

SOCIEDAD CHILENA DE INGENIERÍA HIDRÁULICA
XXVI CONGRESO CHILENO DE INGENIERÍA HIDRÁULICA

**NUEVA METODOLOGÍA PARA LA ESTIMACION DE PMP: INTEGRACION DE LA
TEORIA DE VALORES EXTREMOS Y TECNICAS DE AGRUPAMIENTO**

IGNACIO VILLAVICENCIO A.¹
ÁLVARO OSSANDÓN Á.²
JOAQUIN MEZA A.²
GONZALO SUAZO F.²

RESUMEN

La precipitación máxima probable (PMP) es un parámetro de suma importancia en el diseño de estructuras de gran envergadura, debido a que es considerado como el límite físico superior para la precipitación en un lugar determinado. Por lo tanto, su estimación es clave en el diseño. Por ello, esta investigación se centra en plantear nueva metodología estadística para la estimación de la PMP, donde se utiliza de base la aplicación de una distribución generalizada de valores extremos (GEV) y cuyos parámetros se estiman mediante el método “L-moments”. El método propuesto se centra en el análisis de un coeficiente de frecuencia modificado (ψ), equivalente al coeficiente de frecuencia (K) propuesto por Hershfield (1961), el cual considera la no normalidad de los datos. Con el fin de estudiar la heterogeneidad espacial de los datos, se realiza una regionalización mediante análisis de agrupamiento. Para encontrar el valor máximo de los coeficientes de frecuencia ψ y K, tanto para el análisis regional como general, se utilizan dos distribuciones acotadas. Los resultados indican variaciones en la estimación de la PMP para ambos métodos a través de una regionalización, debido principalmente a la heterogeneidad climática del país. Comparando ambos métodos regionalizados con el límite superior (K=11) establecido para Chile por Stöwahas (1983) se observa que en general los valores de PMP son menores para ambos: Teniendo un -23% de variación promedio para el método propuesto regionalizado y -15% de variación promedio para Herhsfield regionalizado.

¹Alumno, Depto. de Obras Civiles, Universidad Santa María – ignacio.villavicenci@sansano.usm.cl

²Académico, Depto. de Obras Civiles, Universidad Santa María – alvaro.ossandon@usm.cl; joaquin.meza@usm.cl; gonzalo.suazo@usm.cl

1. INTRODUCCIÓN

En el contexto de la ingeniería Civil, es de suma importancia considerar el riesgo de inundaciones al diseñar obras de gran envergadura como son puentes, embalses depósitos de relaves o centrales hidroeléctricas. Las inundaciones representan una amenaza significativa, debido a que se pueden traducir en pérdidas humanas y daños a la propiedad importantes (Salas et al., 2020). Para evitar este tipo de situaciones y asegurar la seguridad de los proyectos, es esencial desarrollar enfoques de la evaluación de riesgo y diseño hidráulico. Esto se puede cuantificar a través de la magnitud y frecuencias de una tormenta de diseño. Para estimar estas tormentas de diseño, la precipitación máxima probable (PMP) es un enfoque que se ha utilizado en varias partes del mundo (Rezacova et al., 2005; Desa M. & Rakhecha, 2007). El fin de este enfoque es implementar un límite teórico para el valor de precipitación utilizada para el diseño de las estructuras anteriormente mencionadas.

Los métodos existentes para la estimación de la PMP se pueden clasificar en dos grandes grupos: Métodos hidrometeorológicos y métodos estadísticos (Salas et al., 2020). Los métodos hidrometeorológicos se basan principalmente en el análisis de los procesos físicos asociados a la precipitación, específicamente la maximización de los parámetros implicados en estos procesos. Sin embargo, su implementación resulta compleja debido a requieren un nivel de experiencia o conocimiento en el área, alto poder de cómputo, además de una amplia cantidad de datos meteorológicos.

Por otro lado, los métodos estadísticos se basan en la trasposición de tormentas desde un punto de vista estadístico. Este enfoque supone un área de tormenta homogénea (un área donde las tormentas tienen características similares), condiciones de estacionariedad y la normalidad de los datos. Estos métodos surgieron a partir de Hershfield (1961), quien utiliza el coeficiente de frecuencia máximo (K_m) para representar la máxima precipitación de una estación. Este coeficiente es estandarizado por su media y desviación estándar y se calculan para la serie de máximas anuales. En dicho estudio, se definió un valor de un coeficiente de frecuencia máximo de un conjunto de estaciones de diferentes partes del mundo, obteniendo un valor de 15. Este valor, al ser fijo, tiende a sobrestimar la PMP en zonas lluviosas y la subestimar en zonas de poca precipitación (Hershfield, 1965). Debido a lo anterior, Hershfield (1965) implementa una modificación al método donde define K_m^{max} como una curva envolvente en función de la media, que logra caracterizar la variabilidad del comportamiento de las estaciones. Desde un punto de vista práctico, este método tiene la ventaja que necesita una limitada cantidad de información y es sencillo de aplicar.

Debido a esto, el método estadístico se ha utilizado en varias partes del mundo, como en la India (Rakhecha & Soman, 1994), Chile (Stöwhas, 1983; Blanlot, 2021), entre otros países (Casas et al., 2008; Wangwongwiroj & Khemngoen, 2019). Dentro de ellos, resalta el trabajo de Stöwhas (1983), que aplico a Chile tanto el trazado de la curva envolvente como la obtención de un K_m^{max} , el que corresponde a un valor 11. Este último valor ha sido ampliamente utilizado en la práctica ingenieril como el límite máximo para el diseño hasta la actualidad. A pesar del amplio uso que tiene el método de Herhsfield, su fundamento se basa en supuestos cuya validez se ha visto cuestionado en el tiempo. Entre estos supuestos se encuentran: (1) Normalidad de los datos, que supone un ajuste normal de las precipitaciones máximas anuales de las estaciones; (2) Homogeneidad espacial, la cual supone un comportamiento uniforme en la totalidad de la cuenca. Sin embargo, esta suposición tampoco es necesariamente valida, especialmente cuando el área de análisis es extensa. Blanlot

(2021) intento abordar la heterogeneidad espacial de los datos mediante métodos de agrupamiento, obteniendo claras sobreestimaciones en algunas zonas de Chile.

Con el objetivo de abordar la no normalidad y heterogeneidad espacial de los datos de precipitaciones máximas, se propone una nueva metodología para la estimación de la precipitación máxima probable (PMP) para un país con geografía y clima tan heterogéneos como el nuestro, mediante la integración de la teoría de valores extremos y técnicas de agrupamiento. Para llevar a cabo esto, se ajustaron funciones de distribución generalizada de valores extremos (GEV) a las series de precipitaciones máximas diarias de 431 estaciones, obteniendo un coeficiente de frecuencia modificado (ψ) que considera la asimetría de los datos para cada estación. Luego, para abordar la heterogeneidad de los datos, se realizó un análisis de agrupamiento utilizando el método K-means. Los resultados obtenidos son comparados con los valores de PMP obtenidos al considerar $K_m^{max} = 11$ (Stöwhas, 1983) y los obtenidos al aplicar K-means los valores de K_m de las estaciones analizadas.

2. METODOLOGÍA

2.1 MÉTODO DE ESTIMACIÓN DE PMP PROPUESTO BASADO EN LA TEORÍA DE VALORES EXTREMOS

El presente estudio se centrará en el enfoque de “Bloque máximo” (Ej: precipitaciones máximas anuales para un periodo definido) para la estimación de la PMP de cada estación. Este enfoque está asociado a la distribución generalizada de valores extremos (GEV), donde para la serie de tiempo de precipitaciones máximas diaria anuales, x , de una estación, la función de probabilidad acumulada queda expresada por:

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

Donde los parámetros propios de la función son la ubicación ($\mu > 0$), escala ($\sigma > 0$) y forma (ξ). Este último parámetro permite clasificar la GEV en tres tipos: (i) Tipo 1 o Gumbel ($\xi = 0$); (ii) Tipo 2 o Fréchet ($\xi > 0$); Tipo 3 o Weibull ($\xi < 0$). La estimación de los parámetros de la GEV para cada estación analizada se utiliza el método de máxima verosimilitud generalizada (GML).

Una vez determinado los parámetros de la distribución, se introduce un nuevo enfoque que conserva la estructura del método de Hershfield (1961), pero utilizando un coeficiente de frecuencia modificado (ψ). Este coeficiente corresponde a una variable transformada a partir de la serie original x , la cual tiene una distribución Gumbel ($\xi = 0$) estandarizada ($\mu = 0, \sigma = 1$) (Katz, 2013), y queda definida como:

$$\psi = [1/\xi] \ln\{1 + \xi [x - \mu]/\sigma\} \quad (2)$$

Luego, su valor máximo (ψ_m) corresponderá al obtenido para el valor máximo de la serie original x , quedando de la siguiente forma:

$$\psi_m = [1/\xi] \ln\{1 + \xi [X_m - \mu]/\sigma\} \quad (3)$$

De esta manera, despejando X_m , y reemplazando ψ_m por un ψ_m^{MAX} (un valor máximo para una región definida), se obtiene la PMP para la estación mediante la siguiente expresión:

$$PMP = (\sigma e^{\psi_m^{MAX}\xi} - \sigma + \mu\xi)/\xi \quad (4)$$

El valor de ψ_m^{MAX} se estima como el límite superior de una función de probabilidad acotada ajustada a la serie de ψ_m de las estaciones correspondientes a una región definida.

2.2 MÉTODO DE HERSHFIELD

Con el fin de comparar los resultados del método propuesto, se utiliza el método de Hershfield, el cual se basa en el coeficiente de frecuencia “K” (Chow, 1951) donde K es definido por la siguiente expresión:

$$K_m = \frac{X_m - \bar{X}_n}{S_n} \quad (5)$$

Donde X_m , representa el máximo valor de la serie, \bar{X}_n y S_n representa la media y desviación estándar de la serie respectivamente. Este término es el coeficiente de frecuencias máximo, y se calcula para una serie de “n” datos, de una estación en específico. En el contexto nacional (Stöwhas, 1983; Blanlot, 2021) han utilizado esta expresión, por lo que se utilizara esta expresión, para generar poder generar una línea base de comparación.

La expresión para estimar la PMP en consecuencia se puede estimar mediante la siguiente expresión:

$$PMP = \bar{X}_n + K_m^{max} S_n \quad (6)$$

En este caso el coeficiente de frecuencia máximo regional (K_m^{max}), representa el máximo valor del coeficiente de frecuencias máximo de una determinada región.

2.3 FUNCIONES DE PROBABILIDAD ACOTADAS

Con el fin de encontrar los máximos valores regionales y para la extensión de estaciones utilizadas, se utilizan dos funciones de distribución acotada: La distribución de Weibull acotada, la cual ya ha sido utilizada para estimar un límite superior de K_m en Chile (Stöwhas, 1983) y la distribución de valores extremos de 4 parámetros (EV4), la cual ha presentado un buen desempeño con extremos hidrológicos (Botero & Francés, 2010). Estas distribuciones se aplican a las series de ψ_m y K_m .

2.4 REGIONALIZACIÓN

Para la delimitación de regiones homogéneas se utiliza un análisis de agrupamiento, el cual es un método del análisis multivariado, que permite reducir una compleja cantidad de información en grupos o clústeres, donde los componentes comparten características. El análisis de agrupamiento es una técnica exploratoria que no utiliza ningún tipo de modelo estadístico para llevar a cabo la clasificación, razón por la cual se clasifica como una técnica de aprendizaje no supervisada. Estos grupos pueden estar o no superpuestos (Chavan & Srinivas, 2017).

En este estudio se testaron 3 métodos de agrupamiento: K-means con distancia euclidiana; K-medoids con distancia euclidiana; y K-medoids con distancia modificada según Wu et al. 2022. Los primeros dos métodos muy populares en diferentes ámbitos (Madhulatha, 2011), mientras el

tercer método se utiliza debido a que incluye la componente espacial en la distancia, lo que resulta útil para una regionalización. Estos métodos se implementaron considerando diferentes números de grupos (entre 4 y 6) y combinaciones de 5 variables (latitud, longitud, coeficiente de asimetría (Cs), coeficiente de variación (Cv), y parámetro de forma) para definir regiones homogéneas del coeficiente de frecuencia. Los parámetros estadísticos se obtienen a partir de la serie de máximos anuales. Para el caso del Cv, se hace una corrección en la media y la desviación estándar, debido a que existe un sesgo que depende de la longitud de registro (Hershfield, 1965). Independiente del método de agrupamiento, es necesario estandarizar las variables en función de su rango de variabilidad, para así evitar problemas con el orden de magnitud de las diferentes variables involucradas en el agrupamiento.

3. ZONA DE ESTUDIO

Para llevar a cabo este trabajo, se utilizan los datos de precipitación acumulada diaria registrados en 879 estaciones meteorológicas al largo de Chile, las cuales están contempladas durante un periodo desde enero de 1900 hasta junio de 2020 (Blanlot, 2021). La base está compuesta por 845 estaciones pertenecientes a la Dirección General de Aguas (DGA) y las 34 estaciones restantes corresponden a la Dirección Meteorológica de Chile (DMC). Para poder generar la serie de precipitaciones máximas diarias anuales de cada estación, se consideran solo los años en los que la precipitación máxima diaria no sea nula (caso posible en algunas estaciones del norte de Chile). En el caso que un año que no tenga registro completo, este puede ser incluido en la serie de máximos diarios anuales si tiene no más de un 33% de datos faltantes o si su precipitación máxima diaria anual supera el cuantil 40 de la serie de máximos diarios anuales. Este filtro ayuda a mitigar el sesgo que podría surgir al utilizar todos los máximos diarios anuales disponibles o al eliminar por completo los años con datos faltantes (Papalexiou & Koutsoyiannis, 2013). Luego, para obtener las series de precipitaciones máximas anuales en 24 horas, se amplifican las series de precipitaciones máximas diarias anuales por un factor de 1.13 (World Meteorological Organization, 2009). Este factor es considerado debido a que el registro diario (desde 8 AM a 8 AM) no necesariamente coincide con el valor máximo ocurrido en 24 horas. Finalmente, se consideran solo las estaciones que tengan series de 25 años o más de extensión temporal, lo que da un total de 431 estaciones. La Figura 1a muestra en color negro las estaciones utilizadas y en rojo las estaciones que tienen menos de 25 años de registro no nulo. Además, en la Figura 1b se visualiza la altura de las distintas estaciones, lo que ayuda a identificar las estaciones cordilleranas y no cordilleranas.

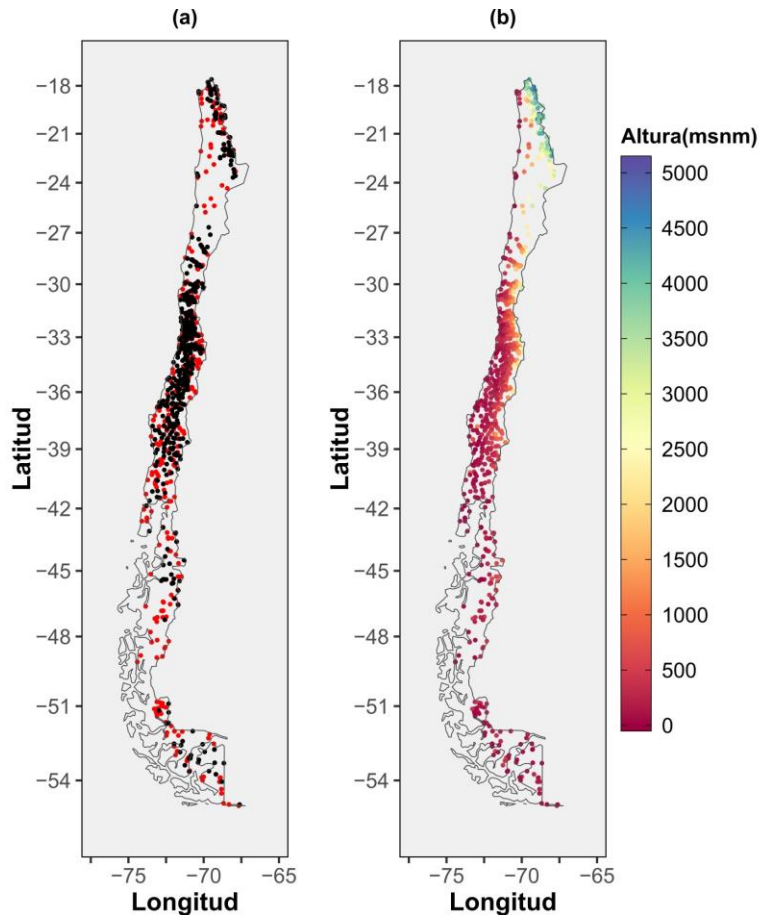


Figura 1. Estaciones compiladas. (a) Las estaciones utilizadas se presentan en color negro y en color rojo las que tienen menos 25 años de registro no nulo; (b) Altura de las estaciones (msnm).

4. RESULTADOS Y DISCUSION

4.1 ANÁLISIS CUALITATIVO DE GRUPOS

Luego de un exhaustivo análisis de los métodos de agrupamiento y variables a considerar, se estima que el método que mejor logra distinguir los grupos de la zona norte y extremo sur del país, es el k-means para 6 grupos. Para determinar los 6 grupos, se utilizaron el coeficiente de variación, longitud y latitud. En el análisis de sensibilidad de las variables se constató que el CV corregido es un factor preponderante en la generación de zonas más consistentes. Además, al integrar las variables de ubicación, como latitud y longitud, se logró diferenciar las estaciones de gran altura y baja altura en el norte grande ($\text{Latitud} > -28.3$), y obtener una regionalización consistente con las macrozonas del país reportadas por Alvarez-Garretón et al., (2018). La Figura 2 muestra la regionalización obtenida, donde se observa una correcta diferenciación entre la zona norte y central del país. La zona norte abarca desde la región de Arica y Parinacota hasta la región de Atacama. Además, dicha zona se encuentra dividida en dos grupos: grupo 1, que corresponde a la región de norte grande cordillerana; y el grupo 2 que está compuesto por la zona no cordillerana del norte grande. Esta separación resulta razonable, ya que a grandes alturas existen otro tipo de comportamiento climático. Referente a los grupos de la zona centro, el grupo 3 abarca desde la región de Coquimbo hasta aproximadamente la región Metropolitana. A pesar de la presencia de estaciones en las zonas montañosas de esta región, las alturas son inferiores en comparación con

las del norte grande, y el CV no caracteriza esta información. Por otro lado, el grupo 4 abarca desde la región de O'Higgins hasta la región de Los Ríos, grupo que se espera tenga una media espacial de la precipitación máxima diaria anual mayor a la del grupo 3. En la parte sur-austral del país, se distinguen dos grupos específicos. El grupo 5 engloba las regiones de Los Lagos y la de Aysén. Mientras que el grupo 6 abarca exclusivamente la región de Magallanes. Como observación general los grupos de la zona centro y sur, tienen una relación directa con las regiones políticas del país, con la salvedad de las estaciones que están en zonas más cercanas a la cordillera en la zona norte. Con esto queda en evidencia la posible relación de 3 variables: Clima (en este caso se limita al análisis los valores de precipitaciones), altura y CV. Lo cual puede justificar el por qué se mezclan algunas estaciones en altura en las zonas limítrofes de los grupos.

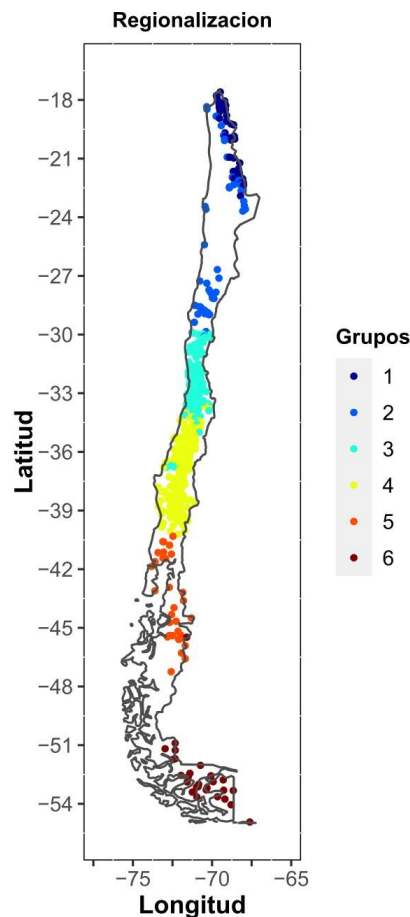


Figura 2. Regionalización usando método de clúster K-mean, para 6 grupos.

4.2 ANÁLISIS DE VARIABLES POR GRUPO

Dado a que las variables más importantes en relación con los métodos analizados son ψ_m y K_m , resulta útil caracterizarlas estadísticamente por grupos y en su totalidad (sin regionalizar). En la Tabla 1 se pueden ver las características de las series de ψ_m y K_m por grupo y total (sin regionalizar). Cuando se examina el CV de estas variables, se observa que en términos generales la dispersión general de ψ_m es menor que la de K_m . Por otra parte, se puede ver que el rango de ψ_m es mayor al de K_m para los grupos 3 a 6, habiendo particularmente una gran diferencia para el

grupo 4 (6.9 para ψ_m y 4.1 para K_m). Valores altos del rango pueden repercutir en el cálculo de PMP de las estaciones que se encuentren en los cuantiles extremos de cada grupo.

Tabla 1. Resumen de K_m y ψ_m para la regionalización y la totalidad de estaciones utilizadas.

		Grupo						Total
		1	2	3	4	5	6	
Cantidad de estaciones		42	49	133	147	31	28	456
ψ_m	Mínimo	3.2	3.2	3.0	2.5	3.5	3.0	2.5
	Máximo	7.3	5.1	7.0	9.4	7.2	6.3	9.4
	Media	4.4	4.1	4.2	4.3	4.5	4.4	4.3
	Desviación	0.891	0.438	0.772	0.942	0.760	0.892	0.826
	Rango	4.1	1.9	4	6.9	3.7	3.3	6.9
	Cs	1.089	0.205	1.071	1.953	1.575	0.665	1.593
	Cv	0.203	0.107	0.184	0.219	0.169	0.203	0.192
K_m	Mínimo	1.7	2.2	1.8	1.4	2.2	1.9	1.4
	Máximo	6.2	5.4	4.8	5.5	4.9	5.7	6.2
	Media	3.1	3.4	2.9	2.9	3.3	3.4	3.0
	Desviación	0.908	0.678	0.639	0.732	0.708	1.005	0.758
	Rango	4.5	3.2	3	4.1	2.7	3.8	4.8
	Cs	1.186	0.661	0.76	0.699	0.629	0.778	0.843
	Cv	0.293	0.199	0.220	0.252	0.215	0.296	0.253

Adicionalmente, la Tabla 2 muestra los valores promedio de los estadígrafos asociados a la precipitación máxima anual en 24 horas de las estaciones calculados para cada grupo y el total. En dicha tabla se destaca que el grupo 2 presenta valores Cs y Cv promedio (1.706 y 0.966, respectivamente) notablemente más altos que el resto de los grupos. Lo anterior implica que este grupo tiene la mayor dispersión y cantidad de valores atípicos en cada serie de máximos, lo que claramente no puede ser capturado por un enfoque que asuma normalidad de los datos. Este atributo si logra ser capturado por el ajuste de una GEV, lo que se ve reflejado en un mayor promedio positivo del parámetro de forma ($\xi = 0.237$) y bajo parámetro de ubicación $\mu = 8.86$. Estos valores bastante distintos a los obtenidos para el resto de los grupos y el análisis global, donde el rango de ξ para el resto de los grupos varía entre 0 y 0.12 y μ es superior a 14 para todo el resto de los grupos.

Tabla 2. Valores promedio por grupo y total de estadígrafos asociados a la precipitación máxima anual en 24 horas de cada estación.

Estadígrafo	Grupo						Total
	1	2	3	4	5	6	
Media (mm)	19.3	15.9	54.2	82.9	63.6	25.9	55.1
Cv	0.546	0.966	0.487	0.321	0.333	0.510	0.481
Cs	1.201	1.706	0.927	0.953	1.266	1.629	1.122
Forma, ξ	0.039	0.237	0.008	0	0.064	0.120	0.047
Ubicación, μ	14.75	8.86	42.28	71.00	54.03	20.07	45.02

Para el caso del grupo 1, al ser estaciones que se encuentran en altura y en la zona norte del país, es una zona donde llueve poco y también tiene precipitaciones atípicas que aumentan el Cs, tiene un Cv promedio similar al del conjunto total de estaciones. Al analizar los grupos restantes, observamos que los grupos 3 y 4, que corresponden a la zona centro norte y centro sur respectivamente, presentan parámetros promedios que se mantienen dentro del rango de variación de los valores del resto de los grupos y no lejanos a los valores globales (Tablas 1 y 2). Sin embargo,

difieren de forma sustancial en cuanto al promedio de precipitación máxima diaria anual, donde el grupo 4 presenta el mayor valor correspondiente a 82.9 mm y alrededor de 30 mm más que el grupo 3 (Tabla 3), lo que es consistente con el régimen de precipitaciones de la zona comprendida entre la región de O'Higgins y Los Ríos, donde en el extremo sur del grupo, generalmente se registra el mayor monto de precipitaciones anuales totales a nivel país. En el caso del grupo 5, se registra la media más alta para ψ_m (4.5), lo que se puede atribuir a sus características climáticas y topográficas y al bajo número de estaciones que lo componen (34 estaciones). Respecto al grupo 6, posee el tercer promedio más bajo de precipitación máxima anual en 24 horas y el K_m (junto al grupo 2) promedio más alto de todos los grupos. Además, se identifica que también es el segundo grupo con el valor más alto de Cs, lo que hace que sea susceptible a valores atípicos, al igual que el grupo 2. Los atributos observados para este grupo pueden ser explicados por la existencia de un clima tipo tundra en algunas zonas de esta región del país, lo que implica zonas frías y de baja cantidad de precipitación.

En general se puede ver los grupos 2 y 6 se destacan como los más atípicos dentro del conjunto de áreas estudiadas. No obstante, esto se encuentra justificado en gran medida debido a la disposición geográficas de las estaciones que lo conforman. Los grupos 2 y 4 son los grupos que más difieren entre sí, lo que se verifica tanto para los parámetros de ubicación y forma de la distribución GEV como para sus parámetros estadísticos análogos (promedio y Cs). Dicho contraste entre estos dos grupos también es consistente con los coeficientes de frecuencia, debido a que estos poseen valores promedio en los dos extremos de los datos. Por lo tanto, esto permite inferir que la regionalización cumplió su objetivo principal que era caracterizar la heterogeneidad del país. Al analizar los valores por grupo de los diferentes parámetros estadísticos, se puede ver una posible relación entre el valor promedio coeficiente de frecuencia (K_m) y el coeficiente de asimetría (Cs). Esta relación, se verifica con una correlación de Spearman significativa de 0.86 considerando un nivel de significancia del 5%. Adicionalmente, esta correlación puede ser corroborada en la Figura S1.

4.3 ESTIMACIÓN DE PMP REGIONAL Y TOTAL

En la Tabla se presentan los K_m^{max} y ψ_m^{max} obtenidos para cada grupo, así como para el conjunto total, empleando dos funciones de distribuciones acotadas. Se observa que existen casos, donde ambas distribuciones dan valores similares, como es en el caso del grupo 1, 2 y 6 de ψ_m^{max} . Adicionalmente, el K_m^{max} para el caso total presenta un valor de 10. Este valor es un poco más bajo que el valor de $K=11$ obtenido por Stöwhas (1983), a pesar de que en su análisis se excluyeron las estaciones del norte grande, mientras que en este análisis si se incluyen dichas estaciones. En base a un análisis gráfico y métricas estadísticas de bondad de ajuste como el Criterio de Información de Akaike (AIC), se utilizan los valores obtenidos de la distribución Weibull para el método de Hershfield (K_m^{max}) y los obtenidos mediante la EVA para el método propuesto (ψ_m^{max}).

Tabla 3. K_m^{max} y ψ_m^{max} calculados usando distribución Weibull y EV4 para la regionalización y la totalidad de estaciones.

	Grupos						Total
	1	2	3	4	5	6	
K_m^{max} Weibull	7.7	8.9	8.2	10.5	7.1	9.1	10
ψ_m^{max} Weibull	8.6	5.7	14.9	11.5	6.4	6.4	29.9
K_m^{max} EV4	7.7	6	5.2	5.9	5.8	6.7	6.5
ψ_m^{max} EV4	8.4	5.3	7.6	9.7	10	6.5	9.6

En la Figura 3 se presenta la comparación de los resultados obtenidos para análisis regional de K_m^{max} y ψ_m^{max} , expresada como la diferencia relativa respecto al valor máximo obtenido mediante un análisis global, que corresponden a $K_m^{max} = 10$ y $\psi_m^{max} = 9.6$, respectivamente. En la Figura 3a se puede ver que para el método de Hershfield regional se obtienen valores similares, pero mayormente diferencias relativas negativas respecto a su análogo global, salvo en el grupo 4, donde se presentan valores de PMP mayores para el análisis regional que el global, producto del alto valor de K_m^{max} regional (10.5). Este tipo de datos, series de precipitaciones máximas anuales en 24 horas, que presentan un grado de no normalidad alto, claramente no pueden ser abordados correctamente por el método de Hershfield, generando una incorrecta estimación de la PMP. Mientras que para el método de propuesto (Figura 3b), se aprecia una clara sobreestimación de la PMP al utilizar ψ_m^{max} uniforme para todo el país. Además, se aprecia que para el grupo 2 aumenta aún más la sobreestimación de la PMP obtenida con un ψ_m^{max} global. Esto último se explica por los altos coeficientes de forma y bajos valores del coeficiente de ubicación (μ) de dicho grupo (Tabla 2), lo que indica que el método propuesto es más idóneo con datos no normales. Además, existen algunas estaciones en el extremo sur en las que la PMP es subestimada utilizando un ψ_m^{max} global, lo que es indicio de que existe un parámetro con igual importancia que ψ_m^{max} para la estimación de PMP. Dicho parámetro corresponde al parámetro de forma, debido a que las estaciones en las que se subestima la PMP poseen un parámetro de forma alto. Por último, independiente del método utilizado para la regionalización de la PMP, los resultados obtenidos resaltan la importancia de la regionalización para la estimación de la PMP.

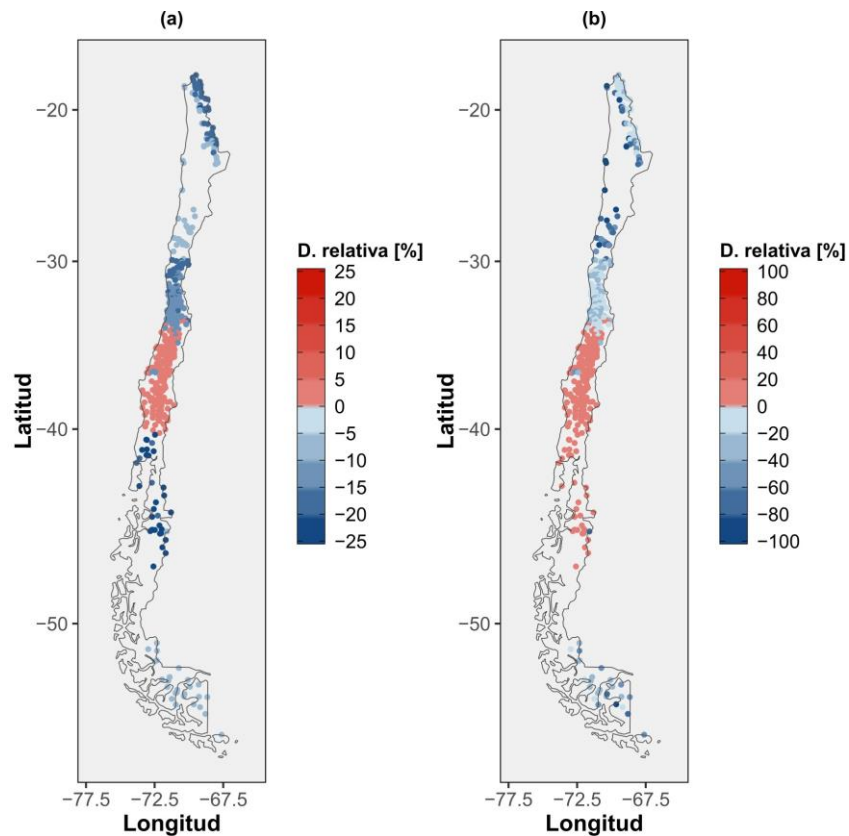


Figura 3. Diferencia relativa [%] de las PMP estimadas utilizando coeficientes de frecuencia máximos regionales respecto al coeficiente de frecuencia máximo de la totalidad de las estaciones para el (a) método de Hershfield (K_m^{max}) y (b) método propuesto (ψ_m^{max}). Las diferencias relativas están expresadas en porcentaje.

Para poder establecer un contraste con la práctica común en Chile para la estimación de la PMP, en la Figura 4 se presentan las distribuciones de la PMP (boxplots) para cada grupo, obtenidas utilizando el coeficiente de frecuencia máximo uniforme ($K=11$) propuesto por Stöwahas (1983), y los K_m^{max} y ψ_m^{max} de los métodos de Hershfield y propuesto regionalizados. En términos de la mediana y rango inter-cuartil, se puede apreciar que para todos los grupos (excepto el grupo 5), el método propuesto exhibe magnitudes de PMP más bajas comparadas a las obtenidas con los métodos de Stöwahas (1983) y Hershfield regionalizado. En cuanto al 95% de la distribución, se aprecia que nuevamente el método propuesto exhibe valores de PMP más bajos para la mayoría de los grupos con excepción de los grupos 4 y 5. Lo que produce que las estimaciones de PMP mediante el método propuesto para algunas estaciones de los grupos 4 y 5 sean mayores a las obtenidas mediante los otros métodos es la existencia de una gran dispersión del parámetro de forma en dichos grupos, una de los parámetros de mayor importancia para la estimación de la PMP. En la Tabla 3, se puede apreciar que el grupo 4 posee un parámetro de forma promedio de 0, lo que implica que pueden existir algunas estaciones con parámetro de forma negativo, lo que producirá estimaciones de PMP sesgadas hacia valores más bajos producto de un ajuste deficiente de los parámetros de la GEV. Esto último, se fundamenta en base a que un parámetro de forma negativo no es correcto para una serie de máximos anuales (Papalexiou & Koutsoyiannis, 2013).

Al analizar los outliers (puntos negros) de las distribuciones del método propuesto, los que en general entregan un mayor valor de PMP respecto a los otros métodos, se obtiene como característica común que todas estas estaciones poseen parámetros de forma altos (>0.3), lo que da cuenta del alta no normalidad de dichas series de datos. Sin embargo, Papalexiou & Koutsoyiannis (2013) reportó que la estimación del parámetro de forma de una GEV puede resultar muy sensible a la longitud del registro de datos, obteniéndose valores más consistentes para series con una longitud mayor a 40 años, que este caso representan aproximadamente un tercio del total de las estaciones, lo que limita la cobertura espacial para la aplicación del método.

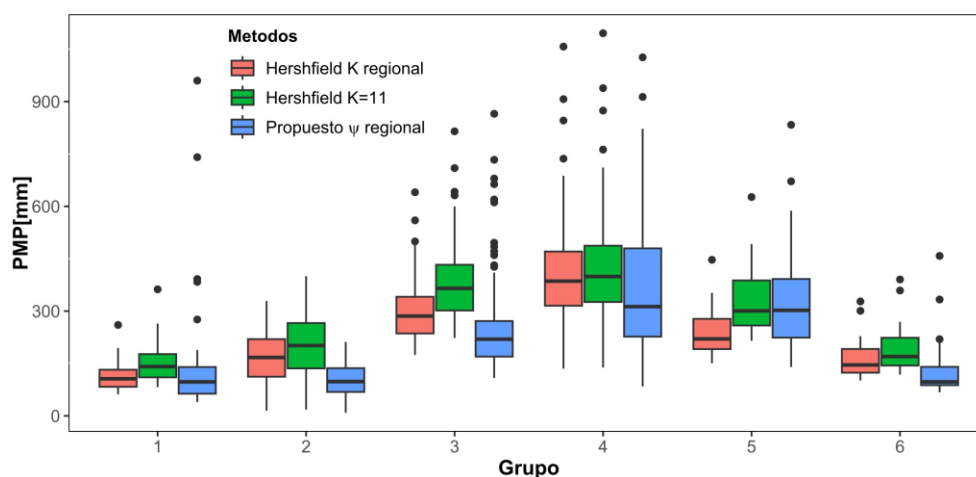


Figura 4. Distribución de la PMP (boxplots) para cada grupo, obtenidas utilizando el coeficiente de frecuencia máximo uniforme ($K=11$; Stöwahas, 1983), y los K_m^{max} y ψ_m^{max} de los métodos de Hershfield y propuesto regionalizados. La barra horizontal dentro de las cajas corresponde a la mediana, el largo de las cajas al rango Inter cuartil (75%), el largo de los bigotes al 95%, y los puntos a outliers. 2 valores de PMP del grupo 4 obtenidos mediante el método propuesto no son presentados en el gráfico, ya que superan los 1000 mm (tienen valores de forma muy altos).

5. CONCLUSIONES Y RECOMENDACIONES

En este artículo, se ha llevado a cabo un análisis regional para la estimación de la precipitación máxima probable (PMP), en el que se han considerado 431 estaciones en el territorio nacional, con un registro mínimo de 25 años, empleando el algoritmo de agrupación K-means. Paralelamente, se ha aplicado la distribución de valores extremos generalizada (GEV) para obtener un coeficiente de frecuencia modificado ψ_m , análogo al planteado por Hershfield (1961), pero que considera la no normalidad de los datos. A partir de la implementación de lo mencionado, se han calculados los valores máximos de ψ_m por cada grupo detectado y para el global de los datos. Esto ha posibilitado comparaciones a nivel global y por grupos que resaltan la relevancia de los métodos utilizados y la regionalización propuesta en la estimación de la PMP en Chile.

Los análisis realizados han destacado la importancia de la regionalización, independientemente del método empleado, especialmente en un país con geografía y clima tan heterogéneos como el nuestro. En cuanto a la comparación entre los dos métodos y sus variantes, se han observado dos conclusiones principales: (1) los valores de precipitación máxima probable (PMP) obtenidos mediante el método propuesto son generalmente inferiores a los calculados con el método de Hershfield (-23%); y (2) el método propuesto entrega valores más bajos de PMP con una variación menor en el norte y extremo sur del país, lo que demuestra que trabaja mejor junto a series con valores atípicos.

Una potencial limitante o desafío a ser abordado en un trabajo futuro corresponde a la posible sensibilidad del parámetro de forma de la GEV de acuerdo con la longitud del registro de datos (Papalexiou & Koutsoyiannis, 2013). Este aspecto reviste una importancia significativa en la región central del país, donde el parámetro de forma tiende a acercarse a cero. Esto podría sugerir que, hasta la fecha de los datos disponibles (hasta 2020), los datos podrían distribuir normal. Sin embargo, la ocurrencia de las tormentas del mes de agosto recién pasado, particularmente en la estación de Curicó, donde la precipitación máxima alcanzó valores cercanos a los 170 mm, superando el máximo histórico de 100 mm, indica un cambio significativo en las condiciones climáticas de la zona. El problema de lo anterior es que con el método de Hershfield, considerar este valor atípico, podría llevar a sobreestimaciones en futuras construcciones, debido a que ese valor aumentaría bastante el promedio y desviación estándar. Este aspecto es especialmente relevante en nuestro país donde no se dispone de registro de datos de gran longitud en la mayoría de las estaciones. A todo lo anterior, además se suma el hecho de los métodos estadísticos para la estimación de la PMP asumen estacionariedad de los datos, supuesto que difícilmente siga siendo válido bajo el escenario de cambio climático actual. Por lo tanto, en trabajos futuros se intentará abordar la corrección de la estimación del parámetro de forma de la GEV en función de la longitud de registro, estimación de la PMP para duraciones mayores a 24 horas, la incorporación de la no estacionariedad en la estimación de la PMP, y realizar su estimación considerando escenarios futuros de cambio climático.

Por último, aun considerando sus limitantes, el método propuesto arroja resultados que corresponde una mejora para la estimación de la PMP comparado al método utilizado tradicionalmente para la estimación de la PMP en Chile, ya que considera la no normalidad y heterogeneidad espacial de los datos. Además, este método se presenta como una herramienta viable para su implementación en otras regiones del mundo, que cuenten con estaciones con registros de precipitación más extensos. Asimismo, se vislumbra que el método propuesto, considerando correcciones para

superar las limitaciones antes mencionadas, tiene un gran potencial de aplicación en Chile, tanto en análisis de riesgo como para el diseño y verificación del diseño de infraestructura de gran envergadura.

AGRADECIMIENTOS

Deseo expresar mi agradecimiento a la Universidad Técnica Federico Santa María, que a través del Proyecto Interno USM 2022 PI_LII_23_06, ha brindado el soporte esencial para esta investigación. Agradezco especialmente a la Dirección de Postgrados y Programas (DPP) por financiar y facilitar mi formación académica. Finalmente, extendiendo mi gratitud a Francisco Blanlot, por su generosidad al proporcionar los datos cruciales para esta investigación y por su invaluable orientación.

REFERENCIAS

- Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., 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>
- Blanlot, F. (2021). Reanálisis y actualización de métodos para la estimación de la PMP en cuencas chilenas. Universidad Técnica Federico Santa María, Valparaíso, Chile. <https://repositorio.usm.cl/handle/11673/53085>
- 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>
- Casas, M. C., Rodríguez, R., Nieto, R., & Redaño, A. (2008). The Estimation of Probable Maximum Precipitation. *Annals of the New York Academy of Sciences*, 1146(1), 291–302. <https://doi.org/10.1196/annals.1446.003>
- Chavan, S. R., & Srinivas, V. V. (2017). Regionalization based envelope curves for PMP estimation by Hershfield method. *International Journal of Climatology*, 37(10), 3767–3779. <https://doi.org/10.1002/JOC.4951>
- 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>
- Desa M., M. N., & Rakhecha, P. R. (2007). Probable maximum precipitation for 24-h duration over an equatorial region: Part 2-Johor, Malaysia. *Atmospheric Research*, 84(1), 84–90. <https://doi.org/10.1016/j.atmosres.2006.06.005>
- 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>
- Hershfield, D. M. (1965). Method for Estimating Probable Maximum Rainfall. *Journal - American Water Works Association*, 57(8), 965–972. <https://doi.org/10.1002/J.1551-8833.1965.TB01486.X>
- Katz, R. W. (2013). Statistical Methods for Nonstationary Extremes, 15–37. https://doi.org/10.1007/978-94-007-4479-0_2
- 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>
- Rakhecha, P. R., & Soman, M. K. (1994). Estimation of probable maximum precipitation for a 2-day duration: Part 2 ? north Indian region. *Theoretical and Applied Climatology*, 49(2), 77–84. <https://doi.org/10.1007/BF00868192>
- Rezacova, D., Pesice, P., & Sokol, Z. (2005). An estimation of the probable maximum precipitation for river basins in the Czech Republic. *Atmospheric Research*, 77(1–4), 407–421. <https://doi.org/10.1016/j.atmosres.2004.10.011>

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)

Stöwhas, L. (1983). Precipitaciones máximas diarias en Chile. In Congreso VI, SOCHID (pp. 455–473).

Wangwongwiroj, N., & Khemngoen, C. (2019). Probable maximum precipitation in tropical zone (Thailand) as estimated by generalized method and statistical method. *International Journal of Climatology*, 39(13), 4953–4966. <https://doi.org/10.1002/joc.6119>

Wu, J., Liu, X., Li, Y., Yang, L., Yuan, W., & Ba, Y. (2022). A Two-Stage Model with an Improved Clustering Algorithm for a Distribution Center Location Problem under Uncertainty. *Mathematics*, 10(14), 2519. <https://doi.org/10.3390/math10142519>

ANEXO

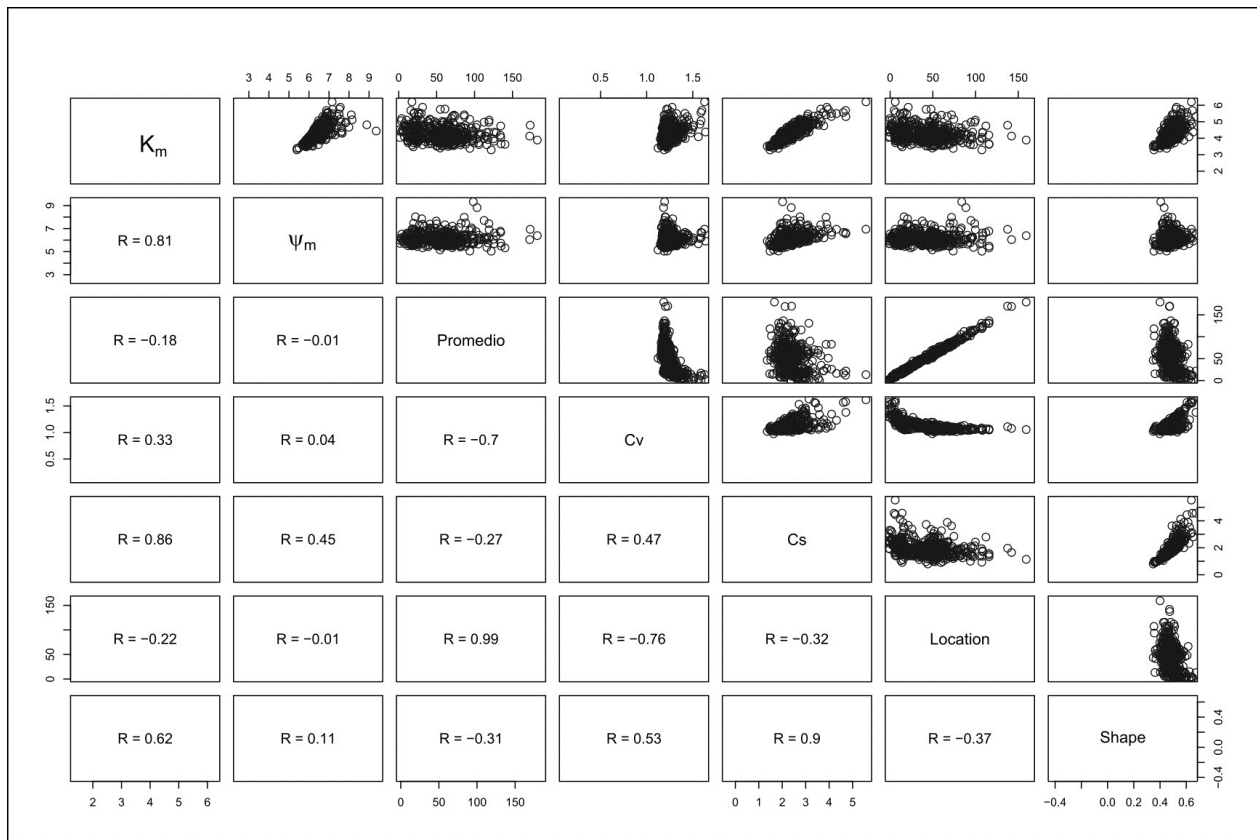


Figura S1. Gráficos de dispersión y correlación de Spearman entre parámetros analizados.