, dado que las series evaluadas constituyen tan sólo una muestra de las múltiples posibles de la población. Atendiendo a la recomendación de Hosking y Wallis (1997), se procedió a una simulación tipo Montecarlo en la que se realizaron 1000 simulaciones de cada una de las funciones de distribución para los cuantiles asociados a los periodos de retorno  $T$  de interés. La selección de la función en cada región se establece en función del valor del error cuadrático medio relativo obtenido a partir de las simulaciones para cada función, la función que proporcione el menor valor es la que se selecciona. De esta forma, se han seleccionado las siguientes funciones: Pearson Tipo 3 (PE3) para las regiones 1, 2, 4, 5; Logística Generalizada (GLO) para la región 3; y Normal Generalizada (GNO) para la región 6. Hay que indicar que habría que estudiar más a fondo porqué la Región 3 presenta un comportamiento tan diferenciado al resto de las regiones de alrededor. Que la Región 6 presente un comportamiento diferenciado al resto parece coherente ya que es la región que con un clima más continental donde las precipitaciones convectivas son más frecuentes. Por otra parte, parece lógico que el resto de las regiones salga la misma función de distribución.

**Tabla 4** | Cuantiles de precipitación regionales adimensionalizados,  $\hat{\beta}(F)$ .

Región	Función	T (años)								
		2	5	10	20	50	100	200	500	1000
Región 1	PE3	0.897	1.228	1.470	1.710	2.024	2.260	2.495	2.805	3.039
Región 2	PE3	0.961	1.194	1.340	1.473	1.638	1.756	1.871	2.018	2.127
Región 3	GLO	0.952	1.190	1.361	1.543	1.813	2.047	2.313	2.719	3.076
Región 4	PE3	0.932	1.251	1.461	1.659	1.907	2.089	2.267	2.498	2.669
Región 5	PE3	0.920	1.289	1.532	1.761	2.050	2.262	2.469	2.737	2.937
Región 6	GNO	0.909	1.299	1.571	1.841	2.203	2.485	2.775	3.175	3.490

Definidas las funciones regionales de frecuencia de cada una de las regiones, se procede a particularizarlas localmente en cada una de las estaciones de todas las regiones. Aplicando el método del Índice de Avenida (ecuación 5 y 8) se obtienen las curvas de cuantiles locales de cada estación, de forma que la precipitación asociada a un periodo de retorno o cuantil, , será igual al índice de precipitación (índice de avenida) de cada estación, asumido como la media de sus observaciones (valor medio de las

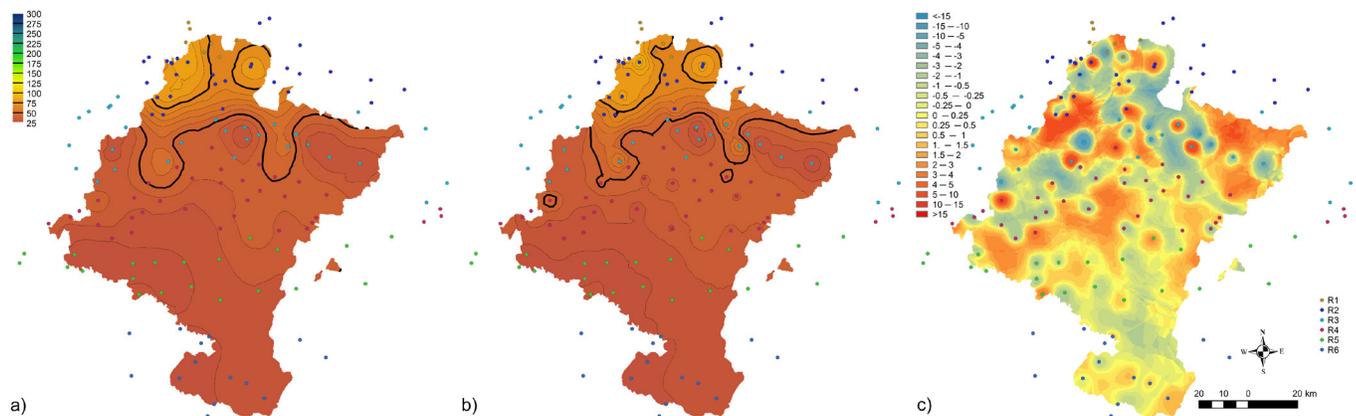
precipitaciones diarias máximas anuales), multiplicado por el cuantil regional adimensional,  $\hat{P}(F)$ , (Tabla 4)

$$\hat{P}_i(F) = I_1^{(i)} \hat{p}(F) \quad (8)$$

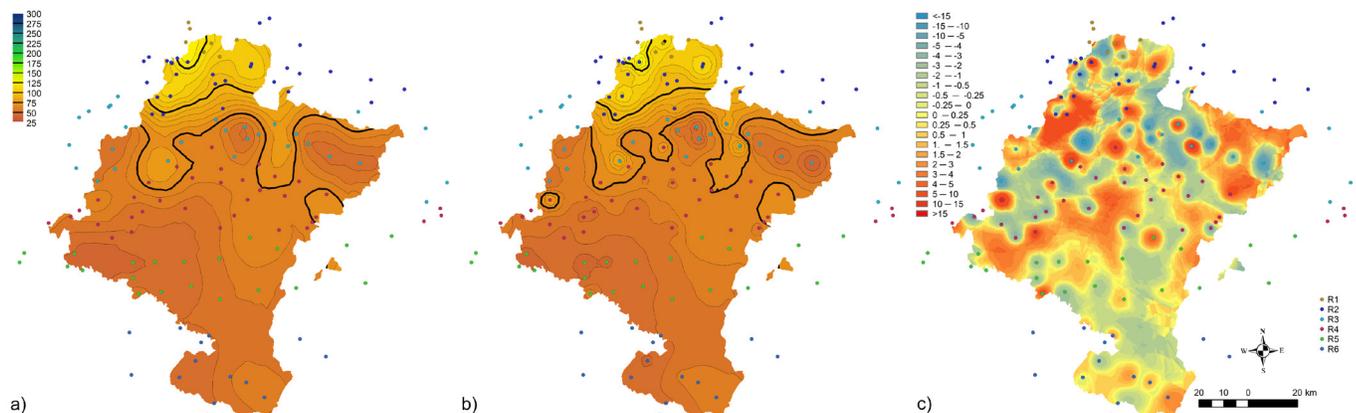
## Representación espacial de las precipitaciones máximas

En las figuras 3, 4, 5 y 6 se muestran los resultados de la interpolación espacial para los métodos de Kriging e IDW y la diferencia entre ellos para los periodos de retorno de 2, 10, 100 y 500 años.

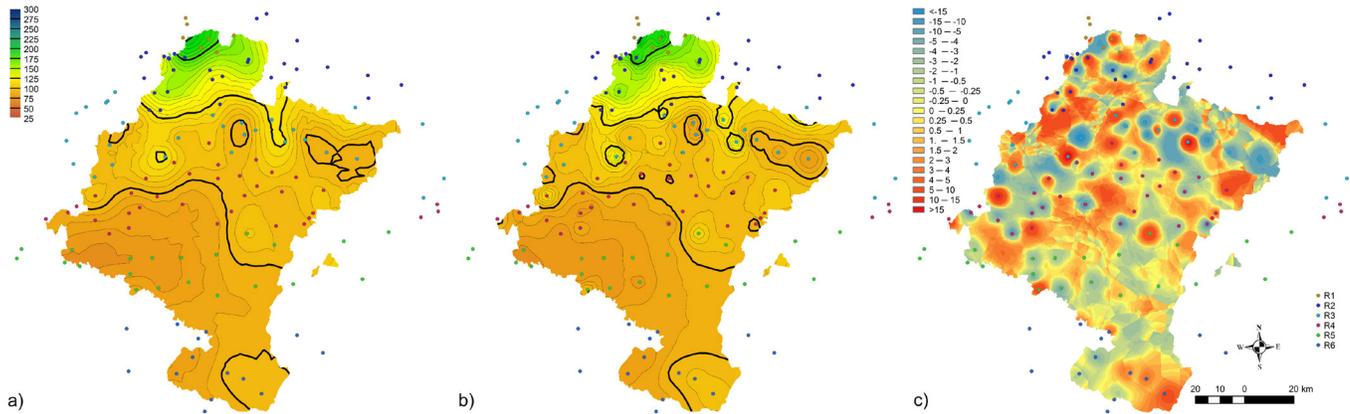
Se observa como ambos métodos de interpolación dan distribuciones semejantes y las mayores diferencias se localizan en las ubicaciones de las estaciones. Esto se debe al ser el método de IDW un interpolador exacto, ajusta el valor de la interpolación en el punto observado al valor observado, a diferencia del método de interpolación de Kriging que el valor de la interpolación en los puntos de observación (estaciones) no tiene por qué coincidir con el valor observado. Esto explica que en la interpolación espacial realizada con IDW se observe un mayor número de mínimos y máximos locales alrededor de las estaciones lo que se ve reflejado en las diferencias entre ambas interpolaciones. El comportamiento se reproduce para todos los periodos de retorno aunque a diferentes escalas. Analizando las diferencias en el conjunto de Navarra la media de las diferencias entre las dos interpolaciones no supera los 0.34 mm para cualquier periodo de retorno (Tabla 4).



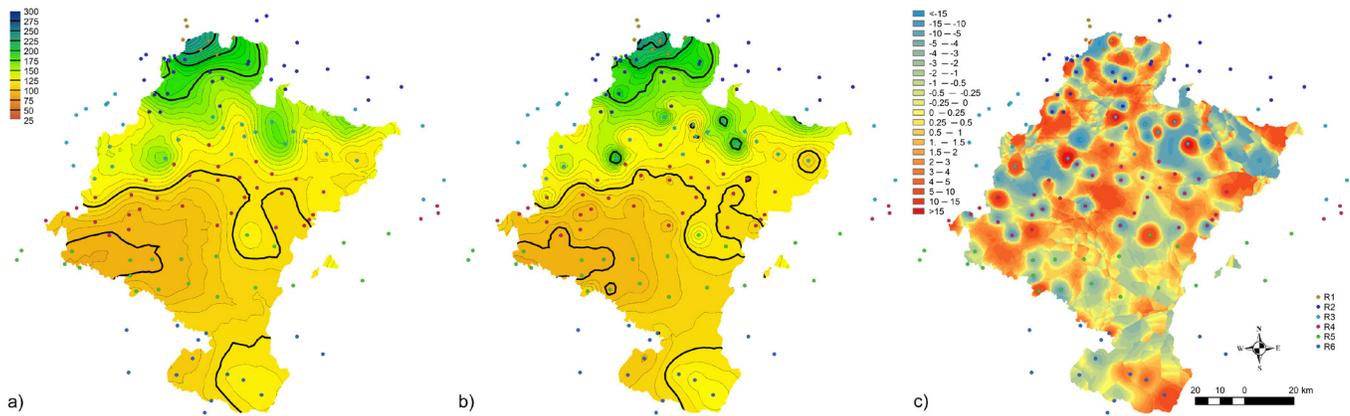
**Figura 3 |** Mapas de las precipitaciones diarias en Navarra obtenidas, para el periodo de retorno de  $T=2$  años y mediante interpolación espacial de kriging (a), IDW (b) y la diferencia entre IDW y Kriging (c).



**Figura 4 |** Mapas de las precipitaciones diarias en Navarra obtenidas, para el periodo de retorno de  $T=10$  años y mediante interpolación espacial de kriging (a), IDW (b) y la diferencia entre IDW y Kriging (c).



**Figura 5 |** Mapas de las precipitaciones diarias en Navarra obtenidas, para el periodo de retorno de T=100 años y mediante interpolación espacial de kriging (a), IDW (b) y la diferencia entre IDW y Kriging (c).



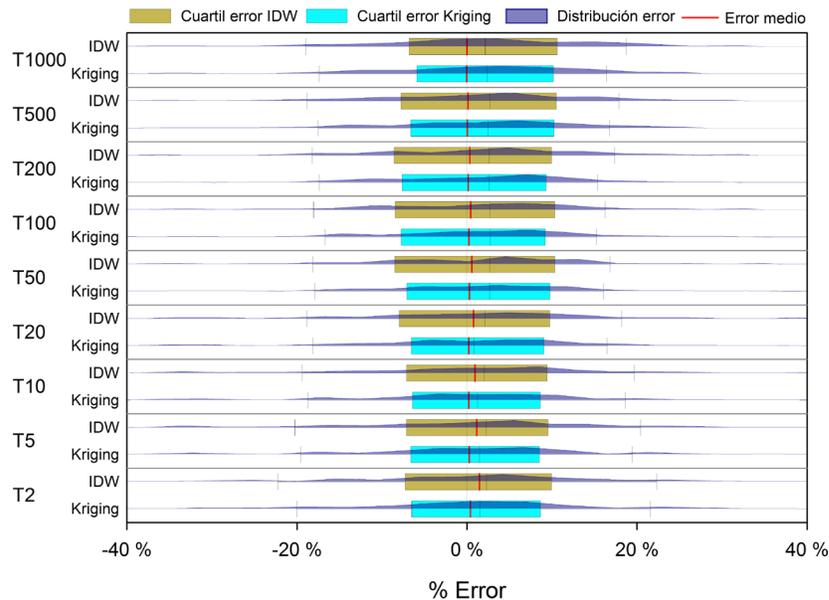
**Figura 6 |** Mapas de las precipitaciones diarias en Navarra obtenidas, para el periodo de retorno de T=500 años y mediante interpolación espacial de kriging (a), IDW (b) y la diferencia entre IDW y Kriging (c).

Para comparar el rendimiento de la interpolación por ambos métodos se utilizó el método “*Leave-one-out Cross-Validation*” (LOOCV) (e.g. Kumari *et al.*, 2016). Este método consiste en quitar una de las estaciones con su respectiva información del mapa e inferir su valor paramétrico a partir del resto, calculando también el error de predicción al mismo tiempo en la ubicación espacial. Por lo tanto, el error de predicción es el promedio de los errores de predicción para cada iteración. En la Tabla 4 también se muestran los resultados del error medio cuadrático (EMC), error medio absoluto y error medio de la validación cruzada. En la Figura 7 se muestra los cuantiles y la distribución de los errores respecto al valor promedio de la observación.

**Tabla 5 |** Media de las diferencias entre la interpolación IDW y Kriging. Error Medio Cuadrático (EMC), Error Medio Absoluto (EMA) y Error Medio (EM) del resultado de la validación cruzada para los diferentes periodos de retorno y métodos de interpolación espacial IDW y Kriging.

T (años)	Media diferencias IDW-Kriging	EMC		EMA		EM	
		IDW	Kriging	IDW	Kriging	IDW	Kriging
2	0.29	74.07	72.71	6.41	6.23	-0.75	-0.21
5	0.33	115.47	112.00	8.14	7.84	-0.76	-0.18
10	0.34	149.98	144.02	9.31	8.90	-0.73	-0.17
20	0.33	188.92	179.67	10.45	9.93	-0.69	-0.18
50	0.25	250.72	232.41	12.04	11.56	-0.60	-0.30
100	0.24	310.27	285.59	13.29	12.79	-0.50	-0.26
500	0.22	538.56	482.54	16.76	16.01	-0.18	-0.07
1000	0.19	710.23	629.92	18.64	17.64	0.01	0.04

Tanto en los valores de los errores de la Tabla 5 como en los cuantiles representados en la Figura 7 se observa como los errores obtenidos mediante el Kriging son menores a los obtenidos con IDW aunque las diferencias son poco significativas. En donde se aprecian más esas diferencias es en el error medio de la validación cruzada (EM). En este caso se puede decir que el Kriging proporciona resultados algo mejores.



**Figura 7 |** Cuantiles y distribución del error de la validación cruzada con respecto al valor medio observado para los diferentes periodos de retorno y los métodos de interpolación espacial IDW y Kriging.

Aunque no hay mucho publicado sobre la comparación entre el rendimiento de Kriging y de IDW, Yao *et al.* (2013) indicaron que la densidad de la red afecta el rendimiento de estas técnicas. Almasi *et al.* (2014), al comparar Kriging con IDW, afirmó que no existe un mejor método, ya que la calidad de la interpolación dependerá de la variación espacial de la información. Varios autores han encontrado que el método de Kriging obtiene mejores resultados en la interpolación de datos que el IDW (e. g. Guodong *et al.*, 2003; Liu y Yang, 2008; Yasrebi *et al.*, 2009 Liu *et al.*, 2016). Sin embargo, hay otros autores que indican lo contrario (e. g. Weber y Englund, 1992; Gotway *et al.*, 1996; Li y Heap, 2011; Yao *et al.*, 2013; Gong *et al.*, 2014). Esto podría ser debido a la gran demanda de datos que el Kriging necesita para proporcionar una estimación fiable, que se refleja en los valores de covarianza (Liu *et al.*, 2016).

## CONCLUSIONES

El Análisis Regional de Frecuencias (ARF) propuesto por Hosking y Wallis (1997) es un método robusto que permite filtrar estaciones estimando la discordancia y estableciendo regiones homogéneas en las que se puede considerar que todas las observaciones se ajustan a la misma función de distribución. Se han establecido 6 regiones homogéneas para las series de precipitación diaria máxima en Navarra. Asimismo se han desechado 6 series por aparecer discordantes en las distintas posibilidades de regiones estudiadas.

Las funciones seleccionadas por el análisis realizado han sido: Pearson Tipo 3 (PE3) para las regiones 1, 2, 4, 5; Logística Generalizada (GLO) para la región 3; y Normal Generalizada (GNO) para la región 6. La función GEV, aunque no ha salido seleccionada en ningún caso, presentó ajustes aceptables en las 6 regiones.

Para obtener los mapas de los cuantiles de precipitación se han aplicado dos métodos de interpolación: Kriging y Distancia Inversa Ponderada (IDW). Se ha observado que el método que menores errores produce mediante el método “*Leave-one-out Cross-Validation*” es el Kriging aunque las diferencias son poco significativas, lo que puede ser debido a la gran densidad de estaciones de precipitación según queda reflejado en la bibliografía.

---

## AGRADECIMIENTOS

Este trabajo de investigación se ha financiado y se ha desarrollado gracias al contrato de investigación “OTRI 2015 021 115”, firmado entre la Universidad Pública de Navarra (UPNA) y la empresa pública Navarra de Infraestructuras Locales S.A. (NILSA), titulado “Sistemas de drenaje urbano sostenible. Bioretención. Caracterización de parámetros locales”. Especial agradecimiento a los técnicos de NILSA responsables de la supervisión del trabajo, Ana Marta Las Heras y Gregorio Berrozpe, por su interés, rigor y seguimiento. Expresar nuestro agradecimiento también al Servicio de Meteorología y Climatología de Navarra perteneciente al Gobierno de Navarra por haber facilitado las series de datos de precipitación, especialmente a su responsable D. Joaquín del Valle de Lersundi, que en todo momento ha manifestado su apoyo incondicional, y a Miren Otazu, técnica de Tragsatec.

---

## REFERENCIAS

- Abdi, A., Hassanzadeh, Y., Ouarda, T.B.M.J. 2017. Regional frequency analysis using Growing Neural Gas network. *Journal of Hydrology*, 550, 92-102. <https://doi.org/10.1016/j.jhydrol.2017.04.047>
- Adamowski, K., Alila, Y., Pilon, P.J. 1996. Regional rainfall distribution for Canada. *Atmospheric Research*, 42, 75-88. [https://doi.org/10.1016/0169-8095\(95\)00054-2](https://doi.org/10.1016/0169-8095(95)00054-2)
- Akkala, A., Devabhaktuni, V., Kumar, A. 2010. Interpolation techniques and associated software for environmental data. *Environmental Progress and Sustainable Energy*, 29(2), 134-141. <https://doi.org/10.1002/ep.10455>
- Almasi A., Ahmad J., Toomanian N., 2014. Using OK and IDW methods for prediction the spatial variability of a horizon depth and OM in soils of Shahrekord, Iran. *Journal of Environmental Research and Management*, 5(8):139-147.
- Álvarez, A.J., Orduña, L.M. 2014. Caracterización del comportamiento estadístico de los caudales máximos estacionales en la España peninsular: Propuesta metodológica para su cálculo. *Ingeniería Civil*, 174, 51-69.
- Álvarez, M., Puertas, J., Soto, B., Díaz-Fierros, F. 1999. Análisis regional de las precipitaciones máximas en Galicia mediante el método del índice de avenida. *Ingeniería Del Agua*, 6(4), 379-386. <https://doi.org/10.4995/ia.1999.2795>
- Ayoade, J.O. 1976. A preliminary study of magnitude of frequency and distribution of intense rainfall in Nigeria. *Hydrological Sciences Journal*, 21(3), 419-421. <https://doi.org/10.1080/02626667609491650>
- Bilham, E.G. 1936. Classification of heavy falls in short periods. *British Rainfall*, 1935, 262-280.
- Blanchet, J., Ceresetti, D., Molinié, G., Creutin, J.D. 2016. A regional GEV scale-invariant framework for Intensity–Duration–Frequency analysis. *Journal of Hydrology*, 540, 82-95. <https://doi.org/10.1016/j.jhydrol.2016.06.007>
- Cannarozzo, M., D’asaro, F., Ferro, V. 1995. Regional rainfall and flood frequency analysis for Sicily using the two component extreme value distribution. *Hydrological Sciences Journal*, 40(1), 19-41. <https://doi.org/10.1080/02626669509491388>
- Caporali, E., Cavigli, E., Petrucci, A. 2006. The index rainfall in the regional frequency analysis of extreme events in Tuscany (Italy). *Environmetrics*, 19, 714-724. <https://doi.org/10.1002/env.949>
- Dalrymple, T. 1960. Flood frequency analyses. U.S. Geol. Surv. Water Supply Pap., no. 1543A, p. 80.

- Delhomme, J.P. 1978. Kriging in the hydrosociences. *Advances in Water Resources*, 1(5), 251-266. [https://doi.org/10.1016/0309-1708\(78\)90039-8](https://doi.org/10.1016/0309-1708(78)90039-8)
- García-Marín A.P. 2007. *Análisis multifractal de series de datos pluviométricos en Andalucía*. Universidad de Córdoba, Servicio de Publicaciones.
- García-Marín, A.P., Estévez, J., Sangüesa-Pool, C., Pizarro-Tapia, R., Ayuso-Muñoz, J.L., Jimenez-Hornero, F.J. 2015. The use of the exponent  $K(q)$  function to delimit homogeneous regions in regional frequency analysis of extreme annual daily rainfall. *Hydrological Processes*, 29, 139-151. <https://doi.org/10.1002/hyp.10284>
- G.N. (Gobierno de Navarra). 2001. Estudio Agroclimático de Navarra. Gobierno de Navarra. <http://meteo.navarra.es/>
- Gong, G., Mattevada, S., O'Bryant, S. 2014. Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. *Environmental Research*, 130, 59-69. <https://doi.org/10.1016/j.envres.2013.12.005>
- Gotway, C., Ferguson, R., Hergert, G., Peterson, T. 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. *Soil Science Society of American Journal*, 60, 1237-1247. <https://doi.org/10.1016/j.envres.2013.12.005>
- Guodong, J., Yancong, L., Wenjie, N. 2003. Comparison between inverse distance weighting method and Kriging. *Journal of Changchun University of Technology*, 24(3), 53-57.
- Hailegeorgis, T.T., Alfredsen, K. 2017. Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology, Regional Studies*: 9, 104-126.
- Halbert, K., Nguyen, C.C., Payrastra, O., Gaume, E. 2016. Reducing uncertainty in flood frequency analyses: A comparison of local and regional approaches involving information on extreme historical flood. *Journal of Hydrology*, 541, 90-98. <https://doi.org/10.1016/j.jhydrol.2016.01.017>
- Hengl, T. 2009. *A Practical guide to Geostatistical Mapping*. [http://spatial-analyst.net/book/system/files/Hengl\\_2009\\_GEOSTATE2c1w.pdf](http://spatial-analyst.net/book/system/files/Hengl_2009_GEOSTATE2c1w.pdf)
- Hosking, J., Wallis J. 1997. *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press.
- Hosking, J. 2015a. Regional Frequency Analysis using L-Moments, Lmom R Package, Version 2.5.
- Hosking, J. 2015b. Regional Frequency Analysis using L-Moments, LmomRFA R Package, Version 3.0-1.
- Kjeldsen, T.R., Smithers, J., Schulze, R. 2002. Regional flood frequency analysis in the KwaZulu-Natal province, South Africa, using the index-flood method. *Journal of Hydrology*, 255(1), 194-211. [https://doi.org/10.1016/S0022-1694\(01\)00520-0](https://doi.org/10.1016/S0022-1694(01)00520-0)
- Kumari, M., Basistha, A., Bakimchandra, O., Singh, C.K. 2016. Comparison of Spatial Interpolation Methods for Mapping Rainfall in Indian Himalayas of Uttarakhand Region. In: Raju JN (ed) *Geostatistical and Geospatial Approaches for the Characterization of Natural Resources in the Environment*. pp. 159-168. Springer International Publishing. [https://doi.org/10.1007/978-3-319-18663-4\\_27](https://doi.org/10.1007/978-3-319-18663-4_27)
- Kysely, J., Picek, J. 2007. Regional growth curve and improved design values of extreme precipitation events in the Czech Republic. *Climate Research*, 33(3), 243- 255. <https://doi.org/10.3354/cr033243>
- Kysely, J., Picek, J., Huth, R. 2007. Formation of homogeneous regions for regional frequency analysis of extreme precipitation events in the Czech Republic, *Studia Geophysica et Geodeatica*, 51(2), 327-344. <https://doi.org/10.1007/s11200-007-0018-3>
- Lee, S.H., Maeng, S.J. 2003. Frequency analysis of extreme rainfall using Lmoments. *Irrigation and Drainage*, 52(3), 219-230. <https://doi.org/10.1002/ird.90>
- Lin G.-F., Chen Lu-H. 2006. Identification of homogeneous regions for regional frequency analysis using the self-organizing maps. *Journal of Hydrology*, 324(1-4), 1-9. <https://doi.org/10.1016/j.jhydrol.2005.09.009>
- Li, J., Heap, A. 2011. Review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. *Ecological Informatics*, 6(3-4), 228-241. <https://doi.org/10.1016/j.ecoinf.2010.12.003>

- Liu, G., Yang, X. 2008. Spatial variability analysis of soil properties within a field. *International Conference on Computer and Computing Technologies in Agriculture*. Wuyishan, China (08-18-2007). 1341-1344 pp. Springer. US. [https://doi.org/10.1007/978-0-387-77253-0\\_75](https://doi.org/10.1007/978-0-387-77253-0_75)
- Liu, J., Doan, C.D., Liong, S.-Y., Sanders, R., Dao, A.T., Fewtrell, T. 2015. Regional frequency analysis of extreme rainfall events in Jakarta. *Natural Hazards*, 75(2), 1075-1104. <https://doi.org/10.1007/s11069-014-1363-5>
- Liu, W., Du, P., Zhao, Z., Zhang, L. 2016. An Adaptive Weighting Algorithm for Interpolating the Soil Potassium Content. *Scientific Reports*, 6, 1-12. <https://doi.org/10.1038/srep23889>
- Malekinezhad, H., Zare-Garizi, A. 2014. Regional frequency analysis of daily rainfall extremes using L-moments approach. *Atmósfera*, 27(4), 411-427. [https://doi.org/10.1016/S0187-6236\(14\)70039-6](https://doi.org/10.1016/S0187-6236(14)70039-6)
- 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>
- Ngongondo, C.S., Xu, C.-Yu, Tallaksen, L.M., Alemaw, B., Chirwa, T. 2011. Regional frequency analysis of rainfall extremes in Southern malwai using index rainfall and L-moments approaches. *Stochastic Environmental Research and Risk Assessment*, 25(7), 939-955. <https://doi.org/10.1007/s00477-011-0480-x>
- Norbiato, D., Borga, M., Sangat, i M., Zanon, F. 2007. Regional frequency analysis of extreme precipitation in the eastern Italian Alps and the August 29, 2003 flash flood. *Journal of Hydrology*, 345(3), 149-166. <https://doi.org/10.1016/j.jhydrol.2007.07.009>
- Oliver, M.A., Webster, R. 1990. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information Systems*, 4(3), 313-332. <https://doi.org/10.1080/02693799008941549>
- Parida, B.P., Kachroo, R.K., Shrestha, D.B. 1998. Regional Flood Frequency Analysis of Mahi-Sabarmati Basin (subzone 3-a) using Index-Flood Procedure with L-moments. *Water Resources Management*, 12(1-2), 1-12.
- Park, J.S., Jung, H.S., Kim, R.S., Oh, J.H. 2001. Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution. *International Journal of Climatology*, 21(11), 1371-1384. <https://doi.org/10.1002/joc.701>
- Pearson, C., Mc Kerchar, A., Woods, R. 1991. Regional flood frequency analysis of western Australian data using L-moments. *National Conference Publication-Institute of Engineers*. Australia. 631-632.
- Schaefer, M.G. 1990. Regional analyses of mean annual precipitation in Washington State. *Water Resources Research*, 26(1), 119-131. <https://doi.org/10.1029/89WR01513>
- Shahzadi, A., Akhter, A.S., Saf, B. 2013. Regional Frequency Analysis of Annual Maximum Rainfall in Monsoon Region of Pakistan using L-moments. *Pakistan Journal of Statistics and Operation Research*, 9(1), 111-136. <https://doi.org/10.18187/pjsor.v9i1.461>
- Smithers, J.C., Schulze, R.E. 2001. A methodology for the estimation of short duration design storms in South Africa using a regional approach based on L-moments. *Journal of Hydrology*, 24(1-2), 42-52. [https://doi.org/10.1016/S0022-1694\(00\)00374-7](https://doi.org/10.1016/S0022-1694(00)00374-7)
- Sveinsson, O.G.B., Salas, J.D., Boes, D.C. 2002. Regional frequency analysis of extreme precipitation in northeastern Colorado and Fort Collins flood of 1997. *Journal of Hydrologic Engineering*, 7(1), 49-63. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2002\)7:1\(49\)](https://doi.org/10.1061/(ASCE)1084-0699(2002)7:1(49))
- Trefry, C.M., Watkins, Jr., Johnson, D. 2005. Regional rainfall frequency analysis for the State of Michigan. *Journal of Hydrological Engineering*, 10(6), 437-449. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2005\)10:6\(437\)](https://doi.org/10.1061/(ASCE)1084-0699(2005)10:6(437))
- Vogel, R.M., Thomas, Jr. W.O., McMahon, T.A. 1993. Flood-flow frequency model selection in southwestern United States. *Journal of Water Resources Planning and Management*, 119(3), 353-366. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1993\)119:3\(353\)](https://doi.org/10.1061/(ASCE)0733-9496(1993)119:3(353))
- Weaver, J.C. 2006. Frequency of annual maximum precipitation in the city of Charlotte and Mecklenburg country, North Carolina, through 2004. *Scientific Investigation Report 2006-5017*. U.S. Department of Interior, U.S.
- Weber, D., Englund, E. 1992. Evaluation and comparison of spatial interpolators. *Mathematical Geology*, 24, 381-391. <https://doi.org/10.1007/BF00891270>

Yang, T., Shao, Q., Hao, Z., Chen, X., Zhang, Z., Xu, C., Sun, L. 2010. Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the pearl river basin, China. *Journal of Hydrology*, 380(3), 386-405. <https://doi.org/10.1016/j.jhydrol.2009.11.013>

Yao, X., Fu, B., Lu, Y., Sun, F., Wang S., Liu, M. 2013. Comparison of Four Spatial Interpolation Methods for Estimating Soil Moisture in a Complex Terrain Catchment. *Plos One*, 8(1), 1-13. <https://doi.org/10.1371/journal.pone.0054660>

Yasrebi, J., Saffari, M., Fathi, H., Karimian, N., Moazallahi, M., Gazni, R. 2009. Evaluation and comparison of ordinary kriging and inverse distance weighting methods for prediction of spatial variability of some soil chemical parameters. *Research Journal of Biological Sciences*, 4(1), 93-102.