

**SOCIEDAD CHILENA DE INGENIERÍA HIDRÁULICA**  
**XXVII CONGRESO CHILENO DE INGENIERÍA HIDRÁULICA**

**EXTENSIÓN Y MEJORA DE UNA METODOLOGÍA BASADA EN LA TEORÍA DE VALORES EXTREMOS PARA LA ESTIMACIÓN DE LA PMP: APLICACIÓN EN CHILE**

**ÁLVARO OSSANDÓN Á.<sup>1</sup>**  
**IGNACIO VILLAVICENCIO A.<sup>2</sup>**  
**JOAQUIN MEZA A.<sup>3</sup>**  
**GONZALO SUAZO F.<sup>4</sup>**

**RESUMEN**

Una estimación precisa de la Precipitación Máxima Probable (PMP) es esencial para el diseño de infraestructura hidráulica resiliente y la reducción del riesgo hidrológico. Los métodos estadísticos tradicionales, como el enfoque de Hershfield, suelen basarse en supuestos que no necesariamente caracterizan correctamente los eventos extremos de precipitación, caracterizados por marcada asimetría. Estas simplificaciones, junto a su implementación estándar con un factor de frecuencia global (K) fijo, pueden inducir sesgos en la estimación de la PMP, cuya magnitud y signo varían según el régimen hidroclicmático, la topografía y la longitud del registro. Este estudio presenta y evalúa una versión mejorada del marco metodológico propuesto por Villavicencio et al. (2023), integrando el modelado de valores extremos mediante la distribución Generalizada de Valores Extremos (GEV) con corrección de sesgo y una regionalización estadística basada en agrupamiento K-means, aplicada a la red nacional de estaciones pluviométricas de Chile continental. Los parámetros de la GEV se estiman mediante L-momentos y se corrigen para series de registro temporal reducido ajustando el parámetro de forma. Posteriormente, el dominio de estudio se divide en zonas homogéneas con base en atributos geográficos y estadísticos. La corrección del sesgo redujo la variabilidad del parámetro de forma, estabilizando las estimaciones de magnitudes extremas de precipitación. Se identificaron seis regiones homogéneas, cada una con patrones distintos de precipitación. Comparado con el método de Hershfield, el enfoque propuesto entrega estimaciones de PMP más consistentes, entendidas como menor incremento relativo y menor dispersión respecto de los máximos históricos por su menor sensibilidad a valores atípicos a escala regional, especialmente en zonas áridas.

---

<sup>1</sup>Profesor Asistente, Depto. de Obras Civiles, Universidad Técnica Federico Santa María – [alvaro.ossandon@usm.cl](mailto:alvaro.ossandon@usm.cl)

<sup>2</sup>Ex Alumno, Depto. de Obras Civiles, Universidad Técnica Federico Santa María – [ignacio.villavicenci@sansano.usm.cl](mailto:ignacio.villavicenci@sansano.usm.cl)

<sup>3</sup>Profesor Asistente, Depto. de Obras Civiles, Universidad Técnica Federico Santa María – [joaquin.meza@usm.cl](mailto:joaquin.meza@usm.cl)

<sup>4</sup>Línea de Servicios de Relaves para las Américas, GHD, Concón, Chile – [Gonzalo.Suazo@ghd.com](mailto:Gonzalo.Suazo@ghd.com)

## 1. INTRODUCCIÓN

El diseño y la seguridad de infraestructuras críticas, como represas, aliviaderos, tranques de relaves y defensas contra inundaciones, dependen en gran medida de evaluaciones precisas de eventos extremos de precipitación. Fallas en estas estructuras pueden tener consecuencias catastróficas, incluyendo pérdidas humanas, daños ambientales y altos costos económicos (Salas et al., 2020). Además, el cambio climático proyecta un aumento en la frecuencia e intensidad de eventos extremos, lo cual incrementa la vulnerabilidad de dichas infraestructuras (Papalexiou & Montanari, 2019).

La Precipitación Máxima Probable (PMP), definida por la Organización Meteorológica Mundial (WMO) como la mayor profundidad de precipitación meteorológicamente posible para una duración y ubicación específicas, se emplea ampliamente para el diseño de obras hidráulicas (WMO, 2009). No obstante, su estimación presenta desafíos significativos, particularmente en regiones con escasa disponibilidad de datos o alta variabilidad climática. A efectos de diseño, la PMP se utiliza como cota superior plausible de precipitación para una duración específica de análisis (ej., 24, 48 o 72 horas). La PMP no tiene período de retorno asociado y funciona como límite teórico que evita sobredimensionar obras al extrapolar cuantiles hacia retornos extremadamente altos. En Chile, Decreto Supremo MOP N.º 50/2015 establece que el diseño de embalses de gran envergadura y embalses de relave debe verificarse frente a la crecida máxima probable (PMF), la cual en la práctica se estima a partir de la PMP.

Los métodos para estimar la PMP se agrupan generalmente en enfoques hidrometeorológicos (Hiraga & Meza, 2025) y estadísticos (Stöwhas, 2016). Aunque los primeros se basan en principios físicos, su aplicación suele requerir grandes volúmenes de datos y experiencia técnica especializada, lo cual puede ser limitante en casos como Chile donde existe escasez de datos. En contraste, los métodos estadísticos, como el de Hershfield (1961), han ganado popularidad por su practicidad y uso de registros históricos. Este método, sin embargo, presenta limitaciones al basarse únicamente en la media y la desviación estándar, ignorando la asimetría característica de las precipitaciones extremas, lo que puede derivar en extrapolaciones poco realistas.

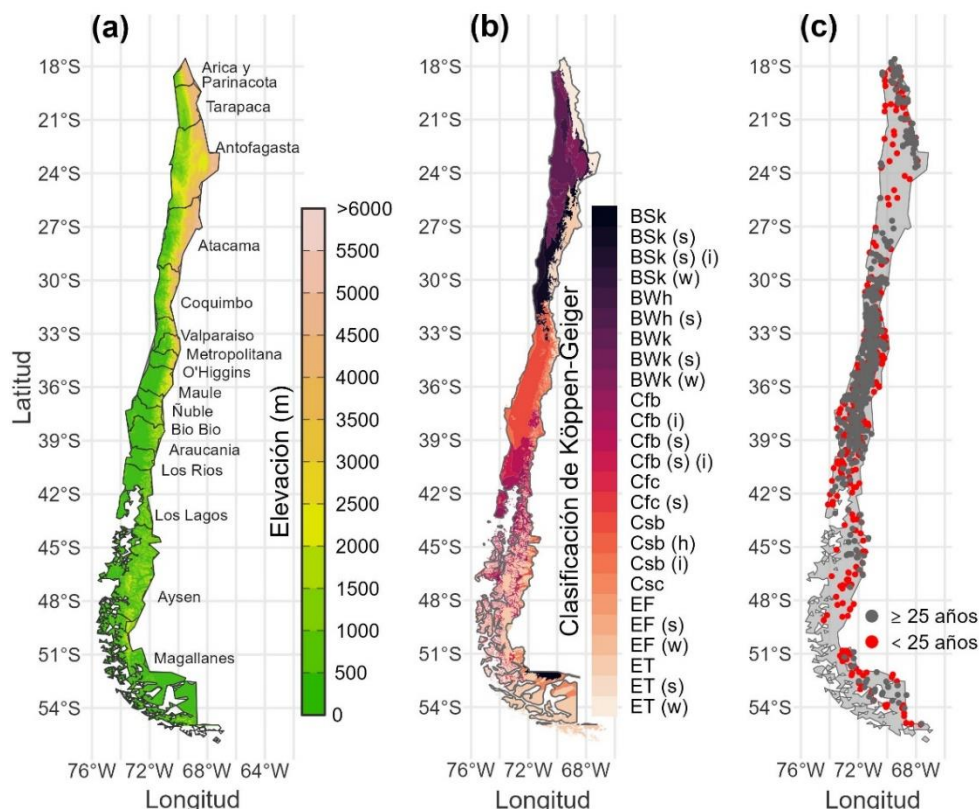
En respuesta a estas limitaciones, Villavicencio et al. (2023) propuso un marco metodológico basado en el uso combinado de la distribución Generalizada de Valores Extremos (GEV) y regionalización estadística. El presente trabajo amplía y mejora dicho enfoque mediante la incorporación de una corrección del parámetro de forma de la GEV para longitudes de registro reducidas y la delimitación de regiones homogéneas mediante técnicas más robustas de agrupamiento. La metodología se aplica a series de precipitación máxima diaria anual de 428 estaciones meteorológicas distribuidas a lo largo del territorio nacional, caracterizado por una alta diversidad hidroclimática y topográfica, lo cual permite evaluar la aplicabilidad y solidez del enfoque propuesto.

## 2. ZONA DE ESTUDIO

Chile se extiende desde aproximadamente los 17.8°S hasta los 55°S, abarcando más de 4.300 km a lo largo de la costa occidental de América del Sur (Figura 1a). Esta extensión latitudinal, combinada con una topografía abrupta compuesta por la Cordillera de la Costa, la Depresión Intermedia y la Cordillera de los Andes (con elevaciones que superan los 6.000 m), genera una notable diversidad de condiciones climáticas y geográficas. Las precipitaciones anuales varían desde menos de 10 mm en el Desierto de Atacama hasta más de 3.800 mm en el extremo sur del

país (Gateño et al., 2024), reflejando distintos regímenes climáticos que presentan desafíos particulares para el análisis hidrológico.

Siguiendo a Alvarez-Garretón et al. (2018), se considera una subdivisión de Chile continental en seis macrozonas: Norte Grande, Norte Chico, Zona Central, Zona Sur, Zona Austral y Patagonia Sur. Esta clasificación captura gradientes climáticos y topográficos relevantes para el análisis de eventos extremos, donde los mecanismos dominantes varían desde tormentas convectivas en el norte hasta sistemas frontales en el sur. Estas macrozonas también coinciden en gran medida con la clasificación climática de Köppen-Geiger (Figura 1b), proporcionando un marco coherente para interpretar los regímenes de precipitación del país.



**Figura 1.** Mapas de Chile continental que muestran: (a) la altitud (en metros) junto con los límites y nombres de las regiones; (b) la clasificación climática de Köppen-Geiger, incluyendo BSk: semiárido frío, BWh: desierto cálido, BWk: desierto frío, Cfb: oceánico templado, Cfc: oceánico subpolar, Csb: mediterráneo de verano cálido, Csc: mediterráneo de verano frío, EF: clima de hielos perpetuos y ET: tundra; y (c) distribución espacial de los datos según la longitud del registro de PMDA (gris para  $\geq 25$  años, rojo para  $< 25$  años).

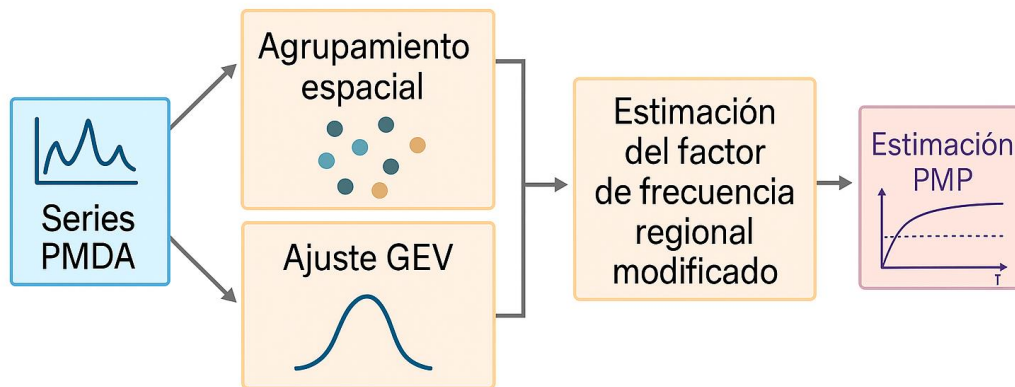
### 3.1 DATOS

Se recopilieron registros diarios de precipitación desde el Centro de Ciencia del Clima y la Resiliencia (CR2), conformando una base de datos de 879 estaciones meteorológicas en todo Chile continental (Figura 1c). Los datos provienen principalmente de dos organismos: la Dirección General de Aguas (DGA) y la Dirección Meteorológica de Chile (DMC). Los registros abarcan desde enero de 1900 hasta junio de 2020, aunque con distintos grados de completitud. Para cada estación, se construyó una serie de precipitaciones máximas diarias anuales (PMDA), seleccionando el valor diario más alto registrado por año. Luego, para obtener las series de

precipitaciones máximas anuales en 24 horas (PMA24), se multiplicaron las series de PMDA por un factor de 1.13 (WMO, 2009), dado que el registro diario (desde 8 AM a 8 AM) no necesariamente coincide con el valor máximo ocurrido en 24 horas. Se conservaron únicamente las estaciones con al menos 25 años de datos, para balancear cobertura espacial y robustez estadística. Para asegurar la calidad de los datos, se aplicaron dos criterios adicionales: (i) se descartaron años con más del 33% de días faltantes; y (ii) si el valor máximo del año superaba el percentil 40 de la serie PMA24, se retuvo el año, aun si tenía datos incompletos. Como resultado, se seleccionaron 428 estaciones con series consistentes y longitudes de registro entre 25 y 99 años.

### 3. METODOLOGÍA

Este trabajo presenta una extensión del enfoque para la estimación de la Precipitación Máxima Probable (PMP) propuesto por Villavicencio et al. (2023), el cual se resume esquemáticamente en la Figura 2. El método consta de tres componentes donde (i) el ajuste de la distribución Generalizada de Valores Extremos (GEV) a las series de PMA24, con corrección del parámetro de forma para registros de corta duración, y (ii) el agrupamiento espacial para delimitar regiones homogéneas se ejecutan en paralelo, ya que este último utiliza exclusivamente atributos independientes del ajuste GEV. Posteriormente, (iii) se estima el factor de frecuencia regional mediante el ajuste de una distribución acotada superior a la serie de factores máximos obtenidos en cada región, para finalmente estimar la PMP para cada estación.



**Figura 2.** Esquema del método propuesto para la estimación de la PMP.

#### 2.1 DISTRIBUCIÓN GENERALIZADA DE VALORES EXTREMOS

La distribución Generalizada de Valores Extremos (GEV) es ampliamente utilizada para modelar máximos de bloques (ej., PMA24). La función de distribución acumulada (FDA) de la GEV se expresa como:

$$F(x) = \exp \left\{ - \left[ 1 + \xi \frac{(x - \mu)}{\sigma} \right]^{-\frac{1}{\xi}} \right\}, \quad 1 + \xi \frac{(x - \mu)}{\sigma} > 0, \sigma > 0 \quad (1)$$

donde  $x$  representa la variable aleatoria (en este caso, PMA24),  $\mu$ ,  $\sigma$  y  $\xi$  son los parámetros de ubicación, escala y forma, respectivamente. Este último parámetro permite clasificar la GEV en tres tipos de distribuciones límites: Gumbel (tipo I,  $\xi = 0$ ), tipo II ( $\xi > 0$ ) y tipo III ( $\xi < 0$ ). Una vez estimados los parámetros, para estimar la PMP se propone el uso de un factor de frecuencia modificado  $K_{EV}$ , el que corresponde a la variable transformada de Gumbel y es definida como:

$$K_{EV} = \frac{1}{\xi} \ln \left( 1 + \xi \frac{(x - \mu)}{\sigma} \right) \quad (2)$$

## 2.2 ESTIMACIÓN Y CORRECCIÓN DE PARÁMETROS

Los parámetros de la distribución GEV (localización  $\mu$ , escala  $\sigma$  y forma  $\xi$ ) se estiman mediante el método de L-momentos, una técnica robusta basada en momentos ponderados por orden que ofrece ventajas frente a métodos tradicionales como máxima verosimilitud, especialmente cuando las series de datos son cortas o contienen valores atípicos (Hosking et al., 1985). Las expresiones para estimar los parámetros de la distribución GEV mediante L-momentos son (Carney, 2016):

$$\xi \approx -7.8590c - 2.9554c^2; \quad \sigma = \frac{-l_2\xi}{\Gamma(1-\xi)(1-2^\xi)}; \quad \mu = l_1 - \frac{\sigma}{\xi} [\Gamma(1-\xi) - 1]; \quad (3)$$

donde, el parámetro  $c$  se define como

$$c = \frac{2l_2}{l_3 + 3l_2} - \frac{\ln(2)}{\ln(3)} = \frac{2}{t_3 + 3} - \frac{\ln(2)}{\ln(3)} \quad (4)$$

$l_1$ ,  $l_2$  y  $l_3$  son los L-momentos muestrales de orden 1 a 3, y  $\Gamma(\cdot)$  corresponde a la función gamma. Adicionalmente, los coeficientes de variación (L-CV), asimetría (L-skewness) y curtosis (L-kurtosis) como las siguientes razones de los L-momentos  $t = l_2/l_1$ ,  $t_3 = l_3/l_2$ , and  $t_4 = l_4/l_2$  (Hosking & Wallis, 1997).

Sin embargo, diversos estudios han reportado que la estimación de  $\xi$  es particularmente sensible a la longitud del registro (e.g., Carney, 2016; Hossain et al., 2022), lo que introduce sesgos sistemáticos en estaciones con menos de  $\sim 40$  años de datos. Para corregir este sesgo, se adopta el enfoque propuesto por Papalexiou & Koutsoyiannis (2013), el cual modela la media y desviación estándar de  $\xi$  como funciones del tamaño muestral  $n$ , bajo el supuesto de que  $\xi$  sigue una distribución normal:

$$\mu_\xi(n) = a_\mu + b_\mu n^{-c_\mu}; \quad \sigma_\xi(n) = a_\sigma + b_\sigma n^{-c_\sigma}; \quad (5)$$

donde  $n$  es la longitud de registro (en años);  $\mu_\xi(n)$  y  $\sigma_\xi(n)$  son la media esperada y la desviación estándar del estimador de  $\xi$  cuando la longitud de registro es  $n$ ;  $a_\mu$ ,  $b_\mu$ ,  $c_\mu$ ,  $a_\sigma$ ,  $b_\sigma$ , y  $c_\sigma$  son los parámetros de una ley de potencia con término constante de forma  $a + bn^{-c}$ ;  $\mu_\xi = a_\mu$  y  $\sigma_\xi = a_\sigma$  representan los valores asintóticos cuando  $n \rightarrow \infty$ . Los parámetros de las ec. (5) se ajustan empíricamente mediante el método de los mínimos cuadrados, utilizando subconjuntos de estaciones con registros largos ( $\geq 40$  años). Para ello, se generan subseries de distinta longitud y se calcula el parámetro de forma ( $\xi$ ) correspondiente mediante la ec. (3). A partir de este procedimiento, se obtiene una versión corregida del parámetro de forma:

$$\hat{\xi}_{corr}(n) = \frac{\sigma_\xi}{\sigma_\xi(n)} \left( \hat{\xi} - \mu_\xi(n) \right) + \mu_\xi \quad (6)$$

donde  $\hat{\xi}$  es el estimador de  $\xi$  obtenido por L-momentos para la subserie de longitud  $n$ ;  $\mu_\xi(n)$  y  $\sigma_\xi(n)$  son las funciones definidas en la ec. (5);  $\hat{\xi}_{corr}$  es el estimador corregido;  $\mu_\xi$  y  $\sigma_\xi$  corresponden a la media y desviación estándar asintóticas de  $\xi$ . Finalmente, una vez calculado  $\hat{\xi}_{corr}$ , se actualizan  $\mu$  y  $\sigma$  para mantener la consistencia interna del ajuste GEV. Esta corrección mejora sustancialmente la estabilidad de las estimaciones en estaciones con longitud de registro reducida, evitando valores negativos de  $\xi$ , los cuales son físicamente poco realistas en contextos de precipitaciones extremas. Para aplicar la corrección de sesgo en el parámetro de forma ( $\xi$ ) en la zona de estudio, se utilizan solamente las estaciones con al menos 40 años de datos (243

estaciones). Cada serie se divide en subseries de entre 10 y 55 años de longitud (en incrementos de 3 años), y para cada longitud de registro se estiman los parámetros GEV mediante L-momentos. Con estos resultados se ajustaron las funciones de corrección de la ecuación (5), que se aplicaron primero para corregir los valores de  $\xi$ , y posteriormente  $\mu$  y  $\sigma$ . El límite superior de 55 años se eligió porque longitudes mayores reducen significativamente el número de subseries generadas, quedando con menos de 100 registros.

### 2.3 REGIONALIZACIÓN

Dada la marcada variabilidad hidrológica de Chile, se utiliza el algoritmo K-means para agrupar las estaciones en grupos homogéneos, elegido por su sencillez, eficiencia y eficacia al particionar conjuntos de datos de alta dimensión minimizando la varianza intragrupo (Hartigan & Wong, 1979; Tuel & Martius, 2022). A partir de un análisis exploratorio se seleccionaron atributos de clasificación la latitud, la elevación y el coeficiente de L-variación (L-CV), por su capacidad para capturar gradientes espaciales y el comportamiento extremo de la precipitación. El número óptimo de grupos  $P$  se determina con el método del codo aplicado a la suma de cuadrados intra-grupo (SCIG): se grafica la SCIG en función de  $P$  y se elige el “codo” de la curva, esto es, el punto a partir del cual aumentar  $P$  produce rendimientos decrecientes en la reducción de la SCIG (Hastie et al., 2009).

### 2.4 DISTRIBUCIÓN EV4 PARA LA ESTIMACIÓN DEL FACTOR DE FRECUENCIA MÁXIMO REGIONAL

Para representar la variabilidad del factor de frecuencia máximo dentro de cada región o grupo  $p$ , se considera la serie

$$K_{EV,p,m} = \{K_{EV,1,m}, K_{EV,2,m}, \dots, K_{EV,j,m}, \dots, K_{EV,n_p,m}\}, \quad p = 1, \dots, P \quad (7)$$

donde  $K_{EV,j,m}$  corresponde al factor de frecuencia modificado máximo observado en la estación  $j$  del grupo  $p$ , obtenido al evaluar la ec. (2) en el máximo valor observado de PMA24 de esa estación, y  $n_p$  es el número de estaciones del grupo  $p$ . Sobre esta serie se ajusta la distribución de valor extremo de cuatro parámetros (EV4; Botero & Francés, 2010), que considera límites inferior y superior. Su FDA está dada por:

$$F(x) = \exp\left(-\left[\frac{g-x}{v(x-a)}\right]^k\right), \quad k > 0; v > 0; a \leq x \leq g \quad (8)$$

donde  $a$  y  $g$  son los límites inferior y superior,  $v$  es el parámetro de escala, y  $k$  el de forma. Todos los parámetros, incluyendo el límite superior  $g$ , se estiman mediante máxima verosimilitud.  $g$  cumple un rol fundamental en la estimación de la PMP, ya que, está estrechamente relacionado con el factor de frecuencia regional máximo, el cual define la intensidad máxima esperada de precipitación en una determinada zona homogénea.

### 2.5 PROCEDIMIENTO DE IMPLEMENTACIÓN

El procedimiento completo para estimar la PMP a partir de una base de datos de series PMA24 en  $M$  estaciones se compone de los siguientes pasos:

- i. **Estimación de parámetros GEV y cálculo del factor de frecuencia modificado.** En cada estación  $j$ , los parámetros  $\mu_j, \sigma_j, \xi_j$  se estiman mediante L-momentos. Luego, el parámetro  $\xi_j$

se corrige en función de la longitud de la serie, como se indica en la sección 2.2. Con estos valores, se calcula la serie de factores de frecuencia modificados ( $K_{EV}$ ) de acuerdo con la ec. (2). Posteriormente, se selecciona el valor máximo  $K_{EV,j,m}$  para cada estación  $j$ , usando el valor máximo observado de PMA24  $x_{j,m}$ .

- ii. **Regionalización mediante K-means.** Se aplica el algoritmo K-means al conjunto de estaciones, usando atributos geográficos (latitud, altitud) y estadísticos (eg., media, varianza L-CV), para agrupar las estaciones en  $P$  regiones o grupos homogéneos. El número óptimo de grupos se determina mediante el método del codo, evaluando SCIG.
- iii. **Estimación del factor de frecuencia regional.** Para cada grupo  $p$ , se considera el conjunto de valores máximos  $K_{EV,p,m} = \{K_{EV,1,m}, K_{EV,2,m}, \dots, K_{EV,n_p,m}\}$ , y se ajusta la distribución EV4 a estos datos. El valor regional máximo ( $K_{EV,p}^{MAX}$ ) se toma como el límite superior  $g$  de la distribución EV4 ajustada para el grupo  $p$ .
- iv. **Cálculo de la PMP en cada estación.** Finalmente, la PMP, en [mm], en cada estación  $j$  del grupo  $p$  se calcula usando la expresión inversa de la variable transformada de Gumbel:

$$PMP_{j,p} = \frac{(\sigma_{j,p} \exp(K_{EV,p}^{MAX} \xi_{j,p}) - \sigma_{j,p} + \mu_{j,p} \xi_{j,p})}{\xi_{j,p}}, \quad j = 1, 2, \dots, n_p; p = 1, 2, \dots, P \quad (9)$$

Esta formulación permite incorporar las características locales de cada estación, manteniendo coherencia regional en la extrapolación de valores extremos. Todos los pasos anteriores fueron implementados en R (R Core Team, 2024). Para el ajuste de la EV4, el método de máxima verosimilitud se implementó utilizando el comando `optim()` de la librería `stats`.

## 2.6 MÉTODO DE HERSHFIELD

Con el fin de comparar los resultados del método propuesto, se utiliza el método de Hershfield (Hershfield, 1961), el cual se basa en el coeficiente de frecuencia  $K$  propuesto por Chow (1951), donde el factor de frecuencia máximo para la estación  $j$  ( $K_{j,m}$ ) es definido como:

$$K_{j,m} = \frac{(x_{j,m} - \bar{x}_j)}{SD_j} \quad (10)$$

donde  $x_{j,m}$  es el máximo observado de PMA24 en la estación  $j$ ,  $\bar{x}_j$ , and  $SD_j$  corresponden a la media y desviación estándar de la serie de PMA24 de la estación  $j$ , calculadas excluyendo el valor máximo de la serie. Para la estimación de la PMP en cada estación se siguen los pasos (ii) a (iv) de la sección 2.5, pero la ec. (8) es remplazada por la siguiente expresión:

$$PMP_{j,p} = \bar{x}_j + K_p^{MAX} SD_j \quad (11)$$

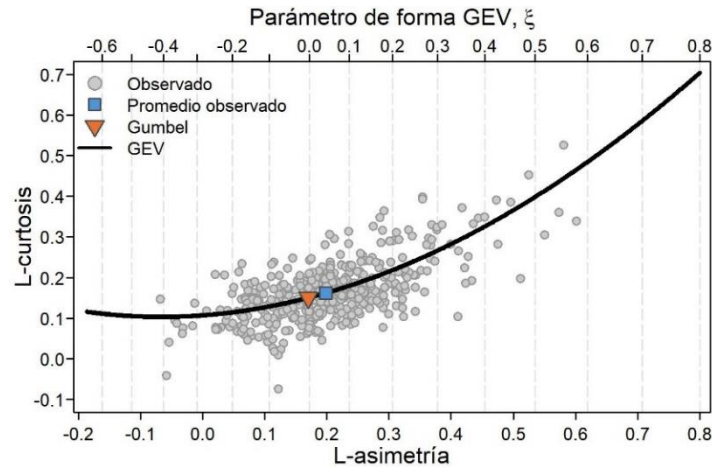
Donde  $K_p^{MAX}$  es el factor de frecuencia máximo regional del grupo  $p$ , obtenido al aplicar el paso (iii) sección 2.5 a la serie de factores de frecuencia máximos  $K_{p,m}$  del grupo  $p$ . En el contexto nacional, Stöwhas (1983) utilizó este enfoque para estimar la PMP; por lo tanto, aquí se utiliza el mismo método para generar una línea base de comparación.

## 4. RESULTADOS

### 4.1 AJUSTE GEV Y CORRECCIÓN DEL PARÁMETRO DE FORMA

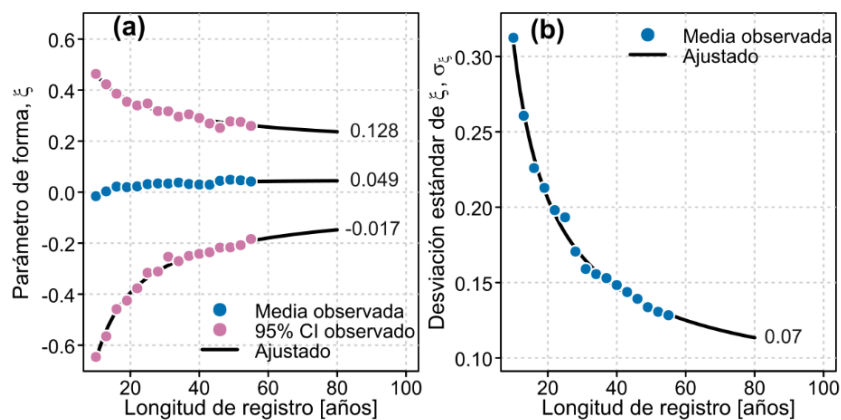
La Figura 3 muestra el diagrama de L-momentos (L-asimetría, L-curtosis), obtenido a partir de las 428 series de PMA24. Para una distribución con forma fija como Gumbel ( $\xi = 0$ ), los pares ( $\tau_3$ ,  $\tau_4$ ) deberían concentrarse en su punto teórico único del plano. En la figura, la nube de puntos se

dispersa ampliamente y no se concentra en ese punto; por el contrario, se distribuye a lo largo de la curva teórica de la familia GEV (línea negra), tal como indica la escala superior que mapea  $\xi$ . Esta separación visible respecto del punto de Gumbel, junto con la variación de  $\xi$ , evidencia alta variabilidad en la forma entre estaciones, por lo que un modelo con forma fija no resulta adecuado para series de extremos, lo que se alinea con lo previamente reportado Koutsoyiannis (2004).



**Figura 3.** Diagrama de L-Momentos ( $\tau_3, \tau_4$ ) para las 428 series de PMA24. Círculos grises: valores por estación. Línea negra: curva teórica de la GEV para distintos  $\xi$  (escala superior). Triángulo naranja: punto teórico de Gumbel ( $\xi = 0$ ).

Por otro lado, la Figura 4 (paneles a y b) muestra la media observada, junto al intervalo de confianza (IC) del 95%, y la desviación estándar del parámetro de forma ( $\xi$ ) de la distribución GEV, en función de la longitud del registro. La figura incluye, además, los ajustes teóricos de la ley de potencia decreciente  $g(n) = a + bn^{-c}$  (línea negra), que serán utilizados corregir el sesgo del parámetro  $\xi$  en función de la longitud del registro (sección 2.2). En todos los casos considerados (media con IC y desviación estándar de  $\xi$ , los datos observados se ajustan a la ley de potencia decreciente y convergen de manera asintótica hacia valores límite a medida que aumenta la longitud del registro; en particular, la media reduce también su variabilidad, en concordancia con lo reportado por Papalexiou & Koutsoyiannis (2013). La Tabla 1 presenta los parámetros estimados para estas curvas.



**Figura 4.** Efecto de la longitud del registro sobre el parámetro de forma ( $\xi$ ) de la distribución GEV. (a) Media observada de  $\xi$  junto al intervalo de confianza (IC) del 95% (puntos) y ajuste  $g(n) = a + bn^{-c}$  (línea negra). (b) Desviación estándar observada de  $\xi$  (puntos) y su ajuste de potencia (línea negra). Las estadísticas

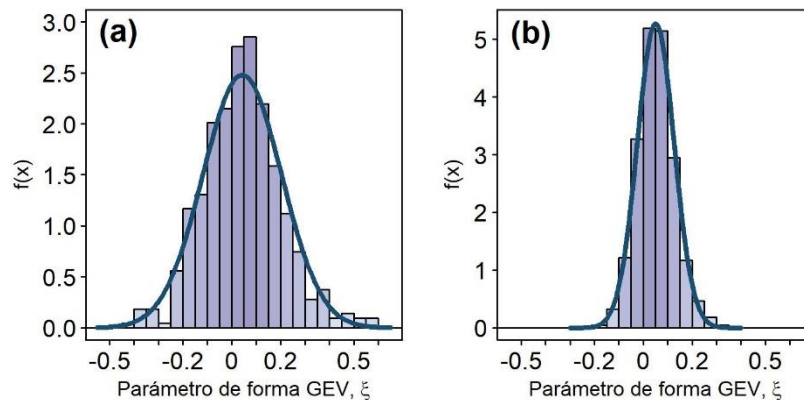
se calcularon a partir de subseries de 10 a 55 años, en incrementos de 3 años, generadas desde estaciones con  $\geq 40$  años de datos (243 estaciones).

**Tabla 1.** Parámetros estimados para las curvas potenciales ajustadas ( $g(n) = a + bn^{-c}$ ) en la Figura 4.

Estadístico	$a$	$b$	$c$
Media	0.049	-1.243	1.295
95% IC inferior	-0.017	-3.754	0.767
95% IC superior	0.128	1.170	0.543
Desviación estándar	0.070	1.596	0.824

Con base en estos resultados, se aplicó una corrección de sesgo al parámetro de forma de las 428 estaciones consideradas en este estudio, utilizando las siguientes funciones en la ec. (6):  $\mu_\xi = 0.049$ ,  $\sigma_\xi = 0.07$ ,  $\mu_\xi(n) = 0.049 - 1.243 \cdot n^{-1.295}$  y  $\sigma_\xi(n) = 0.07 + 1.596 \cdot n^{-0.824}$ . Luego, los parámetros de localización y escala se recalcularon según la ec. (3).

La Figura 5 muestra como aplicación de la corrección de sesgo redujo significativamente la variabilidad del parámetro  $\xi$ , estabilizando su media en torno a 0.05 y disminuyendo su desviación estándar aproximadamente a la mitad (de 0.16 a 0.08). Los parámetros  $\mu$  y  $\sigma$  actualizados tras la corrección presentaron con cambios menores comparados a sus valores iniciales. Esta mejora en la estabilidad de  $\xi$  es clave para evitar estimaciones extremas no realistas en estaciones con registros cortos. Además, se verificó el cumplimiento del supuesto de normalidad en ambas distribuciones mediante el test de Shapiro-Wilk, sin evidencias de desviaciones sistemáticas respecto de la normalidad, lo que respalda la validez del método de corrección aplicado.

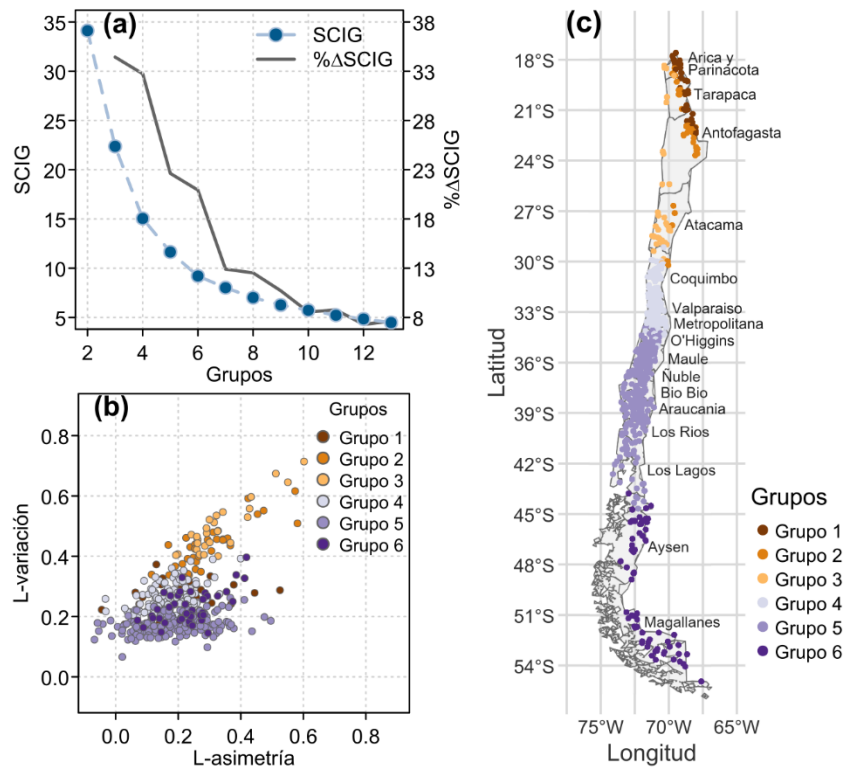


**Figura 5.** Distribución empírica del parámetro de forma ( $\xi$ ) para las 428 estaciones: (a) antes y (b) después de aplicar la corrección de sesgo. Barras: histograma; línea: distribución normal ajustada.

## 4.2 AGRUPAMIENTO ESPACIAL

La Figura 6a, la suma de cuadrados intragrupo (SCIG) decrece al aumentar el número de grupos, pero no aparece un “codo” nítido. Adoptamos seis grupos como solución óptima por tres razones complementarias: (i) a partir de  $P = 6$  las reducciones relativas de SCIG son marginales ( $\% \Delta SCIG \leq 10\%$ ) frente al aumento de complejidad; (ii) con  $P > 6$  se observa en el mapa (Figura 6c) un traslape espacial creciente y la aparición de grupos muy pequeños, lo que dificulta la coherencia regional; y (iii) la nube L-CV vs. L-asimetría (Figura 6b), común en clasificación hidrológica (Viglione et al., 2007), muestra que  $P = 6$  captura de forma parsimoniosa los patrones hidroclimáticos principales sin generar particiones redundantes. Al analizar en conjunto las Figura 6b y 6c se observa una coherencia clara con las regiones hidroclimáticas descritas en la Figura 1b:

se concentran valores altos en el norte árido (grupos 2 y 3), valores bajos en la zona centro-sur (grupos 4 y 5) y aparece una superposición entre los grupos 1 y 6 en áreas de clima tipo tundra del norte andino y la Patagonia austral.



**Figura 6.** Evaluación y resultados del agrupamiento. (a) Suma de cuadrados intra-grupo (SCIG) y su cambio porcentual en función del número de grupos (%ΔSCIG); (b) dispersión de coeficientes L-variación vs. L-asimetría, coloreados según su grupo; y (c) distribución espacial de los grupos en el área de estudio.

La Tabla 2 refuerza la diferenciación de regímenes de precipitación capturada por los seis grupos definidos. Por ejemplo, el grupo 2 presenta la media más baja de PMA24 (16.98 [mm]), junto con una alta variabilidad relativa ( $CV \approx 0.86$ ) y la mayor asimetría (1.62), lo que refleja eventos de precipitación de corta duración, intensos, pero poco frecuentes, típicos de zonas áridas. En contraste, el grupo 5 muestra la media más alta (92.41 [mm]) y la menor variabilidad relativa ( $CV \approx 0.32$ ), consistente con un régimen más uniforme, característico de climas templado-húmedos. Estas diferencias respaldan la efectividad del enfoque de agrupamiento aplicado.

**Tabla 2.** Estadísticos descriptivos de los registros de PMA24 para cada grupo.

Grupo	1	2	3	4	5	6
N° de estaciones	34	33	31	127	162	41
Media	23.07	16.98	23.38	60.36	92.41	42.64
Coefficiente de variación (CV)	0.49	0.86	0.94	0.48	0.32	0.45
Asimetría	1.09	1.62	1.56	0.88	1	1.47

### 4.3 ESTIMACIÓN DEL FACTOR DE FRECUENCIA MÁXIMO REGIONAL

La Tabla 3 presenta estadísticas descriptivas de los factores de frecuencia máximos estimados con los métodos propuesto ( $K_{EV,m}$ ) y de Hershfield ( $K_m$ ) para las 428 estaciones (paso (i), sección 2.5), agrupadas según los clústeres definidos previamente. Para el método propuesto, los valores medios

de  $K_{EV,m}$  varían entre 4.11 y 4.48, con rangos de 2.28 a 7.19. En todos los grupos se observa asimetría positiva, destacando los grupos 2 y 6 (zonas de clima tipo tundra) por registrar valores más elevados. En el método de Hershfield, aunque las medias son similares (entre 3.36 y 4.65), se observa una mayor variabilidad intragrupo, con máximos de hasta 12.92 (grupo 2) y 10.89 (grupo 6), superando incluso el valor máximo de 11 definido por Stöwhas (1983); todos los grupos presentan también asimetría positiva, pero vuelven a resaltar los grupos 2 y 6, en particular el 2, por sus niveles significativamente más altos. Esta combinación de alta dispersión y fuerte asimetría evidencia la sensibilidad del método de Hershfield a valores atípicos y su menor robustez frente a distribuciones sesgadas, especialmente en regiones con registros extremos o escasos.

**Tabla 3.** Estadísticas de los factores de frecuencia máximos ( $K_{EV,m}$  y  $K_m$ ) por grupo.

Grupo		1	2	3	4	5	6
N° de estaciones		34	33	31	127	162	41
$K_{EV,m}$	Min	2.76	3.41	3.22	2.77	2.28	2.89
	Max	6.42	6.15	5.26	6.13	6.62	7.19
	Media	4.22	4.46	4.27	4.11	4.14	4.48
	Asimetría	0.54	0.90	-0.12	0.64	0.46	0.92
$K_m$	Min	1.83	2.61	2.41	1.93	1.50	2.12
	Max	7.77	12.92	6.27	8.15	9.41	10.89
	Media	3.65	4.65	4.15	3.36	3.55	4.39
	Asimetría	1.19	2.56	0.14	1.70	1.64	1.60

Para cada uno de los seis grupos, se ajustó, mediante máxima verosimilitud, la distribución EV4 a la serie de factores de frecuencia máximos del método propuesto ( $K_{EV,m}$ ) y el método de Hershfield ( $K_m$ ); en ambos casos, el parámetro  $g$  estimado representa el factor de frecuencia máximo regional. La Tabla 4 resume estos valores. En el enfoque propuesto, los factores regionales máximos varían entre 5.5 (grupo 3) y 8.9 (grupo 2), en general coherentes con la asimetría reportada en la Tabla 3: los grupos con mayor asimetría del factor tienden a mostrar  $g$  más alto, destacando los grupos 2 y 5; el grupo 6 presenta asimetría elevada pero un  $g$  intermedio. En cambio, el método de Hershfield exhibe una mayor variabilidad entre grupos, con un valor extremo de 26.1 en el grupo 2, muy por sobre el umbral de 11 sugerido por Stöwhas (1983) y superior al resto de los grupos. Además, en Hershfield todos los grupos presentan asimetría positiva, sobresaliendo nuevamente el grupo 2, lo que evidencia la alta sensibilidad de este enfoque a valores atípicos y su limitada robustez frente a distribuciones sesgadas, especialmente en regiones con climas extremos o registros escasos.

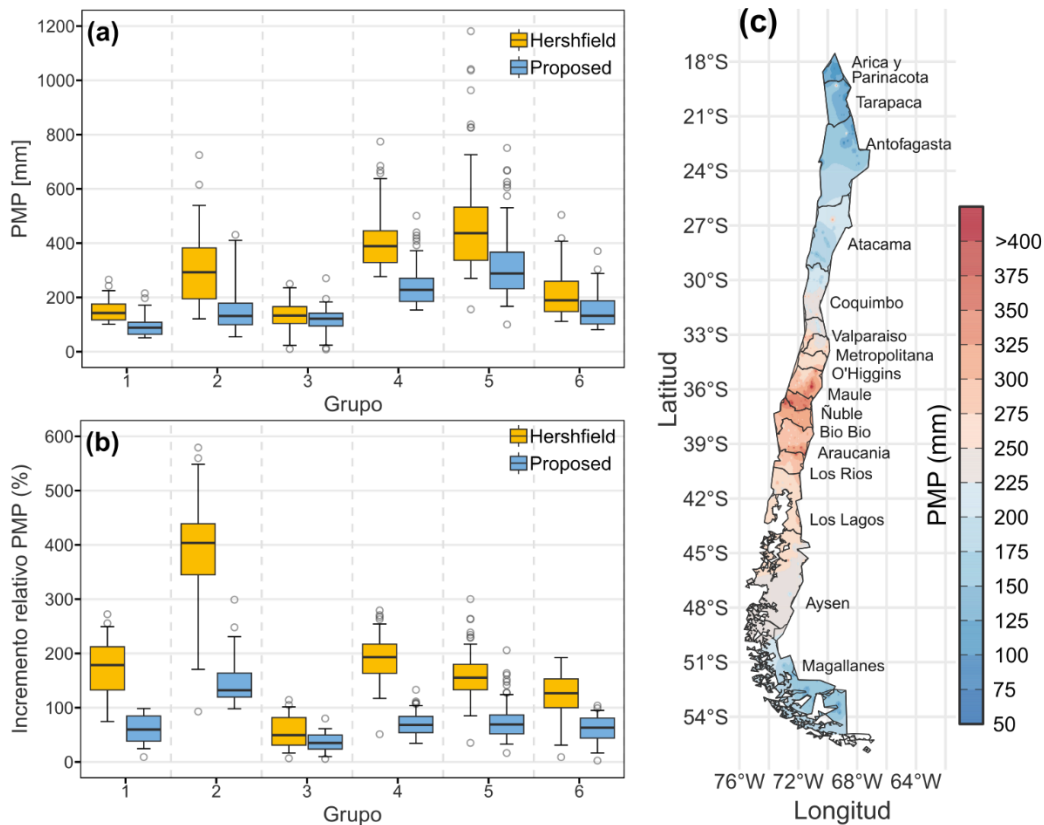
**Tabla 4.** factores de frecuencia máximos regionales de Hershfield ( $K_p^{MAX}$ ) y propuesto ( $K_{EV,p}^{MAX}$ ).

Grupo	1	2	3	4	5	6
$K_p^{MAX}$	13.2	26.1	6.8	13.7	14.1	12.1
$K_{EV,p}^{MAX}$	7.3	8.9	5.5	7.8	8.5	7.4

#### 4.4 ESTIMACIÓN DE LA PMP

Usando los valores de los factores de frecuencia máximos regionales de cada método (propuesto y Hershfield) se estimó la PMP en las 428 estaciones (paso (iv), sección 2.5). Los resultados se muestran en la Figura 7a. En todos los grupos, el método de Hershfield entrega PMP más altas y dispersas que el enfoque propuesto, con diferencias especialmente marcadas en los grupos 2, 4 y 5. Esta mayor variabilidad está directamente relacionada con la alta sensibilidad del método tradicional a valores extremos, lo que se evidencia en el notable rango de valores y asimetrías

reportados en la Tabla 3 y con los factores de frecuencia máximos regionales elevados de la Tabla 4. En particular, el grupo 2 presenta para el método de Hershfield, una mediana de PMP por encima de 290 [mm] y casos que superan los 700 [mm], en concordancia con su  $K_p^{MAX}$  excepcionalmente alto. No obstante, estas magnitudes resultan poco realistas para el norte de Chile, donde los eventos intensos suelen ser de muy corta duración y cobertura espacial limitada; ello sugiere que el enfoque de Hershfield tiende a sobredimensionar la PMP en regiones con eventos extremos aislados y registros escasos, mientras que el método propuesto reduce la dispersión y entrega estimaciones más estables.



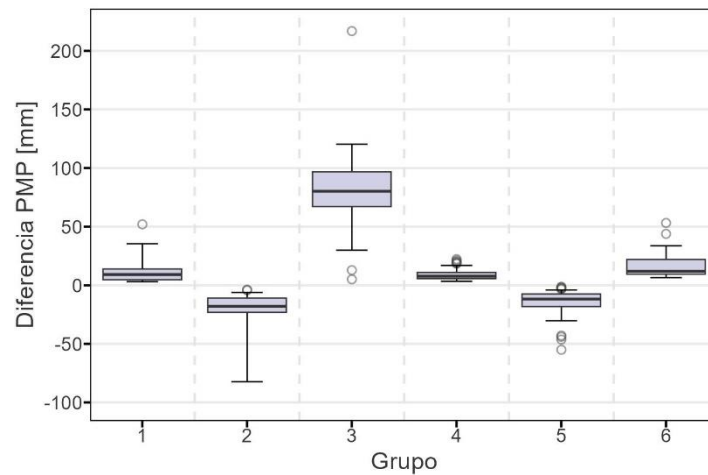
**Figura 7.** (a) Comparación de las estimaciones de la Precipitación Máxima Probable (PMP) mediante el método de Hershfield (amarillo) y el método propuesto (azul) para los distintos grupos. (b) Incremento relativo de la PMP respecto al valor máximo observado de PMA24 (%), para ambos métodos. (c) Distribución espacial de la PMP en Chile estimada con el método propuesto, interpolada mediante IDW.

La Figura 7b complementa estos resultados mostrando el incremento relativo de la PMP respecto al valor máximo observado de PMA24. En todos los grupos, el método propuesto presenta incrementos más moderados y consistentes, típicamente inferiores al 100%, lo que es coherente con factores de amplificación por humedad inferiores a 2, reportados por métodos hidrometeorológicos (Hiraga & Meza, 2025). En contraste, el enfoque de Hershfield muestra incrementos significativamente más altos, destacando nuevamente el grupo 2, con valores que en algunos casos superan el 500%. Estos resultados reflejan las limitaciones del método tradicional en contextos con alta asimetría o escasez de datos, y refuerzan la necesidad de utilizar enfoques más robustos y adaptados regionalmente, como el aquí propuesto.

La Figura 7c muestra la distribución espacial de la PMP estimada mediante el enfoque propuesto, interpolada mediante el método de ponderación por distancia inversa (IDW). Esta representación

revela un gradiente latitudinal coherente, con valores bajos en el norte árido y aumentos progresivos hacia el sur más húmedo, en concordancia con los patrones hidroclimáticos conocidos de Chile.

La Figura 8 cuantifica el efecto de usar un factor de frecuencia único para todo el territorio nacional frente a factores de frecuencia regionales definiendo de  $\Delta PMP = PMP(K_{EV}^{MAX} = 8.1) - PMP(K_{EV,p}^{MAX})$ . Así, valores positivos indican PMP mayores bajo el valor fijo y valores negativos indican PMP menores, ambos respecto del esquema regional. Se observan medianas aproximadamente de +10–15 [mm] en los grupos 1, 4 y 6, -18 [mm] en el grupo 2, +80 [mm] en el grupo 3 y -12 mm en el grupo 5. La dispersión es baja en la mayoría de los grupos, salvo en el grupo 3, que presenta la mayor variabilidad y algunos atípicos sobre los 200 [mm]; el grupo 2 muestra algunos valores atípicos negativos. El signo de  $\Delta PMP$  coincide con la relación entre el factor fijo y el regional de la Tabla 4: cuando  $K_{EV,p}^{MAX}$  del grupo es menor que 8,1 (grupos 1, 3, 4 y 6) las diferencias son positivas; cuando es mayor (grupos 2 y 5) son negativas. Estos resultados subrayan que aplicar un factor de frecuencia único puede inducir diferencias importantes según la región, reforzando la necesidad de enfoques regionalizados que representen con mayor fidelidad la variabilidad espacial de los eventos extremos (p. ej., WMO, 2009; Bracken et al., 2015).



**Figura 8.** Diferencias en las estimaciones de PMP al emplear un valor fijo de  $K_{EV}^{MAX} = 8.1$  para todo el dominio, frente a valores específicos por grupo  $K_{EV,p}^{MAX}$ .

## 5. CONCLUSIÓN

Este estudio presenta un nuevo enfoque estadístico para la estimación de la Precipitación Máxima Probable (PMP), integrando la distribución Generalizada de Valores Extremos (GEV) con una regionalización basada en agrupamiento K-means. Una de sus principales innovaciones es la corrección por sesgo del parámetro de forma ( $\xi$ ), lo que permite mejorar la estabilidad de las estimaciones, especialmente en estaciones con registros cortos o incompletos y con alta proporción de años secos o con PMA24 muy baja, condiciones que generan marcada asimetría en las series. Para la estimación del factor de frecuencia máximo regional de cada grupo, se utilizó el límite superior  $g$  de la distribución acotada EV4

La metodología fue aplicada a una base de datos de 428 estaciones distribuidas en Chile continental, caracterizado por una alta diversidad hidroclimática. Los resultados indican que el método propuesto entrega estimaciones de PMP más coherentes y consistentes a nivel regional en comparación con el enfoque tradicional de Hershfield. Este último exhibe alta sensibilidad a valores

excepcionalmente altos en grupos áridos (p. ej., grupo 2), con factores de frecuencia máximos desproporcionados y elevada asimetría, mientras que el método propuesto mitiga dichos efectos y acota de manera más estable los parámetros regionales (véanse Figuras 7a-7b y Tablas 3-4).

Una contribución destacada es que los valores de PMP estimados mediante el enfoque propuesto presentan incrementos más moderados respecto a los máximos observados de PMA24, típicamente inferiores al 100%, en concordancia con los factores de amplificación por humedad reportados por métodos hidrometeorológicos (Hiraga & Meza, 2025). En cambio, el método de Hershfield muestra incrementos mucho mayores y variables, superando el 500% en zonas como el grupo 2, lo que evidencia su limitada robustez frente a distribuciones sesgadas o registros extremos. Además, se identificaron factores de frecuencia máximos empíricos ( $K_m$ ) que superan el límite de  $K = 11$  establecido por Stöwhas (1983), alcanzando valores de hasta 12.92, lo que refuerza la necesidad de adoptar métodos más robustos.

Por otro lado, el uso de factores de frecuencia máximos regionales ( $K_{EV,p}^{MAX}$ ) representa de mejor forma las diferencias hidroclimáticas entre regiones. Comparado con el uso de un valor fijo nacional ( $K_{EV}^{MAX} = 8.1$ ) —adoptando la convención  $\Delta PMP = PMP(K_{EV}^{MAX} = 8.1) - PMP(K_{EV,p}^{MAX})$ , de modo que  $\Delta PMP > 0$  indica valores mayores con el factor fijo— se observan diferencias sistemáticas por grupo: discrepancias positivas reducidas en los grupos 1, 4 y 6, diferencias negativas en los grupos 2 y 5, y el mayor desplazamiento y dispersión en el grupo 3, con casos cercanos a 200 [mm]. Estos resultados refuerzan la importancia de emplear enfoques regionalizados.

En conjunto, la combinación de modelado GEV corregido por sesgo, el uso de distribuciones con límite superior y la regionalización mediante K-means ofrece un marco metodológico sólido, replicable y físicamente coherente para la estimación de la PMP. Esta aproximación mejora la representatividad espacial de las estimaciones y constituye una alternativa más realista para el diseño de infraestructura hidráulica en contextos de alta variabilidad hidroclimática.

Finalmente, considerando la evidencia de cambios en la PMP bajo condiciones de cambio climático (p. ej., Visser et al., 2022) y la heterogeneidad espacial de los extremos (Bracken et al., 2015), recomendamos revisar y actualizar las estimaciones de PMP de manera periódica (por ejemplo, cada 10 años), incorporando nuevos registros observados y productos reanálisis, y manteniendo un enfoque regionalizado.

## AGRADECIMIENTOS

Esta investigación fue financiada por el proyecto ANID FONDECYT de Iniciación N.º 11250238 y por el proyecto interno de investigación PI\_LIR\_24\_01 de la Universidad Técnica Federico Santa María.

## REFERENCIAS

- Alvarez-Garretón, C., Mendoza, P. A., Pablo Boisier, J., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., et al. (2018). The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies-Chile dataset. *Hydrology and Earth System Sciences*, 22(11), 5817–5846. <https://doi.org/10.5194/HESS-22-5817-2018>
- Botero, B. A., & Francés, F. (2010). Estimation of high return period flood quantiles using additional non-systematic information with upper bounded statistical models. *Hydrology and Earth System Sciences*, 14(12), 2617–2628. <https://doi.org/10.5194/hess-14-2617-2010>
- Bracken, C., Rajagopalan, B., Alexander, M., & Gangopadhyay, S. (2015). Spatial variability of seasonal extreme precipitation in the western United States. *Journal of Geophysical Research*, 120(10), 4522–4533. <https://doi.org/10.1002/2015JD023205>

- Carney, M. C. (2016). Bias Correction to GEV Shape Parameters Used to Predict Precipitation Extremes. *Journal of Hydrologic Engineering*, 21(10). [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001416](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001416)
- Chow, V. Te. (1951). A general formula for hydrologic frequency analysis. *Eos, Transactions American Geophysical Union*, 32(2), 231–237. <https://doi.org/10.1029/TR032I002P00231;PAGEGROUP:STRING:PUBLICATION>
- Gateño, F., Mendoza, P. A., Vásquez, N., Lagos-Zúñiga, M., Jiménez, H., Jerez, C., et al. (2024). Screening CMIP6 models for Chile based on past performance and code genealogy. *Climatic Change* (Vol. 177). <https://doi.org/10.1007/s10584-024-03742-1>
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, 28(1), 100. <https://doi.org/10.2307/2346830>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. <https://doi.org/10.1007/978-0-387-84858-7>
- Hershfield, D. M. (1961). Estimating the Probable Maximum Precipitation. *Journal of the Hydraulics Division*, 87(5), 99–116. <https://doi.org/10.1061/JYCEAJ.0000651>
- Hiraga, Y., & Meza, J. (2025). Extreme precipitation modeling and Probable Maximum Precipitation (PMP) estimation in Chile. *Journal of Hydrology: Regional Studies*, 58(August 2024), 102274. <https://doi.org/10.1016/j.ejrh.2025.102274>
- Hosking, J. R. M., & Wallis, J. R. (1997). *Regional Frequency Analysis: An approach based on L-moments*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511529443>
- Hosking, J. R. M., Wallis, J. R., & Wood, E. F. (1985). Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics*, 27(3), 251–261. <https://doi.org/10.1080/00401706.1985.10488049>
- Hossain, I., Khastagir, A., Aktar, M. N., Imteaz, M. A., Huda, D., & Rasel, H. M. (2022). Comparison of estimation techniques for generalised extreme value (GEV) distribution parameters: a case study with Tasmanian rainfall. *International Journal of Environmental Science and Technology*, 19(8), 7737–7750. <https://doi.org/10.1007/s13762-021-03693-5>
- Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records / Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations. *Hydrological Sciences Journal*, 49(4), 591–610. <https://doi.org/10.1623/HYSJ.49.4.591.54424>
- Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, 49(1), 187–201. <https://doi.org/10.1029/2012WR012557>
- Papalexiou, S. M., & Montanari, A. (2019). Global and Regional Increase of Precipitation Extremes Under Global Warming. *Water Resources Research*, 55(6), 4901–4914. <https://doi.org/10.1029/2018WR024067>
- R Core Team. (2024). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- Salas, J. D., Anderson, M. L., Papalexiou, S. M., & Frances, F. (2020). PMP and Climate Variability and Change: A Review. *Journal of Hydrologic Engineering*, 25(12), 03120002. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0002003](https://doi.org/10.1061/(ASCE)HE.1943-5584.0002003)
- Stöwahas, L. (1983). Precipitaciones máximas diarias en Chile. In *VI Congreso Chileno de Ingeniería Hidráulica* (pp. 455–473). Santiago, Chile.
- Stöwahas, L. (2016). *Fundamentos de Hidrología aplicada*. Valparaíso: Sello editorial USM.
- Tuel, A., & Martius, O. (2022). Subseasonal Temporal Clustering of Extreme Precipitation in the Northern Hemisphere: Regionalization and Physical Drivers. *Journal of Climate*, 35(11), 3537–3555. <https://doi.org/10.1175/JCLI-D-21-0562.1>
- Viglione, A., Laio, F., & Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, 43(3). <https://doi.org/10.1029/2006WR005095>
- Villavicencio, I., Ossandón, Á., Meza, J., & Suazo, G. (2023). Nueva metodología para la estimación de PMP: Integración de la teoría de valores extremos y técnicas de agrupamiento. In SOCHID (Ed.), *XXVI Congreso Chileno De Ingeniería Hidráulica* (pp. 1–14). Valdivia.
- Visser, J. B., Kim, S., Wasko, C., Nathan, R., & Sharma, A. (2022). The Impact of Climate Change on Operational Probable Maximum Precipitation Estimates. *Water Resources Research*, 58(11), e2022WR032247. <https://doi.org/10.1029/2022WR032247>
- WMO. (2009). *Manual on estimation of Probable Maximum Precipitation (PMP) (WMO-No. 1045)*. Geneva, Switzerland: World Meteorological Organization.