UNIVERSIDAD TECNICA FEDERICO SANTA MARIA

Repositorio Digital USM **https://repositorio.usm.cl**

Tesis USM TESIS de Pregrado de acceso ABIERTO

2021-05

REANÁLISIS Y ACTUALIZACIÓN DE MÉTODOS PARA LA ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE EN CUENCAS CHILENAS

BLANLOT MUCKE, FRANCISCO JOSÉ

https://hdl.handle.net/11673/53085

Repositorio Digital USM, UNIVERSIDAD TECNICA FEDERICO SANTA MARIA

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE OBRAS CIVILES
VALPARAÍSO, CHILE

“REANÁLISIS Y ACTUALIZACIÓN DE MÉTODOS PARA LA ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE EN CUENCAS CHILENAS”

FRANCISCO JOSÉ BLANLOT MUCKE

MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

PROFESOR GUÍA: DR. RAÚL FLORES AUDIBERT

Comisión Revisora:

DR. ÁLVARO OSSANDÓN

Mayo, 2021

Resumen

La precipitación máxima probable cumple un rol muy importante en la seguridad e integridad de la población, puesto que la magnitud de estos extremos teóricos determina las capacidades de almacenamiento y evacuación de obras hidráulicas de gran envergadura, ubicadas comúnmente aguas arriba de ciudades importantes. En Chile y en el mundo se han realizado múltiples estudios, que proponen metodologías para estimar la magnitud de los eventos extremos de manera estadística o física, en función de la información que se disponga. En el presente estudio se genera una base de datos actualizada que contempla la totalidad de estaciones meteorológicas públicas disponibles en Chile. A partir de ella, se recalibran y actualizan las metodologías estadísticas propuestas en estudios previos. Debido a la variabilidad de la precipitación a lo largo del territorio nacional, se evalúa también el comportamiento de ellas a nivel regional. Esto último se lleva a cabo mediante análisis de cluster, lo que permite proponer distintas alternativas de regionalización. Si bien los resultados de la actualización y regionalización validan y complementan lo expuesto en estudios anteriores, el comportamiento no estacionario de las estadísticas producto del cambio climático, y la incertidumbre asociada al no cumplimiento de la hipótesis de homogeneidad climática, sugiere investigar alternativas para la incorporación de estos efectos en las estimaciones.

Abstract

The Probable Maximum Precipitation accomplishes a very important role in the safety and integrity of the hydraulic infrastructure during events of extreme rainfall, fact that´s related to the safety and health of the population. Multiples studies have been realized in Chile and the whole world, which propose methodologies to estimate the magnitude of extreme events in a statistical or physical way, depending on the amount and quality of the available information. This study is an updated database which includes the largest number of meteorological stations available in Chile. Based on it, the previous methodologies proposed were recalibrated and updated, also including their behavior at local level. The latter is carried out through data cluster analysis, this allows proposing different regionalization alternatives. The non-stationary behavior of the statistics resulting from climate change, and the uncertainty associated with the non-compliance with the hypothesis of climate homogeneity, suggests investigating alternatives for the incorporation of these effects in the estimates. All these while the results of the update and regionalization validate and complement what was stated in previous studies.

Índice

1 Introducción 7

2 Objetivos 8

3 Estado del arte 9

3.1 Gestación de tormentas 9
3.2 Precipitación máxima probable 10
3.2.1 Enfoque directo: métodos hidrometeorológicos 11
3.2.2 Enfoque indirecto: métodos estadísticos 16

4 Metodología 28

4.1 Compilación de las bases de datos 28
4.2 Análisis de frecuencia y teoría de valores extremos 30
4.3 Análisis de cluster 31
4.4 Aplicación del método de Hershfield,1961 a la realidad chilena 34
4.5 Implementación computacional del método de Stowhas, 1983 36
4.6 Formación de conglomerados y análisis a nivel regional 38

5 Resultados 41

5.1 Resultados de la actualización de los métodos 41
5.1.1 Máximo coeficiente de frecuencia (Hershfield, 1961) 41
5.1.2 Envolventes de precipitación máxima adimensionalizada 42
5.2 Resultados del análisis regional 44
5.2.1 General 44
5.2.2 Obtención de la PMP a nivel regional 54

6 Discusión 57

6.1 Sobre la actualización de los métodos 57
6.1.1 A partir de la PMP estimada en cada estación 57
6.1.2 De los resultados obtenidos bajo los distintos escenarios 62
6.2 Sobre la regionalización 64
6.2.1 Evaluaciones 64
6.2.2 Del análisis de clúster para definir las regiones 66
6.2.3 De la validez de cada método a nivel regional 67
6.3 Incertidumbre en la estimación de la PMP 68

7 Conclusiones 73

8 Referencias 76

9 Anexo A: Códigos 80

9.1 Python 80
9.1.1 Filtrado y depurado de la BBDD fuente 80
9.1.2 Generación de estadísticos para recalibración de M. Hershfield 81
9.1.3 Generación de matrices con parámetros para análisis de cluster 89
9.1.4 Generación de resumen muestral, y corrección de promedios y desviaciones estándar por longitud de registro y evento máximo 97

9.2 Matlab 99
9.2.1 Código Análisis de Cluster y obtención de valores límites (Matlab) 99
10 Anexo B: Figuras complementarias 104
10.1 Comparación entre envolventes de escenarios A, A2, B y Stow,1983 104
10.2 Envolventes intermedias de M. Stowhas 105

Índice de Figuras

Figura 1: Factores de ajuste para el promedio y desviación estándar de la serie de máximos anuales, por efectos del máximo evento observado (World Meteorological Organization, 1986) 18
Figura 2: Factores de ajuste del promedio y desviación estándar de la serie de máximos anuales, por efectos de longitud de registro (World Meteorological Organization, 1986) .. 19
Figura 3: Factor de ajuste de la PMP por número de observaciones por intervalo de medición (Weiss,1964) 20
Figura 4: $Kmmáx$ para distintas duraciones de tormenta y promedios de precipitación máxima anual (World Meteorological Organization, 1986; D. M. Hershfield, 1965) 21
Figura 5: Factor de reducción de la PMP por efectos de duración de la tormenta (Huff, 1967) y área de la cuenca para EEUU occidental (United States Weather Bureau, 1960) .. 22
Figura 6: Coeficientes máximos de frecuencias en función del promedio de precipitaciones máximas anuales y coeficientes de variación (Stowhas, 1983). 27
Figura 7: Distribución espacial de las estaciones meteorológicas compiladas en la base de datos 29
Figura 8: Ajustes de distribuciones generalizadas de valores extremos (GEV) a las series de coeficientes de frecuencia máximos, escenarios A, A2 y B 41
Figura 9: Envolventes intermedias y envolventes dobles para precipitación máxima adimensionalizada y centrada - Escenario B 42
Figura 10: Actualización de las envolventes de precipitación máxima adimensionalizada y centrada, respecto de los estadígrafos CV-P con y sin estaciones en altura 43
Figura 11: Solución 1, alterativa de regionalización mediante Análisis de Clúster 46
Figura 12: Solución *2*, alterativa de regionalización mediante Análisis de *cluster* 47
Figura 13: Validación interna de la Solución 1 - Silhouette width index 50

Figura 14: Validación visual de los clusters; dispersión de la precipitación acumulada anual, altitud y máxima histórica (Solución 1) 51
Figura 15: Validación interna de la Solución 2 - Silhouette width index 52
Figura 16: Determinación del número óptimo de clusters mediante distintos índices (Solución 2) 52
Figura 17: Validación visual de los clusters; dispersión de la altitud, precipitación acumulada anual y máxima histórica (Solución 2) 53
Figura 18: Validación visual de los clusters; dispersión de la precipitación acumulada anual, altitud y máxima histórica (Solución 2) 53
Figura 19: Envolventes de Stowhas en cada región - Solución 1 55
Figura 20: Envolventes de precipitación máxima probable y precipitaciones máximas históricas en Chile, dónde ET representa la envolvente trazada con el total de estaciones y SEC la envolvente trazada excluyendo estaciones de régimen nival. 60

Índice de Tablas

Tabla 1:Cantidad de estaciones utilizadas en escenarios A, A2 y B 36
Tabla 2: Características de la muestra según (Stowhas, 1983) y escenarios A, A2 y B 36
Tabla 3: Parámetros de la distribución generalizada de valores extremos ajustada a cada escenario 41
Tabla 4: Parámetros que definen las relaciones envolventes Pm(P) de figura 10- Escenario B 43
Tabla 5: Características del Análisis de Clúster realizado 44
Tabla 6: Estaciones eliminadas en las soluciones 1 y 2 45
Tabla 7: Máximos coeficientes de frecuencia obtenidos al aplicar el método de Hershfield a cada Clúster – Solución 1 54
Tabla 8: Máximos coeficientes de frecuencia obtenidos al aplicar el método de Hershfield a cada Clúster – Solución 2 54
Tabla 9: Parámetros de funciones envolventes regionalizadas, método de Stowhas 2020 .. 56
Tabla 10: Casos en los que la PMP calculada utilizando M. Stowhas es (no es) menor que la obtenida vía M.Hershfield, considerando las envolventes Pm-P actualizadas y k=0.967 58

Tabla 11: Estaciones en las que los máximos históricos sobrepasan la PMP estimada vía método Stowhas, considerando las envolventes anteriores (Stowhas, 1983) y actualizadas (Escenario B). 60
Tabla 12: Estaciones consideradas en el cálculo de la PMP a nivel regional 64
Tabla 13: Resultados de estimaciones de la PMP en las 5 estaciones de mayor longitud de registro, según métodos de Hershfield y Stowhas con envolventes globales y regionales. . 65
Tabla 14: Diferencias promedio a nivel regional entre estimaciones regionales y globales, de las 5 estaciones con mayor longitud de registro de cada región 65

1 Introducción

La aproximación tradicional al problema de determinar caudales para el diseño de obras hidráulicas ha sido el uso de análisis de frecuencia de estadísticas de caudales o precipitaciones, con el propósito de determinar las magnitudes de las variables hidrológicas en estudio en términos de una determinada probabilidad de excedencia. Esta metodología adquiere gran incertidumbre para el caso de eventos de gran magnitud, por la inevitable necesidad de asumir una función de densidad de frecuencia y extrapolar estas funciones a períodos de retorno, que muchas veces exceden considerablemente la longitud del período estadístico base. (Stowhas B, 1983)

Ante la inminente necesidad de proyectar las grandes obras hidráulicas cada vez más seguras, ha surgido el criterio de diseñar en términos de la denominada "crecida (precipitación) máxima probable", concepto que pretende definir el valor máximo que una crecida (precipitación) puede alcanzar cuando se maximizan todas las variables y parámetros que intervienen en su gestación. Para ello existen básicamente dos metodologías: la primera, conocida como "método hidrometeorológico", establece que una tormenta será máxima, cuando se maximice el contenido de humedad de la atmósfera y la velocidad de convergencia de dicha humedad, hacia el lugar donde se ha centrado la tormenta (Stowhas B, 1983). En la práctica, este método es complejo y requiere disponer de una red de estaciones meteorológicas con información de precipitación, temperatura, humedad y velocidad del viento, a nivel horario; limitando la aplicabilidad del método. El segundo método, propuesto por Hershfield (1965), estima la PMP a partir de una envolvente superior, en función de la precipitación media, trazada respecto de los máximos coeficientes de frecuencia que hayan sido medidos en diversas regiones.

Stowhas (1983) adaptó y aplicó el método de Hershfield a la realidad chilena, proponiendo un método estadístico de simple aplicación, para determinar la precipitación máxima probable en Chile, método que es muy utilizado en consultorías tradicionales de ingeniería. Aun cuando han pasado casi 40 años desde que se propuso el método, éste no se ha actualizado considerando las nuevas estadísticas pluviométricas, ni las tendencias asociadas al cambio climático.

En este trabajo se busca actualizar los parámetros y valores envolventes propuestos por Stowhas (1983); utilizando la estadística disponible a la fecha, un mayor número de estaciones pluviométricas, y nuevas técnicas estadísticas/probabilísticas. Además, mediante análisis de clúster, se busca establecer, de ser posible, una zonificación con distintas envolventes y coeficientes para estimar la PMP en una determinada región.

2 Objetivos

El objetivo general de la presente memoria es actualizar la metodología para determinar la precipitación máxima probable en Chile basada en métodos estadísticos/probabilísticos, que resulte en un método de fácil aplicación en consultoría en Ingeniería Civil. Dentro de los objetivos específicos se tiene:

1. Generar una base de datos de estadística de precipitación actualizada (al año 2020), que incorpore el mayor número posible de estaciones pluviométricas y años de registro.
2. Contrastar los valores de los coeficientes presentados por Stowhas, L. (1983) para la determinación de las PMP en Chile (obtenidos a partir de las estadísticas de 190 Estaciones Meteorológicas), con los coeficientes obtenidos de aplicar la misma metodología a las estadísticas disponibles a la fecha.
3. Efectuar un análisis regional para el trazado de las envolventes de precipitación, proponiendo, de ser posible, una zonificación con los distintos coeficientes para determinar la PMP en Chile.
5. Estudiar posibles variaciones a la metodología propuesta por Hershfield y posteriormente modificada por Stowhas, en términos de herramientas estadísticas/probabilísticas del estado del arte.

3 Estado del arte

3.1 Gestación de tormentas

Cuando la circulación de masas de aire húmedo se ve obstaculizada por la presencia de barreras orográficas naturales (cadenas montañosas) dispuestas perpendicularmente a la dirección del viento, el aire se ve obligado a ascender, enfriándose hasta condensar. Este proceso da origen a precipitaciones del tipo **orográficas**, lo que a su vez explica que en las vertientes a barlovento de las montañas la precipitación es bastante mayor que a sotavento de las mismas (Stowhas, 2016).

Si bien las precipitaciones orográficas puras suelen generar sólo lloviznas, las tormentas en general son consecuencia de la combinación de este mecanismo con algún otro, que puede ser ciclónico o convectivo.

Las precipitaciones **ciclónicas** se producen por la presencia de una zona de baja presión (ciclón), que provoca la convergencia de aire hacia ese punto a través de un movimiento en espiral, debiendo el aire ascender y enfriarse hasta condensar. En general estas tormentas son de duración prolongada y cubren uniformemente una gran extensión.

Por otra parte, existen precipitaciones de tipo **convectivo,** cuyo mecanismo generador del ascenso de la masa de aire es de origen térmico. El ascenso de aire genera nubes del tipo cúmulus (con de gran desarrollo vertical), dando origen a lluvias de gran intensidad y corta duración, típicamente observadas en zonas tropicales o períodos calurosos en zonas templadas.

En Chile, salvo las precipitaciones altiplánicas (invierno boliviano) y algunas precipitaciones de verano en la cordillera, que son del tipo convectivo, las principales precipitaciones son de origen ciclónico-frontal.

Los frentes, que se generan normalmente sobre el océano Pacífico, son desplazados por los vientos que en esas regiones predominan en dirección Oeste – Este (hacia la costa y el territorio de Chile), provocando la mayoría de las precipitaciones desde la III Región hacia el Sur.

El desplazamiento sucesivo de un frente caliente seguido de uno frío en un lapso de uno a dos días debiera en principio generar dos períodos de mal tiempo, separados por algunas horas de tiempo inestable, aún cuando en la práctica, los frentes calientes tienden a pasar desapercibidos. Al alcanzar los frentes la zona continental, se hace presente el efecto orográfico debido a la presencia de la cordillera de la Costa y cordillera de Los Andes, que obligan a las masas a ascender aún más, provocando un aumento en las precipitaciones a barlovento de las montañas, y su diminución a sotavento, generando en definitiva, una distribución bastante más irregular de las precipitaciones que la que correspondería a un fenómeno ciclónico – frontal puro (Stowhas, 2016).

Por otra parte, el desplazamiento anual en el sentido norte – sur del Ecuador térmico (provocado por la inclinación del eje terrestre), provoca a su vez el desplazamiento estacional de los frentes de mal tiempo, generándose el clima característico de Chile. La zona norte, al encontrarse permanentemente bajo predominio de condiciones anti ciclonales, es de carácter desértico, mientras que la zona central presenta una clara distribución de precipitaciones que se concentran en los meses de invierno. La zona sur, en cambio, se mantiene permanentemente bajo la influencia del frente polar, lo que explica que las precipitaciones sean bastante más parejas entre el invierno y el verano (Stowhas, 2016).

3.2 Precipitación máxima probable

La precipitación máxima probable (PMP) se define como la mayor cantidad teórica de precipitación, de una duración dada, que es físicamente (meteorológicamente) posible, sobre una cuenca hidrográfica determinada y para un período del año dado, sin tener en cuenta las tendencias climáticas que se producen a largo plazo (World Meteorological Organization, 1986). La finalidad de la estimación de la PMP es poder determinar la crecida máxima probable (CMP) para un proyecto dado, en una cuenca determinada.

Para estimar dicha magnitud, actualmente existen básicamente dos enfoques metodológicos; el primero de ellos, conocido como método directo (o estadístico), adopta un enfoque basado en la zona de la cuenca hidrográfica, estimando directamente la PMP para la zona de recogida de un proyecto dado, en una cuenca determinada. El segundo

enfoque metodológico, conocido como método indirecto (o hidrometeorológico), utiliza un enfoque basado en la zona de la tormenta, extensa y con condiciones hidrológicas homogéneas; determinando primero la PMP para cada familia de isoyetas que rodean la zona de tormenta (considerando distintas duraciones y distintas áreas) y luego convirtiéndola en la PMP de la zona del proyecto en particular.

3.2.1 Enfoque directo: métodos hidrometeorológicos

Debido a las diferentes características que pueden adquirir las tormentas, según sea el mecanismo que provoca la condensación, para una duración dada de ellas será generalmente el mismo tipo de precipitación el que origina las tormentas críticas que tienden a producir la PMP en el lugar (Stowhas, 1983).

Las precipitaciones mas desfavorables serán cortas, intensas y de carácter local si son del tipo convectivo; y de larga duración y sobre zonas extensas, si son del tipo ciclónicas. A las características anteriores se les superpone el efecto que tienen las barreras orográficas, en los procesos de elevación, enfriamiento, condensación y posterior precipitación de la humedad atmosférica. Este último efecto frecuentemente adquiere más importancia que las características ciclónicas o convectivas, creando precipitaciones del tipo orográfico, de propiedades distintas dependiendo de las condiciones geomorfológicas del lugar.

Para entender el problema de manera general, conviene analizar primero los principales factores o condiciones meteorológicas que determinan la magnitud de una tormenta no orográfica, siendo evidente que la PMP resultará de la maximización de todas ellas.

(i) **Contenido de humedad de la atmósfera**: mientras mayor sea el contenido de agua que sea capaz de almacenar la atmósfera, mayor será la cantidad de agua que podrá precipitar.

(ii) **Velocidad de condensación**: La intensidad con que el agua atmosférica pueda precipitar queda determinada por la intensidad de condensación (paso del agua del estado gaseoso al líquido), que depende principalmente de la velocidad vertical de los movimientos de las masas de aire húmedo.

(iii) **Convergencia de humedad**: Dado que el contenido de agua que puede contener la atmósfera rara vez excede un par de centímetros (condiciones de saturación), para que una precipitación se mantenga en el tiempo, es necesaria la reposición continua de aire húmedo proveniente otra fuente de humedad. Esto explica que la magnitud de una tormenta esté condicionada por la velocidad de convergencia de aire húmedo, que a su vez viene dada por la velocidad del viento.

Teniendo en consideración que, lo que indirectamente los distintos métodos hidrometeorológicos buscan, es maximizar las variables anteriormente descritas; dependiendo de qué tanta información estadística o expertiz se tenga, se pueden utilizar diversos métodos para estimar la PMP (Schreiner & Riedel, 1978).

El **método local** (World Meteorological Organization, 1986) calcula la PMP a partir de la máxima tormenta observada en la cuenca en estudio, lo que indudablemente acarrea la desventaja de requerir muchos años de estadística.

El **método de transposición de tormentas,** como su nombre lo indica, transpone a la zona en estudio una tormenta de gran duración, ocurrida en una zona adyacente. El proceso se centra en dos aspectos; uno de ellos es verificar la probabilidad de transposición de dicha tormenta a través de 3 vías propias del método, y el otro es estudiar el alcance de la tormenta y la capacidad de transportarse hacia la zona en la que se desea estimar la PMP (Paulhus & Gilman, 1953).

Por otra parte, el **método combinado** combina dos o más tormentas ocurridas en áreas locales. Basándose en principios de meteorología sinóptica y pronósticos del tiempo, se forman secuencias de tormentas artificiales de larga duración, lo que centra el trabajo en seleccionar combinaciones, estrategias de combinación y demostraciones lógicas de las secuencias de combinación. Este método es muy utilizado en cuencas de gran tamaño, y requiere un alto expertiz en el área de la meteorología (Beauchamp et al., 2013).
(Meteorología sinóptica**: Estudia los fenómenos meteorológicos en tiempo real, basándose en las observaciones realizadas a la misma hora y anotadas sobre mapas geográficos con el objeto de predecir el estado del tiempo futuro)

Otro método hidrometeorológico es el **método inferencial** (World Meteorological Organization, 1986), que, mediante una generalización de la estructura espacial tridimensional de una tormenta en la zona de diseño, crea una ecuación física simplificada de la tormenta, que contiene los factores meteorológicos más influyentes en su gestación. Acorde a la información disponible del viento, el método puede utilizar un modelo de convergencia o un modelo laminar. El modelo de convergencia asume afluencia de humedad hacia el centro de la tormenta (en todas direcciones); mientras que el modelo laminar asume que la afluencia ocurre a lo largo de un plano inclinado con flujo laminar, aumentando en ambos casos la magnitud de la humedad hasta producirse la saturación. Si bien este método requiere robustas estadísticas meteorológicas medidas en altura, es aplicable a cuencas con áreas desde cientos hasta miles de kilómetros cuadrados.

El **método generalizado,** por su parte, involucra agrupar la información de tormentas observadas en lluvias del tipo ciclónicas y orográficas. Este proceso permite obtener las relaciones precipitación – duración – área (PDA) producidas vía transposición de tormentas, junto con su distribución espacial (isoyetas elípticas concéntricas) y temporal (single-peak maps yetograms). Si bien este método es útil para determinar la PMP en zonas extensas y meteorológicamente homogéneas, requiere de una importante cantidad de largos registros con información de tormentas de características similares, medidas en la zona en estudio. Esto le quita aplicabilidad al método, y lo convierte en un método lento y costoso. No obstante, es aplicable a cuencas orográficas de hasta 13.000 km2 y no-orográficas de 72.000 km2, con lluvias de hasta 72 hrs (Casas et al., 2011).

Otro método hidrometeorológico frecuentemente utilizado para estimar la PMP en cuencas de gran tamaño es el **método de simulación de tormentas basado en crecidas históricas** (World Meteorological Organization, 1986)**.** Se les llama crecidas históricas a eventos de períodos de retorno entre 100 y 1000 años, cuya magnitud es mayor a los eventos típicamente observados. El método es capaz de reproducir una tormenta que puede potencialmente haber creado dicha crecida histórica, mediante la generación de un modelo hidrológico de la cuenca. Está basado inherentemente en la información (frecuentemente incompleta) de distribución espacial y temporal de la crecida, considerando además el

hecho de que los modelos meteorológicos sinópicos y de pronóstico de tiempo son aplicables al período histórico en que la crecida histórica se produjo.

Las tormentas que generan eventos históricamente extraordinarios (crecidas históricas) pueden ser consideradas tormentas de alta eficiencia, por lo que, si ellas pueden ser reconstruidas incorporando la maximización de su contenido de humedad, se obtiene la PMP producida por la CMP. Pese a que el método es muy bueno para estimar crecidas en cuencas de gran tamaño, es aplicable solo cuando se dispone de hidrogramas de crecidas y los registros de las tormentas que los generaron, medidos en las cercanías de la sección de diseño.

Los métodos anteriormente descritos pueden ser revisados en detalle en el "Manual of estimation of Probable Maximum Precipitation" de la World Meteorological Organization (2009) . No obstante, la aplicación de ellos exige en principio disponer de una red de estaciones con información meteorológica como la temperatura, humedad y velocidad del viento, en lo posible a nivel horario, requerimiento que en chile limita enormemente la aplicabilidad de ellos.

Otra alternativa para abordar la estimación de la PMP, es mediante la **modelación numérica de ríos atmosféricos** (Toride et al., 2019; Hiraga et al., 2021). Lo que estos modelos buscan es la maximización de humedad a lo largo de la trayectoria de flujos (ríos) atmosféricos, cuyo comportamiento es modelado mediante modelos climáticos numéricos (NWM; Numerical Weather Models). El enfoque basado en los NWM no depende de las hipótesis tradicionales, que suponen una relación lineal entre la precipitación y la cantidad de agua precipitable; y además, no presenta limitaciones en la extensión espacial de la zona de estimación, siempre y cuando los recursos computacionales lo permitan (Hiraga et al., 2021).

Los NWM utilizan frecuentemente el método de maximización de humedad (también llamado maximización de humedad relativa - RHM), que consiste en elevar la humedad relativa hasta el 100% a partir de las condiciones iniciales y las condiciones de borde. Toride et al. (2019) apunta a mejorar esa aproximación introduciendo criterios de integración del vapor de agua transportado (IVT), sosteniendo que saturar la totalidad de la

región al interior de las fronteras pueden generar condiciones atmosféricas poco realistas, que no necesariamente maximizan la precipitación sobre la cuenca, debido a factores como la estructura y estabilidad de la tormenta, y a la topografía.

El procedimiento para estimar la PMP propuesto por Toride et al. (2019) se basa, a grandes rasgos, en 5 pasos. El primero de ellos es determinar, en base a precipitaciones históricas y temperaturas promedio en la cuenca en estudio, los años objetivos de lluvia; para luego, en base a una escala de categorías definidas por valores umbrales de IVT (Ralph et al., 2019), identificar los eventos de río atmosféricos (AR) históricos que generaron las precipitaciones máximas sobre la cuenca en estudio. En seguida se maximiza el agua precipitable acumulada en los eventos AR identificados, mediante la optimización simultánea de la posición y humedad atmosférica sobre el área de interés. Hecho lo anterior, se sustituye cada evento AR histórico por el correspondiente evento maximizado, para formar de este modo, la secuencia de eventos de precipitación maximizada (secuencia de precipitación de diseño) durante la temporada de interés. Finalmente, acumulando la secuencia de diseño sobre el área de drenaje se obtiene la estimación de la máxima probable.

Las estimaciones de PMP vía modelación numérica de ríos atmosféricos anteriormente expuestas (Toride et al., 2019; Hiraga et al., 2021), utilizan para sus simulaciones, condiciones iniciales y de contorno obtenidas del Sistema de Investigación de Predicción del Clima (CFSR; Saha et al. 2010). Este último es, un set de datos generados por modelos de circulación general océano-atmosféricos, a partir de información de radiosondas, aviones, observaciones superficiales y observaciones satelitales.

Si bien se piensa que este último enfoque se aproxima de mejor manera al problema en estudio, este es de difícil aplicación para la realidad chilena, donde la mayoría de las veces se requieren estimaciones rápidas, ante una inminente falta de información para abordar el problema.

3.2.2 Enfoque indirecto: métodos estadísticos

Los métodos estadísticos corresponden a la segunda familia de métodos y enfocan su análisis en el área de la tormenta, más que en una cuenca en particular.

La esencia de estos métodos es la transposición de tormentas, con la salvedad que en vez de transponer una cantidad especifica de precipitación, lo hacen de manera abstracta (estadística), suponiendo condiciones meteorológicas homogéneas.

La primera versión del método estadístico, propuesto por Hershfield (1961) en Estados Unidos, calcula la PMP a partir de una extensa red de estaciones meteorológicas, utilizando conjuntamente análisis de frecuencia y análisis regional generalizado. Si bien el procedimiento difiere del análisis de frecuencia tradicional, es aplicable a cuencas cuya área sea menor a 1.000 km2 (World Meteorological Organization, 1986).

Las dos diferencias principales entre el método original propuesto por Hershfield (1961a) y el análisis de frecuencia tradicional, es que el primero se focaliza en una zona extensa (más que en una determinada estación o cuenca), y que involucra el trazado de una envolvente (más que determinar el extremo estadístico).

Los procedimientos estadísticos para estimar la PMP son muy útiles para estimaciones rápidas, y pueden ser utilizados cuando se dispone de suficiente información de las precipitaciones, o en su defecto, cuando no se dispone de información meteorológica como velocidades del viento, humedad, puntos de rocío o temperaturas, en lo posible a nivel horario (World Meteorological Organization, 1986). La no existencia en Chile de registros con las variables meteorológicas mencionadas obliga a utilizar alguno de los métodos indirectos, calibrados estadísticamente para la realidad local.

Método de Hershfield y su evolución en el tiempo

La metodología propuesta por Hershfield en 1961 se basa en el concepto coeficiente de frecuencia "K" (Chow, 1951), que adimensionaliza la precipitación para un período de retorno "t" dado (X_t), en función del promedio $\overline{X_n}$ y desviación S_n de una serie de "n" máximos anuales.

$$K = \frac{X_t - \overline{X_n}}{S_n}$$

Ecuación 1: definición de coeficiente de frecuencia

Así, al sustituir el máximo valor registrado en cada estación (X_m) por X_t, el máximo coeficiente de frecuencia K_m corresponde al número de desviaciones estándar que se debería agregar al promedio (de la serie de máximos anuales), para obtener la máxima histórica (X_m).

Hershfield postula que el coeficiente de frecuencia máximo K_m es la variable estadística que se comporta siguiendo una distribución de valor extremo, siendo el máximo de ellos ($K_m^{máx}$) la magnitud que la variable alcanza con 0% de probabilidad de excedencia. Este último valor, permite determinar la PMP de una estación.

$$X_m = \overline{X_n} + K_m S_n$$
$$PMP = \overline{X_n} + K_m^{máx} \; S_n$$

Ecuación 2: Expresión para calcular la PMP según Método de Hershfield

El primer estudio (Hershfield,1961) se basa en registros de precipitaciones acumuladas en 24 horas en 2700 estaciones meteorológicas ubicadas en Estados Unidos, entregando un valor límite del coeficiente de frecuencia ($K_m^{máx}$) igual a 15.

Hasta ese entonces Hershfield pensaba que dicho valor era independiente de la magnitud de las tormentas, evidenciándose luego (Hershfield 1961, b) que el valor planteado sobreestimaba y subestimaba la PMP, en zonas lluviosas y zonas áridas, respectivamente.

Ajuste del promedio y desviación por evento máximo observado

Las tormentas extremas son eventos que se pueden catalogar como “de extraña o rara ocurrencia” puesto que poseen períodos de retorno de al menos 500 años. En las series registradas estas magnitudes corresponden a outliers, y tienen un efecto apreciable en el promedio y desviación de series de máximos anuales (World Meteorological Organization, 1986).

Basándose en las series con longitud de registro variable, Hershfield (1961b) se dedicó a estudiar la magnitud de dicho efecto, evidenciando que esta es mayor en registros cortos que en registros de mayor duración. En base a lo anterior, presentó factores para ajustar el promedio y desviación estándar de cada serie, en función de la magnitud del máximo evento observado (World Meteorological Organization, 1986).

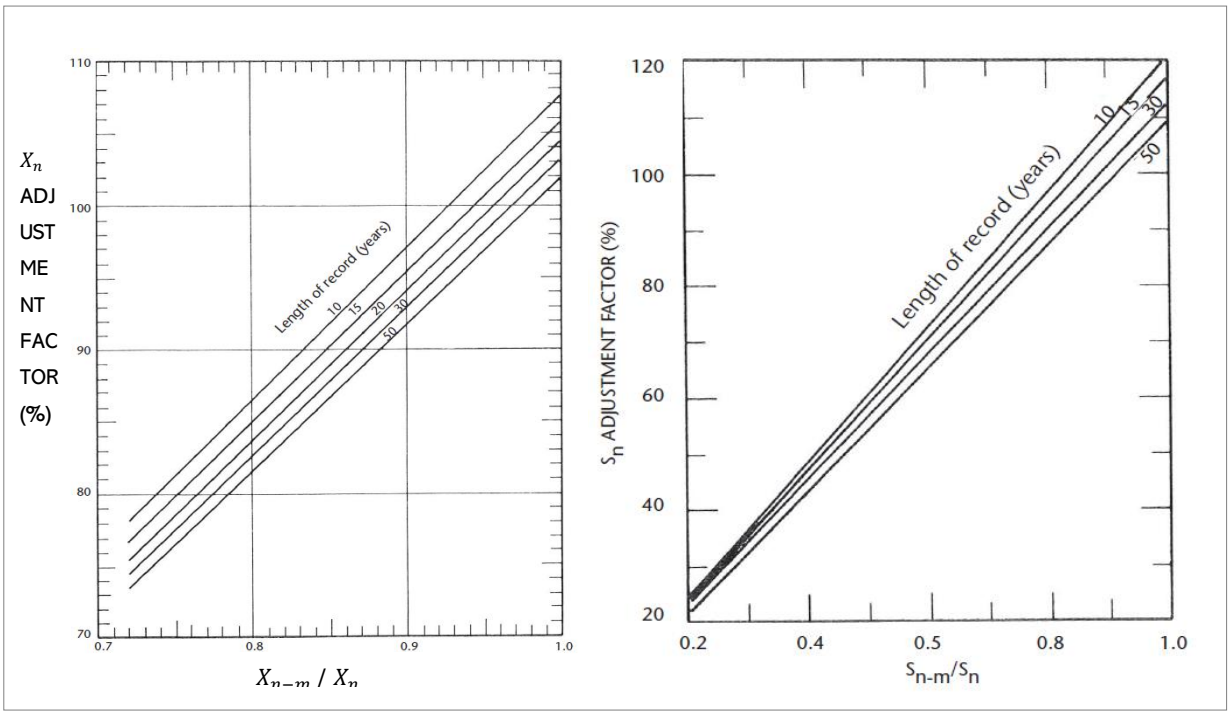


Figura 1: Factores de ajuste para el promedio y desviación estándar de la serie de máximos anuales, por efectos del máximo evento observado (World Meteorological Organization, 1986)

Por otra parte, resulta razonable pensar que los valores del promedio y desviación estándar tienden a aumentar a medida que aumenta la longitud del registro (tamaño de la muestra). Lo anterior, debido a que las distribuciones de frecuencia de eventos extremos están sesgadas hacia la derecha, lo que aumenta la posibilidad de encontrar un máximo a medida que aumenta la longitud del registro (World Meteorological Organization, 1986).

Para dar respuesta a ello (Hershfield ,1961b) se incorpora un segundo factor de corrección al promedio y desviación estándar, por efectos de longitud de registro, si es que este es menor a 50 años.

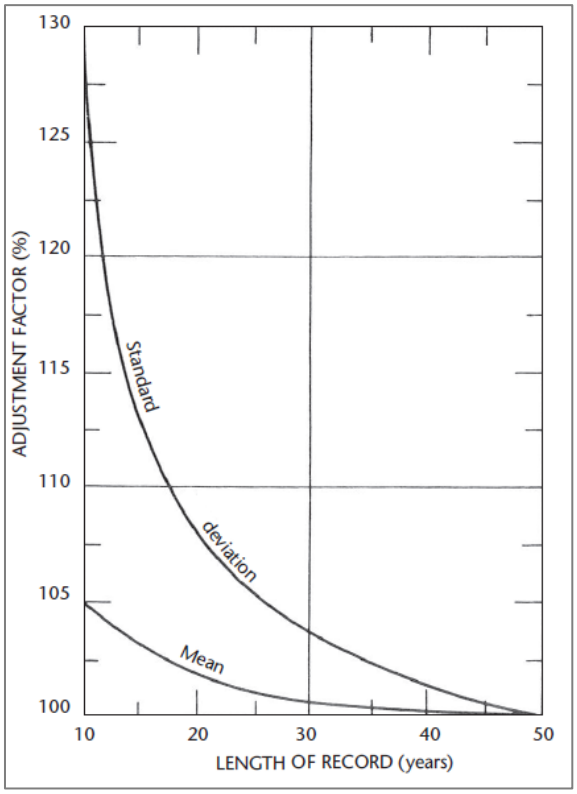


Figura 2: Factores de ajuste del promedio y desviación estándar de la serie de máximos anuales, por efectos de longitud de registro (World Meteorological Organization, 1986)

Para estimar de manera rápida la PMP en una estación determinada, el promedio y desviación estándar corregido(a) ($\overline{X_c}$, S_c) de la serie de precipitaciones máximas anuales por efecto de la longitud del registro (N) y la presencia del valor máximo observado pueden calcularse según las expresiones presentadas en la ecuación 3,

$$\overline{X_c} = \overline{X_N}\,(1 + 0.143\,e^{-0.105\,N})\left((1.05 - 0.0008N)\frac{\overline{X_{N-1}}}{\overline{X_N}} + 3.9 * 10^{-5}(N-37)^2 + 0.002\right)$$

$$S_c = S_N\left(0.993 + 0.307\,e^{-0.258\,(N-10)^{0.699}}\right)\left((1.09 + 0.223\,e^{-0.07\,N})\frac{S_{N-1}}{S_N} + 0.008\right)$$

Ecuación 3: Correcciones para promedio ($\overline{X_N}$) y desviación estándar (S_N) por efectos de longitud de estadística y presencia de eventos extremos (Stowhas, 2016), basadas en ábacos presentados por Hershfield, 1961b (World Meteorological Organization, 1986).

dónde $\overline{X_{N-1}}$ y S_{N-1} representan el promedio y desviación estándar sin considerar el máximo valor de la serie.

Si bien para efectos del análisis estadístico, los máximos anuales son obtenidos a partir de los valores máximos registrados en la ventana temporal de interés (5 min, 1- 6- 24 horas) definida por horarios (límites de los intervalos) de medición fijos, resulta evidente pensar que el máximo precipitado bajo una determinada duración no coincida con el valor registrado en un intervalo fijo de igual duración; así como también que dicha diferencia sea sensible al variar número de observaciones realizadas en el intervalo de medición (World Meteorological Organization, 1986).

Es por ello que estudios de registros pluviométricos de miles de años de estadística (Hershfield, 1961; Weiss, 1964; Miller, 1964) indican que para tormentas de entre 1 y 24 horas, cuyo valor se obtuvo de una única medición en un intervalo fijo, la PMP puede ajustarse por un valor de 1.13; mientras que valores máximos registrados de manera horaria en intervalos de 6 y 24 horas, requieran incrementar la PMP en sólo 1.02 y 1.01, respectivamente.

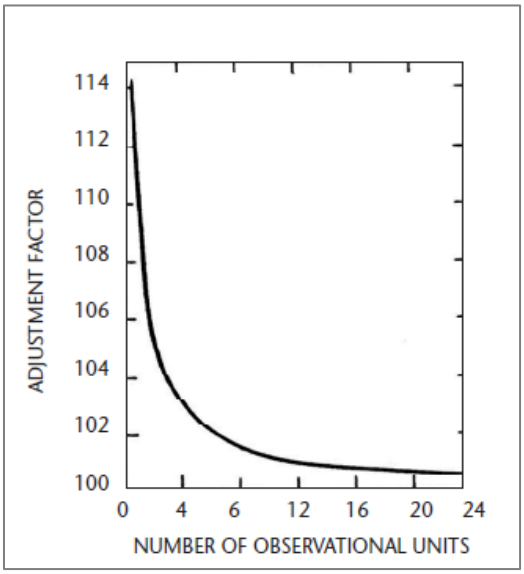


Figura 3: Factor de ajuste de la PMP por número de observaciones por intervalo de medición (Weiss,1964)

Años más tarde, Hershfield (D. M. Hershfield, 1965) actualiza el procedimiento, proponiendo una nueva alternativa para determinar el coeficiente de frecuencia límite.

En dicho estudio, Hershfield (1965) grafica geográficamente los coeficientes de frecuencia máximos de cada estación, calculados siguiendo el procedimiento tradicional. Pese a que

se aprecia que los promedios y desviaciones estándar en estaciones vecinas son en general bastante similares, las diferencias de magnitud en las precipitaciones máximas producen coeficientes de frecuencia máximos que no presentan patrones sistemáticos, postulándose que ellos ocurren de manera aleatoria.

En busca de una relación de dependencia estadística entre los estadísticos, Hershfield grafica los coeficientes máximos de frecuencia (Km) respecto de su promedio (Xn-1), sugiriendo que existe una envolvente superior para el estadístico Km, que decrece al aumentar la magnitud de lluvia. En el proceso, Hershfield postula que para un promedio determinado, el crecimiento de la envolvente debe parar en un valor definido, de manera tal que no exceda la relación envolvente de precipitación mundial (Jennings, 1950).

Las envolventes sugeridas por Hershfield en su última versión son ampliamente utilizadas en Estados Unidos hasta el día de hoy.

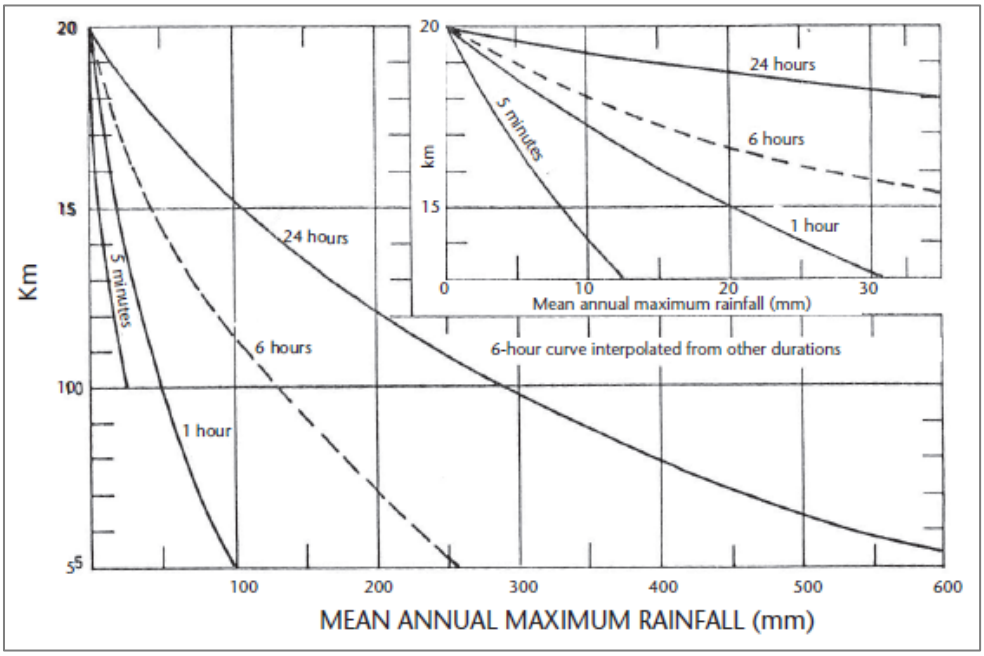


Figura 4: $K_m^{máx}$ para distintas duraciones de tormenta y promedios de precipitación máxima anual (World Meteorological Organization, 1986; D. M. Hershfield, 1965)

Distribución temporal y espacial de la PMP

El procedimiento propuesto por Hershfield fue desarrollado para obtener magnitudes puntuales de la PMP, asociadas a duraciones de tormentas que vienen dadas por la ventana temporal de los registros estadísticos que definieron las magnitudes límites del máximo coeficiente de frecuencia. En la práctica, interesa no solo conocer la magnitud de la PMP a nivel puntual, sino también la magnitud promedio de esta sobre una determinada área (World Meteorological Organization, 1986).

Para dar respuesta a ello, existen 2 tipos de relaciones precipitación-área. La primera de ellas asume que la PMP ocurre cuando la tormenta se centra en el área afectada; mientras que la segunda, asume que el centro de la tormenta está desplazado, y, por ende, afecta total o parcialmente el área de interés.

Múltiples estudios señalan que existen diversas variantes a estas dos relaciones (Court, 1961; United States Weather Bureu,1960) y pese a que dependen fuertemente de las condiciones locales de las tormentas, se pueden utilizar las relaciones PDA (Precipitación – Duración - Área) promedio, obtenidas de un análisis general de múltiples tormentas ocurridas en el Oeste de Estados Unidos.

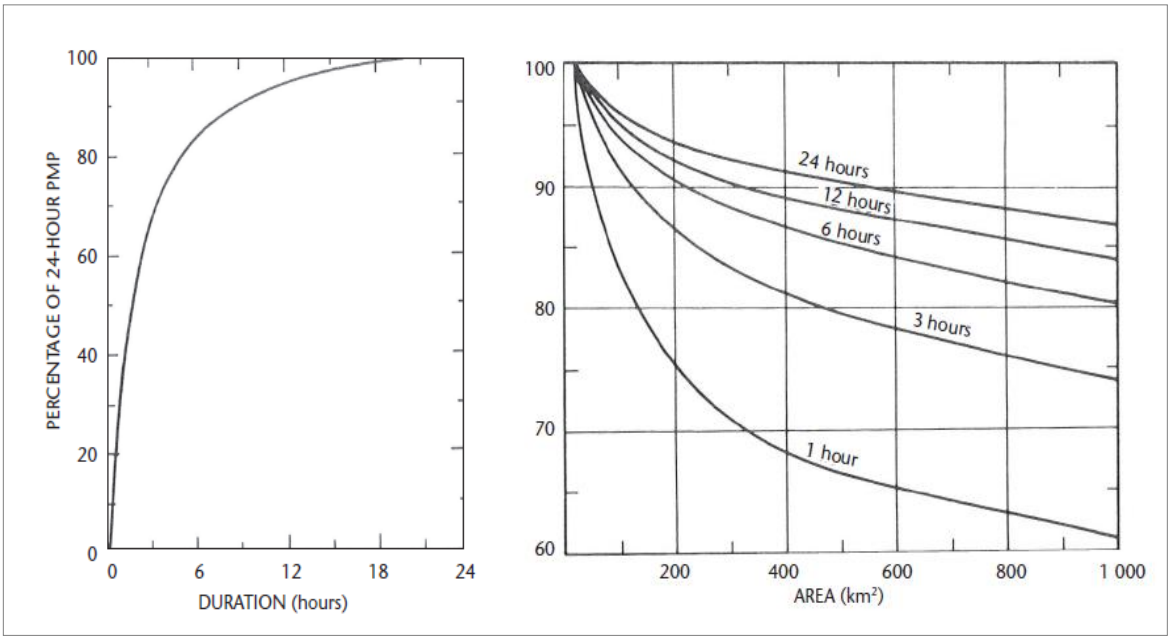


Figura 5: Factor de reducción de la PMP por efectos de duración de la tormenta (Huff, 1967) y área de la cuenca para EEUU occidental (United States Weather Bureau, 1960)

Método alternativo

Posteriormente se realiza en Chile (Stowhas, 1983), un estudio de recopilación, procesamiento y análisis de la información de precipitaciones máximas diarias en términos estadísticos y probabilísticos, con el propósito de reunir antecedentes respecto a los máximos factores de frecuencia y cuestionar así la aplicabilidad del método de Hershfield, además de proponer una metodología de cálculo modificada a partir de las condiciones locales.

En el estudio en cuestión, titulado "Precipitaciones máximas diarias en chile", se sometió a un análisis de frecuencia la serie de 190 valores correspondientes a los coeficientes máximos de frecuencia de cada estación en estudio, postulando distribuciones de valores extremos tipo I (Gumbel) y tipo III (Weibull) acotada superiormente. El análisis se lleva a cabo incluyendo y excluyendo el valor máximo medido en el cálculo del promedio y la desviación estándar de cada serie, lográndose mejores ajustes para la serie que si incluye el valor máximo medido. Un test x^2 acepta ambas distribuciones extremas, resultando ajustarse mejor la distribución acotada de Weibull. De acuerdo con estos resultados, el valor límite máximo de los coeficientes de frecuencia, incluyendo en su cálculo el mayor valor registrado, sería de k=11.02. Así, si se acepta la distribución abierta de Weibull, el valor k=11.02 tendría una probabilidad de excedencia de 1 en 2.600.000, lo que quiere decir que un valor del coeficiente de frecuencia k=11 aparece como una estimación bastante conservadora de la PMP en una determinada estación (Stowhas, 1983).

En dicha oportunidad, Stowhas L. (1983) indica que un aspecto que merece análisis es el ajuste de la curva envolvente trazada por Hershfield, la cual sostiene la existencia de un coeficiente de variación constante o al menos de aleatoriedad independiente, existiendo evidencias tanto de variaciones regionales del coeficiente de variación de los registros pluviométricos como correlaciones positivas entre las precipitaciones máximas registradas y las precipitaciones máximas medias para determinadas duraciones de tormentas (Hershfield, 1962), que podrían ser utilizadas para el trazado más confiable de envolventes superiores para la estimación de la PMP.

Con el propósito de establecer dependencias entre el coeficiente de frecuencia máximo y los parámetros de las estadísticas que permitieran establecer los criterios de maximización y el trazado de envolventes, se intentó correlacionar los valores de los coeficientes de frecuencia máximos con los promedios, desviaciones estándar y coeficientes de variación de las estadísticas. De ello se evidencia la inexistencia de correlación entre el coeficiente de frecuencia máximo y los estadígrafos de las muestras, aún cuando se observa una aparente disminución de los valores máximos alcanzados por el coeficiente de frecuencia a medida que aumenta el valor de los estadígrafos (Stowhas, 1983).

Con el objetivo de dilucidar la posible influencia de los estadígrafos de las muestras en la tendencia de los coeficientes de frecuencia, se somete a la información a un análisis de varianza. Para normalizar la distribución de los coeficientes de frecuencia, se trabaja con los logaritmos de ellos, clasificándose previamente los estadígrafos en 10 intervalos de clase según su magnitud.

El análisis realizado deja en evidencia que los promedios de cada intervalo de clase definido no serían significativamente diferentes entre sí, no detectando ninguna relación entre estos y los coeficientes de frecuencia. De igual forma, pruebas F de varianza indican que las desviaciones estándar de los coeficientes de frecuencia no serían significativamente diferentes al clasificar las muestras en función de los promedios y desviaciones estándar de las estadísticas base (Stowhas, 1983). No obstante, al clasificar las muestras en función de sus coeficientes de variación, se observa que las varianzas de los coeficientes de frecuencia bajos serían mayores que las varianzas de los coeficientes de frecuencia provenientes de estaciones con coeficiente de variación mayor, lo que permite a Stowhas (1983) concluir que sólo en el caso de los coeficientes de variación podría haber argumentos de índole estadístico para trazar envolventes decrecientes a los máximos coeficientes de frecuencia.

Incorporando la definición del coeficiente de variación (Cv), se puede reescribir el coeficiente de frecuencia máximo como:

$$C_V = \frac{S_n}{\bar{P}}$$

Ecuación 4: Coeficiente de variación

$$K_{max} = \frac{P_{max} - \bar{P}}{\bar{P}} \frac{1}{C_V}$$

Ecuación 5: Máximo coeficiente de frecuencia en términos del coeficiente de variación

La presencia del coeficiente de variación en el denominador de la ecuación 5 podría explicar la posible tendencia decreciente de las envolventes superiores de los coeficientes máximos de frecuencia en función del Cv, ya que grandes valores de Km podrían estar asociados a pequeños valores de Cv, más que a grandes precipitaciones.

Con la finalidad de cuantificar y comparar la magnitud de los eventos extremos en función del coeficiente de variación y precipitación máxima media, Stowhas (1983) introduce la definición de la precipitación máxima centrada y adimensionalizada,

$$P_m = \frac{P_{max} - \bar{P}}{\bar{P}}$$

Ecuación 6: Precipitación máxima anual adimensionalizada y centrada (Stowhas, 1983)

constatando existencia de dependencia entre ella y los estadígrafos precipitación máxima media y coeficiente de variación.

Teniendo en cuenta las relaciones límites de dependencia entre el coeficiente de variación y la media de precipitaciones máximas anuales, y entre la precipitación máxima adimensionalizada y el coeficiente de variación, se trazan 2 curvas envolventes (Pm-Cv y Cv-$\bar{P}$), que involucran de manera cruzada las 3 variables en análisis ($Cv, Pm, \bar{P}$).

De esta forma, si se pretende que los resultados sean consistentes entre sí, se traza una doble envolvente para la precipitación máxima adimensionalizada, en función de la precipitación máxima anual media. Lo anterior asegura un valor envolvente en función del coeficiente de variación, ya que han respetado las relaciones más extremas entre el Cv y $\bar{P}$ que muestra el total de las estaciones chilenas analizadas (Stowhas, 1983).

Es interesante notar que en el estudio realizado por Stowhas (1983) las estaciones que definen las envolventes de las relaciones coeficiente de variación – promedio de precipitaciones máximas anuales, corresponden a estaciones cordilleranas o en altura, que

podrían poseer componentes nivales que pudieran hacerlas no representativas de las condiciones a menor altitud.

En efecto, las curvas envolventes a las precipitaciones máximas adimensionales, en función de la precipitación media, se pueden representar por la expresión 7,

$$P_m(\bar{P}) = \frac{P_{max} - \bar{P}}{\bar{P}} = \Phi(\bar{P})$$

Ecuación 7: Envolvente precipitación máxima adimensionalizada – precipitación media

resultando en consecuencia, las relaciones envolventes a los coeficientes de frecuencia máximos, como una familia de curvas que tienen como parámetro el coeficiente de variación.

$$K_m^{máx} = \frac{\Phi\,(\bar{P})}{C_V}$$

Ecuación 8: Envolventes de los coeficientes de frecuencia máximos

Las magnitudes de los coeficientes de frecuencia máximos obtenidos por Stowhas,1983 para los distintos coeficientes de variación y precipitación media (obtenidas a partir del trazado de la doble envolvente y la ecuación 8) son contrastados (figura 6) con los propuestos por Hershfield,1965 para distintas precipitaciones medias (figura 1).

De despejar en la ecuación 7 el máximo valor registrado (P_{max}), se obtiene una expresión alternativa para calcular la PMP a partir la relación envolvente entre los máximos eventos registrados y las precipitaciones máximas medias (Stowhas, 1983).

$$P_{max} = \bar{P}(1 + \Phi(\bar{P}))$$

Ecuación 9: Ecuación alternativa para calcular la PMP en Chile, según método de Stowhas

A modo de verificación y validación del método, la figura 6 muestra el trazado de esta curva, incorporando los valores máximos muestrales de las estadísticas chilenas y los valores más extremos que resultan de estadísticas internacionales (Hershfield,1962)

Si se evalúa en la última expresión las dos relaciones envolventes $\Phi(\bar{P})$ resultantes de considerar y no considerar la presencia de estaciones cordilleranas, se pueden obtener las

expresiones resultantes del procedimiento propuesto por Stowhas,1983, pudiéndose estimar la PMP en estaciones cordilleranas y no cordilleranas, respectivamente.

$$PMP_1 = \overline{P_c}\ (\ 4 + 3.8\ e^{-0.0069\ \overline{P_c}^{1.141}})$$

$$PMP_2 = \overline{P_c}\ (\ 3.5 + 3.65\ e^{-0.0076\ \overline{P_c}^{1.102}})$$

Ecuación 10: Alternativa adicional para estimación de la PMP en estaciones cordilleranas (PMP_1) y no cordilleranas (PMP_2) (Stowhas, 2016).

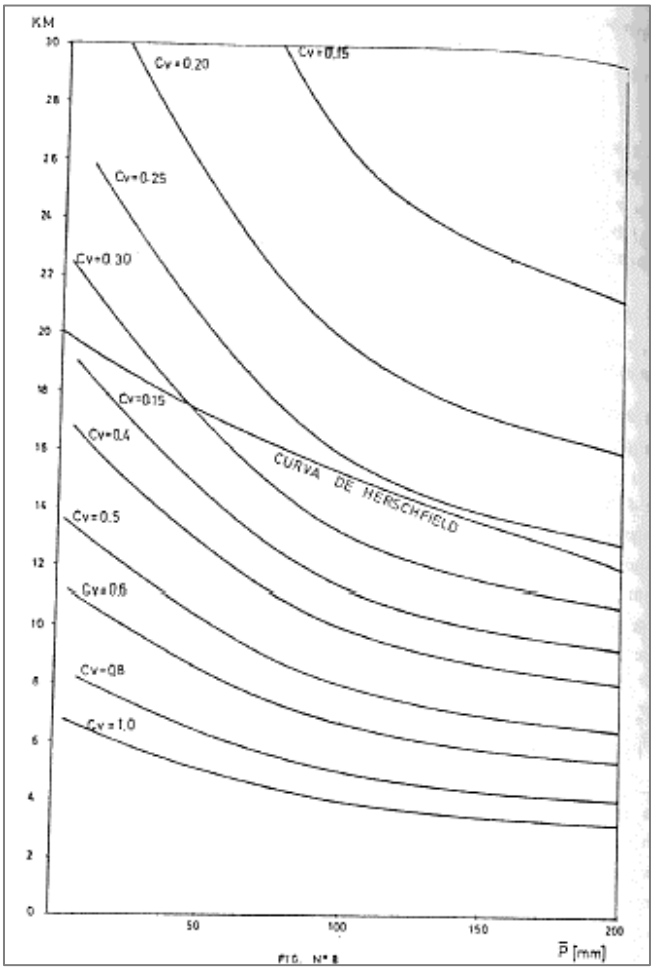


Figura 6: Coeficientes máximos de frecuencias en función del promedio de precipitaciones máximas anuales y coeficientes de variación (Stowhas, 1983).

4 Metodología

Para dar cumplimiento a los objetivos, se recopilan y analizan estadística y probabilísticamente las precipitaciones máximas diarias en Chile, a fin de reunir antecedentes respecto a los máximos coeficientes de frecuencia y aplicabilidad de los métodos estadísticos.

4.1 Compilación de las bases de datos

Fuente de los datos

Para el desarrollo del estudio, se trabaja con los datos de precipitación acumulada diaria registrados en 879 estaciones meteorológicas a lo largo Chile, durante el período Enero de 1900 a Junio de 2020. Estas estadísticas fueron recolectadas de sistemas informáticos de diversas fuentes y compilados por el equipo de Datos y Cómputos del "(CR)2: Center of Climate and Resilience Research", para fines de investigación y docencia. Estos datos no sustituyen los datos originales provistos por las instituciones responsables de los mismos.

En este caso, de las 879 estaciones meteorológicas estudiadas, 845 pertenecen a la Dirección General de Aguas, mientras que las 34 las restantes pertenecen a Dirección Meteorológica de Chile. Los datos fueron descargados y generados en servidores CR2 desde las páginas web de las instituciones mencionadas, utilizando scripts Python y base de datos mySql.

Descarga de archivos del servidor y filtrado de BBDD para aplicar Hershfield-Stowhas

La base de datos para el estudio en cuestión fue descargada directamente desde el servidor de CR2, y posteriormente depurada y ordenada mediante algoritmos de manejo de archivos de texto implementadas en Python Pandas.

A grandes rasgos, lo que la rutina desarrollada hace es, entre otras cosas, estructurar una matriz que contenga las precipitaciones acumuladas diarias en milímetros, de tantas columnas como estaciones existan, y tantas filas como días hayan transcurridos desde la primera observación.

(BBDD**: Base de datos)

Resulta bastante útil trabajar con el complemento “Pandas” de “Python”, porque este permite el manejo de la información a través de cuadros de datos (data-frames), los cuales permiten almacenar las estadísticas como números dentro de una matriz, y a su vez, asignar índices de columna (códigos de la estación meteorológica) y fila (fecha de la observación) a cada dato.

La información no relevante para la recalibración del máximo coeficiente de frecuencia registrado en Chile (Stowhas, 1983), pero si relevante para el posterior análisis regional, es almacenada en una nueva base de datos. En esta se resumen características como el nombre, código, ubicación, altitud, latitud, longitud, fecha de inicio y fin de las observaciones, y cuenca y subcuenca a la que pertenece.

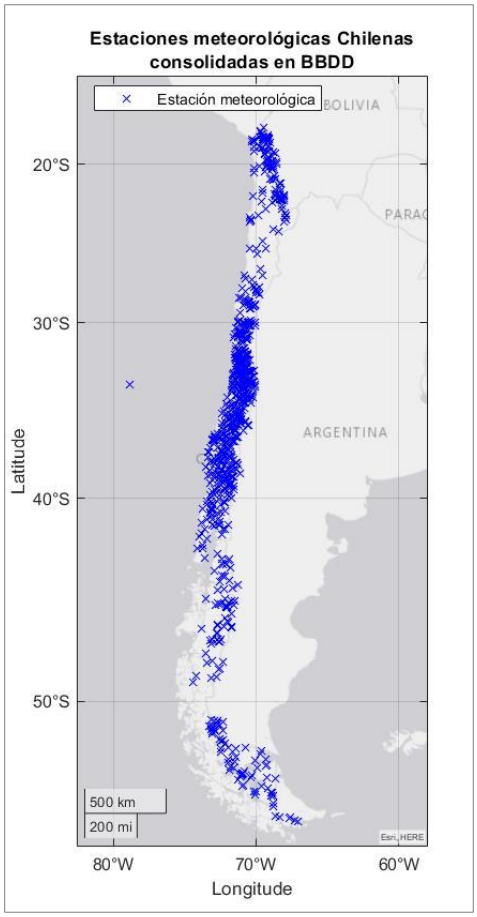


Figura 7: Distribución espacial de las estaciones meteorológicas compiladas en la base de datos

En el anexo C se presenta un resumen de las estaciones complidas en la base de datos, junto con los estadígrafos más importantes de cada una de ellas.

Dado que la PMP se calcula ya sea física o estadísticamente en base a los eventos máximos registrados, con la finalidad de validar los extremos estadísticos (outliers), se ordenan de manera decreciente las máximas precipitaciones registradas en 24 horas, analizando detalladamente la distribución temporal (a nivel diario) de los 20 eventos de mayor magnitud.
Esto permite por un lado eliminar valores erróneos, y por otro, sosteniéndose en información de medios de prensa e informes históricos de catástrofes naturales, validar la ocurrencia de los extremos superiores de la muestra.

4.2 Análisis de frecuencia y teoría de valores extremos

Los datos hidrológicos pueden presentarse en forma continua o discreta, pero para su análisis estadístico deben ser discretizados. Es común que variables esencialmente continuas sean discretizadas tomando valores promedios o acumulados en un intervalo de tiempo, así como también considerando los valores extremos (máximos o mínimos) o aquellos que excedan un cierto umbral, en una ventana temporal periódica.

Las series de duración completa incluyen toda la información disponible respecto a una variable hidrológica, y si bien tienen importantes aplicaciones ingenieriles, no resulta apropiado someterlas a un análisis de frecuencia, porque poseen una fuerte dependencia temporal y porque incluyen mucha información irrelevante, especialmente en estudios en que solo interesan los valores críticos o extremos que puede tomar la variable en cuestión.

Para el desarrollo de análisis de frecuencia, se seleccionan subconjuntos de las series de duración completa, que pueden ser series de duración parcial o series de valores extremos.
Las series de duración parcial son construidas considerando todos los valores de la variable que excedan (o complementariamente, no logran exceder) la magnitud de un cierto valor umbral base; mientras que las series de valor extremo se construyen a partir de los valores

máximos o mínimos que toma la variable en un período de tiempo previamente establecido (Stowhas, 2016).

En estos términos, es conceptualmente lícito aplicar la teoría de valores extremos a la serie de datos conformada por los máximos coeficientes de frecuencia, ya que esta proporciona distribuciones de frecuencia límites aplicables a este tipo de variables.

La distribución generalizada del valor extremo (GEV) posee función de distribución acumulada:

$$F(x) = e^{-(1+\frac{\xi(x-\mu)}{\sigma})^{-\frac{1}{\xi}}}$$

Ecuación 11: Función de distribución de probabilidad - distribución generalizada de valores extremos

Dónde x es la variable aleatoria (coeficiente de frecuencia) y ξ, σ y μ parámetros de forma, escala y posición, respectivamente.

De la distribución generalizada de valores extremos se desprenden 3 casos, que pueden ser GEV Tipo I (Distribución Gumbel), GEV Tipo II (Distribución Fréchet) o GEV Tipo III (Distribución Weibull), cuando el parámetro de forma “k” es igual, mayor o menor que cero, respectivamente.

4.3 Análisis de cluster

El análisis de cluster (o de conglomerados) corresponde a una técnica de aprendizaje no supervisada, utilizada para encontrar agrupaciones y patrones naturales en un conjunto finito de datos. Corresponde a una técnica estadística multivariante, que busca agrupar los objetos o individuos tratando de lograr la máxima homogeneidad en cada grupo, y la mayor diferencia entre grupos (Amat, 2017).

Puesto que la utilización del AC implica un desconocimiento de la clasificación de los datos, es necesario emplear varias metodologías de análisis, con el fin de contrastar los resultados.

Tipos de análisis, métodos y medidas de asociación

A grandes rasgos existen dos tipos de análisis de cluster. El primero, conocido como análisis jerárquico, configura grupos con estructura arbolascendente, de forma que los clusters de niveles más bajos son englobados por clusters de niveles superiores.

Por otra parte, se conocen como análisis no jerárquicos a aquellos que asignan los casos o grupos diferenciados, sin que unos dependan de otros; pudiendo configurar a su vez, grupos disjuntos o solapados.

Todos los métodos de clustering tienen una cosa en común, y es que para poder llevar a cabo las agrupaciones necesitan definir y cuantificar la similitud entre las observaciones. La característica que hace el clustering un método adaptable a escenarios muy diversos es que puede emplear cualquier tipo de distancia, lo que permite escoger la más adecuada en cuestión. En conclusión, se deben tomar tres decisiones antes de iniciar el análisis, y son; (i) seleccionar las variables relevantes para identificar (caracterizar) los grupos, (ii) elegir la medida de proximidad entre los individuos y (iii) seleccionar el criterio para agrupar los individuos en conglomerados (Amat, 2017).

Validación de los clusters

Los métodos de clustering tienen la característica de ser capaces de encontrar agrupaciones en cualquier set de datos, independientemente de que realmente existan o no dichos grupos en la población de la que proceden las observaciones. Es por ello que en el proceso de validación se evalúa la veracidad de los grupos obtenidos mediante tres etapas, que son; (i) el estudio de la tendencia de clustering, (ii) la elección del número óptimo del clustering y (iii) el estudio de calidad/significancia de los clusters generados.

El estudio de tendencia consiste en evaluar si hay indicios de que realmente existe algún tipo de agrupación en los datos. Esto puede llevarse a cabo de manera visual ("Visual Assessment of Cluster Tendency"), o mediante test estadísticos como el "Hopkins statistic test".

Determinar el número óptimo de grupos es uno de los pasos más complejos, sobre todo en los métodos no jerárquicos, en los que se debe especificar de antemano el número de grupos antes de poder ver los resultados.

Dado que no existe una única forma de averiguar el número óptimo de grupos, y a que el proceso es bastante subjetivo (depende en gran medida del tipo de clustering empleado), se han desarrollado estrategias como lo son el "Elbow method", el "average silhouette method", el "GAP statistic", que ayudan en el proceso de selección de este input.
En general, es común que cada estrategia entregue cantidades óptimas distintas, por lo que se sugiere utilizar los tres y decidir en conjunto.

Una vez seleccionado el numero adecuado de clústers y aplicado el algoritmo de clustering pertinente, se debe evaluar la calidad de estos. Esta evaluación puede ser de manera interna (empleando únicamente información interna del proceso para evaluar la bondad de las agrupaciones) o externa (combinan los resultados del clustering con información externa. Esta última puede ser empleada únicamente si se conoce el verdadero grupo al que pertenece cada observación.

Para la validación interna, se verifica la homogeneidad (también llamada cohesión o compactness) y la distancia entre clústers (similitud intra-clusters). Los índices más frecuentemente empleados para la validación interna de clusters combinan tanto medidas de homogeneidad como de separación, puesto que estas se relacionan inversamente entre sí (a mayor número de clústers la homogeneidad aumenta, pero la separación disminuye). Los más utilizados son el "Silhouette width" y el "Dunn index" (Razali et al., 2018), los cuales se detallan comúnmente en la literatura.

Estandarización de los datos

El AC es un método multivariado sensible al uso de datos no estandarizados, más aún en el caso en estudio, que la medida de similitud es la distancia euclidiana al cuadrado. Tanto (Peña, 2002) como (Castellarin et al., 2001) expresan que la falta de estandarización genera inconsistencia entre las soluciones clúster, dado que la medida de distancia entre objetos se ve afectada por el rango numérico de variación entre variables.

Si cada una de las variables se mide en escalas diferentes, el orden de las similitudes puede variar con sólo un cambio de escala en una de ellas (Castro Heredia et al., 2012). Existen diversos métodos de estandarización, que pueden ser vistos en (Rao & Srinivas, 2008) y (Lin & Chen, 2006).

Para efectos del análisis realizado, se convierte cada puntuación (o parámetro) x_{ij} de los datos originales en un valor estandarizado $\widehat{x_{ij}}$, con media cero y desviación estándar uno. Esto se consigue, restando la media x_j y dividiendo por la desviación estándar σ_j del parámetro en cuestión:

$$\widehat{x_{ij}} = \frac{x_{ij} - \overline{x_j}}{\sigma_j}$$

Ecuación 12 : Estandarización de variables

4.4 Aplicación del método de Hershfield,1961 a la realidad chilena

Con el fin de conocer la variación del límite superior del máximo coeficiente de frecuencia de la precipitación máxima diaria anual en Chile, se implementa una rutina computacional, que realiza el siguiente procedimiento:

Procedimiento general

1.- Para efectos del estudio, se asume que la precipitación máxima diaria anual corresponde al máximo valor registrado en las precipitaciones acumuladas en 24 horas, durante un año. Para ello, el algoritmo realiza en cada estación un remuestreo a nivel anual, generando una nueva base de datos que contiene dichos máximos (precipitación máxima diaria anual).

2.- Se calcula el promedio, la desviación estándar y el coeficiente de variación a la series de precipitación máxima diaria anual, en cada estación.

3.- Se obtiene la precipitación máxima histórica o máxima “precipitación máxima diaria anual” en cada estación, y con dicho valor se calcula el coeficiente de frecuencia máximo de la estación.

4.- A la serie de valores de los coeficientes de frecuencia máximos de las estaciones en estudio (solo a aquellas que cumplen con los requisitos estipulados en el análisis), se les ajusta un modelo de distribución de probabilidad de valores extremos.

5.- Se realizan pruebas bondad de ajuste y se valida la representatividad del modelo de distribución de probabilidad ajustado a la serie de valores.
6.- Dado que se conoce la distribución de probabilidad de la variable estadística "coeficiente máximo de frecuencia", se obtiene el valor que esta alcanza bajo una probabilidad de excedencia del 100%.

Los puntos 4, 5 y 6 se llevan a cabo con el software matemático Matlab.

Escenarios de cálculo

Con la finalidad de reproducir los parámetros determinantes de la PMP actualmente usados en consultoría, se repite, bajo condiciones de análisis similares, el procedimiento para lo obtención de ellos (Stowhas, 2016). Esto se logra aplicando un filtro a las estaciones meteorológicas disponibles, y no considerando en el análisis aquellas cuya longitud del registro sea menor a 15 años, aquellas que registren solo mediciones nulas, y aquellas que se ubiquen en el Norte Grande (se considera el Río Copiapó como límite sur).

La exclusión de las estaciones del Norte Grande puede sustentarse en el hecho de que en sus estadísticas pluviométricas predominen valores nulos (Stowhas B, 1983), y en que las características climáticas y los mecanismos de gestación de las tormentas son considerablemente distintas al resto del territorio.

El análisis mencionado se realiza primero como línea base, considerando solamente las estadísticas disponibles hasta el 1983 (Escenario A); y posteriormente, bajo el mismo procedimiento y bajo las mismas condiciones, considerando las estadísticas disponibles al año 2020 (Escenario B).

Con la intención de justificar eventuales diferencias entre los resultados de los Escenarios A y B, se decide incorporar otro escenario de cálculo (Escenario A2), que considera solamente las 146 estaciones existentes a la fecha de publicación del estudio (1983), pero su registro estadístico hasta el 2020 (no incorpora estaciones nuevas).

Tabla 1:Cantidad de estaciones utilizadas en escenarios A, A2 y B

Escenario	A – A2	B
Estaciones Totales	879	879
1) # Est. Norte Grande	160	160
2) # Est. Mediciones nulas	464	6
3) # Est. Registro <15 años	703	277
(1) ∩ (2) ∩ (3)	733	389
# Est. Utilizadas	146	490

Tabla 2: Características de la muestra según (Stowhas, 1983) y escenarios A, A2 y B

	Características de la muestra			
	Stow.L, 1983 - SOCHID	Escenario A	Escenario A2	Escenario B
Cant de Estaciones	190	146	146	490
Total de años	6504	4145	8752	19424
mínimo de años	15	15	15	15
máximo de años	67	66	102	102
media de años	34.3	28.4	59.9	39.6
σ de años	13.3	10.5	16.5	18.1
Máx P_{max}^{24} [mm]	521.7	455	455	487

4.5 Implementación computacional del método de Stowhas, 1983

El método de Stowhas se implementa computacionalmente para un conjunto genérico de estadígrafos (mismo algoritmo para los distintos escenarios o clusters), siendo evidente que las envolventes resultantes no serán las mismas en los distintos escenarios o grupos.

Para ello se realiza el siguiente procedimiento:

1.- Con la intención de establecer una posible dependencia de los estadígrafos promedio (P), desviación estándar (σ) y coeficiente de variación (Cv) con los coeficientes de frecuencia máximos (Km) de la serie de máximos anuales, se grafica la dispersión de los pares Km-P, Km- σ y Km-CV, no evidenciándose dependencia.

2.- Para cada estación del escenario o Clúster en análisis, se calcula la precipitación media máxima centrada y adimensionalizada (Pm).

3.- Se grafican los pares CV-P, Pm-CV y Pm-P, y se identifica la estación asociada a los pares que más se alejan del conjunto de datos, y que podrían determinar una curva que "envuelva" superiormente todos ellos.

4.- Si la estación que se escapa de los datos posee condiciones climáticas distintas a las del conjunto de datos (estación muy austral o estación en altura, donde pudiese predominar un régimen de precipitación nival, o estaciones insulares como Juan Fernández), se identifica la misma, permitiendo que esta quede fuera de la envolvente.

5.- De manera gráfica, tanto para los pares CV-P como para los pares Pm-P, se traza una curva envolvente, que como su nombre indica, acota superiormente al conjunto de estaciones.

6.- Con la finalidad de parametrizar las curvas trazadas, se ajusta un polinomio de grados "n" a algunos puntos característicos de esa curva trazada "manualmente".

7.- Para conocer la función envolvente Pm(P), se evalúa el polinomio Pm (CV) en la relación CV(P), obteniéndose como resultado una función que considera las relaciones máximas de los estadígrafos Pm-CV-P. A esta última relación, en el contexto del método de Stowhas, se le conoce como envolvente doble.

8.- Se verifica que todos los pares Pm-P queden por debajo de la envolvente doble, permitiendo que esta sea sobrepasada únicamente por las estaciones que se identificaron como "no representativas", en el punto 4.

9.- Finalmente, con la finalidad de trazar una segunda envolvente-doble que no considere las estaciones en altura, se reduce la curva CV-P, viendo si eventualmente existe una nueva relación, menor y paralela a la original, que pudiese excluir únicamente las estaciones en altura (estaciones con régimen nival).

10.- Realizando la evaluación de esta nueva relación CV*(P) en la relación Pm (Cv), se verifica que la nueva envolvente doble Pm*(P) siga acotando superiormente la totalidad del conjunto.

4.6 Formación de conglomerados y análisis a nivel regional

Con la intención de efectuar un análisis regional para el trazado de las envolventes de precipitación, se busca generar un mapa de zonificación con los distintos coeficientes para determinar la PMP en Chile.

Para ello, se busca encontrar un patrón natural en el conjunto de estaciones, mediante análisis de Cluster (o de conglomerados). Esta técnica estadística multivariante busca agrupar los objetos o individuos, tratando de lograr la máxima homogeneidad en cada grupo, y la mayor heterogeneidad entre ellos.

A grandes rasgos, el procedimiento empleado para lograr la(las) mejor(es) formas de agrupación de las estaciones, se consigue respondiendo a 4 preguntas que surgen en el proceso de implementar computacionalmente la metodología de análisis. Estas son: (i) ¿qué parámetros son los representativos para caracterizar la PMP en una determinada estación?; (ii) ¿cuán incidente es cada uno de ellos a la hora de decidir a qué grupo (o región) pertenece la estación?; (iii) ¿cómo caracterizar la similitud y diferencia entre grupos?; y finalmente, (iv) ¿cuántos grupos (o regiones) es pertinente formar?.

Para dar respuesta a dichas interrogantes, es necesario iterar y probar distintos escenarios, tomando decisiones que permiten converger a una familia reducida de soluciones, consistentes con la realidad climática del país.

Primera decisión: Qué parámetros utilizar

De los distintos métodos estudiados para determinar la PMP, se observa que, si bien se conoce la física del problema y los factores que en teoría influyen en la ocurrencia del evento extremo, en la práctica no existen registros estadísticos que permitan sostenerse en ellos para realizar un análisis a nivel global. No obstante, con los datos disponibles en la base de datos consolidada, se pueden calcular parámetros que indirectamente estén correlacionados con las variables hidrológicas que describen la física del problema.

Resulta razonable pensar que la temperatura mínima, media y máxima diaria, el contenido de humedad de la atmósfera, la velocidad del viento y las precipitaciones máximas

registradas, son determinantes de un clima predominante, y en este, los eventos extremos ocurren bajo mecanismos de gestación del mismo tipo, repitiéndose en un mismo lugar.
Esto permite postular los parámetros latitud, longitud, altitud, precipitación diaria máxima registrada y precipitación media anual (módulo pluviométrico) de cada estación, como parámetros representativos para aplicar el análisis de Cluster.

Para implementar computacionalmente el AC, la ubicación (latitud, longitud y altitud) de las estaciones se obtiene directamente de la base de datos consolidada, mientras que el módulo pluviométrico es calculado a partir del "data frame" que contiene los registros de precipitaciones diarias. De dichos valores, mediante rutinas implementadas en Python, se determina la precipitación mensual de un año determinado, y de la suma de estas se obtiene la precipitación anual. Del promedio de estas últimas, para un período en lo posible de 30 años, se obtiene el módulo pluviométrico (o precipitación media anual) de la estación en cuestión.

Segunda decisión: Cómo comparar los parámetros entre sí

Para asignar un mismo nivel de importancia a cada uno de los parámetros representativos de las condiciones climáticas en una determinada estación, se adimensionaliza (estandariza) cada parámetro. Esto debido a que la similitud entre objetos (estaciones) se expresa en términos de una magnitud (medida de distancia), que, en caso de trabajar con datos no estandarizados, se ve afectada por el rango numérico de variación o escala de cada una de las variables.

Tercera decisión: Selección de medidas de similitud entre grupos y métodos de agrupación

Los métodos de clustering, para poder llevar a cabo las agrupaciones, necesitan definir y cuantificar la similitud entre las observaciones. Para ello, dentro del contexto del clustering, se emplea el término distancia como medida de cuantificación de la similitud o diferencia entre observaciones.

La característica que hace del clustering un método adaptable a escenarios muy diversos es que puede emplear cualquier tipo de distancia, lo que permite al investigador escoger la más adecuada para el estudio en cuestión. Para efectos del estudio, se realiza el AC iterando

con las métricas distancia Coseno, distancia Manhattan, distancia Cityblock, distancia de correlación y distancia euclidiana al cuadrado.

Por otra parte, se realiza el AC considerando primero el método de aglomeración k-means y luego el k-medoids, obteniéndose distintos resultados de agrupación.

Cuarta decisión: Cuántas regiones formar

Esta última decisión es una de las más sencillas, ya que existen una serie de índices y pruebas que permiten validar interna y externamente los resultados del análisis de Clúster.

La validación interna evalúa que tan buena es la estructura del resultado sin necesidad de información ajena, a través de medidas de homogeneidad, cohesión ("compactness") y similitud.

Además, el número óptimo de clusters entregado por los distintos tests de validación interna, es complementado con criterios cómo la conveniencia de tener muchos o pocos grupos distintos para efectos de posteriormente estimar la PMP, y cuál sería el mínimo razonable de estaciones que deben existir en una región, para que la estimación del máximo coeficiente de frecuencia sea estadísticamente consistente (aplicabilidad del método de Hershfield).

De esta última decisión surge la interrogante de si incorporar o no en el análisis las estaciones del Norte Grande (no consideradas en Stowhas B, 1983); y también, de evaluar la alternativa de eliminar estaciones en altura y estaciones de la Patagonia Sur, donde generalmente las precipitaciones son de régimen nival.

El análisis presentado anteriormente se lleva a cabo en el Software matemático MatLab, utilizando las funciones y aplicaciones para describir, analizar y modelar datos, contenidas en el paquete de herramientas "Statistics and Machine Learning Toolbox".

5 Resultados

5.1 Resultados de la actualización de los métodos

5.1.1 Máximo coeficiente de frecuencia (Hershfield, 1961)

Ajustando una distribución generalizada de valores extremos a la serie de valores formada por los máximos coeficientes de frecuencia de cada escenario, se obtiene el valor actualizado del coeficiente, para la estimación de la PMP vía método de Hershfield. Los parámetros de forma, escala y posición de las distribuciones de probabilidad, y los valores límites del coeficiente máximo de frecuencia son presentados en la figura 8 y tabla 3, respectivamente.

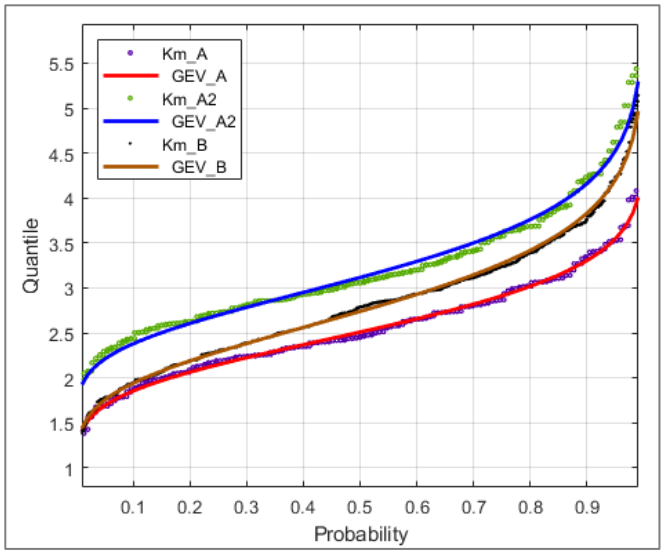


Figura 8: Ajustes de distribuciones generalizadas de valores extremos (GEV) a las series de coeficientes de frecuencia máximos, escenarios A, A2 y B

Tabla 3: Parámetros de la distribución generalizada de valores extremos ajustada a cada escenario

Escenario	A	A2	B
Modelo ajustado	Distribución Generalizada de Valores Extremos		
Log - Verosimilitud	-122.14	-150.84	-537.14
Media	2.55	3.21	2.82
Varianza	0.32	0.515	0.5688
Parámetros del ajuste			
k (forma)	-0.1667	-0.066	-0.088
σ (escala)	0.5248	0.605	0.651
μ (posición)	2.3238	2.89	2.51
Km máx	**5.45**	**11.21**	**9.67**

5.1.2 Envolventes de precipitación máxima adimensionalizada

La figura 9 permite observar la justificación de la envolvente trazada a partir de la dispersión de los pares precipitación máxima anual, coeficiente de variación de las precipitaciones máximas anuales, y precipitaciones máximas adimensionalizadas y centradas, en cada estación. En la figura en cuestión, se identifican también aquellas estaciones (Cameron, Río Rubens, Lago O'Higgins) que producto de las razones ya comentadas, se excluyen del análisis.

Por otra parte, la figura 10 muestra los resultados de las envolventes intermedias CV(P), Pm (CV) y las doble envolventes resultantes de la evaluación de dichas funciones, tanto para lo presentado en estudios anteriores (Stowhas, 1983) como para el escenario de actualización.

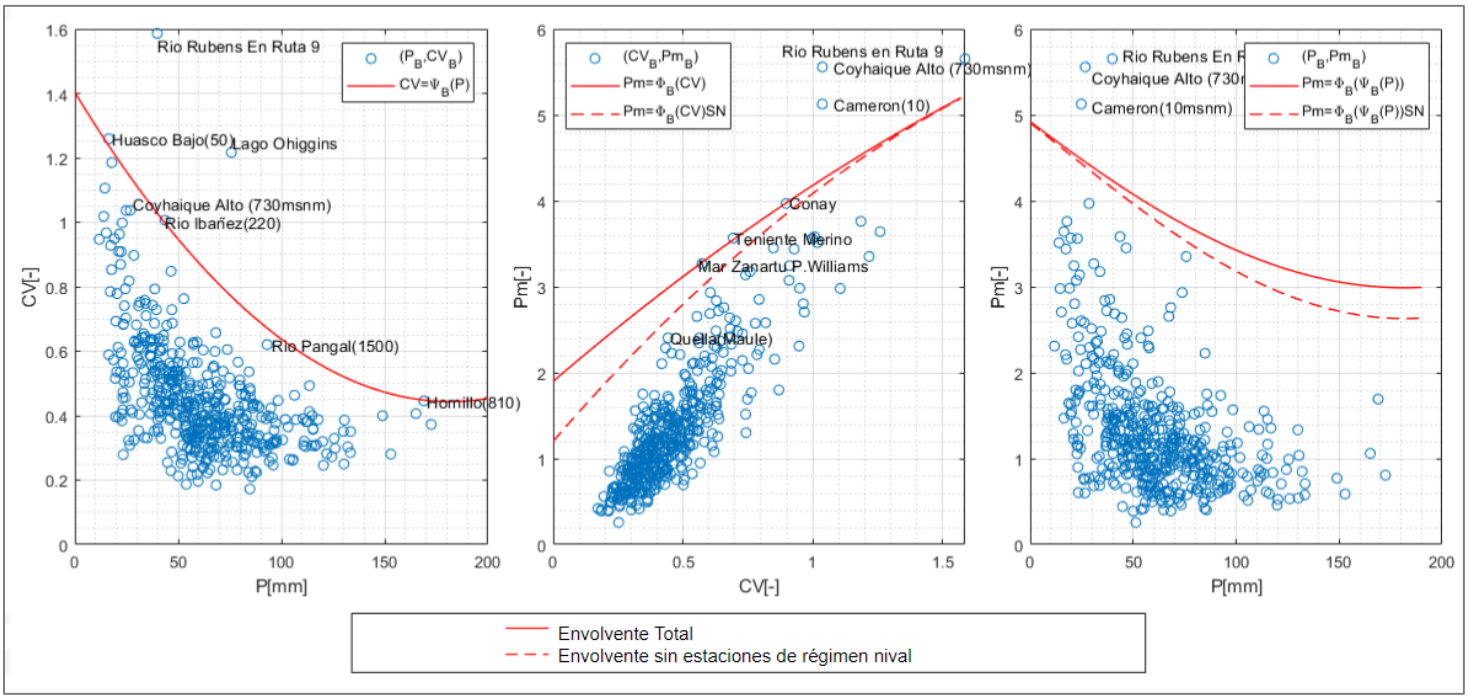


Figura 9: Envolventes intermedias y envolventes dobles para precipitación máxima adimensionalizada y centrada - Escenario B

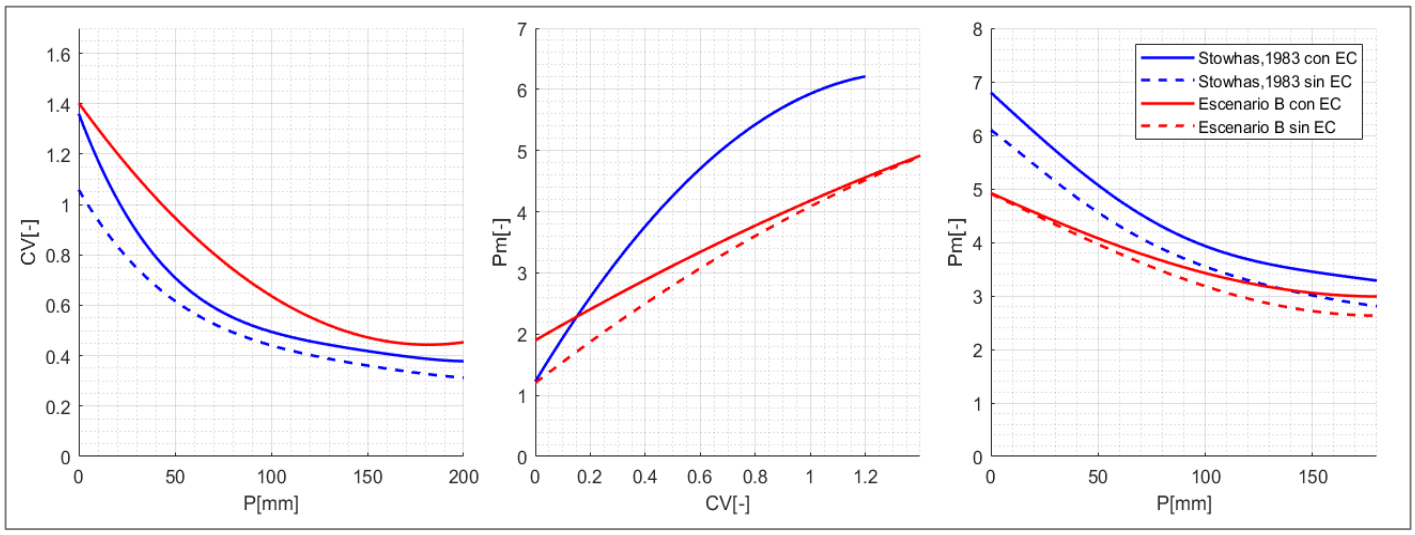


Figura 10: Actualización de las envolventes de precipitación máxima adimensionalizada y centrada, respecto de los estadígrafos CV-P con y sin estaciones en altura

Las expresiones que definen las relaciones actualizadas para el cálculo de la PMP vía método de Stowhas, pueden ser obtenidas a través de la ecuación 13, con los parámetros presentados en la tabla 4.

$$PMP\ (P_c) = \left[a_1\left(b_1{P_c}^2 + b_2P_c + b_3\right)^2 + a_2\left(b_1{P_c}^2 + b_2P_c + b_3\right) + a_3\right]\ P_c + P_c$$

Ecuación 13: Expresiones actualizadas para la obtención de la PMP vía método de Stowhas, a partir de la precipitación máxima anual corregida (P_c)

Tabla 4: Parámetros que definen las relaciones envolventes Pm(P) de figura 10- Escenario B

Envolvente	Parámetro					
	a1	a2	a3	b1	b2	b3
Total	-0.313	2.594	1.899	2.9.E-05	-0.011	1.403
Sin Est. Cord	-0.603	3.487	1.202	2.9.E-05	-0.011	1.403

De las curvas y valores presentados se puede evidenciar que, tanto el valor del máximo coeficiente de frecuencia, como la envolvente Pm(P) propuesta por Stowhas, se ven reducidos al actualizar los métodos con las estadísticas disponibles a la fecha. Esto indica a primera vista, que podría existir una tendencia de disminución en el valor de la PMP, siempre y cuando no se considere que los promedios y desviaciones estándar de las precipitaciones máximas anuales en cada estación también podrían haber sufrido un cambio

(recordar que la PMP es función, también, del promedio y desviación estándar de las precipitaciones máximas anuales, registradas en la estación en la que se desea realizar la estimación).

5.2 Resultados del análisis regional

5.2.1 General

Del proceso iterativo de análisis de las distintas técnicas y escenarios, y de una revisión bibliográfica exhaustiva, se concluye que la técnica de análisis que más se adecúa a la naturaleza de las variables y a las características de la muestra, es la presentada en la tabla 5.

Tabla 5: Características del Análisis de Clúster realizado

Tipo de Análisis	No jerárquico
Método de agrupación	k-medianas
Parámetros	Latitud [°]
	Longitud [°]
	Altitud [msnm]
	Módulo pluviométrico [mm]
	Precipitación máxima diaria histórica [mm]
Medida de asociación	Distancia Euclidiana
Número de Clústers	4 y 5

En el proceso resulta atractivo estudiar 2 posibles soluciones. Si bien ambas son obtenidas bajo el mismo método de clustering, la diferencia en la cantidad de estaciones consideradas permite postular soluciones con distinto número de clusters.

La primera solución no considera en el análisis aquellas estaciones que no cumplen con los criterios establecidos por Stowhas (1983), que dicen relación principalmente con la exclusión de las estaciones cuyo régimen de precipitaciones predominante suele ser del tipo nival (criterios C3, C4 y C5 en tabla 6).

La segunda solución propuesta, solo elimina las estaciones cuya longitud de registro sea menor a 15 años, o aquellas en las que la totalidad de las mediciones registradas son nulas (criterios C1 y C2).

Tabla 6: Estaciones eliminadas en las soluciones 1 y 2

Criterio de eliminación	Condición	Solución 1	Solución 2
		879	879
C1: Mediciones nulas	sum (Pmax)=0	6	6
C2: Registro <15 años	count(t)<15	277	277
C3: Norte cordillerano	Lat >-25° & alt>1500msnm	114	0
C4: En altura	alt > 3000msnm	93	0
C5: Región de Magallanes	Lat < -50°	80	0
	C1 ∩ C2 ∩ C3 ∩ C4 ∩ C5	401	279
	# Est. utilizadas	**478**	**600**

Solución 1

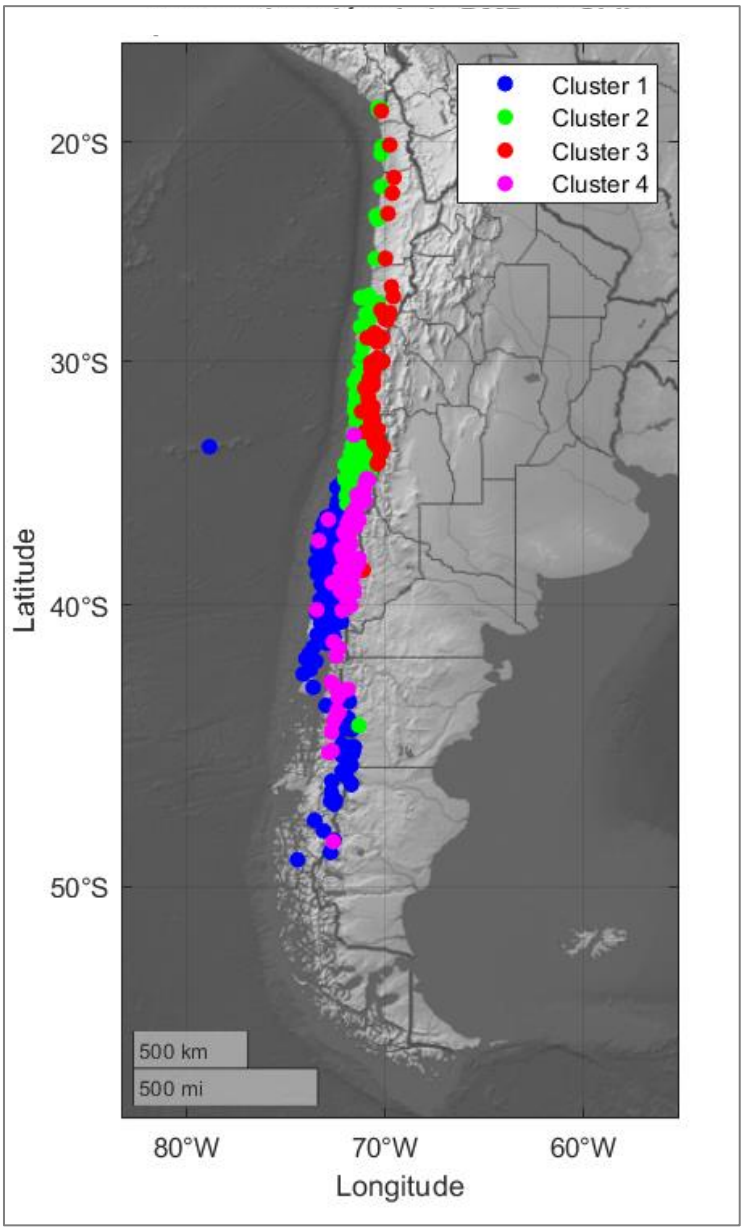


Figura 11: Solución 1, alterativa de regionalización mediante Análisis de Clúster

La figura 11 muestra una alternativa de zonificación para el cálculo de la PMP, que divide Chile continental en 4 regiones; en el sentido Este-Oeste se presentan como 2 franjas, que a su vez son delimitadas en el sentido Norte-Sur por el paralelo 35°S. En el Anexo 11 se presenta un listado las estaciones y la macro-zona a la que cada una pertenece.

Solución 2

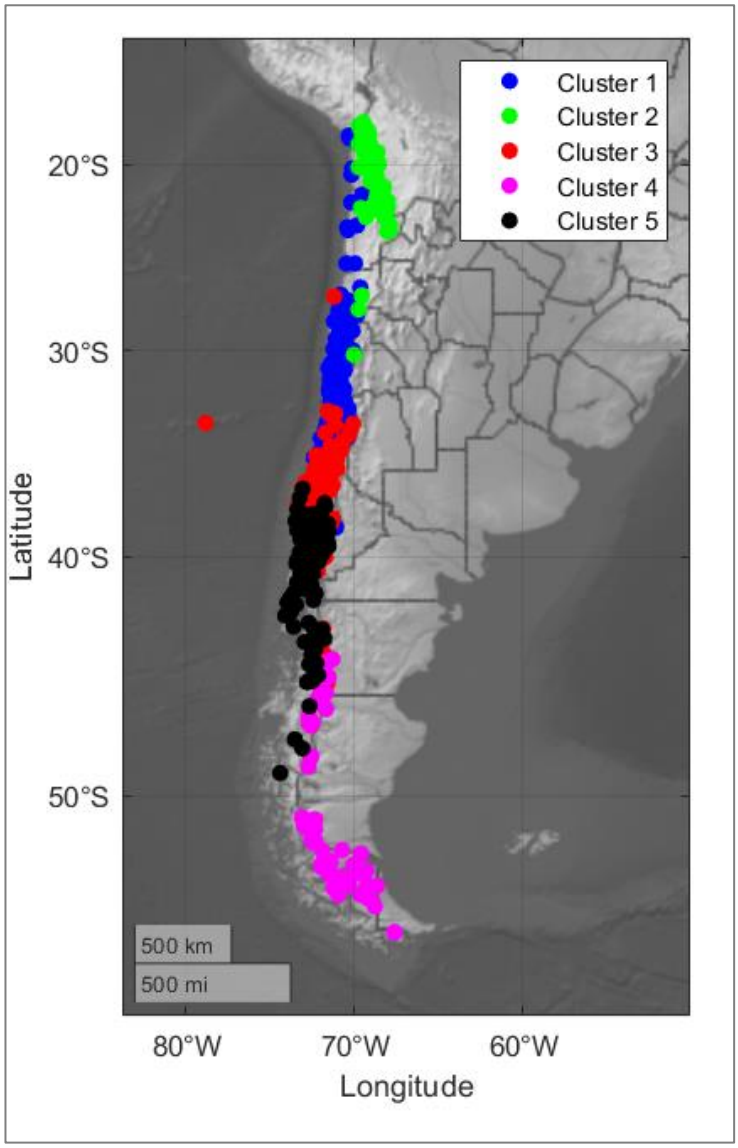


Figura 12: Solución *2*, alterativa de regionalización mediante Análisis de *clu*ster

La figura 12 muestra una segunda alternativa de zonificación, que a diferencia de a anterior, divide Chile continental en 5 regiones en sentido Norte-Sur.

En los acápites que siguen se discute la validez de las distintas envolventes y coeficientes máximos de frecuencia, para la correcta estimación de la PMP en cada una de las zonas.

Las figuras 13 y 17 presentan los valores del índice Silhouette with, que cuantifica que tan buena es la asignación de una estación meteorológica (elemento) a una determinada zona (grupo). Valores altos son indicativos de que la observación se ha asignado al cluster correcto, valores cercanos a cero indican que la observación se encuentra en un punto intermedio; por último, valores negativos indican una posible asignación incorrecta. Por las características de la muestra, existen algunas estaciones que distan prácticamente de todos los grupos, debiendo la técnica de análisis empleada (que tiene un número predeterminado de clusters) asignarla al grupo o zona que menos difiera.

Por otra parte, las figuras 14 y 18 muestran los resultados del cálculo de un conjunto de ratios, que permiten indicar cual es el número óptimo de grupos a formar según las características de la muestra. Si bien cada uno de los ratios presentados se interpreta de una forma distinta, en conjunto pueden validar el número de clústers elegido para cada una de las soluciones propuestas.

El índice "average silhouette" maximiza la media de los "silhouette coeficients" o índices silueta (figuras 13 y 17). Este coeficiente cuantifica que tan buena es la asignación que se ha hecho de una observación a un determinado cluster, siendo valores altos un indicativo de que la observación se ha asignado al cluster correcto. El método "average silhouette" considera como número óptimo de clústers aquel que maximiza la media del "silhouette coeficient" de todas las observaciones. Para las observaciones de las soluciones propuestas (Solución 1 y Solución 2) el número óptimo de clústers resulta ser, en ambos casos, 2.

El estadístico gap compara, para diferentes valores de k, la varianza total intra-cluster; observada frente al valor esperado acorde a una distribución uniforme de referencia (Amat, 2017). La estimación del número óptimo de clusters se consigue seleccionando el valor k con el que se maximiza el estadístico gap, es decir, encontrando el valor de k con el que se consigue una estructura de clusters lo más alejada posible de una distribución uniforme aleatoria. (**k: número de clusters)

Para identificar el número óptimo de clusters mediante el estadístico Gap, lo que se hace es identificar el menor de los k para el que el estadístico se aleja menos de una desviación

estándar del valor Gap del siguiente k. Según el método señalado, el número óptimo de clusters para las soluciones 1 y 2 es 4 y 5, respectivamente.

El índice CalinskiHerabaesz (también conocido como criterio de relación de varianza) es una medida de qué tan similar es un objeto a su propio grupo (cohesión) en comparación con otros grupos (separación). El índice estima la cohesión en función de las distancias desde los puntos de datos en un grupo a su centroide de grupo; y la separación a partir de la distancia de los centroides de grupo respecto del centroide global. Un valor más alto del índice CH significa que los conglomerados son densos y están bien separados, no existiendo un valor de corte "aceptable" (Amat, 2017).

Para efectos de las soluciones propuestas en el estudio, en ambos casos se evidencia que mientras mayor es el número de clusters, los grupos se vuelven menos densos y poco separados unos de otros, siendo prudente formar, en ambos casos, no más de 5 regiones.

El índice de DaviesBouldin se interpreta similar al CalinskiHerabaesz, con la diferencia que valores pequeños del índice DB indican clústeres compactos, y cuyos centros estos bien separados los unos de los otros. Consecuentemente el número de clústeres que minimiza el índice DB se toma como el óptimo. Para efectos de las soluciones propuestas, el índice en cuestión no entrega información clara, lo que puede deberse al método de agrupación utilizado (k-medianas).

La información proporcionada por los distintos ratios, es luego contrastada con la inspección visual de los grupos, estableciéndose que es razonable formar 4 y 5 clusters para las soluciones 1 y 2, respectivamente.

Finalmente, las figuras 15, 16, 19 y 20 permiten validar de manera visual los resultados del análisis. De ellas se puede observar que los grupos formados consideran estaciones con altitudes, módulos pluviométricos y precipitaciones máximas históricas acotados en intervalos distintos para los diferentes grupos, siendo también evidente que las estaciones pertenecientes a un determinado grupo están ubicadas geográficamente de forma contigua o compacta. Esto último permite posteriormente definir barreras físicas o geográficas, que delimiten cada una de las zonas.

Validación de la Solución 1

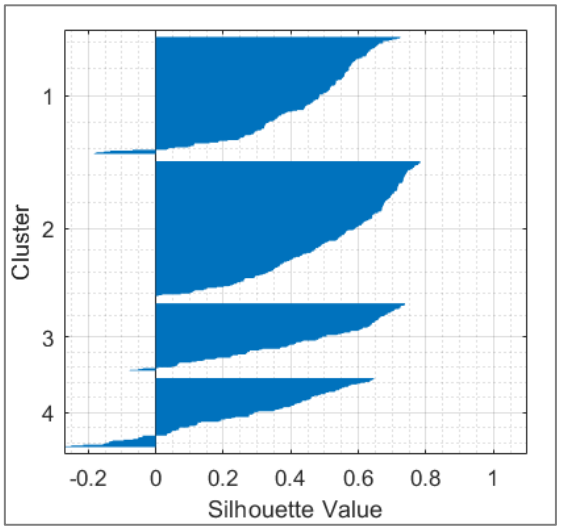


Figura 13: Validación interna de la Solución 1 - Silhouette width index

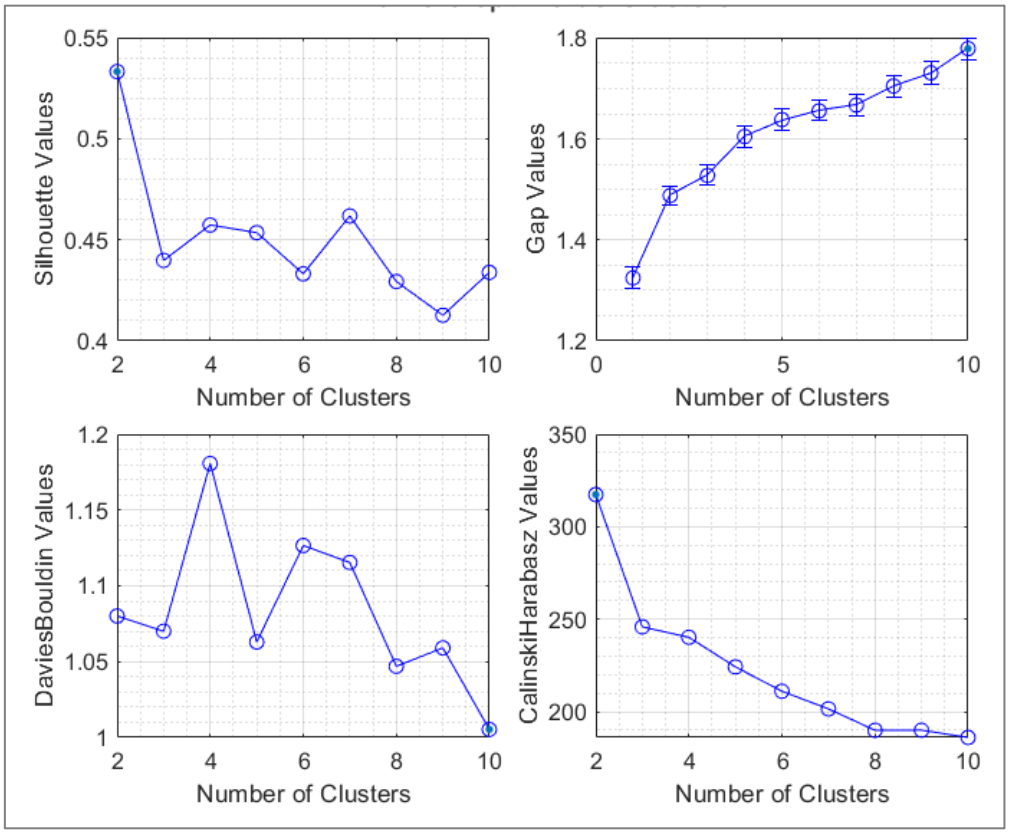


Figura 14: Determinación del número óptimo de clusters mediante distintos índices (Solución 1)

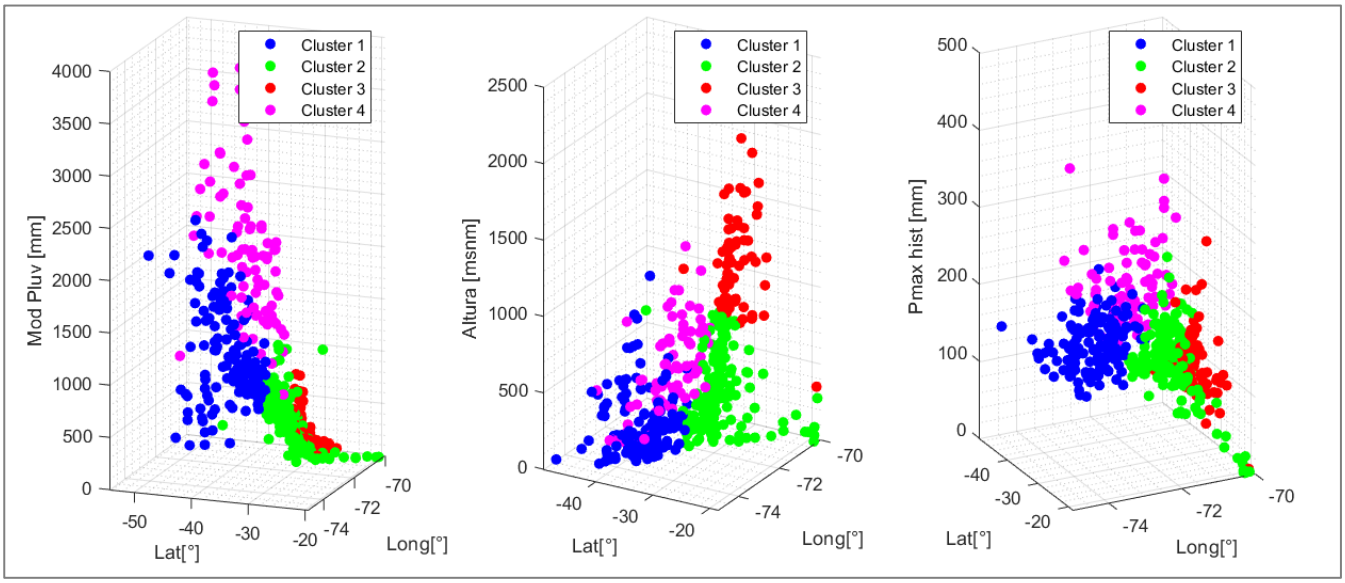


Figura 15: Validación visual de los clusters; dispersión de la altitud, precipitación acumulada anual y máxima histórica (Solución 1)

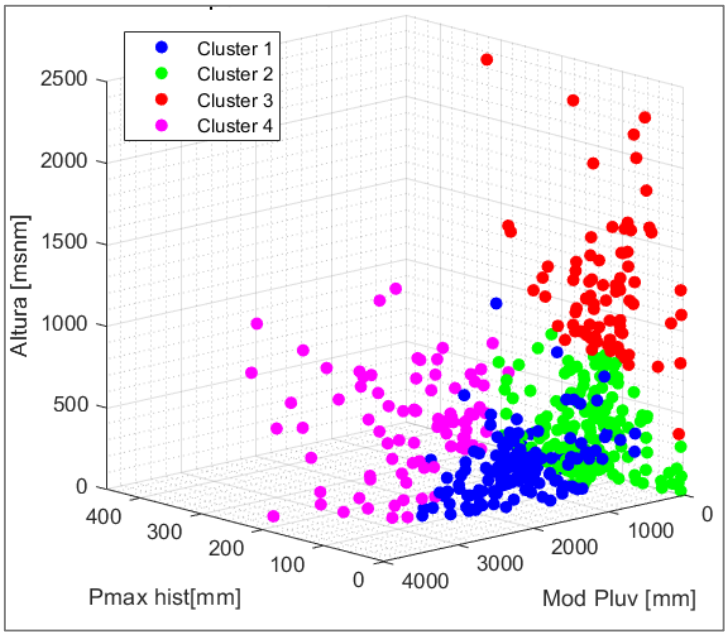


Figura 16: Validación visual de los clusters; dispersión de la precipitación acumulada anual, altitud y máxima histórica (Solución 1)

Validación de la Solución 2

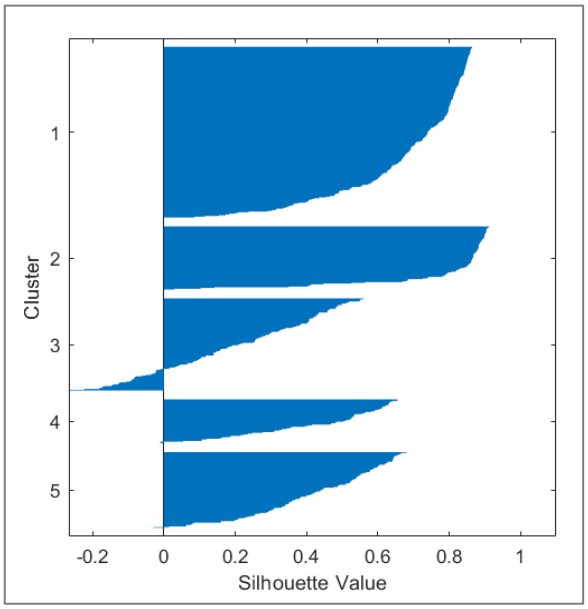


Figura 14: Validación interna de la Solución 2 - Silhouette width index

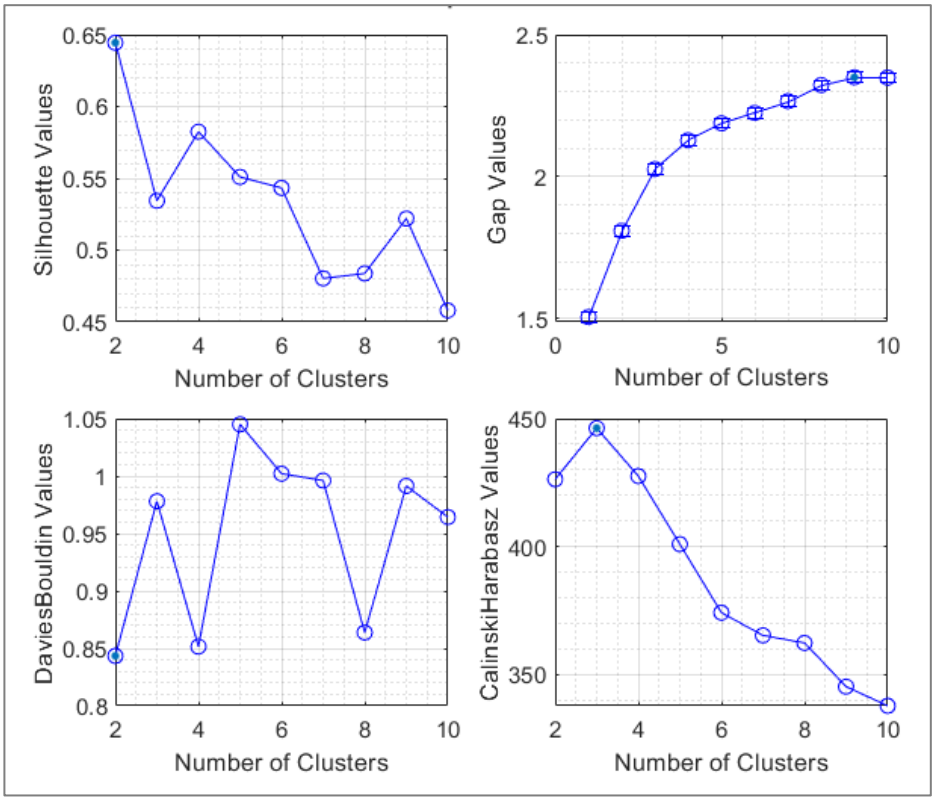


Figura 18: Determinación del número óptimo de clusters mediante distintos índices (Solución 2)

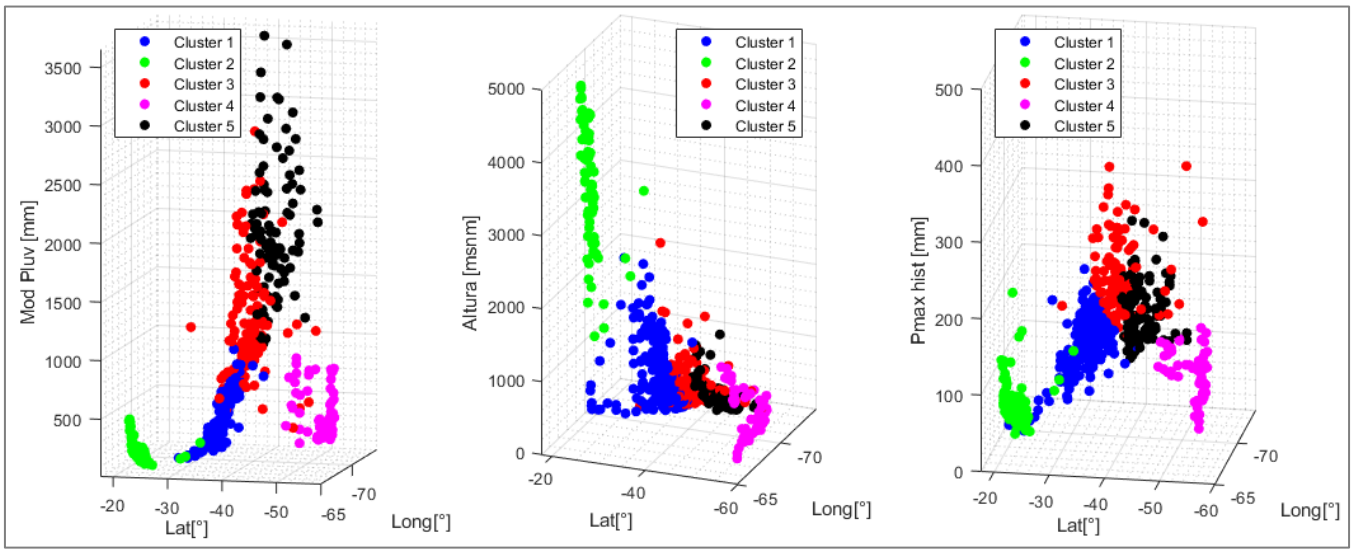


Figura 15: Validación visual de los clusters; dispersión de la altitud, precipitación acumulada anual y máxima histórica (Solución 2)

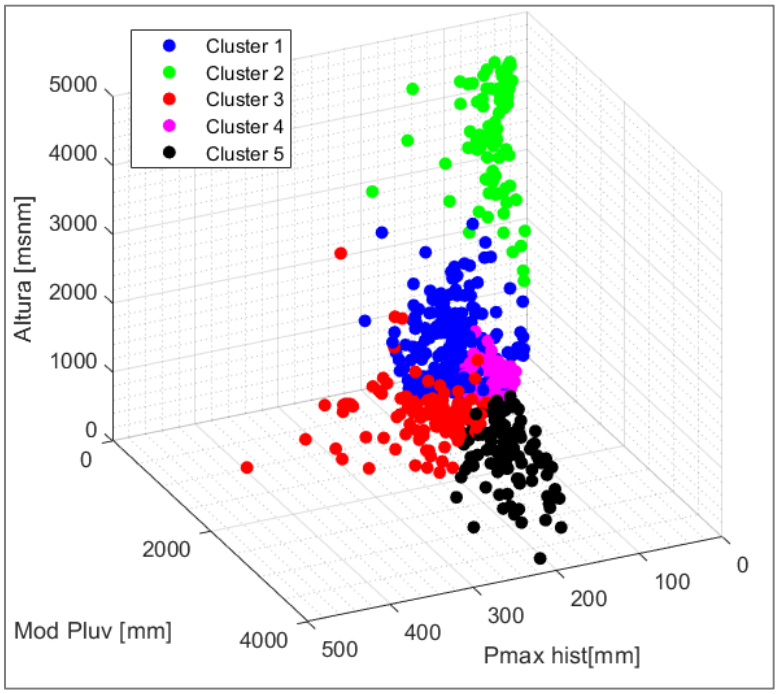


Figura 20: Validación visual de los clusters; dispersión de la precipitación acumulada anual, altitud y máxima histórica (Solución 2)

5.2.2 Obtención de la PMP a nivel regional

A cada uno de los grupos formados mediante el análisis descrito se les aplica los métodos de Hershfield y Stowhas, obteniéndose los coeficientes máximos de frecuencia y las doble-envolventes presentadas a continuación.

Aplicación del método de Hershfield a nivel regional

Tabla 7: Máximos coeficientes de frecuencia obtenidos al aplicar el método de Hershfield a cada Clúster – Solución 1

Grupo	Cant. Estaciones	Km máx
Norte Cordillerano	83	9.8
Norte	166	10.3
Sur	144	8.6
Sur Cordillerano	85	25.5
Todos juntos	478	10.3

Tabla 8: Máximos coeficientes de frecuencia obtenidos al aplicar el método de Hershfield a cada Clúster – Solución 2

Grupo	Cant. Estaciones	Km máx
Cluster 1	230	10.8
Cluster 2	85	22.3
Cluster 3	124	8.2
Cluster 4	59	36.9
Cluster 5	102	10.5
Todos juntos	600	12.4

Las tablas 7 y 8 muestran las cantidades de estaciones por región y los valores de los máximos coeficientes de frecuencia resultantes al aplicar el método de Hershfield (1961) a cada una de ellas.

Del análisis regional de la primera solución propuesta se evidencia que tanto la macrozona Norte como la Sur-Poniente presentan valores similares entre sí, siendo también bastante parecidos cada uno de ellos al valor obtenido al considerar todo Chile como una única gran región.

Lo anterior no ocurre en la región Sur-Oriente (cordillerana), donde los altos módulos pluviométricos y las altas precipitaciones máximas históricas dan origen a un valor límite del coeficiente de frecuencia considerablemente mayor.

De los valores límites resultantes para los coeficientes de frecuencia de la segunda solución propuesta, no se pueden obtener conclusiones a-priori.

Aplicación del Método de Stowhas a nivel regional

Con la finalidad de conocer las relaciones envolventes Pm(P) en cada una de las zonas propuestas, se repite el procedimiento planteado por Stowhas para el trazado de la doble envolventes, a partir de las relaciones límites entre los pares coeficiente de variación - promedio de precipitación máxima anual, y precipitación máxima anual adimensionalizada - coeficiente de variación.

Las estaciones que determinan las relaciones límites, y aquellas que por motivos justificados fueron excluidas del análisis a nivel regional, son presentadas en el Anexo B – 10.2. Los resultados del proceso en cuestión son presentados conjuntamente, en la figura 21.

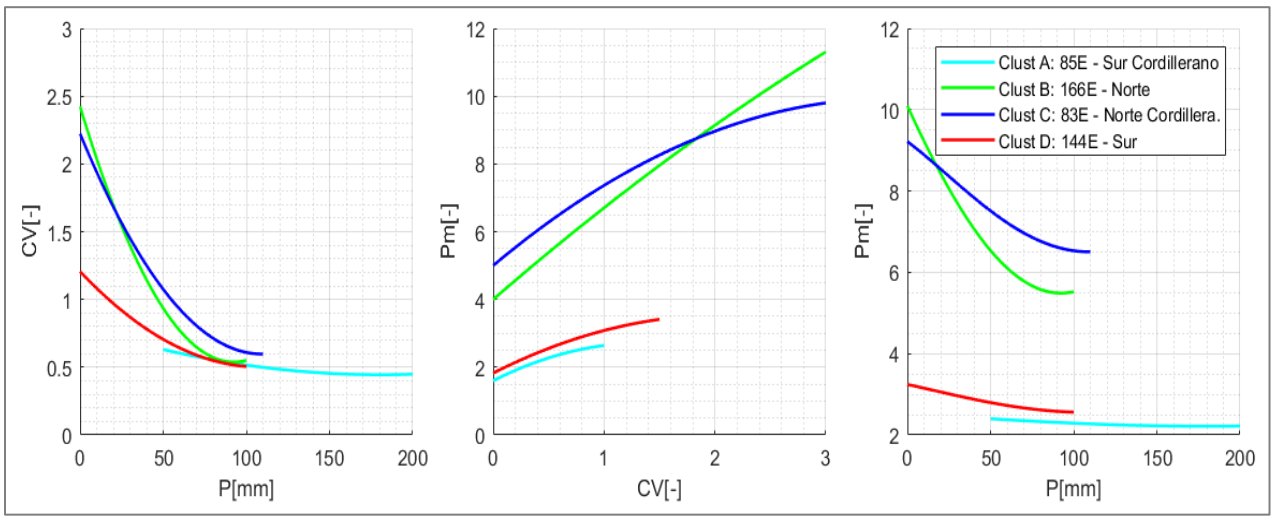


Figura 21: Envolventes de Stowhas en cada región - Solución 1.

Las envolventes trazadas en la figura 21 fueron parametrizadas mediante polinomios cuadráticos, cuyas constantes son presentadas en la tabla 9. Con ello, la PMP puede ser obtenida directamente, remplazando estos en la ecuación 13.

Tabla 9: Parámetros de funciones envolventes regionalizadas, método de Stowhas 2020

Parámetro	Región			
	Sur cordillerano	Norte	Norte cordillerano	Sur
a_1	-0.616	-0.134	-0.3827	-0.4008
a_2	1.6559	2.8381	2.748	1.6576
a_3	1.6	4.0	5.0	1.8264
b_1	1.09E-05	2.21E-04	1.36E-04	6.00E-05
b_2	-0.0039	-0.0408	-0.0298	-0.013
b_3	0.8	2.4251	2.2241	1.2064

6 Discusión

6.1 Sobre la actualización de los métodos

6.1.1 A partir de la PMP estimada en cada estación

Comparación entre métodos de Hershfield y Stowhas, según publicación 1983

En el primer estudio realizado en Chile (Stowhas, 1983) se calculó la PMP en cada una de las 190 estaciones disponibles a la fecha, evidenciándose que solamente en el 11% de ellas las estimaciones mediante el método de Stowhas son mayores que las obtenidas vía método de Hershfield; resultando en consecuencia, estimaciones obtenidos vía envolventes-dobles, menores en el 89% de los casos. Esto, trazando las envolventes dobles a partir del conjunto completo de estaciones disponibles (dicho de otra forma, considerando las estaciones cordilleranas en el análisis).

Según lo presentado en dicho estudio, si se excluyen las estaciones en altura en el proceso de obtención de las envolventes, los casos en los que el método de Hershfield entrega resultados mayores aumentan del 89% al 92%, siendo menor la PMP calculada vía Hershfield solo en el 8% restante.

Evaluación de los métodos sin calibrar (1983) bajo la estadística actual

Con la finalidad de evaluar si se mantiene la fracción de estaciones en las que el método alternativo (M. Stowhas) sub-estima la PMP, se estudian los resultados obtenidos a partir de ambos métodos, con las estadísticas disponibles en la actualidad (no se varían las envolventes ni el coeficiente de frecuencia presentados en 1983). De ello se obtiene que, de las 490 estaciones meteorológicas, en el 72% de los casos el método de Hershfield entrega resultados más conservadores, mientras que en el 28% restante las estimaciones vía método de Stowhas resultan menores.

Estimaciones con las estadísticas actuales

Para poder sacar conclusiones respecto de una eventual tendencia de sobreestimación del método de Hershfield, se obtiene la PMP en cada estación en estudio, considerando el coeficiente de frecuencia actualizado (k=9.67) y las nuevas envolventes Pm-P del método de Stowhas (Escenario B).

El número de estaciones en las que un método entrega valores mayores o menores que el otro, en cada uno de los casos (con y sin estaciones cordilleranas en el análisis), es presentado en la tabla 10.

Tabla 10: Casos en los que la PMP calculada utilizando M. Stowhas es (no es) menor que la obtenida vía M.Hershfield, considerando las envolventes Pm-P actualizadas y k=0.967

	Env. Total	Sin Est. Cord
PMP St > PMP H:	101 (21%)	79 (16%)
PMP St < PMP H:	389 (79%)	411 (84%)
TOTAL:	490 (100%)	490 (100%)

Del análisis en cuestión se observa que en el 79% de las estaciones utilizadas para la recalibración de los métodos, el método de Stowhas entrega estimaciones menos conservadoras, aumentando dicha fracción en un 5% al no considerar las estaciones con precipitaciones de régimen nival (cordilleranas o en altura) en el trazado de las envolventes.

Estaciones en altura

Resulta interesante estudiar la incidencia que tienen en el trazado de las envolventes las estaciones meteorológicas con régimen de precipitaciones predominantemente nival. Para intentar dilucidar dicho efecto, se estima la PMP en cada una de las estaciones con las envolventes dobles resultantes del procedimiento presentados por Stowhas, L., considerando y no considerando las estaciones en altura (de régimen nival o cordilleranas) en el trazado de estas mismas.

Del análisis anterior, se evidencia que la envolvente sin estaciones de régimen nival entrega estimaciones que están entre el 1% y el 9% por debajo de aquellas estimadas a partir de la envolvente que considera la totalidad de ellas, disminuyendo la PMP en promedio un 3% al utilizar la envolvente que excluye las estaciones cordilleranas en el trazado.

De manera general, resulta evidente que al actualizarse los parámetros de ambas metodologías se produce una disminución en los valores de PMP obtenida, resultando también interesante evaluar cual método sufre mayor variación tras la actualización, y con ello intentar concluir respecto del nivel de confianza de cada uno.

Del cuestionamiento se evidencia que al actualizar el valor del máximo coeficiente de frecuencia (Hershfield,1961) la PMP sufre una disminución del 11%, en promedio. Al actualizar la envolvente de Stowhas, se observa que la nueva envolvente entrega estimaciones que son en promedio un 19% (Envolvente total) y 12% (Envolvente sin estaciones en altura) menores a las obtenidas a partir de lo presentado en estudios anteriores (Stowhas, 1983).

Comparación entre los métodos

Si se evalúan las diferencias entre los resultados obtenidos mediante los distintos métodos, para 84% (o 79%) de las estaciones en que Hershfield entrega mayores estimaciones de la PMP, se evidencia que independientemente de cuál envolvente se utilice, los valores entregados por el método de Hershfield son en promedio un 25% mayores a los estimados mediante el método de Stowhas.

Por el contrario, si se estudia la diferencia entre las PMPs obtenidas bajo los distintos métodos para el 16% (o 21%) de las estaciones en que el método de Stowhas entrega menores estimaciones, se observa que las diferencias son, en promedio, del orden del 16%.

Lo evidenciado en el análisis en cuestión es consistente con lo expuesto en estudios anteriores, donde se expresa que la similitud climática y orográfica de las costas americanas en el Pacífico Norte y Sur, y algunas estimaciones de la PMP hechas en Chile en base al método hidrometeorológico (M. Nenadovich – Harza, 1981, 1982) parecían indicar que el método de Hershfield tendería a sobreestimar la PMP en Chile, lo que se traduciría en un sobredimensionamiento de las obras, o al menos en un nivel de riesgo tan remoto, que signifique un innecesario encarecimiento de ellas (Stowhas, 1983).

Comparación entre las estimaciones y los máximos históricos

Con la finalidad de validar la actualización de los parámetros que definen las nuevas envolventes y el nuevo máximo coeficiente de frecuencia, resulta ventajoso contrastar la magnitud de los máximos eventos históricos con la PMP estimada bajo los distintos métodos, en cada una de las 490 estaciones en análisis.

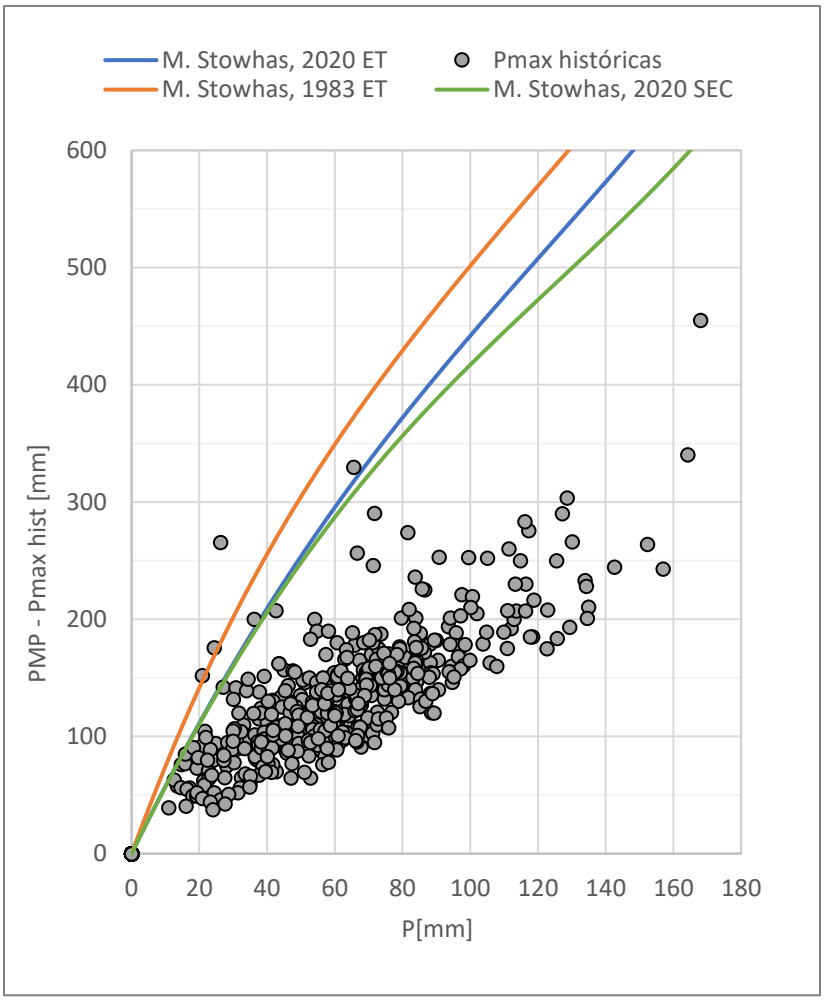


Figura 22: Envolventes de precipitación máxima probable y precipitaciones máximas históricas en Chile, dónde ET representa la envolvente trazada con el total de estaciones y SEC la envolvente trazada excluyendo estaciones de régimen nival.

Tabla 11: Estaciones en las que los máximos históricos sobrepasan la PMP estimada vía método Stowhas, considerando las envolventes anteriores (Stowhas, 1983) y actualizadas (Escenario B).

PMP vía	Stowhas,1983 Env. Total	Stowhas, 2020 Env. Total	Stowhas, 2020 Env sin est. Nivales
Estaciones PMP < Pmax histórica	Coyhaique Alto Cameron Rio Rubens	Coyhaique Alto Rio Ibañaez en desemb. Cameron Lago O'higgins Rio Rubens	Coyhaique Alto Rio Ibañaez en desemb. Cameron Lago O'higgins Rio Rubens
Total casos	3	5	5

De la comparación entre la máxima precipitación histórica registrada en cada estación (figura 22) y la PMP estimada vía Hershfield en dicho lugar, se observa que, en 2 de las 490 estaciones utilizadas para recalibrar los métodos, la precipitación máxima histórica sobrepasa a la PMP estimada. Las estaciones en las que esto ocurre son “Río Rubens” y “Cameron”, ambas ubicadas en Tierra de Fuego.

Del análisis de las condiciones meteorológicas predominantes en dicha región se puede concluir que las tormentas extremas en estas estaciones suelen ser con precipitaciones del tipo nivales, lo que estadísticamente se traduce en promedios de precipitación máxima anual (pluviales) muy bajos, y en consecuencia, en una sub-estimación de la PMP. Esto le quita aplicabilidad al método de Hershfield, siendo no recomendable utilizarlo en zonas con estas características.

Siguiendo con la misma lógica y en busca de definir la aplicabilidad de los métodos estadísticos como medio de estimación de la PMP, se contrastan los máximos históricos con las estimaciones obtenidas vía método de Stowhas, en cada una de las estaciones en estudio.

Del análisis resulta que 5 de las 490 estaciones presentan estimaciones puntuales menores a los máximos históricos registrados, lo que obliga a estudiar cada caso en particular, y de esta forma intentar buscar una explicación a la sub-estimación.

Los resultados son consistentes con el procedimiento presentado para el trazado de las envolventes (Inciso 4.5.- Implementación computacional del método de Stowhas), ya que las 5 estaciones en las que se subestima la PMP (presentadas en la tabla 11) están fuera de la envolvente trazada. Ellas son; “Coyhaique Alto”, “Río Ibáñez en Desembocadura”, “Cameron”, “Lago O’Higgins en Villa O’Higgins” y “Río Rubens”. La exclusión de dichas estaciones en el proceso de definición de la envolvente se sustenta en las características climáticas de la zona a la que pertenecen. En ellas predominan temperaturas bajas en los meses de invierno, siendo tormentas del tipo nival los eventos extremos más frecuentes.

Lo evidenciado en el proceso de validación permite reafirmar lo presentado por Stowhas (1983), quien sostiene que las estaciones que definen la envolvente muchas veces

corresponden a estaciones cordilleranas o en altura, con posibles componentes nivales que pudieran hacerlas no representativas de las condiciones a menor altitud. Aun cuando la cantidad de estaciones en altura es insuficiente para analizar este efecto, se puede trazar una segunda envolvente despreciando estos valores, que pudiera ser más representativa de estaciones bajas (Stowhas, 1983).

6.1.2 De los resultados obtenidos bajo los distintos escenarios

A lo largo del estudio se aplica la metodología de Hershfield,1961 a diversos escenarios, que consideran no solo distintos períodos históricos, sino también distintas estaciones meteorológicas.

En el proceso de actualización, se evalúa el coeficiente de frecuencia resultante del conjunto de estaciones que cumplen los criterios establecidos en (Stowhas, 1983), definiéndose los escenarios A, A2 y B, que consideran entre 146 y 490 estaciones.

Al comparar el coeficiente de frecuencia y la envolvente Pm(P) resultantes del escenario A (de comparación) con los presentados por Stowhas (1983), se observa que los resultados del escenario de comparación son del orden de la mitad (11 v/s 5.45) de los presentados anteriormente. Dado que el estudio previo no presenta el detalle de las estaciones consideradas, no se puede saber cuántas de esas 190 estaciones están contenidas en las 146 estaciones del presente estudio, atribuyendo la diferencia de los resultados únicamente a las diferencias de la estadística base.

Al contrastarse los coeficientes de frecuencia y las doble envolventes del escenario A con los del escenario A2, se observa ambos parámetros (o curvas) aumentaron su magnitud. Si bien los dos escenarios consideraron las mismas 146 estaciones, el segundo de ellos (A2) tiene 37 años más de estadística. Resulta lógico pensar que en dicho período adicional se registraron precipitaciones máximas mayores, y que producto de ello, las estimaciones esperadas aumenten.

Entre los resultados obtenidos en el escenario de actualización (Escenario B) y los presentados por Stowhas (1983) se puede realizar la misma comparación. Si bien las estaciones consideradas en el escenario de actualización cumplen con los mismos criterios

del estudio, la cantidad de estaciones aumentó a más del doble (de 190 a 490), extendiéndose también las estadísticas de ellas en 37 años. Lo anterior se tradujo en una leve disminución tanto del coeficiente de frecuencia (de 11 a 9.67) como de la envolvente Pm(P), considerándose estos resultados, por ahora, como los “los parámetros y envolventes recalibradas”.

Adicionalmente, en el intento de generar envolventes de validez regional, se incorporan y quitan del análisis nuevas estaciones, obteniéndose también en dicho proceso, el coeficiente de frecuencia para una envolvente global, que integra todas las regiones juntas.

La primera propuesta de regionalización (Solución 1) abarca un total de 478 estaciones, que quedarían acotadas por un coeficiente Km de 10.3; mientras que la segunda alternativa (Solución 2), con un total de 600 estaciones, entrega un valor máximo de 12.4. Las 122 estaciones no consideradas en la primera solución, pero si consideradas en la segunda, corresponden a estaciones con un alto componente nival, puesto que se ubican a más de 3000msnm, o en la región de magallanes. Lo anterior permite afirmar que las estaciones meteorológicas con condiciones hidrológicas nivales tienden a elevar el valor máximo de la serie de coeficientes de frecuencias.

6.2 Sobre la regionalización

6.2.1 Evaluaciones

Con la intención de dilucidar el efecto que tiene la utilización de envolventes regionales en el cálculo de la PMP, se realizan estimaciones bajo ambas metodologías, en las 5 estaciones con mayor longitud de registro, de cada región propuesta.

Las estaciones consideradas en el análisis son las presentadas en la tabla 12.

Tabla 12: Estaciones consideradas en el cálculo de la PMP a nivel regional

Región	Código	Nombre	alt [msnm]	L[años]
Sur cordillerano	7103003	Los Quees	663	102
	7320002	Armerillo	492	96
	7350006	Bullileo Embalse	600	90
	8130003	Fundo Atacalco	740	89
	8317003	Cerro El Padre	346	77
Sur	6027003	La Rufina	743	91
	5740004	Carmen De Las Rosas	165	90
	5748003	Cerrillos De Leyda	182	88
	4723002	Huintil	650	83
	6056003	Rapel	16	80
Norte cordillerano	5403006	Riecillos	1290	91
	4713004	San Agustin	1050	90
	5414004	Resguardo Los Patos	1220	80
	4502005	Hurtado	1100	77
	4511003	Las Ramadas	1380	77
Norte	9404002	Los Laureles	260	77
	8318002	Quilaco	231	72
	9102003	Lumaco	60	72
	360019	Carriel Sur Concepción.	12	71
	400009	Canal Bajo Osorno Ad.	61	71

Las PMP estimadas vía método de Hershfield considerando una única gran zona (PMP Global) y de manera local (PMP regional), se llevan a cabo con los coeficientes máximos de frecuencia presentados en las tablas 3 y 7, respectivamente. Lo mismo para las estimaciones vía método de Stowhas, que consideran los parámetros de las envolventes globales y locales, presentados en las tablas 4 y 9, respectivamente.

Tabla 13: Resultados de estimaciones de la PMP en las 5 estaciones de mayor longitud de registro, según métodos de Hershfield y Stowhas con envolventes globales y regionales.

Región	Código	Pc	Sc	M. Hershfield, 2020			M. Stowhas, 2020		
				PMP Global	PMP Regional	Dif [%]	PMP Global	PMP Regional	Dif [%]
Sur cordillerano	7103003	127.2	75.5	857	2052	**58%**	531	414	**-28%**
	7320002	190.2	89.0	1051	2459	**57%**	757	613	**-23%**
	7350006	142.7	54.3	667	1526	**56%**	582	462	**-26%**
	8130003	133.9	45.2	571	1287	**56%**	553	435	**-27%**
	8317003	109.9	36.6	463	1042	**56%**	475	360	**-32%**
Sur	6027003	103.8	40.6	496	453	**-10%**	455	369	**-23%**
	5740004	71.4	45.6	512	463	**-11%**	340	261	**-30%**
	5748003	58.6	31.3	361	327	**-10%**	290	219	**-32%**
	4723002	47.6	32.2	359	325	**-11%**	243	182	**-34%**
	6056003	71.7	33.6	397	361	**-10%**	341	262	**-30%**
Norte cordillerano	5403006	73.5	40.2	462	467	**1%**	348	580	**40%**
	4713004	52.9	32.3	365	369	**1%**	266	446	**40%**
	5414004	49.8	28.1	322	325	**1%**	253	424	**40%**
	4502005	37.7	29.2	320	323	**1%**	199	336	**41%**
	4511003	57.4	33.2	379	383	**1%**	285	476	**40%**
Norte	9404002	85.3	28.2	358	377	**5%**	391	557	**30%**
	8318002	84.1	28.7	362	381	**5%**	387	550	**30%**
	9102003	60.8	26.3	315	333	**5%**	299	431	**31%**
	360019	76.2	38.0	444	469	**5%**	358	507	**29%**
	400009	58.0	23.1	281	297	**5%**	287	417	**31%**

Los resultados presentados en la tabla 13 pueden ser expresados en términos medios (tabla 14). De ellos se evidencia que la regionalización no produce el mismo efecto en los dos métodos. En algunos clusters las estimaciones se ven reducidas (aumentadas) al aplicar regionalmente un método y aumentadas (reducidas) al aplicar el otro. Lo anterior permite por un lado, cuestionar la validez de los métodos a nivel regional, y por otro, poner en duda la idoneidad de la técnica utilizada para generar las regiones.

Tabla 14: Diferencias promedio a nivel regional entre estimaciones regionales y globales, de las 5 estaciones con mayor longitud de registro de cada región

Región	M. Hershfield	M. Stowhas
Sur cordillerano	57%	-27%
Sur	-10%	-30%
Norte cordillerano	1%	40%
Norte	5%	30%

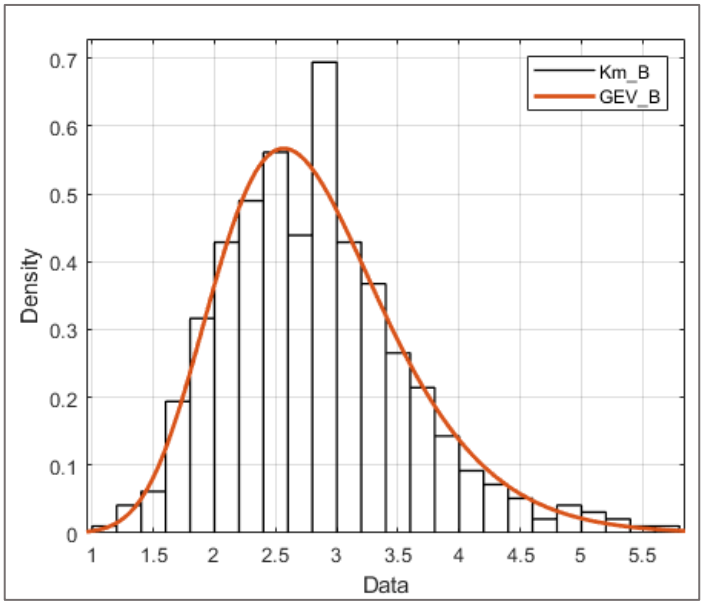


Figura 23: Histograma de los coeficientes de frecuencia y función de densidad de probabilidad de la distribución generalizada de valores extremos – Escenario B (actualización)

Con la intención de dilucidar si los datos (coeficientes de frecuencia) provienen de distintas poblaciones (funciones de probabilidad) se obtiene el histograma de los coeficientes de frecuencia del escenario B (de actualización). Este es presentado en la figura 23, junto con la distribución de probabilidad (GEV) ajustada a los datos. De la figura en cuestión se desprende que los datos podrían pertenecer a un mismo grupo, puesto que el histograma no presenta, de manera evidente, un comportamiento multi modal.

6.2.2 Del análisis de clúster para definir las regiones

Diversos estudios han utilizado el análisis de clúster como técnica para delimitar o establecer zonas meteorológicamente homogéneas (Rao & Srinivas, 2008; Lin & Chen, 2006), ya sea con la finalidad de trazar envolventes regionales para el cálculo de PMP (Chavan & Srinivas, 2017), inferir eventos hidrológicos en una zona con baja densidad de estaciones (Vera & Mora, 2013) o agrupar individualmente estaciones atípicas (Castro Heredia et al., 2012).

De los estudios revisados para llevar a cabo el análisis regional en Chile (Rao & Srinivas, 2008;Chavan & Srinivas, 2017;Castro Heredia et al., 2012), resulta pertinente ahondar en el

realizado en India (Chavan & Srinivas, 2017), puesto que dicho estudio también tiene como objetivo principal establecer zonas de PMP homogéneas, y construir en ellas envolventes mediante el método de Hershfield. En dicho estudio llama la atención el hecho de que se trabaje con productos grillados. Esto significa que cada una de las observaciones (1951-2012) fue extendida a puntos equiespaciados (con una resolución de 0.25°), mediante la estrategia de "interpolación ponderada por el inverso de la distancia" (Shepard,1986). Lo anterior tiene una gran ventaja práctica, ya que "suaviza" los datos.

Así como se evidencian diferencias en la disposición espacial de la información fuente entre el presente estudio y el realizado en India, también se pueden destacar diferencias en los métodos de agrupación empleados. En el caso del presente estudio, el método de agrupación K-medianas entrega mejores resultados que los sugeridos por Chavan & Srinivas (2017), quienes emplean algoritmos de agrupación Global K-means y Fuzzy C-means. Como atributos del análisis de cluster utilizan tanto parámetros de ubicación geográfica (latitud, longitud, altitud) como parámetros determinantes de la PMP (promedio de precipitaciones máximas anuales), incorporando también un indicador de dependencia estacional (día promedio de ocurrencia de la precipitación máxima anual). A excepción del indicador de dependencia estacional, los parámetros empleados en ambos estudios son similares.

En resumen, se puede concluir que si bien el análisis de clúster puede ser muy útil para determinar patrones y agrupar datos con características similares, este debe ser utilizado como medio exploratorio, puesto que los resultados no deben sustituir nunca lo que la descripción visual muestre.

6.2.3 De la validez de cada método a nivel regional

Según lo expuesto en el estudio realizado en India (Chavan & Srinivas, 2017), el hecho de trabajar con una única envolvente puede llevar a sobreestimar la PMP en un gran número de casos, sobre todo en situaciones en las que las tormentas extremas varían significativamente a lo largo del territorio geográfico o área en estudio.

En las regiones propuestas, no se observa la existencia de relación entre las envolventes trazadas mediante el método de Stowhas y los valores de los máximos coeficientes de frecuencia, obtenidos aplicando el procedimiento de Hershfield, 1961.

No se encontraron publicaciones en las que se estudie el comportamiento o validez regional del método de Hershfield en su primera versión (Hershfield, 1961) que es el que desarrolló Stowhas (1983) en su publicación, y el que se desarrolló, en consecuencia, en el presente estudio. No obstante, varios estudios (Chavan & Srinivas, 2017; Lee & Singh, 2020) han trazado envolventes locales para establecer relaciones entre el máximo coeficiente de frecuencia y los promedios máximos anuales, lo que es equivalente a aplicar localmente el método de Hershfield en su última versión (Hershfield, 1965).

En base a lo anterior es importante reforzar, que si se pretende trazar envolventes con validez regional, estas deberían obtenerse siguiendo lo expuesto por Hershfield en 1965, puesto que ni la primera versión del método (D. Hershfield, 1961) ni el método de Stowhas presentaron resultados regionales concluyentes.

6.3 Incertidumbre en la estimación de la PMP

Cambio climático

Las actividades humanas del último siglo han provocado un aumento de la temperatura global (Cheng & Aghakouchak, 2014), que ha elevado la capacidad de contención de humedad atmosférica en alrededor de un 7% por cada °C, efecto que se ha visto directamente reflejado en las precipitaciones (Kunkel et al., 2013; Trenberth, 2011).

Mayores contenidos de vapor de agua pueden llevar a mayores intensidades de precipitación (Berg et al., 2013). De esta forma, los aumentos de la temperatura y contenidos de humedad en la atmósfera deberían aumentar los valores de la PMP o la precipitación extrema esperada (Kunkel et al., 2013).

Mientras que algunos estudios (Papalexiou & Koutsoyiannis, 2006) cuestionan el límite superior de la precipitación teórica, otros (Lagos, 2012) afirman que las proyecciones de cambio climático cambiarán dichos límites, siendo necesario incorporar los efectos del cambio climático en su estimación (más que cuestionar la existencia de la PMP).

Las proyecciones de cambio climático en Chile (Lagos, 2012) muestran no solo un aumento en la PMP, sino también en la crecida asociada a dicho evento. El aumento esperado en la crecida máxima probable se debe principalmente a un aumento en la temperatura los días de ocurrencia del evento extremo, lo que afecta tanto el contenido de humedad precipitable como la magnitud del área de escorrentía.

Según lo presentado por Lagos (2012), un aumento en la temperatura media implica una elevación de la isoterma cero (o línea de nieves), lo que, en cuencas pluvio nivales, conlleva un aumento del área de captación (área pluvial). A raíz de lo anterior sugiere utilizar, para cuencas de régimen hidrológico mixto, el método hidrometeorológico; debiéndose incorporar (en el caso que exista información de temperatura) el aumento de área producto de la elevación de la isoterma cero. Este último efecto puede ser incorporado mediante un análisis de frecuencia de las relaciones temperatura – altura de la línea de nieves, asignando a la variable aleatoria en cuestión (temperatura) un período de retorno de 1000 años.

No estacionariedad

Se dice que una serie temporal es estacionaria cuando sus estadígrafos se mantienen constantes a lo largo del tiempo, es decir, su media y varianza no presentan una tendencia, ni son función del tiempo.

La relación de no estacionariedad climática en series (de tiempo) de precipitaciones extremas puede asociarse tanto con variabilidad climática como con cambio climático (Zhang et al., 2010), existiendo distintos métodos paramétricos y no paramétricos para abordar dicho efecto (Yilmaz & Perera, 2014)

Dentro de los distintos métodos para identificar la presencia de condiciones de no estacionariedad, uno de los más comunes es el método de los estadísticos móviles. A grandes rasgos, este se basa en determinar los primeros y segundos momentos (media y desviación estándar) de las distribuciones de probabilidad de la muestra bajo ventanas temporales móviles, evaluándose eventuales tendencias en la variación de ellos (D'Arcangeli, 1997; Merabtene, T., 2016). Por otra parte, otros estudios (Gocic, 2013) hacen referencia al ajuste de una recta a la serie de tiempo en cuestión, sugiriendo encontrar

la inclinación (m) y posición (n) de ella a través de la resolución de un problema de optimización. Siguiendo en la misma línea, Ricchetti (2018) sugiere incorporar el efecto de la no estacionariedad en el análisis hidrometeorológico de precipitaciones extremas mediante métodos como el “Change Point Anlaysis”, el “Test de Mann-Kendall” o los Diagramas de Hurst”, los cuales pueden estudiarse con mayor detalle en (Ricchetti, 2018).

Dicho lo anterior, se puede concluir que la determinación de condiciones de no estacionariedad y las posibles tendencias en los estadígrafos de las 879 estaciones del presente estudio, implicaría, por un lado, la agrupación de ellas a nivel de cuenca o zona meteorológicamente homogénea, y por otro, posiblemente, el relleno estadístico de la información contenida en ellas (las estadísticas se encuentran incompletas).

Si bien el cuestionamiento de la hipótesis de estacionariedad climática abre una nueva línea investigativa y sugiere estudiar posibles tendencias en la información fuente, múltiples estudios realizados en chile (D’Arcangeli, 1997; Falvey & Garreaud, 2009; Garreaud et al., 2017; Franco, 2018) ya han enfocado sus esfuerzos en demostrar lo anterior.

D’Arcangeli (1997) es uno de los primeros en describir un cambio considerable en algunas variables hidrometeorológicas de Chile. Años más tarde un estudio realizado en el departamento de Geofísica de la Universidad de Chile (DGF, 2006) evidencia, para un gran número de estaciones en Chile central, tendencias en las series de precipitaciones. En esta publicación, se constata una evolución de las tendencias de la media móvil por latitud y tiempo, acentuándose las tendencias a la baja en los últimos años.

Un estudio de tendencia realizado a las series de temperaturas de los últimos 25 años de estaciones chilenas y argentinas (Garreaud et al. ,2017) ha reportado los impactos de la Mega Sequía que ha afectado a nuestro país entre los años 2010 a 2015, entendiendo que el fenómeno no tiene precedentes y que los cambios del clima han intensificado, tanto en magnitud como frecuencia, las sequías en Chile.

Lo anteriormente mencionado es consistente con lo recientemente expuesto por Lagos et al. (2021), quienes han mostrado evidencia respecto a la falta de homogeneidad en las series de precipitación de todo el país, indicando que el supuesto intrínseco de estacionariedad en

diseño hidrológico de obras debe ser revisado. En el estudio en cuestión se aprecia que en sectores costeros del norte y centro del país existe un incremento en el promedio de las precipitaciones máximas diarias anuales, al igual que en algunas estaciones del Altiplano y gran parte de las estaciones la macrozona Austral. En términos de desviación estándar, un 44% de la población muestra tendencias significativas de cambio (para períodos de media móvil de 10 años), concentrándose los mayores aumentos en las macrozona sur y Austral.

Las características no estacionarias en series de precipitaciones han sido evidenciadas tanto en chile como en el resto del mundo, dando cabida a estudios que cuantifican la magnitud de dicho efecto.

Cheng & Aghakouchak (2014) evaluó el efecto de la no estacionariedad tanto en las curvas IDF como en los eventos extremos, evidenciando que asumir la hipótesis de no estacionariedad puede generar subestimaciones sustanciales. En el estudio en cuestión se aprecia que el efecto de considerar condiciones no-estacionarias aumenta en las tormentas de menor duración, lo que implica, por un lado, que los eventos cortos (1 a 3 hrs) se vean intensificados considerablemente; y por otro, que se pueda despreciar este efecto en eventos de larga duración (48 hrs en adelante).

El estudio (Cheng & Aghakouchak, 2014) evidencia además que la incertidumbre aumenta con el período de retorno, independientemente de la duración de la tormenta. Esto tiene como consecuencia que el efecto de no considerar la no-estacionariedad aumente en mayor medida en eventos extremos, cuyo período de retorno es siempre mayor. En dicho contexto, asumir la hipótesis de estacionalidad climática puede llevar a subestimaciones de los valores extremos de hasta un 60%, lo que incrementa el riesgo de inundaciones y el riesgo de falla en los sistemas de infraestructura hidráulica (Cheng & Aghakouchak, 2014)

Incertidumbre

Un estudio realizado en la cuenca del río Brazos (Texas, EEUU) intentó cuantificar la incertidumbre asociada a la estimación de la PMP (Lee & Singh, 2020), evaluando el efecto que tienen en la estimación factores como la topografía, la no estacionariedad de los datos, o el método de estimación empleado.

Del estudio se concluyó que la incertidumbre en la estimación de la PMP es más sensible a las estadísticas existentes y a la selección del método, que, a las diferencias climáticas de una determinada zona, dentro de una misma cuenca.

El estudio anterior, si bien refuerza el efecto que tiene la no consideración de la no-estacionariedad de los datos en la estimación de la PMP, no permite extrapolar a otros estudios la baja incidencia de la topografía en las estimaciones, puesto que ello responde a las características hidrológicas de esa cuenca en particular.

A través del centro de documentación del "servicio de evaluación de impacto ambiental" se accedió a algunas memorias de cálculo de los embalses mas importantes de Chile, evidenciándose que, en la gran mayoría de ellos se ha estimado la PMP aplicando análisis de frecuencia tradicional, a series de duración completa, con períodos de retorno de 10.000 años. Lo anterior deja en evidencia, que, pese a que los métodos estudiados poseen un sin número de inconsistencias y fuentes de error, siguen siendo más adecuados que lo que mayoritariamente se utiliza en Chile, hasta el día de hoy.

7 Conclusiones

El estudio de reanálisis y actualización de métodos para estimar la PMP en Chile consideró tanto la evaluación del comportamiento de los métodos actualizados con las estadísticas actuales, como la valoración de ellos sobre zonas geográficas más acotadas (regiones). En el proceso se investigó que dicen los estudios respecto de los distintos factores que pueden inducir incertidumbre y error en la estimación de la PMP, concluyéndose, en base a ellos, que estos provienen principalmente del no cumplimiento de las dos hipótesis fundamentales de los métodos empleados, que dicen relación con la homogeneidad meteorológica de la zona en la que se aplican los métodos, y comportamiento estacionario de las estadísticas.

Habiendo estimado la PMP en la totalidad de las estaciones meteorológicas en estudio bajo los distintos métodos, independientemente de si estos hayan sido actualizados o no; se puede concluir que el método de Hershfield,1961 entrega en Chile resultados mayores en al menos un 80% de los casos, siendo esto concordante con lo expuesto en estudios anteriores (Stowhas, 1983). A su vez, se concluye que la primera versión del método de Hershfield (1961) tiende a subestimar la PMP en aquellas zonas donde las precipitaciones predominantes son del tipo nival, sugiriéndose no utilizarlo en este tipo de situaciones.

En el presente estudio se evidencia que el máximo coeficiente de frecuencia es tan sensible a la longitud estadística como a las estaciones que se seleccionen para el análisis. Así como las 490 estaciones utilizadas para actualizar el estudio de Stowhas (1983) presentaron un coeficiente máximo de 9.67, las 478 y 600 estaciones utilizadas para las distintas alternativas de zonificación entregaron coeficientes de frecuencia 10.3 y 12.4, respectivamente. De los resultados anteriores se concluye que no existe una relación de dependencia entre la cantidad de estaciones empleadas y el valor límite del máximo coeficiente de frecuencia. No obstante, resulta evidente que las estaciones con precipitaciones predominantemente nivales (estaciones del norte cordillerano, en altura, o en la región de magallanes) tienden a aumentar el valor del máximo coeficiente de frecuencia. El mismo efecto se aprecia en las envolventes de Stowhas, concluyéndose, en consecuencia, que la consideración de las estaciones con régimen nival en el análisis tienden a producir, en términos generales, una sobre- estimación de la PMP.

Existe consenso en que la seguridad e integridad de la infraestructura hidráulica durante precipitaciones extremas está estrechamente relacionada con la seguridad y salud de la población, aguas abajo de las principales obras de almacenamiento. En vista de lo anterior, a pesar de que los métodos actualizados con las nuevas estadísticas sugieren que la PMP tendería a disminuir, la no consideración de factores como la no estacionariedad estadística, en un inminente contexto de cambio climático, obligan a ser criteriosos al reducir los coeficientes de los distintos métodos

Al actualizar el valor del máximo coeficiente de frecuencia replicando las condiciones establecidas por Stowhas en 1983, se observa que este se ve reducido de **11 a 9.67**. Si bien el nuevo valor se sustenta en 2.6 veces más estaciones meteorológicas y en casi 3 veces más años de registro, pareciera este último no ser el más apto para la realidad Chilena. Lo anterior se debe a que se contemplaron en análisis estaciones con una gran componente nival, aun cuando se eliminaron las estaciones del norte grande.

En consecuencia, se sugiere utilizar, para estimar la PMP (en 24 horas) mediante el método de Hershfield, un coeficiente de frecuencia actualizado con el valor de **Km= 10.3**, habiéndose obtenido este excluyendo del análisis las estaciones con comportamiento nival (no solo las estaciones en altura y del norte grande).

De la actualización del método alternativo propuesto por Stowhas se puede concluir que las envolventes dobles también se ven reducidas. No obstante, es importante comentar que el hecho que los coeficientes y las envolventes disminuyan no garantiza en lo más mínimo que la PMP en un determinado lugar disminuya, porque esta es función también del promedio de los máximos anuales, y de las varianzas de estos.

Si bien en el estudio se probaron distintas técnicas y medidas de asociación para agrupar las estaciones y establecer así regiones homogéneas, se puede concluir que las soluciones propuestas, en el contexto de la PMP, son insatisfactorias. Lo anterior dado que en ambas soluciones existe al menos una región en la que los coeficientes de frecuencia se disparan, y no se dispone de información suficiente como para validar dichos coeficientes.

Con base en la revisión bibliográfica y en las múltiples iteraciones, se sugiere continuar en la línea del trabajo realizado, proponiéndose:

1.- Aplicar la metodología de Hershfield,1965 a las regiones ya determinadas mediante el análisis de clúster, evaluando la variación de dichos resultados respecto a los propuestos en el estudio.

2.- Reevaluar las regiones establecidas mediante análisis de clúster, utilizando las estadísticas grilladas presentadas por el "(CR)2: Center of Climate and Resilience Research"; lo que a su vez implicaría chequear previamente que los productos grillados presentados mantienen la media y varianza de los datos observados.

3.- Evaluar la opción de complementar la base de datos generada con estadísticas privadas, como compañías generadoras eléctricas (Endesa, Colbún, etc) y mineras, entre otras.

4.- Evaluar los coeficientes y envolventes estableciendo regiones de manera manual, ya sea con base en mapas climáticos, características de vegetación, temperatura, humedad, clima estacional, etc; poniendo énfasis en que el método de Hershfield es más sensible a las estaciones meteorológicas utilizadas en el trazado de la envolvente, que a la longitud de las estadísticas empleadas.

5.- Estudiar el comportamiento no estacionario de las estadísticas presentadas, y evaluar o proponer una forma sencilla de incorporar este efecto en la estimación de la PMP.

8 Referencias

Amat, J. R. (2017). *Clustering & heatmaps : aprendizaje no supervisado*. https://www.cienciadedatos.net/documentos/37_clustering_y_heatmaps

Beauchamp, J., Leconte, R., Trudel, M., & Brissette, F. (2013). Estimation of the summer-fall PMP and PMF of a northern watershed under a changed climate: PMP and PMF Estimation Under Climate Change. *Water Resour. Res.* https://doi.org/10.1002/wrcr.20336

Berg, P., Moseley, C., & Haerter, J. O. (2013). Strong increase in convective precipitation in response to higher temperatures. *Nature Geoscience*, *6*(3), 181–185. https://doi.org/10.1038/ngeo1731

Casas, M. C., Rodríguez, R., Prohom, M., Gázquez, A., & Redaño, A. (2011). Estimation of the probable maximum precipitation in Barcelona (Spain). *International Journal of Climatology*, *31*(9). https://doi.org/10.1002/joc.2149

Castro Heredia, L. M., Carvajar Escpbar, Y., & Ávila Díaz, Á. (2012). Análisis Clúster como Técnica de Análisis Exploratorio De Registros Múltiples En Datos Meteorológicos. *Ingeniería de Recursos Naturales y Del Ambiente*, *11*, 11–20. https://www.redalyc.org/articulo.oa?id=231125817001

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

Cheng, L., & Aghakouchak, A. (2014). Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. *Scientific Reports*, *4*, 1–6. https://doi.org/10.1038/srep07093

Chow, V. Te. (1951). A general formula for hydrologic frequency analysis. *Eos, Transactions American Geophysical Union*, *32*(2). https://doi.org/10.1029/TR032i002p00231

D'Arcangeli, V. R. (1997). Variación temporal de algunos factores meteorológicos en Chile. *Memoria Para Optar Al Título de Ingeniero Civil, Departamento de Ingeiería Civil, Universidad de Chile*.

DGF, U. de C. (2006). Estudio de la variabilidad climática en Chile para el siglo XXI.

Informe Para CONAMA, 63.

Falvey, M., & Garreaud, R. N. (2009). Regional cooling in a warming world: Recent temperature trends in the southeast Pacific and allong the west coast of subtropical South America. *Journal of Geophysical Research*. https://doi.org/D04102

Garreaud, R., Alvarez-Garreton, C., JBarichivich, J. ., Boisier, D., Christie, M., Galleguillos, C., LeQuesne, C., McPhee, J., & Zambrano-Bigiarini, M. (2017). The 2010-2015 mega drought in Central Chile: Impacts on regional hydroclimate and vegetation. *Hydrology and Earth System Sciences, EGU*.

Gocic, M. & S. T. (2013). Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical test in Serbia. *Global and Planetary Change*, *100*(172–182).

Hershfield, D. (1961). Estimating the Probable Maximum Precipitation. *Journal of the Hydraulics Division*, *87*(5).

Hershfield, D. M. (1965). Method for Estimating Probable Maximum Rainfall. *Journal - American Water Works Association*, *57*(8). https://doi.org/10.1002/j.1551-8833.1965.tb01486.x

Hiraga, Y., Iseri, Y., Warner, M. D., Frans, C. D., Duren, A. M., England, J. F., & Kavvas, M. L. (2021). Estimation of Long-duration Maximum Precipitation during a winter season for large basins dominated by Atmospheric Rivers using a Numerical Weather Model. *Journal of Hydrology*, *598*, 126224. https://doi.org/10.1016/j.jhydrol.2021.126224

Jennings, A. H. (1950). WORLD'S GREATEST OBSERVED POINT RAINFALLS. *Monthly Weather Review*, *78*(1). https://doi.org/10.1175/1520-0493(1950)078<0004:wgopr>2.0.co;2

Kunkel, K. E., Karl, T. R., Easterling, D. R., Redmond, K., Young, J., Yin, X., & Hennon, P. (2013). Probable maximum precipitation and climate change. *Geophysical Research Letters*, *40*(7), 1402–1408. https://doi.org/10.1002/grl.50334

Lagos, M. (2012). Impacto del cambio climático en eventos extremos y análisis de la vulnerabilidad de algunas obras hidráulicas en Chile. In X. Vargas (Ed.), *Memoria para optar al título de ingeniero civil*.

Lagos, M., Pinto, D., Torrez, L., & Rondanelli, R. (2021). ¿Son las precipitaciones

extremas estacionarias en Chile?. *XXV Congreso Chileno de Ingeniería Hidráulica, October*.

Lee, K., & Singh, V. P. (2020). Analysis of uncertainty and non-stationarity in probable maximum precipitation in Brazos River basin. *Journal of Hydrology*, *590*(April), 125526. https://doi.org/10.1016/j.jhydrol.2020.125526

Lin, G. F., & Chen, L. H. (2006). Identification of homogeneous regions for regional frequency analysis using the self-organizing map. *Journal of Hydrology*, *324*(1–4). https://doi.org/10.1016/j.jhydrol.2005.09.009

Merabtene, T., M. S. & A. S. (2016). Assessment of seasonal and anual rainfall trends and variability in Sharjah city, UAE. *Hindawi Publishing Corporation, Advances in Meteorology*. https://doi.org/6206238

Papalexiou, S. M., & Koutsoyiannis, D. (2006). A probabilistic approach to the concept of Probable Maximum Precipitation. *Advances in Geosciences*, *7*. https://doi.org/10.5194/adgeo-7-51-2006

Paulhus, J. L. H., & Gilman, C. S. (1953). Evaluation of probable maximum precipitation. *Eos, Transactions American Geophysical Union*. https://doi.org/10.1029/TR034i005p00701

Rao, A. R., & Srinivas, V. (2008). Regionalization of Watersheds by hybrid-cluster analysis. *Journal of Hydrology*, *318*(December), 37–56. https://doi.org/10.1007/978-1-4020-6852-2

Ricchetti, F. (2018). Incorporación del efecto de la no estacionariedad en el análisis de series hidrometeorológicas de precipitación diaria y temperatura media diaria; valores extremos y mensuales en Chile central. *Memoria Para Optar Al Título de Ingeniero Civil, Universidad de Chile*.

Schreiner, L. C., & Riedel, J. T. (1978). Probable maximum precipitation estimates, United States east of the 105th meridian. *Department of Commerce, National Oceanic and Atmospheric Administration.*

Stowhas, L. (1983). Precipitaciones máximas diarias en Chile. *VI Congreso Nacional SOCHID*, 455–473.

Stowhas, L. (2016). *Fundamentos de Hidrologia Aplicada* (R. F. A. & A. O. A. (eds.)). Editorial USM.

Toride, K., Iseri, Y., Warner, M. D., Frans, C. D., Duren, A. M., England, J. F., & Kavvas, M. L. (2019). Model-based probable maximum precipitation estimation: How to estimate the worst-case scenario induced by atmospheric rivers? *Journal of Hydrometeorology*, *20*(12), 2383–2400. https://doi.org/10.1175/JHM-D-19-0039.1

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, *47*(1–2). https://doi.org/10.3354/cr00953

Vera, J. A. L., & Mora, R. D. (2013). Un método para el análisis de frecuencia regional de lluvias máximas diarias: Aplicación en los Andes Bolivianos. *Ingeniare*, *21*(1), 111–124. https://doi.org/10.4067/s0718-33052013000100010

World Meteorological Organization. (1986). *22. Manual for Estimation of Probable Maximum Precipitation, Operational Hydrology Report 1, 2nd edition, Publication 332* (Issue 1045).

Yilmaz, A. G., & Perera, B. J. C. (2014). Extreme Rainfall Nonstationarity Investigation and Intensity–Frequency–Duration Relationship. *Journal of Hydrologic Engineering*, *19*(6). https://doi.org/10.1061/(asce)he.1943-5584.0000878

Zhang, X., Wang, J., Zwiers, F. W., & Groisman, P. Y. (2010). The influence of large-scale climate variability on winter maximum daily precipitation over North America. *Journal of Climate*, *23*(11). https://doi.org/10.1175/2010JCLI3249.1

9 Anexo A: Códigos

9.1 Python

9.1.1 Filtrado y depurado de la BBDD fuente

```
# 1.- Importar librerias
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib as plt
import pylab as pl

#2.- Abrir BBDD original
path='cr2_prDaily_2020.txt'
base0 = pd.read_csv(path)

#3.- Generar BBDD auxiliar que contenga la información que se eliminará
base_info=base0.iloc[0:14,:].copy()
base_info2=base_info.T
base_info2.columns=['institucion','fuente','nombre','altura','latitud','l
ongitud','codigo_cuenca','nombre_cuenca','codigo_sub_cuenca','nombre_sub_
cuenca','inicio_observaciones','fin_observaciones','cantidad_observacione
s','inicio_automatica']
base_info2=base_info2.drop(['codigo_estacion'],axis=0)
base_info2.to_csv('BBDDinfo_depurada.csv',index=True,header=True)

#Generación de serie de tiempo
listado_fechas=pd.date_range(start='1903/01/01',end='2020/06/06',freq='D'
)

#4.- Generar BBDD que contenga información depurada
base1=base0.drop(list(range(0,14)),axis=0)
base1=base0.drop(list(range(0,1109)),axis=0
base1['fechas']=listado_fechas
base1=base1.set_index('fechas')
base1=base1.drop('codigo_estacion',axis=1)
base_filtrada=base1.copy()
base_filtrada.to_csv('BBDDdepurada2.csv',index=True,header=True)
```

9.1.2 Generación de estadísticos para recalibración de M. Hershfield

Escenario A

```
# 1.- Cargar librerias
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib as plt
import pylab as pl

#2.- Abrir BBDD de la información de las estaciones
path1='BBDDinfo_depurada.csv'
info_est = pd.read_csv(path1,index_col='Unnamed: 0')
path2='BBDDdepurada2.csv'
Pmax_dia = pd.read_csv(path2, parse_dates=['fechas'], index_col='fechas')

#Asignar codigos de estaciones como STR a indices del DF con información
codigos=Pmax_dia.columns.tolist()
info_est['Cod Estacion']=codigos
info_est=info_est.set_index('Cod Estacion')

#2.1- Genera BBDD con ubicación de las estaciones para graficar en matlab
ubicacion=pd.concat([info_est['latitud'],info_est['longitud']],axis=1)
ubicacion.to_csv('BBDD_ubicacion.csv',index=True,header=True)

#2.2- Identificar estaciones del Norte grande
Lat=-28.3
ubicacion2=ubicacion.T
nomb_est=ubicacion2.columns.tolist()
norte_grande=list()
for estacion in nomb_est:
    if ubicacion2[estacion]['latitud']>=Lat:
        norte_grande.append(estacion)

#Genera un DF CON Pmax ANUAL
Pmax_anual=Pmax_dia.resample('Y').max()
Pmax_anualA=Pmax_anual['1903/01/01':'1983/12/31'].copy()
```

```
#2.3- Identifica estaciones con mediciones nulas
nomb_est=Pmax_anualA.columns.tolist()
est_nulas=list()
for estacion in nomb_est:
    if Pmax_anualA[estacion].sum()==0:
        est_nulas.append(estacion)

#2.4- Eliminar Estaciones según criterios establecidos por Stowhas, 1983
est_def=set(est_nulas+est_1med+norte_grande)
nomb_est=Pmax_anualA.columns.tolist()
base2=Pmax_anualA.copy()
for estacion in est_def:
    base2=base2.drop(estacion,axis=1)

#3.- Análisis estadístico  a BBDD Con Pmáx anuales
promedios=base2.mean()
desvest=base2.std()
cv=desvest/promedios
km=(base2-promedios)/desvest
km_max=km.max()
pmaxadim=(base2.max()-base2.mean())/base2.mean()

#4.-Consolidar una BBDD con las estadísticas
estadisticas=pd.concat([km_max,promedios,desvest,cv,pmaxadim],axis=1)
estadisticas.columns=['Km_max','Prom','Desvest','CV','Pm_adim']
estadisticas.to_csv('BBDD_estadisticosA2.csv',index=True,header=True)
```

Escenario B

```
# 1.- Cargar librerias
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib as plt
import pylab as pl

#Generar Matriz con coordenadas de las estaciones
#2.- Abrir BBDD de la información de las estaciones
path1='BBDDinfo_depurada.csv'
info_est = pd.read_csv(path1,index_col='Unnamed: 0')
path2='BBDDdepurada2.csv'
Pmax_dia = pd.read_csv(path2, parse_dates=['fechas'], index_col='fechas')

#Reemplazo longitud isla de pascua
info_est=info_est.replace(info_est['longitud'].min(),info_est['longitud']
.mean())

#Asignar codigos de estaciones como STR a indices del DF con información
codigos=Pmax_dia.columns.tolist()
info_est['Cod Estacion']=codigos
info_est=info_est.set_index('Cod Estacion')

#Genera un DF CON Pmax ANUAL
Pmax_anual=Pmax_dia.resample('Y').max()
Pmax_anualB=Pmax_anual['1903/01/01':'2020'].copy()

#3.- Identifica estaciones que no se consideran en el análisis
#3.1- Identificar estaciones del Norte grande
Lat=-28.3
latitud=info_est['latitud']
norte_grande=list()
for est in codigos:
    if latitud[est]>=Lat:
        norte_grande.append(est)
#print(norte_grande)
print('Est NG en B:',len(norte_grande))
```

```
#3.2- Identifica estaciones con mediciones nulas
est_nulas=list()
for est in codigos:
    if Pmax_anualB[est].sum()==0:
        est_nulas.append(est)
#print(est_nulas)
print('Est nulas en B:',len(est_nulas))

#3.3- Identificar estaciones con T_min años de medición
T_min=15
est_1med=list()
for est in codigos:
    if Pmax_anualB[est].count()<T_min:
        est_1med.append(est)
print('Est t<15 en B:',len(est_1med))

print('Est en B:',879-len(set(norte_grande+est_nulas+est_1med)))
print('Est_def en B:',len(set(norte_grande+est_nulas+est_1med)))

#4.- Eliminar Estaciones con mediciones deficientes de la BBDDmax_anuales
est_def=set(est_nulas+est_1med+norte_grande)
est_def_AC=set(est_nulas+est_1med)
nomb_est=Pmax_anualB.columns.tolist()
base2=Pmax_anualB.copy()

#5.- Borra del DF la estaciones que no cumplen los filtros NG+med0+t<15
for est in est_def:
    base2=base2.drop(est,axis=1)
base2.shape

#6.- Análisis estadístico a BBDD Con Pmáx anuales
pmax=base2.max()
promedios=base2.mean()
desvest=base2.std()
cv=desvest/promedios
km=(base2-promedios)/desvest
km_max=km.max()
```

```
pmaxadim=(base2.max()-base2.mean())/base2.mean()

#7.- Consolidar una BBDD con las estadísticas
estadisticas=pd.concat([km_max,pmax,promedios,desvest,cv,pmaxadim],axis=1
)
estadisticas.columns=['Km_max','Pmax','Prom','Desvest','CV','Pm_adim']
estadisticas.to_csv('BBDD_estadisticosB2.csv',index=True,header=True)
```

Escenario A2

```
# 1.- Cargar librerias
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib as plt
import pylab as pl

#Generar Matriz con coordenadas de las estaciones
#2.- Abrir BBDD de la información de las estaciones
path1='BBDDinfo_depurada.csv'
info_est = pd.read_csv(path1,index_col='Unnamed: 0')
path2='BBDDdepurada2.csv'
Pmax_dia = pd.read_csv(path2, parse_dates=['fechas'], index_col='fechas')

#Reemplazo longitud isla de pascua
info_est=info_est.replace(info_est['longitud'].min(),info_est['longitud']
.mean())

#Asignar codigos de estaciones como STR a indices del DF con información
codigos=Pmax_dia.columns.tolist()
info_est['Cod Estacion']=codigos
info_est=info_est.set_index('Cod Estacion')

#3.-Genera un DF CON Pmax ANUAL
Pmax_anual=Pmax_dia.resample('Y').max()
Pmax_anualC=Pmax_anual['1903/01/01':'2020'].copy()
Pmax_anualA=Pmax_anual['1903/01/01':'1983'].copy()


#4- Identificar estaciones de escenario A
Lat=-28.3
latitud=info_est['latitud']
norte_grande=list()
for est in codigos:
    if latitud[est]>=Lat:
        norte_grande.append(est)
len(norte_grande)
```

```
print('Est NG en A:',len(norte_grande))

est_nulasA=list()
for est in codigos:
    if Pmax_anualA[est].sum()==0:
        est_nulasA.append(est)
print('Est nulas en A:',len(est_nulasA))

T_min=15
est_1medA=list()
for est in códigos:
    if Pmax_anualA[est].count()<T_min:
        est_1medA.append(est)
print('Est t<15 en A:',len(est_1medA))

print('Est en A:',879-len(set(norte_grande+est_nulasA+est_1medA)))
est_def_A=set(norte_grande+est_nulasA+est_1medA)


#5.- Verificar que estaciones en A2 sean las mismas que en A
print('Est NG en C:',len(norte_grande))

est_nulasC=list()
for est in codigos:
    if Pmax_anualC[est].sum()==0:
        est_nulasC.append(est)
print('Est nulas en C:',len(est_nulasC))

T_min=15
est_1medC=list()
for est in codigos:
    if Pmax_anualC[est].count()<T_min:
        est_1medC.append(est)
print('Est t<15 en C:',len(est_1medC))
print('Est en C:',879-len(set(norte_grande+est_nulasC+est_1medC)))

#6.- Eliminar Estaciones que no se consideran en el analisis
nomb_est=Pmax_anualC.columns.tolist()
```

```
base2=Pmax_anualC.copy()
for est in est_def_A:
    base2=base2.drop(est,axis=1)
base2.shape

#7.- Análisis estadístico  a BBDD Con Pmáx anuales
pmax=base2.max()
promedios=base2.mean()
desvest=base2.std()
km=(base2-promedios)/desvest
km_max=km.max()
pmaxadim=(base2.max()-base2.mean())/base2.mean()

#8.- Consolidar una BBDD con las estadísticas
estadisticas=pd.concat([km_max,pmax,promedios,desvest,cv,pmaxadim],axis=1
)
estadisticas.columns=['Km_max','Pmax','Prom','Desvest','CV','Pm_adim']
estadisticas.to_csv('BBDD_estadisticosC.csv',index=True,header=True)
```

9.1.3 Generación de matrices con parámetros para análisis de cluster

Solución 1

```
# 1.- Cargar librerías
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib as plt
import pylab as pl

#Generar Matriz con coordenadas de las estaciones
#2.- Abrir BBDD de la información de las estaciones
path1='BBDDinfo_depurada.csv'
info_est = pd.read_csv(path1,index_col='Unnamed: 0')
path2='BBDDdepurada2.csv'
Pmax_dia = pd.read_csv(path2, parse_dates=['fechas'], index_col='fechas')

#Reemplazo longitud isla de pascua
info_est=info_est.replace(info_est['longitud'].min(),info_est['longitud']
.mean())

#Asignar codigos de estaciones como STR a indices del DF con información
codigos=Pmax_dia.columns.tolist()
info_est['Cod Estacion']=codigos
info_est=info_est.set_index('Cod Estacion')

#Cálculo del módulo pluviométrico
Pacum_mensual=Pmax_dia.resample('M').sum()           #DF que almacena las
precipitaciones acumuladas mensuales
Pacum_anual=Pacum_mensual.resample('Y').sum()        #DF que almacena las
precipitaciones acumuladas anuales
mod_pluv=list()
completo=list()
dias_med=list()
for i in range(len(codigos)):
      inicio=datetime.strptime(info_est['inicio_observaciones'][str(codig
os[i])], '%Y-%m-%d')
```

```
    fin=datetime.strptime(info_est['fin_observaciones'][str(codigos[i])],
'%Y-%m-%d')
    comp=Pmax_dia[inicio:fin][str(codigos[i])].count()/(fin-
inicio).days*100
    mod=Pacum_anual[inicio:fin][str(codigos[i])].mean()
    dias_med.append((fin-inicio).days)
    completo.append(comp)
    mod_pluv.append(mod)
info_est['Modulo_pluv']=mod_pluv
info_est['Fraccion_completa']=completo


#2.2- Elimina Est Cluster Norte Grande y estaciones en altura
Lat=-25                                                  #Latitud límite
Alt1=1500
Alt2=3000
latitud=info_est['latitud']
altura=info_est['altura']
norte_grande1=list()
norte_grande2=list()
for est in codigos:
    if latitud[est]>=Lat and altura[est]>=Alt1:
        norte_grande1.append(est)

for est in codigos:
    if altura[est]>=Alt2:
        norte_grande2.append(est)
norte_grande=set(norte_grande1+norte_grande2)

Lat_min=-50
patagonia=list()
for est in códigos:
    if latitud[est]<=Lat_min:
        patagonia.append(est)

print('C1- Estaciones Lat > -25° y alt >1500m :',len(norte_grande1))
print('C2- Estaciones alt >3000m              :',len(norte_grande2))
print('(C1+C2) Estaciones a borrar            :',len(norte_grande))
```

```
print('C3- Estaciones Patagonia              :',len(patagonia))

#Genera un DF CON Pmax ANUAL
Pmax_anual=Pmax_dia.resample('Y').max()
Pmax_anualB=Pmax_anual['1903/01/01':'2020'].copy()

#Identifica estaciones con mediciones nulas
est_nulas=list()
for est in codigos:
    if Pmax_anualB[est].sum()==0:
        est_nulas.append(est)
print('C4- Est nulas en B2:',len(est_nulas))

#Identificar estaciones con T_min años de medición
T_min=15
est_1med=list()
for est in codigos:
    if Pmax_anualB[est].count()<T_min:
        est_1med.append(est)
print('C5- Est t<15 en B2 :',len(est_1med))
print('Est_defB2                                        set(C1+C2+C3+C4+C5)
:',len(set(est_1med+est_nulas+norte_grande1+norte_grande2+patagonia)))
print('Est  B2                                                    :',879-
len(set(est_1med+est_nulas+norte_grande1+norte_grande2+patagonia)))

#Eliminar Estaciones con mediciones deficientes de la BBDDmax_anuales
est_def=set(est_nulas+est_1med+norte_grande1+norte_grande2+patagonia)
nomb_est=Pmax_anualB.columns.tolist()
base2=Pmax_anualB.copy()
for est in est_def:
    base2=base2.drop(est,axis=1)

#Crea DF con info_cluster y elimina est def
info_cluster=pd.concat([info_est['Modulo_pluv'],info_est['Fraccion_comple
ta'],info_est['latitud'],info_est['longitud'],info_est['altura']],axis=1)
for est in est_def:
    info_cluster=info_cluster.drop(est,axis=0)
```

```
#Análisis estadístico  a BBDD Con Pmáx anuale
pmax=base2.max()
promedios=base2.mean()
desvest=base2.std()
cv=desvest/promedios
km=(base2-promedios)/desvest
km_max=km.max()
pmaxadim=(base2.max()-base2.mean())/base2.mean()

#Consolidar una BBDD con las estadísticas
estadisticas=pd.concat([km_max,pmax,promedios,desvest,cv,pmaxadim],axis=1
)
estadisticas.columns=['Km_max','Pmax','Prom','Desvest','CV','Pm_adim']
estadisticas.to_csv('BBDD_estadisticos_AC_sol1.csv',index=True,header=Tru
e)
analisis_cluster=pd.concat([estadisticas,info_cluster],axis=1)
analisis_cluster.to_csv('BBDDcluster_AC_sol1.csv',index=True,header=True)

#Estandarización de los parámetros
#Generar Matriz con coordenadas de las estaciones
#10.- Abrir BBDD de la información de las estaciones
path='BBDDcluster_AC_sol1.csv'
parametros = pd.read_csv(path,index_col='Cod Estacion')

#11.- Estandarización
Km_max_std2=(parametros['Km_max']-
parametros['Km_max'].mean())/parametros['Km_max'].std()
Pmax_std2=(parametros['Pmax']-
parametros['Pmax'].mean())/parametros['Pmax'].std()
Prom_std2=(parametros['Prom']-
parametros['Prom'].mean())/parametros['Prom'].std()
Desvest_std2=(parametros['Desvest']-
parametros['Desvest'].mean())/parametros['Desvest'].std()
CV_std2=(parametros['CV']-parametros['CV'].mean())/parametros['CV'].std()
Pm_adim_std2=(parametros['Pm_adim']-
parametros['Pm_adim'].mean())/parametros['Pm_adim'].std()
Modulo_adim_std2=(parametros['Modulo_pluv']-
parametros['Modulo_pluv'].mean())/parametros['Modulo_pluv'].std()
```

```
latitud_std2=(parametros['latitud']-
parametros['latitud'].mean())/parametros['latitud'].std()
longitud_std2=(parametros['longitud']-
parametros['longitud'].mean())/parametros['longitud'].std()
altura_std2=(parametros['altura']-
parametros['altura'].mean())/parametros['altura'].std()

#12.- Guardar datos estandarizados
parametros_estandarizados2=pd.concat([Km_max_std2,Pmax_std2,Prom_std2,Des
vest_std2,CV_std2,Pm_adim_std2,Modulo_adim_std2,latitud_std2,longitud_std
2,altura_std2],axis=1)
parametros_estandarizados2.to_csv('cluster_std2_sol1.csv',index=True,head
er=True)
```

Solución 2

```
# 1.- Cargar librerias
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib as plt
import pylab as pl

#Generar Matriz con coordenadas de las estaciones
#2.- Abrir BBDD de la información de las estaciones
path1='BBDDinfo_depurada.csv'
info_est = pd.read_csv(path1,index_col='Unnamed: 0')
path2='BBDDdepurada2.csv'
Pmax_dia = pd.read_csv(path2, parse_dates=['fechas'], index_col='fechas')

#Reemplazo longitud isla de pascua
info_est=info_est.replace(info_est['longitud'].min(),info_est['longitud']
.mean())

#Asignar codigos de estaciones como STR a indices del DF con información
codigos=Pmax_dia.columns.tolist()
info_est['Cod Estacion']=codigos
info_est=info_est.set_index('Cod Estacion')
```

```
#Cálculo del módulo pluviométrico
Pacum_mensual=Pmax_dia.resample('M').sum()           #DF que almacena las
precipitaciones acumuladas mensuales
Pacum_anual=Pacum_mensual.resample('Y').sum()        #DF que almacena las
precipitaciones acumuladas anuales

mod_pluv=list()
completo=list()
dias_med=list()

for i in range(len(codigos)):
      inicio=datetime.strptime(info_est['inicio_observaciones'][str(codig
os[i])], '%Y-%m-%d')
    fin=datetime.strptime(info_est['fin_observaciones'][str(codigos[i])],
'%Y-%m-%d')
    comp=Pmax_dia[inicio:fin][str(codigos[i])].count()/(fin-
inicio).days*100
    mod=Pacum_anual[inicio:fin][str(codigos[i])].mean()
    dias_med.append((fin-inicio).days)
    completo.append(comp)
    mod_pluv.append(mod)
info_est['Modulo_pluv']=mod_pluv
info_est['Fraccion_completa']=completo

#Genera un DF CON Pmax ANUAL
Pmax_anual=Pmax_dia.resample('Y').max()
Pmax_anualB=Pmax_anual['1903/01/01':'2020'].copy()

#Identifica estaciones con mediciones nulas
est_nulas=list()
for est in codigos:
    if Pmax_anualB[est].sum()==0:
        est_nulas.append(est)
print('C1- Est nulas en AC   :',len(est_nulas))

#Identificar estaciones con T_min años de medición
T_min=15
```

```
est_1med=list()
for est in codigos:
    if Pmax_anualB[est].count()<T_min:
        est_1med.append(est)
print('C2- Est t<15 en AC    :',len(est_1med))
print('Est_def_AC set(C1+C2) :',len(set(est_1med+est_nulas)))
print('Est AC                :',879-len(set(est_1med+est_nulas)))

#Eliminar Estaciones con mediciones deficientes de la BBDDmax_anuales
est_def=set(est_nulas+est_1med)
nomb_est=Pmax_anualB.columns.tolist()
base2=Pmax_anualB.copy()
for est in est_def:
    base2=base2.drop(est,axis=1)

#Crea DF con info_cluster y elimina est def
info_cluster=pd.concat([info_est['Modulo_pluv'],info_est['Fraccion_comple
ta'],info_est['latitud'],info_est['longitud'],info_est['altura']],axis=1)
for est in est_def:
    info_cluster=info_cluster.drop(est,axis=0)

#Análisis estadístico  a BBDD Con Pmáx anuales
pmax=base2.max()
promedios=base2.mean()
desvest=base2.std()
cv=desvest/promedios
km=(base2-promedios)/desvest
km_max=km.max()
pmaxadim=(base2.max()-base2.mean())/base2.mean()

#Consolidar una BBDD con las estadísticas
estadisticas=pd.concat([km_max,pmax,promedios,desvest,cv,pmaxadim],axis=1
)
estadisticas.columns=['Km_max','Pmax','Prom','Desvest','CV','Pm_adim']
estadisticas.to_csv('BBDD_estadisticos_AC_sol2.csv',index=True,header=Tru
e)
analisis_cluster=pd.concat([estadisticas,info_cluster],axis=1)
analisis_cluster.to_csv('BBDDcluster_AC_sol2.csv',index=True,header=True)
```

```
#Estandarización de los parámetros
#Generar Matriz con coordenadas de las estaciones
#10.- Abrir BBDD de la información de las estaciones
path='BBDDcluster_AC_sol2.csv'
parametros = pd.read_csv(path,index_col='Cod Estacion')

#11.- Estandarización
Km_max_std2=(parametros['Km_max']-
parametros['Km_max'].mean())/parametros['Km_max'].std()
Pmax_std2=(parametros['Pmax']-
parametros['Pmax'].mean())/parametros['Pmax'].std()
Prom_std2=(parametros['Prom']-
parametros['Prom'].mean())/parametros['Prom'].std()
Desvest_std2=(parametros['Desvest']-
parametros['Desvest'].mean())/parametros['Desvest'].std()
CV_std2=(parametros['CV']-parametros['CV'].mean())/parametros['CV'].std()
Pm_adim_std2=(parametros['Pm_adim']-
parametros['Pm_adim'].mean())/parametros['Pm_adim'].std()
Modulo_adim_std2=(parametros['Modulo_pluv']-
parametros['Modulo_pluv'].mean())/parametros['Modulo_pluv'].std()
latitud_std2=(parametros['latitud']-
parametros['latitud'].mean())/parametros['latitud'].std()
longitud_std2=(parametros['longitud']-
parametros['longitud'].mean())/parametros['longitud'].std()
altura_std2=(parametros['altura']-
parametros['altura'].mean())/parametros['altura'].std()

parametros_estandarizados2=pd.concat([Km_max_std2,Pmax_std2,Prom_std2,Des
vest_std2,CV_std2,Pm_adim_std2,Modulo_adim_std2,latitud_std2,longitud_std
2,altura_std2],axis=1)
parametros_estandarizados2.to_csv('cluster_std2_sol2.csv',index=True,head
er=True)
```

9.1.4 Generación de resumen muestral, y corrección de promedios y desviaciones estándar por longitud de registro y evento máximo

```
# 1.- Cargar librerias
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib as plt
import pylab as pl
import math

#Generar Matriz con coordenadas de las estaciones
#2.- Abrir BBDD de la información de las estaciones
path1='BBDDinfo_depurada.csv'
info_est = pd.read_csv(path1,index_col='Unnamed: 0')
path2='BBDDdepurada2.csv'
Pmax_dia = pd.read_csv(path2, parse_dates=['fechas'], index_col='fechas')
codigos=Pmax_dia.columns.tolist()
info_est['Cod Estacion']=codigos
info_est=info_est.set_index('Cod Estacion')

#3.-Genera un DF CON Pmax ANUAL
Pmax_anuales=Pmax_dia.resample('Y').max()
Pmax_anuales_desc=Pmax_anuales['1903/01/01':'2020'].copy()

#4.-Cálculo del módulo pluviométrico
Pacum_mensual=Pmax_dia.resample('M').sum()
Pacum_anual=Pacum_mensual.resample('Y').sum()
mod_pluv=list()
completo=list()
dias_med=list()
for i in range(len(codigos)):
      inicio=datetime.strptime(info_est['inicio_observaciones'][str(codig
      os[i])], '%Y-%m-%d')
      fin=datetime.strptime(info_est['fin_observaciones'][str(codigos[i])
      ], '%Y-%m-%d')
```

```
    comp=Pmax_dia[inicio:fin][str(codigos[i])].count()/(fin-
inicio).days*100
    mod=Pacum_anual[inicio:fin][str(codigos[i])].mean()
    dias_med.append((fin-inicio).days)
    completo.append(comp)
    mod_pluv.append(mod)
info_est['Modulo_pluv']=mod_pluv
info_est['Fraccion_completa']=completo

#5.-Compilación de CSV con descripción de la BBDD
Pmax_max=Pmax_anuales.max()
Pmax_mean=Pmax_anuales.mean()
Pmax_std=Pmax_anuales.std()
descripcion=pd.concat([info_est['nombre'],Pmax_max,Pmax_mean,Pmax_std],ax
is=1)
descripcion['Modulo pluv']=mod_pluv
descripcion['años registro']=Pmax_anuales.count()
descripcion.columns=['nombre','Pmax_max','Pmax_mean','Pmax_std','Modulo_p
luv','años_registro']
descripcion.to_csv('BBDD_descripcion2.csv',index=True,header=True)

#6.-Cálculo de STDn-1 y Xn-1 (sin el máximo)
#Promedio sin máximo
Pmax_mean_1=(Pmax_anuales.sum()-Pmax_max)/(Pmax_anuales.count()-1)
#Desvest sin máximo
dif_cuad=(Pmax_anuales-Pmax_mean_1)**2
dif_cuad_1=dif_cuad.sum()-(Pmax_anuales.max()-Pmax_mean_1)**2
Pmax_std_1=(dif_cuad_1/Pmax_anuales.count()-2)**0.5

#7.-Compilación de CSV con Pc y Sc
#Compilación de CSV con Pc y Sc
estadigrafos=pd.concat([Pmax_mean,Pmax_std,Pmax_mean_1,Pmax_std_1],axis=1
)
estadigrafos['años registro']=Pmax_anuales.count()
estadigrafos.columns=['Pmax_mean','Pmax_std','Pmax_mean_1','Pmax_std_1','
años_registro']
estadigrafos.to_csv('BBDD_estadigrafos_CSC2_modificado.csv',index=True,he
ader=True)
```

9.2 Matlab

9.2.1 Código Análisis de Cluster y obtención de valores límites (Matlab)

```
% Forma de la matríz BBDDcluster2
% | Etiqueta | Km_max |Pmax| P_prom | Desves | CV | Pm_adim | Mod pluv |%
Comp| Lat | Long | alt |
estaciones=BBDDcluster2;
% | Etiqueta | Km_max |Pmax| P_prom | Desves | CV | Pm_adim | Mod pluv |
Lat | Long | alt |
k=clusterstd22(:,[3 8  9 10 11]); %selecciona los parámeros a considerar
en el AC
nk=4;
[idk1,centroides_k]=kmedoids(k,nk,'dist','sqEuclidean');  %'sqeuclidean'

%Contar la cantidad de elementos por cluster
tamanos=zeros(nk,1);
jf1=size(idk1);
jf=jf1(1);
for i=1:nk
    contador=0;
    for j=1:jf
        if idk1(j)==i
            contador=contador+1;
        end
        if j==jf
            tamanos(i,1)=contador;
        end
    end
end

%M. HERSHFIELD
%Formar matriz que extrae k_max de la BBDD
k_max=zeros(jf,nk);
prom=zeros(jf,nk);
desv=zeros(jf,nk);
frac_comp=zeros(jf,nk);
for i=1:nk
    for j=1:jf
        if idk1(j)==i
            k_max(j,i)=estaciones(j,2);
            prom(j,i)=estaciones(j,4);
            desv(j,i)=estaciones(j,5);
            frac_comp(j,i)=estaciones(j,9);
        end
    end
end
```

```
%Extraer info de matriz a vectores
cluster1=zeros(tamanos(1,1),4);
cluster2=zeros(tamanos(2,1),4);
cluster3=zeros(tamanos(3,1),4);
cluster4=zeros(tamanos(4,1),4);
%cluster5=zeros(tamanos(5,1),4);

for i=1:nk
    cont=1;
    for j=1:jf
        if k_max(j,i)~=0
            if i==1
                cluster1(cont,1)=k_max(j,i);
                cluster1(cont,2)=prom(j,i);
                cluster1(cont,3)=desv(j,i);
                cluster1(cont,4)=frac_comp(j,i);
                cont=cont+1;
            end
            if i==2
                cluster2(cont,1)=k_max(j,i);
                cluster2(cont,2)=prom(j,i);
                cluster2(cont,3)=desv(j,i);
                cluster2(cont,4)=frac_comp(j,i);
                cont=cont+1;
            end
            if i==3
                cluster3(cont,1)=k_max(j,i);
                cluster3(cont,2)=prom(j,i);
                cluster3(cont,3)=desv(j,i);
                cluster3(cont,4)=frac_comp(j,i);
                cont=cont+1;
            end
            if i==4
                cluster4(cont,1)=k_max(j,i);
                cluster4(cont,2)=prom(j,i);
                cluster4(cont,3)=desv(j,i);
                cluster4(cont,4)=frac_comp(j,i);
                cont=cont+1;
            end
            if i==5
                cluster5(cont,1)=k_max(j,i);
                cluster5(cont,2)=prom(j,i);
                cluster5(cont,3)=desv(j,i);
                cluster5(cont,4)=frac_comp(j,i);
                cont=cont+1;
            end
        end
    end
end
```

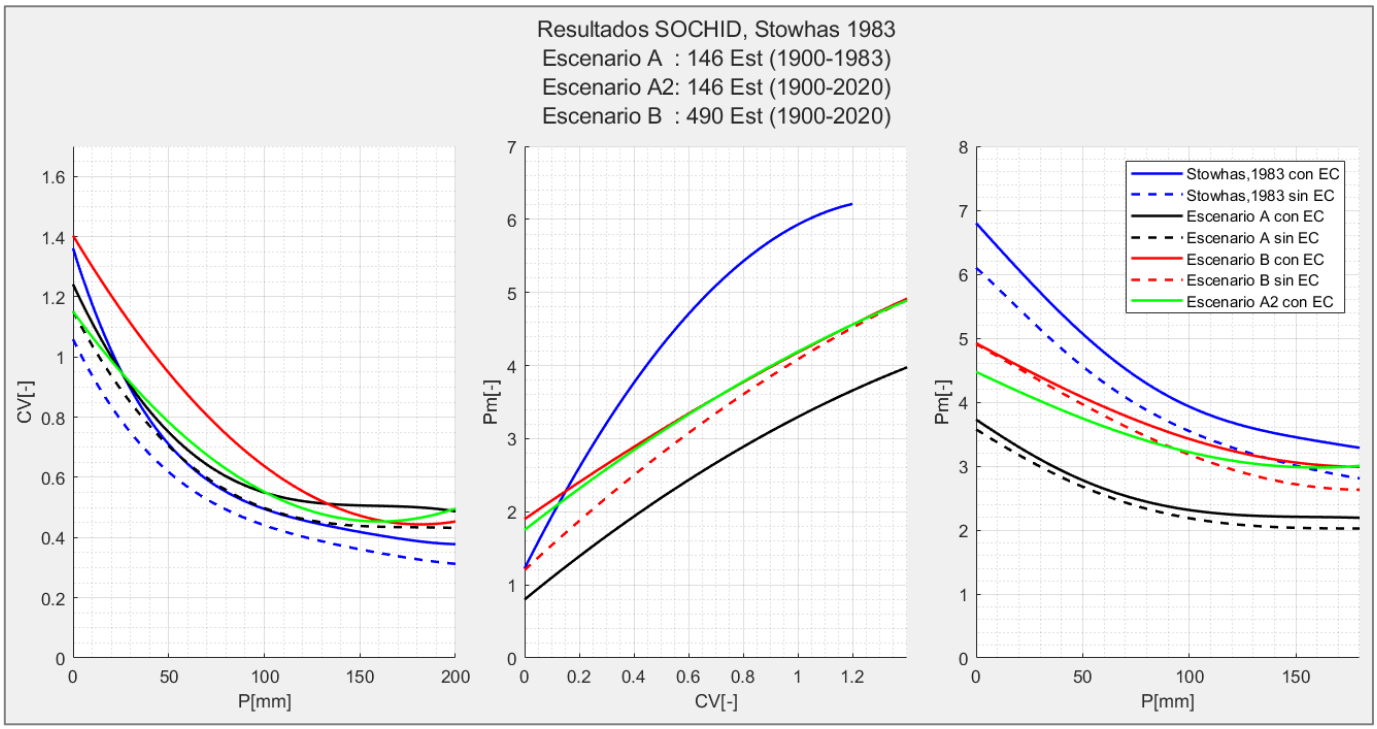
```
%Ajustes GEV a km_max de cada cluster
km1=gevfit(cluster1(:,1));
km2=gevfit(cluster2(:,1));
km3=gevfit(cluster3(:,1));
km4=gevfit(cluster4(:,1));
%km5=gevfit(cluster5(:,1));

%Obtener km_max con Pexc =1 considerando 1 solo grupo
km_todo=gevfit(estaciones(:,2));
max_km_todo=gevinv(0.9999999999999999,km_todo(1),km_todo(2),km_todo(3));

%Obtener km_max con Pexc =1 considerando en cada cluster
max_km=[gevinv(0.9999999999999999,km1(1),km1(2),km1(3));
        gevinv(0.9999999999999999,km2(1),km2(2),km2(3));
        gevinv(0.9999999999999999,km3(1),km3(2),km3(3));
        gevinv(0.9999999999999999,km4(1),km4(2),km4(3))];
        %gevinv(0.9999999999999999,km5(1),km5(2),km5(3))];


%Retorna Km_max(Pexc=1) del cluster y cantidad de estaciones en él
info=[tamanos,max_km]
max_km_todo

%Gráfico de estaciones sobre mapa de Chile
figure(11)
geobasemap grayterrain
title({[' Zonificación propuesta  ']
    ['para estimación de la PMP en Chile']})
hold on
estilos={'bo','go','ro','mo','ko'};
for i=1:nk
    clust = find(idk1==i);

geoscatter(estaciones(clust,10),estaciones(clust,11),estilos{i},'filled')
;
    legend('Cluster 1','Cluster 2','Cluster 3','Cluster 4')
end
hold off

%Gráfico de correlación de parámetros
figure(20)
estilos={'bo','go','ro','mo','ko','co'};

tiledlayout(1,3)
nexttile
title('M_{pluv} [mm] v/s Lat-Long')
hold on
for i=1:nk
    clust = find(idk1==i);
```

```
scatter3(estaciones(clust,10),estaciones(clust,11),estaciones(clust,8),es
tilos{i},'filled');
    legend('Cluster 1','Cluster 2','Cluster 3','Cluster 4')
end
hold off

nexttile
title('Altura [msnm] v/s Lat-Long')
hold on
for i=1:nk
    clust = find(idk1==i);

scatter3(estaciones(clust,10),estaciones(clust,11),estaciones(clust,12),e
stilos{i},'filled');
    legend('Cluster 1','Cluster 2','Cluster 3','Cluster 4')
end
hold off

nexttile
title('Pmax [mm] v/s Lat-Long')
hold on
for i=1:nk
    clust = find(idk1==i);

scatter3(estaciones(clust,10),estaciones(clust,11),estaciones(clust,3),es
tilos{i},'filled');
    legend('Cluster 1','Cluster 2','Cluster 3','Cluster 4')
end
hold off

figure (15)
title('M_{pluv} v/s P_{max} [mm] v/s altura[msnm]')
hold on
for i=1:nk
    clust = find(idk1==i);

scatter3(estaciones(clust,3),estaciones(clust,8),estaciones(clust,12),est
ilos{i},'filled');
    legend('Cluster 1','Cluster 2','Cluster 3','Cluster 4')
end
hold off



%Validación del AC
%Elección del número óptimo de clusters
%Indice de silhouette, índice de Davis Bouldin , Método GAP

S = evalclusters(k,'kmeans','silhouette','klist',[1:10]);
```

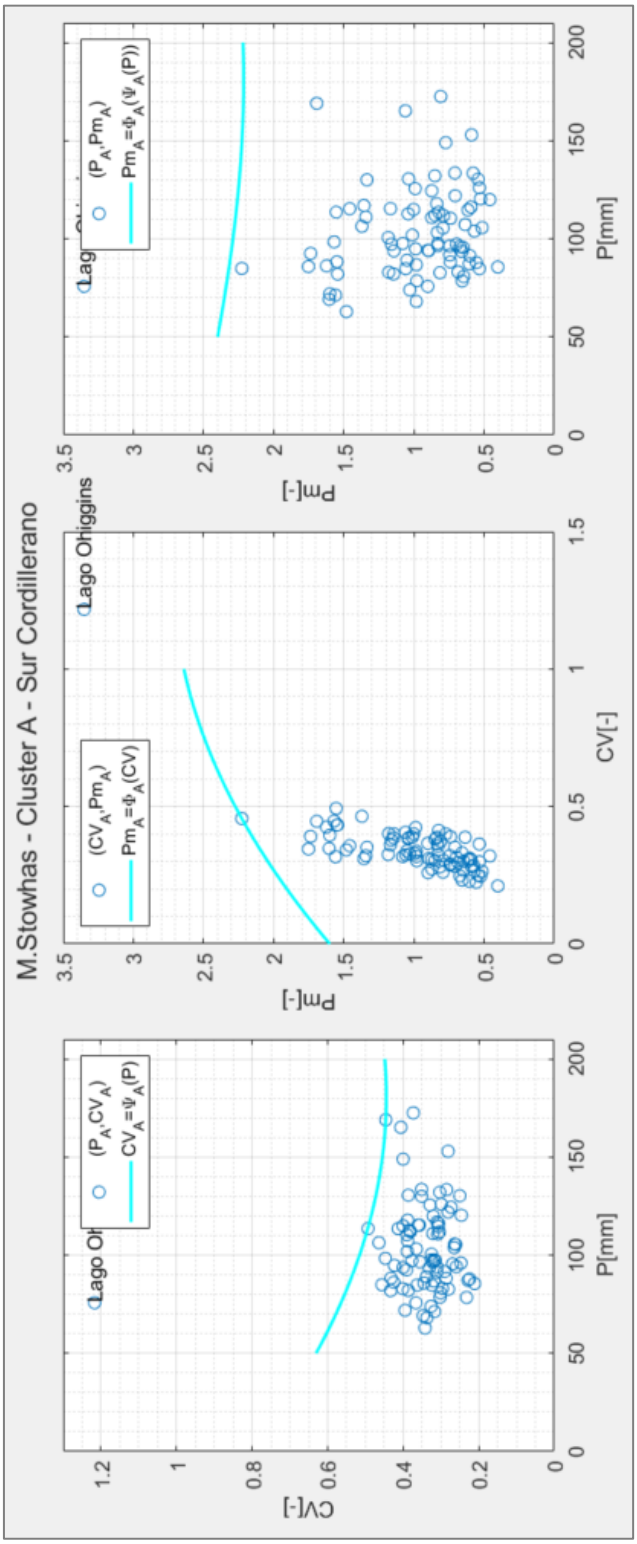
```
G = evalclusters(k,'kmeans','Gap','klist',[1:10]);
DB = evalclusters(k,'kmeans','DaviesBouldin','klist',[1:10]);
CH = evalclusters(k,'kmeans','CalinskiHarabasz','klist',[1:10]);

figure (12)
t=tiledlayout(1,4);
title(t,'Número óptimo de Clusters')
```

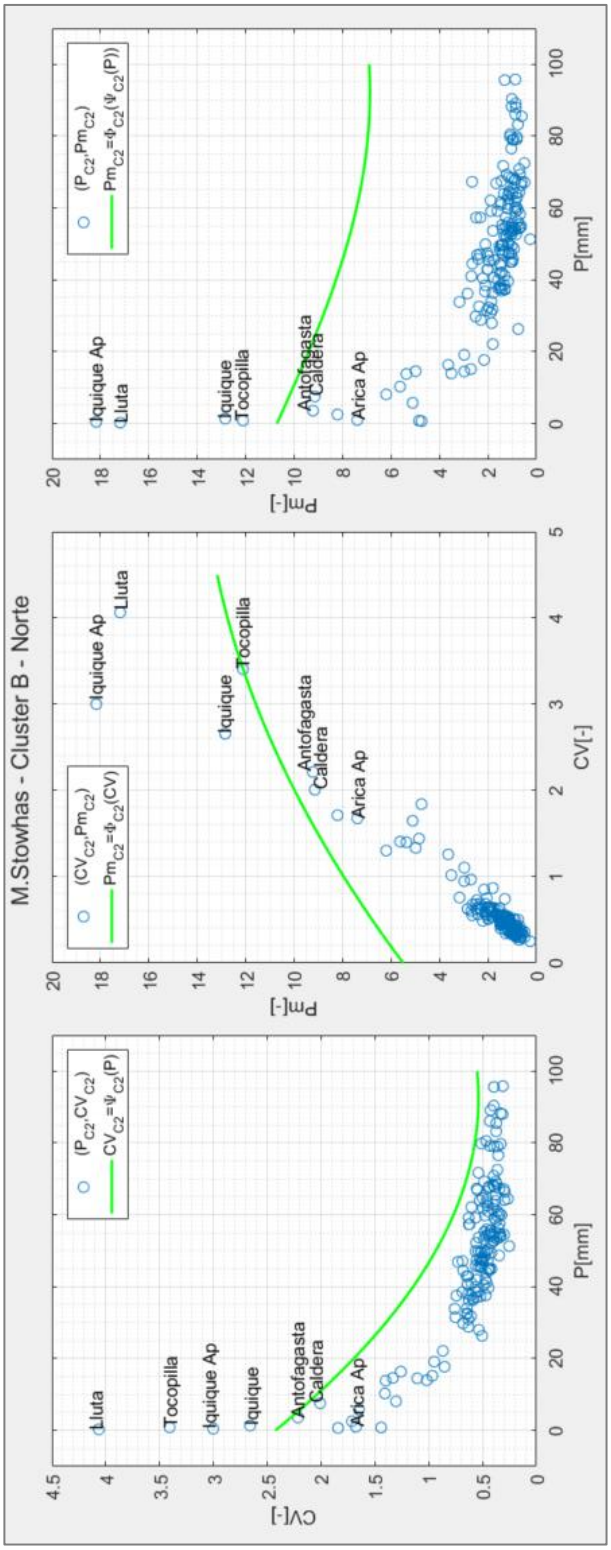
10 Anexo B: Figuras complementarias

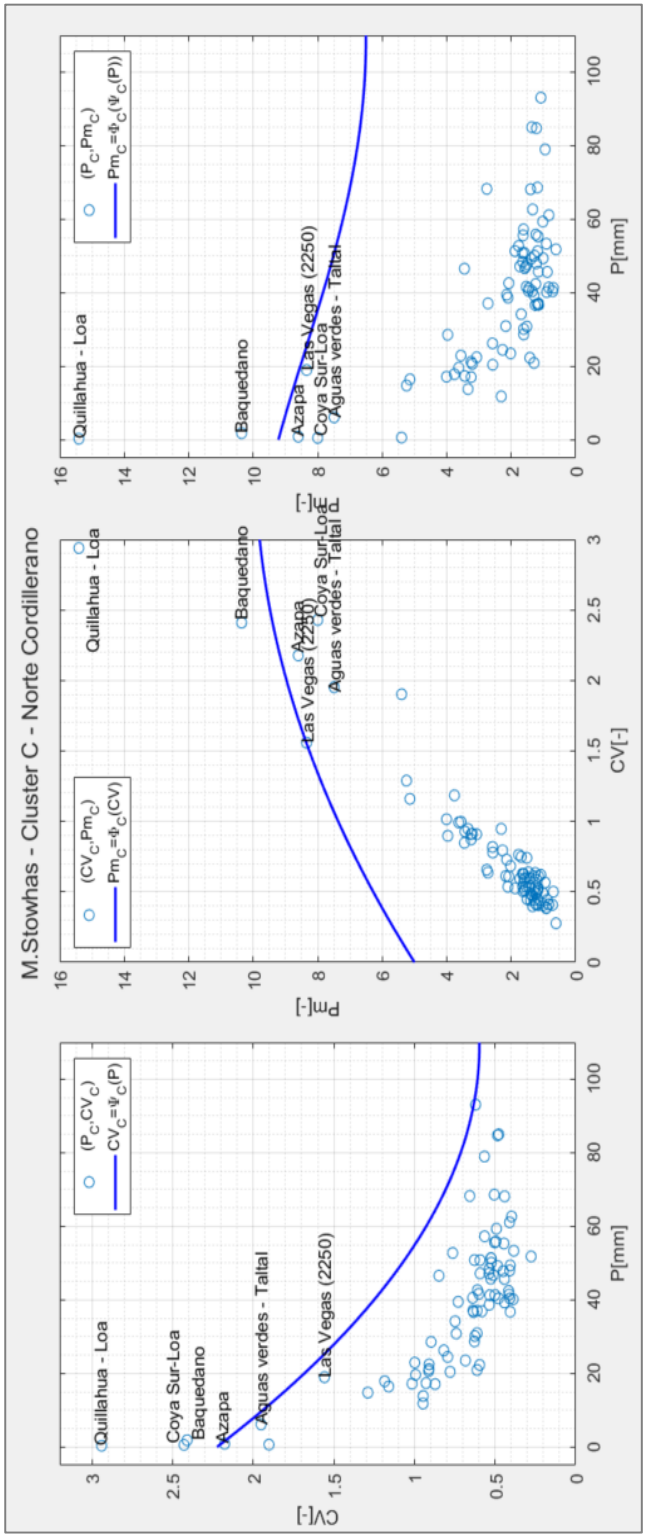
10.1 Comparación entre envolventes de escenarios A, A2, B y Stow,1983

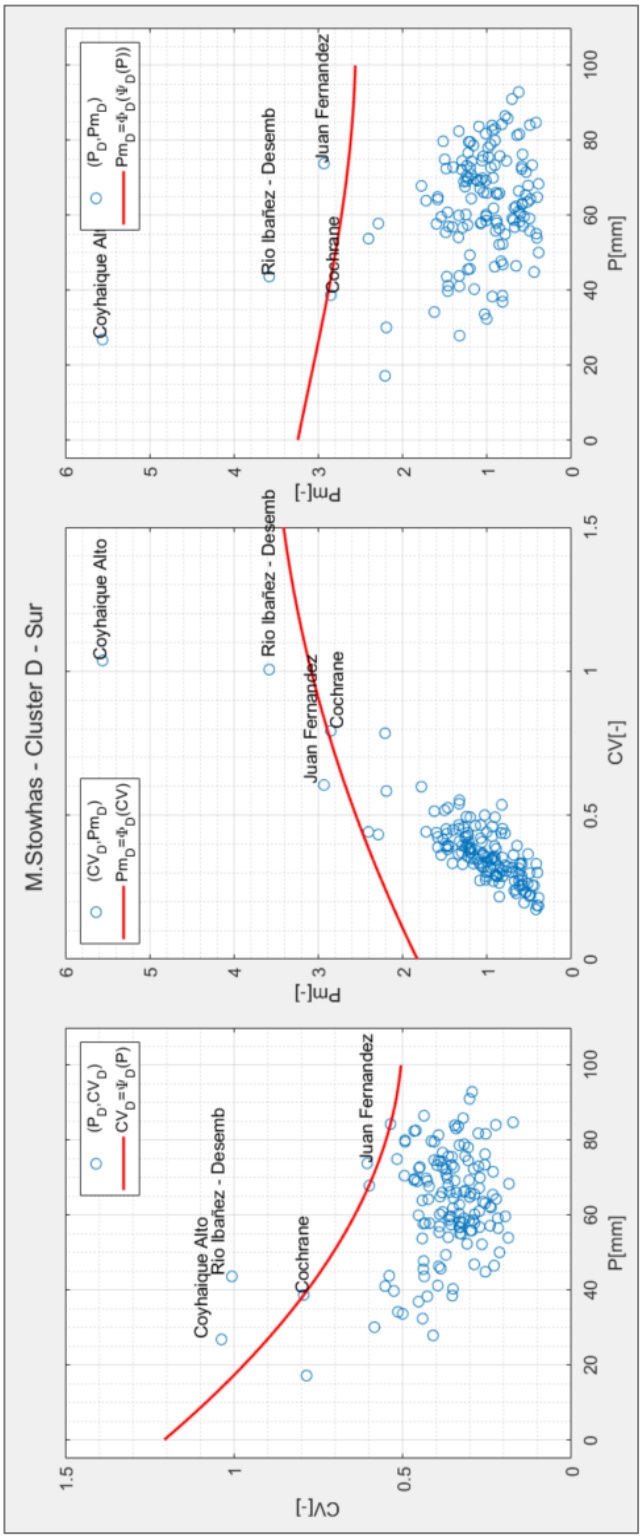


10.2 Envolventes intermedias de M. Stowhas




M.Stowhas - Cluster B - Norte
(P_C2,CV_C2)
CV_C2=Ψ_C2(P)
CV[-]
P[mm]
Lluta
Tocopilla
Iquique Ap
Iquique
Antofagasta
Caldera
Arica Ap
(CV_C2,Pm_C2)
Pm_C2=Φ_C2(CV)
Pm[-]
CV[-]
(P_C2,Pm_C2)
Pm_C2=Φ_C2(Ψ_C2(P))
Pm[-]
P[mm]


M.Stowhas - Cluster C - Norte Cordillerano
(P_C,CV_C)
CV_C=Ψ_C(P)
CV[-]
P[mm]
Quillahua - Loa
Coya Sur-Loa
Baquedano
Azapa
Aguas verdes - Taltal
Las Vegas (2250)
(CV_C,Pm_C)
Pm_C=Φ_C(CV)
Pm[-]
CV[-]
(P_C,Pm_C)
Pm_C=Φ_C(Ψ_C(P))
Pm[-]
P[mm]


M.Stowhas - Cluster D - Sur
(PD,CVD)
CVD=ΨD(P)
Coyhaique Alto
Rio Ibañez - Desemb
Cochrane
Juan Fernandez
CV[-]
P[mm]
(CVD,PmD)
PmD=ΦD(CV)
Coyhaique Alto
Rio Ibañez - Desemb
Juan Fernandez
Cochrane
Pm[-]
CV[-]
(PD,PmD)
PmD=ΦD(ΨD(P))
Coyhaique Alto
Rio Ibañez - Desemb
Cochrane
Juan Fernandez
Pm[-]
P[mm]