



Organización
de las Naciones Unidas
para la Educación,
la Ciencia y la Cultura



Programa
Hidrológico
Intergubernamental

Aqua-LAC

Miembro del Programa Hidrológico Intergubernamental
para América Latina y el Caribe

Comparación de precipitaciones diarias máximas anuales medidas en forma discreta y continua en zonas del centro sur de Chile

Comparison of annual maximum daily rainfall measured in a discrete and continuous way in areas of south-central Chile

Esteban Andrés Celis Recabal ^{1*}

Recibido: 21/09/2021

Aceptado: 17/12/2022

*Autor de correspondencia

Resumen

Este estudio se realizó con el fin de analizar si existen diferencias significativas entre mediciones de precipitación en 24 horas, realizadas con lapsos discretos (de 8 AM a 8 AM), y lapsos continuos de 24 horas. Para esto se trabajó con datos de 7 estaciones pluviográficas ubicadas en la Región del Maule, zona central de Chile (Potrero Grande, Penco, Talca, Colorado, Melozal, Parral y Bullileo Embalse) de las cuales se obtuvieron precipitaciones máximas anuales en lapsos continuos y discretos. Se obtuvo información de registros continuos y discretos con longitudes de datos variables (entre 28 y 39 años). Ambos registros se sometieron a análisis estadísticos con medidas de tendencia central y dispersión, gráficos de Violín, Análisis de Concordancia de Bland y Altman y comparación de medianas con la Prueba U de Mann-Whitney. Mediante la prueba U de Mann-Whitney, se logra determinar que solo una estación (Bullileo Embalse) presenta diferencias estadísticamente significativas entre el método discreto y el método continuo, con un 95% de confiabilidad. Pero, con un $\alpha=0.1$ o un 90% de confiabilidad, esta conclusión se invierte porque 6 de las 7 estaciones presentarían diferencias significativas. Lo anterior determina que hay que ser muy cuidadoso a la hora de elegir la significancia, por sus implicancias en la generación de los datos para diseños hidrológicos e hidráulicos.

Palabras clave: Intensidad diaria máxima anual, medición de precipitación.

Abstract

In this study, it was analyzed if there are significant differences between 24-hour precipitation measurements, performed with discrete lapses (from 8 AM to 8 AM), and continuous 24-hour lapses. For this, we worked with data from 7 pluviographic stations located in the Maule Region (Potrero Grande, Penco, Talca, Colorado, Melozal, Parral and Bullileo Embalse) from which maximum annual precipitation was obtained in continuous and discrete periods. Information was obtained from continuous and discrete records with variable data lengths (between 28 and 39 years). Both records were subjected to statistical analysis with measures of central tendency and dispersion, Violin plots, Bland and Altman Analysis of Agreement and comparison of medians with the Mann-Whitney U Test. Finally, using the Mann-Whitney U test, it is possible to determine that only one station (Bullileo Embalse) presents statistically significant differences between the discrete method and the continuous method, with 95% reliability. But, with $\alpha = 0.1$ or 90% reliability, this conclusion is inverted because 6 of the 7 stations would present significant differences. The above determines that it is necessary to be very careful when choosing significance, due to its implications in the generation of data for hydrological and hydraulic designs.

Keywords: Annual maximum daily intensity, precipitation measurement.

1. INTRODUCCIÓN

El conocimiento sobre el comportamiento temporal y espacial de las precipitaciones es un elemento de suma importancia para el desarrollo humano, pues entrega información útil en el área de la hidrología, que es clave en la planificación estratégica de recursos y para diversas áreas que van desde el urbanismo a la agricultura (Campos-Aranda, 2010; Burn, 2014). Una herramienta que permite analizar la cantidad de agua caída es la intensidad de la precipitación, que se define como la variación de precipitación en un diferencial de tiempo, la cual generalmente se representa en mm/h. Esta cantidad puede ser relacionada con la frecuencia y duración de un evento de manera estadística, dando origen a las Curvas Intensidad-Duración-Frecuencia (IDF). Para elaborar estas curvas IDF se precisa de información de intervalos de precipitación de al menos una hora e idealmente en lapsos menores con el fin de poder analizar el comportamiento de las intensidades, dado que estas poseen una mayor probabilidad de alcanzar mayores valores en la medida que la duración del fenómeno de precipitación sea menor en el tiempo (Chow et al., 1998). Es decir, las curvas IDF corresponden a una familia de curvas del tipo exponencial negativo y en donde cada curva corresponde a un diferente periodo de retorno. Lo expuesto en el anterior párrafo, obliga a pensar que las curvas IDF solo pueden ser construidas con registros continuos de precipitación en el tiempo. No obstante, diversos autores e instituciones han propuesto diferentes metodologías para intentar aproximarse a la construcción de curvas IDF en zonas en que no cuentan con registros continuos de precipitación. Una de estas propuestas es la realizada por Pizarro et al. (2015, 2018) en donde propuso un índice de tormenta que puede ser aplicado a estaciones con registros cada 24 horas (estaciones pluviométricas). El método plantea que en las estaciones pluviográficas o con registros continuos, es posible obtener un cociente entre la intensidad de lluvia para una duración y un periodo de retorno dado y la intensidad para 24 horas y el mismo periodo de retorno, cociente que se denomina índice de tormenta y puede ser aplicado a estaciones pluviométricas. Ahora bien, para la elaboración del índice de tormenta se utilizan como datos base los registros continuos de las máximas intensidades anuales de diferentes duraciones, los cuales provienen de una estación pluviográfica. El índice de tormenta que se consigue es posteriormente aplicado a datos

provenientes de una estación pluviométrica donde solo se cuenta con información de 24 horas. Esta aplicación permite realizar una extrapolación de las curvas IDF. Sin embargo, los datos de la estación pluviométrica han sido registrados a lo largo del tiempo de una forma discreta, es decir cada 24 horas (de 8 a 8). Por otra parte, los valores de las estaciones pluviográficas son obtenidos de manera continua. Entonces surge la pregunta; ¿la serie de valores capturada de 8AM a 8AM, diferirá mucho si la serie es capturada en cualquier lapso manteniendo las 24 horas? Y esa es la interrogante que pretende abordar este estudio.

2. METODOLOGÍA

2.1 Selección de estaciones pluviográficas

Los datos base para la elaboración del estudio provienen de estaciones pluviográficas aportadas por la Dirección General de Aguas de Chile (DGA), las cuales cuentan con registros desde el año 1971 las más antiguas, hasta el año 2009. Después de esta fecha no se logró obtener datos debido a que el método de registro de precipitaciones cambió de pluviógrafos tradicionales a pluviógrafos digitales. Es importante destacar que la información pluviográfica está almacenada en registros del servidor que para tales efectos posee el Centro Tecnológico de Hidrología Ambiental (CTHA) de la Universidad de Talca, Chile. Las estaciones pluviográficas, por medio de bandas de papel milimetrado, han registrado la totalidad de eventos de precipitación en forma continua. Tradicionalmente la extracción de información de precipitación en distintos lapsos se ha obtenido de manera análoga, lo cual resulta una tarea tediosa cuando se cuenta con una gran cantidad de años de registro en conjunto con varias estaciones. Debido a esto, en el marco del proyecto FONDEF D08I1054, (Fondo Nacional de Desarrollo Científico y Tecnológico de Chile), se desarrolló un lector de bandas pluviográficas el cual, mediante un software, identifica la traza registrada por el pluviógrafo en la banda y genera una tabla de datos de tres columnas: fecha, hora y altura de precipitación (mm). Las estaciones pluviográficas utilizadas se encuentran distribuidas dentro de la Región del Maule, zona centro sur de Chile. En la Tabla 1 se muestran las estaciones de las cuales se pudieron obtener los datos desde el CTHA, en conjunto con las coordenadas UTM, su altitud y periodo de registro. Con las coordenadas UTM de las estaciones, en conjunto con

información cartográfica, en específico con un modelo digital de elevación de la Región del Maule, un mapa de las zonas climáticas de Chile según Koeppen (escala 1:500.000) extraída del portal de

Infraestructura de Datos Geospaciales de Chile (IDE Chile) y, un mapa político de Chile, se realizaron, mediante QGIS, cartografías que describen el lugar de las estaciones ya mencionadas.

Tabla 1. Estaciones utilizadas

n	Estación	Coordenadas UTM		Altitud (m)	Periodo de Registro
		N	E		
1	P. Grande	6104860	309175	450	1971-2009
2	Pencahue	6081931	242659	110	1974-2009
3	Talca	6075755	262463	110	1982-2009
4	Colorado	6053778	295308	470	1971-2009
5	Melozal	6037743	248556	90	1971-2009
6	Parral	5991450	245651	160	1974-2009
7	Bullileo. E.	5981671	283173	600	1971-2009

2.2 Tratamiento de la información

En base a la información obtenida desde el CTHA, fue posible obtener archivos de información pluviográfica continua, en archivo de texto (extensión txt). Cada uno de los archivos contenía información de cada estación. Dicha información debió ser dividida en diferentes intervalos temporales, derivado del altísimo volumen de información presente. De esa manera fue posible que el archivo fuese leído por un programa de manejos de datos en formato de

archivos separados con coma (extensión csv); en este caso el programa utilizado fue Excel. Ahora bien, los archivos transformados a Excel son expresados todos en una misma columna, por ello se procedió a separarlos mediante la función de Excel “Texto en columnas”, el que divide el contenido en varias columnas, seleccionando como caracteres divisorios los guiones y las comas, quedando como se muestra en la Figura 1.

	A	B	C	D	E
1	Año	Mes	Día	Hora	PP
39728	1971	5	21	6:30	98,62
39729	1971	5	21	6:35	98,95
39730	1971	5	21	6:40	98,99
39731	1971	5	21	6:45	99,04
39732	1971	5	21	6:50	99,06
39733	1971	5	21	6:55	99,07
39734	1971	5	21	7:00	99,09
39735	1971	5	21	7:05	99,10
39736	1971	5	21	7:10	99,12
39737	1971	5	21	7:15	99,12
39738	1971	5	21	7:20	99,12
39739	1971	5	21	7:25	102,99
39740	1971	5	21	7:30	105,63
39741	1971	5	21	7:35	107,00
39742	1971	5	21	7:40	107,13
39743	1971	5	21	7:45	107,13
39744	1971	5	21	7:50	107,29
39745	1971	5	21	7:55	107,46
39746	1971	5	22	8:00	107,46
39747	1971	5	22	8:05	107,46
39748	1971	5	22	8:10	107,46

Figura 1. Ejemplo de datos separados

2.3 Cálculo de las precipitaciones diarias

Los datos obtenidos a partir de la transformación a Excel permiten que la información se exprese en forma de pluviograma (precipitación acumulada en el tiempo). De esta manera Excel permite procesar la información estableciendo diferenciales de precipitación cada 24 horas. Dicho de otra manera, se realizó una resta entre los datos de precipitación con intervalos casi continuos de 24 horas, obteniendo en una columna anexa los datos de precipitación diarias que pueden obtenerse de una serie de información continua. Con el objetivo de verificar que no existiesen datos faltantes mediante el proceso de digitalización, se realizó otra columna de valores horarios desplazada en 5 minutos y se comparó con la que ya se había generado precedentemente. Si no había diferencias, se asumía el dato registrado como valor válido. Al contrario, si se generaban diferencias se revisaba la información con el fin de determinar si había alguna incongruencia en los datos. Es preciso señalar que nunca se encontraron diferencias.

2.4 Captura de la información en 24 horas en horario fijo y continuo

Con la matriz de datos ya lista se procedió a utilizar la metodología de tablas dinámicas para Excel. Así, en el campo de "Filas" se seleccionaron los años y en el campo de "Valores" se seleccionó el máximo de las precipitaciones cada 24 horas. Con ello se obtuvo el máximo valor anual de un evento de precipitación ocurrido en una duración de 24 horas, en cualquier horario, como se muestra en la primera tabla de la Figura 2. Por otra parte, se le agregó el tiempo a la misma tabla dinámica en el campo de "Filtros", obteniendo de esta manera solo las precipitaciones registradas entre las 8:00 AM y las 8AM del día siguiente, como se puede ver en la Figura 2. De esta forma se obtuvo el máximo de precipitación diaria ocurrida en dicho horario, simulando la medición de precipitaciones que se hace de manera tradicional mediante pluviómetros.

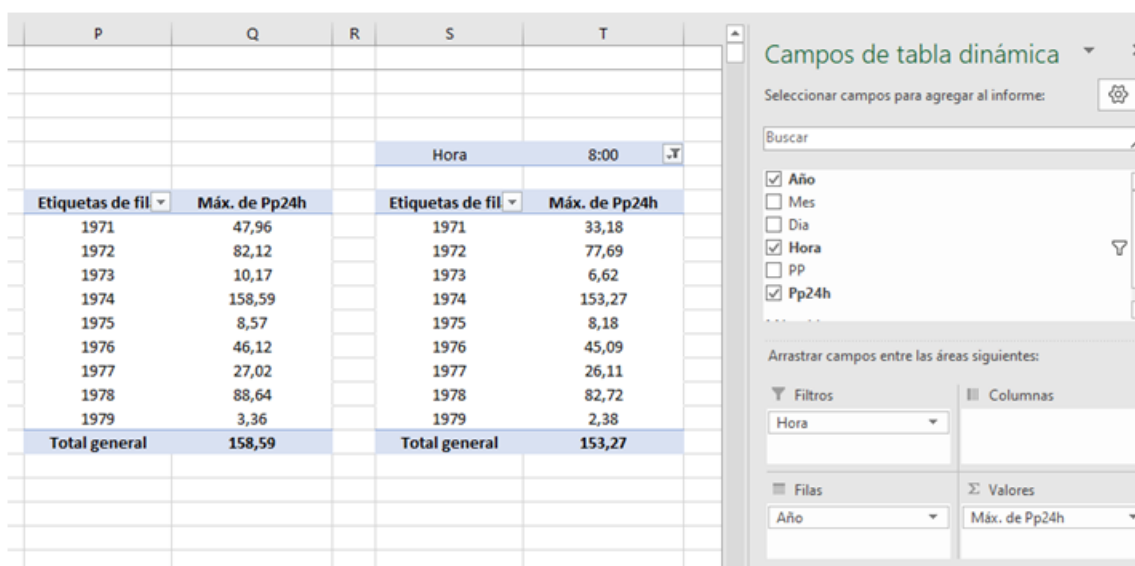


Figura 2. Ejemplo de Tablas dinámicas utilizadas.

Es importante señalar que, debido a errores de diversa naturaleza que se presentaron, es que hubo años en distintas estaciones en donde los valores de precipitación no fueron registrados o bien fueron mal registrados. Por ello, en los datos diarios donde la precipitación en 24 horas fue menor a 5 mm, se decidió borrarlos del registro, pues definían un evidente error. El porcentaje de años con error no

superó el 3% del total de años (2.7%). Con esta información ya ordenada se obtuvieron dos series de datos para cada estación: una de la precipitación diaria máxima anual en horario continuo (PDMAC), y la otra de la precipitación diaria máxima anual en un horario discreto, específicamente de 8:00 a 8:00 (PDMAD). Estas series de datos fueron las utilizadas para el posterior análisis estadístico.

2.5 Análisis estadístico

Se utilizaron diversos análisis estadísticos y matemáticos para el análisis de los datos. Adicionalmente se debe señalar que generalmente, los datos hidrológicos tienen un comportamiento no paramétrico, es decir, no presentan normalidad en sus datos, hecho que obliga a utilizar pruebas no paramétricas. Por lo tanto, los datos fueron analizados en base a lo siguiente.

$$\bar{x} = \frac{\sum x_i}{n}$$

Donde $x_1, x_2, x_3, \dots, x_n$ son las observaciones de la variable y n el número total de observaciones.

Se debe tener en cuenta que la media es un valor sensible a observaciones atípicas o extremas. Un valor alejado del resto puede generar un cambio importante en el valor de la media. La mediana se presenta como una opción al cálculo de la media y tiene la ventaja de que no es sensible a registros atípicos o extremos. Para un conjunto de datos, el valor de la mediana se consigue de forma que deja la misma cantidad de observaciones a su izquierda que

2.6 Medidas de tendencia central (media y mediana)

La media aritmética es una medida de tendencia central muy utilizada. Se interpreta como el promedio de los datos y se construye de forma que interviene la totalidad de los datos observados. Se calcula a través de la siguiente forma:

a su derecha (Moncho, 2014). Aunque podría existir una infinidad de valores que cumplirán este requisito para un grupo de observaciones, la forma habitual de cálculo garantiza que la mediana será única para una serie de datos. Así, en primer lugar, los datos se ordenan de menor a mayor. Después de esto, se calculará el rango de la mediana (r_{md}), que informará sobre la posición que debe ocupar esta, de la siguiente forma:

$$r_{md} = \frac{n + 1}{2}$$

Finalmente, dentro de los datos ordenados la mediana será el dato en específico que ocupa la posición r_{md} , en el caso que n sea impar. Ahora, si el número total de datos fuera par, para obtener la mediana se calcularía la semisuma entre los dos datos centrales.

2.7 Medidas de dispersión (desviación estándar y coeficiente de variación)

La desviación típica o estándar es una medida resumen de la distancia de cada dato a la media (Moncho, 2014). Se define la varianza como la sumatoria de las diferencias entre cada dato y su

media al cuadrado, dividido por el número de datos en el caso de una población. La varianza está expresada, por tanto, en unidades al cuadrado de la variable. Para conseguir una medida en las mismas unidades que la variable original se extrae la raíz cuadrada, obteniéndose la denominada desviación típica o estándar. El coeficiente de variación se obtiene dividiendo la desviación estándar (S) por la media (\bar{x}); si este se multiplica por 100, podrá interpretarse como el porcentaje de variabilidad de los datos para los que se calcula.

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n} \quad S = \sqrt{S^2} \quad CV = \frac{S}{\bar{x}}$$

2.8 Diagrama de violín

Un diagrama de violín (Hintze y Nelson, 1998), es básicamente una combinación de un gráfico de caja y un histograma suavizado. Un gráfico de caja (Figura 3) es un método utilizado para representar gráficamente una serie de datos numéricos, exponiendo 4 características principales de una variable; centro, extensión, asimetría y valores

atípicos. Por su parte el histograma (Figura 4) se obtiene al graficar frecuencias relativas de las clases contra sus respectivos intervalos; este se utiliza ya que puede hacer evidentes los patrones existentes dentro de un conjunto de datos (Canavos, 1987). Hintze y Nelson (1998) señalan que individualmente, los gráficos de caja proveen resúmenes sucintos de datos. La combinación sinérgica del diagrama de caja y el histograma permite que gran parte de la

información de cada uno se muestre en un diagrama (Figura 5). Esta estructura de gráfico único permite que las comparaciones de factores de distribución de

varias variables sean más fáciles de hacer y a esto se denomina gráfico de violín.

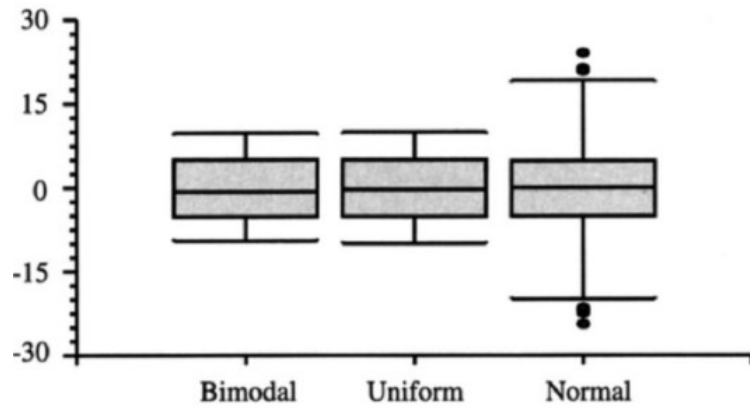


Figura 3. Ejemplo de gráficos de cajas para distribuciones conocidas (Hintze y Nelson, 1998)

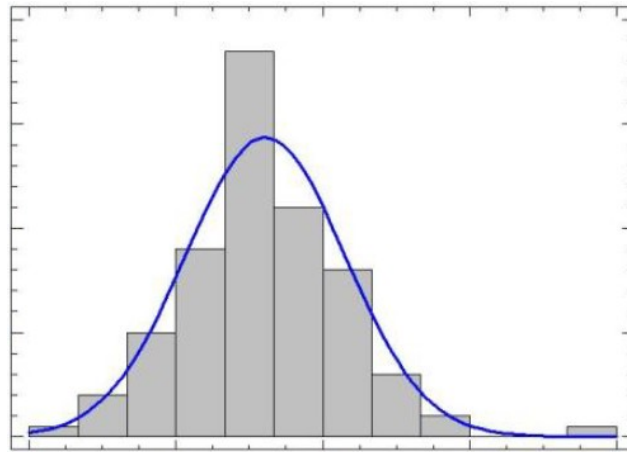


Figura 4. Ejemplo de histograma con curva de suavización.

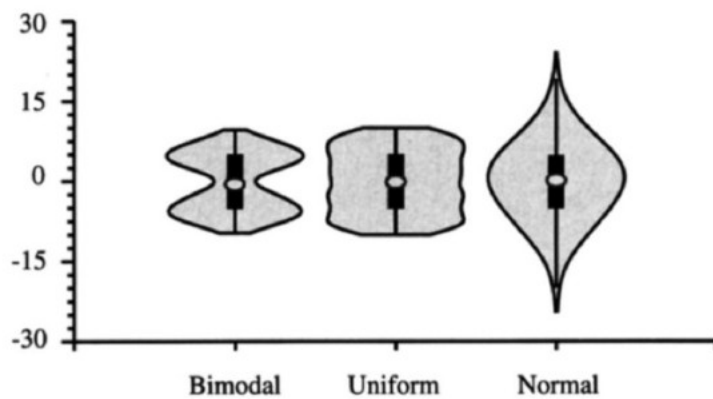


Figura 5. Ejemplo de gráficos de violín para distribuciones conocidas

2.9 Análisis de Concordancia de Bland y Altman

Análisis de Concordancia de Bland y Altman (ACBA). Carrasco y Jover (2004) señalan que a través de esta prueba se puede establecer si los modelos de estimación son intercambiables, es decir, si concuerdan lo suficiente. Para su cálculo, se obtienen las diferencias promedio (dp) entre el valor

$$LC = dp \pm 1,96 \cdot DS$$

Así, la relación valor real versus valor estimado con un dp en valor absoluto lo más cercano a cero, será la de mejor ajuste. Si llegase a existir una igualdad o una mínima diferencia del valor dp, el mejor ajuste corresponderá al que presente un menor valor DS y sus límites de concordancia sean más estrechos (Bland y Altman, 1999). Entonces, de no existir error

real y el estimado. Esta diferencia representa el error sistemático, y la varianza de las diferencias (DS) da cuenta de la imprecisión. De esta forma, se obtienen los límites de concordancia del 95%. Este límite de concordancia (LC) está definido por la expresión:

sistemático, los puntos que representan el valor de la diferencia de los datos se distribuirán de forma aleatoria a uno y otro lado de la recta, lo que corresponde a la diferencia cero entre medidas, como muestra la figura 6. En esta, la línea de color rojo representa la media de las diferencias y las líneas azules representan los límites de confianza.

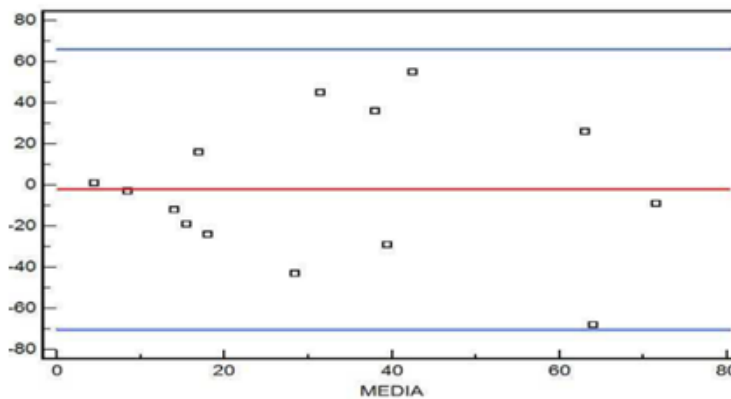


Figura 6. Ejemplo gráfico de la prueba de concordancia de Bland y Altman

2.10 Prueba U de Mann-Whitney

La prueba U de Mann-Whitney es uno de los procedimientos más usados para demostrar si dos poblaciones tienen la misma distribución (UNESCO, 2013). Se utiliza para series no paramétricas, por lo tanto, no es necesario analizar los supuestos de normalidad e igualdad de varianzas (Lind et al, 2001). Canavos (1987) indica que el equivalente no paramétrico de la prueba t de Student es la prueba U de Mann Whitney. Se basa en la combinación de 2 series, con n1 y n2 observaciones respectivamente, para formar un solo conjunto de n1+n2 observaciones. Luego estas observaciones se arreglan en orden creciente de magnitud, y se asigna un rango a cada observación en la secuencia. Si la mayor de las

muestras tiene más de 20 observaciones, se considera una muestra grande. En su defecto, si tiene 20 o menos observaciones se considera una muestra pequeña (Lind et al, 2001). Para determinar si dos poblaciones presentan o no la misma distribución, la prueba U de Mann Whitney establece las siguientes hipótesis:

H0: No existen diferencias significativas entre las medianas de los grupos, con un nivel de confianza del 95%.

Ha: Existen diferencias significativas entre las medianas de los grupos, con un nivel de confianza del 95%.

Debido a que las muestras de este estudio todas tienen 20 o más observaciones, a continuación, se presenta el procedimiento utilizado para las muestras grandes, propuestas por Mendenhall et al. (1997):

Para $N > 20$

Paso 1: Se unen ambas muestras n_1 y n_2 para formar un solo conjunto (n_1+n_2) datos.

$$U_A = n_1 \cdot n_2 + \frac{(n_1(n_1 + 1))}{2} + T_A$$

$$U_B = n_1 \cdot n_2 + \frac{(n_2(n_2 + 1))}{2} - T_B$$

Donde n_1 y n_2 son el número de datos para las muestras A y B, respectivamente; y T_A y T_B se refieren a la suma de los rangos para las muestras A y B, respectivamente.

Paso 4: Se realiza una aproximación a una distribución normal de la siguiente manera:

$$Z = \frac{U - E(U)}{\sqrt{V(U)}}$$

donde

U = es el valor mayor seleccionado entre U_A y U_B ;

$$E(U) = \frac{n_1 \cdot n_2}{2} \quad V(U) = \frac{n_1 \cdot n_2 + (n_1 + n_2 + 1)}{12}$$

Paso 5: Se busca el valor correspondiente a la probabilidad (valor P) asociada al Z calculado y de acuerdo con el nivel de significancia escogido.

Paso 6: Para $\alpha < 0,05$, los valores críticos son $\pm 1,96$. Por lo tanto, si el valor P calculado cae dentro de esa región, se acepta la hipótesis nula.

3. RESULTADOS Y DISCUSIÓN

3.1 Cartografía

Las estaciones utilizadas para este estudio se muestran en la Figura 7, la que proyecta la ubicación geográfica de las estaciones bajo un modelo de elevación del terreno. Así se observa que las estaciones Potrero Grande (1), Colorado (4) y Bullileo Embalse (7), están ubicadas en las cercanías de la precordillera. Por otra parte, las estaciones de Talca (3), Melozal (5) y Parral (6) se ubican en la depresión intermedia, quedando solo la estación de Péncahue (2) en las cercanías de la cordillera de la costa, hacia el secano interior. Por otro lado, la Figura 8 define la ubicación de las estaciones bajo la clasificación climatológica de Köppen-Geiger (1936), la que clasifica diferentes climas basándose en dos factores: precipitación y temperatura (Seoáñez Calvo,

Paso 2: Se ordenan los datos en orden ascendente y se les asigna un rango de 1 a (n_1+n_2).

Paso 3: Se calculan los estadísticos U_A y U_B y se selecciona el mayor de ellos.

2001). Dentro de la región existen dos climas: Clima de tundra, ubicado en la cordillera, y el clima mediterráneo, que abarca gran parte de la región. Este último se subdivide en 4 subtipos dependiendo de la temperatura y cantidad de precipitación. De importancia son entonces el clima mediterráneo de lluvia invernal y el clima mediterráneo de lluvia invernal con influencia de altura, pues dentro de estos tipos climáticos se encuentran las estaciones. Así, las estaciones Péncahue (2), Talca (3), Melozal (5) y Parral (6) se ubican dentro del clima mediterráneo de lluvia invernal y las estaciones Potrero Grande (1), Colorado (4) y Bullileo Embalse (7) en una zona de transición entre clima mediterráneo de lluvias invernal y el clima mediterráneo de lluvia invernal de altura.

Región del Maule

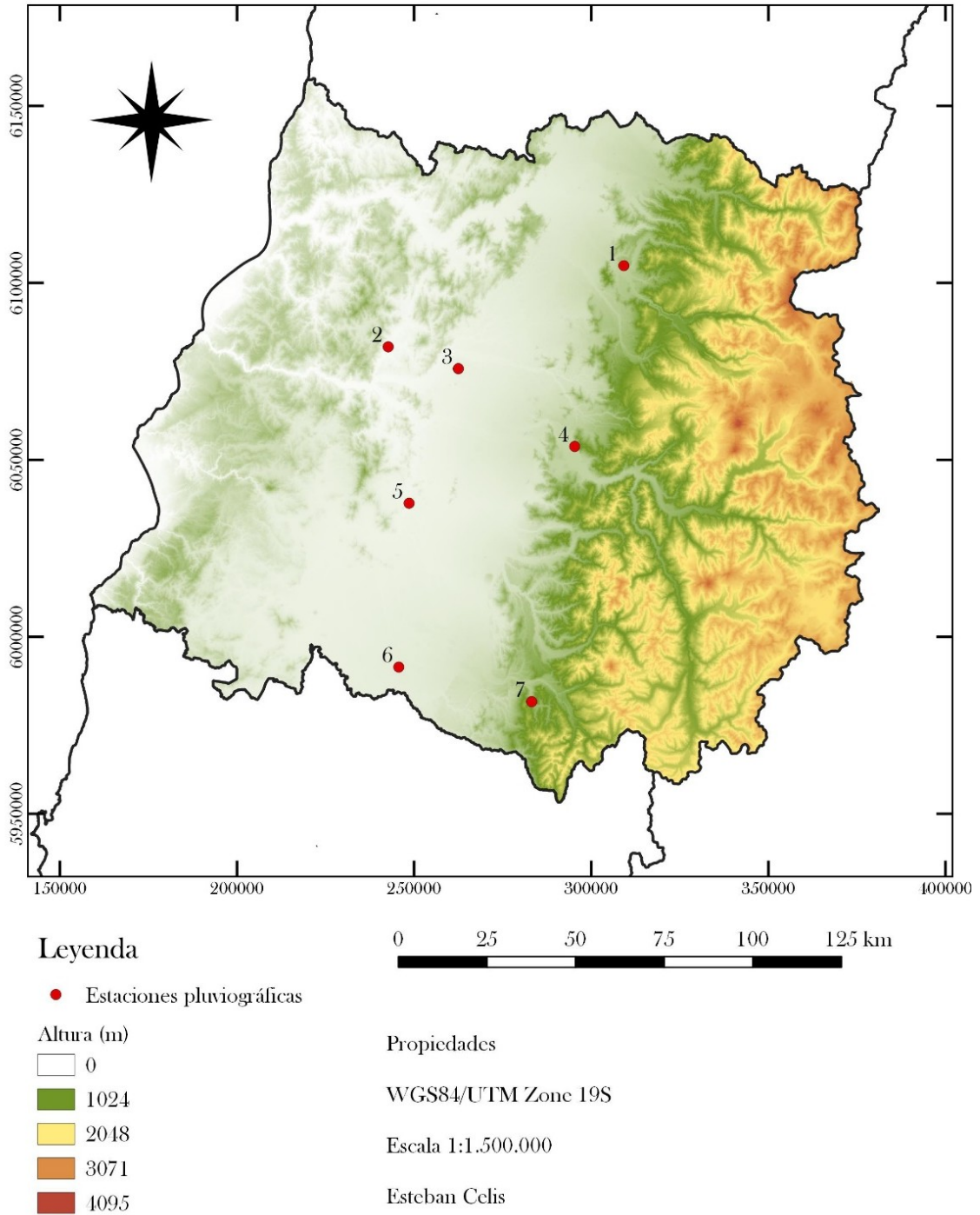
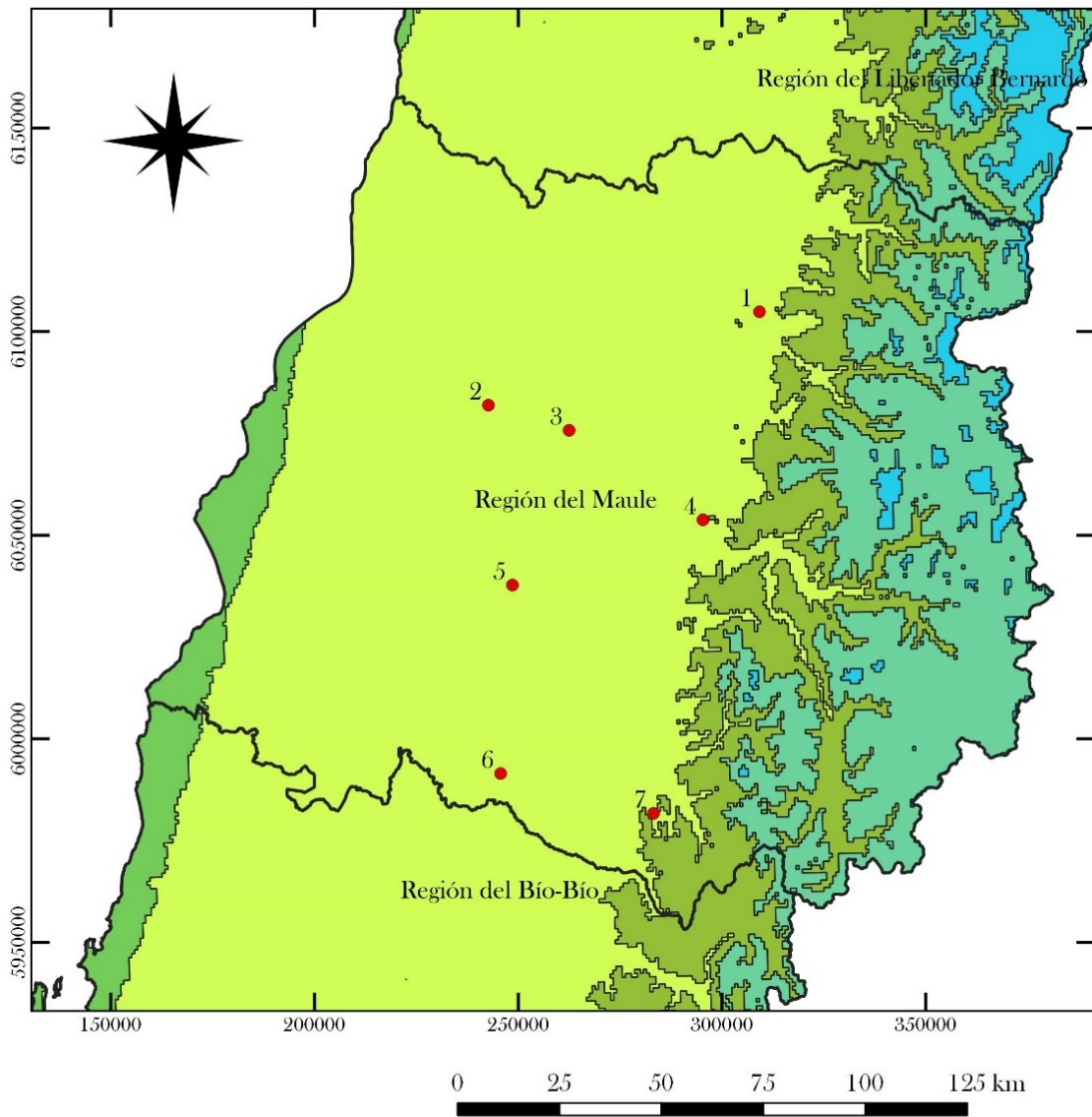


Figura 7. Mapa con las estaciones utilizadas, detalle de altitud

Región del Maule



Leyenda

● Estaciones

Clima según Köppen

- Clima de tundra de lluvia invernal
- Clima mediterráneo de lluvia invernal
- Clima mediterráneo de lluvia invernal de altura
- Clima mediterráneo de lluvia invernal e influencia costera
- Clima mediterráneo frío de lluvia invernal

Propiedades

WGS84/UTM Zone 19S

Escala 1:1.500.000

Esteban Celis

Figura 8. Mapa con las estaciones utilizadas. Detalle clasificación climática según Köppen-Geiger (Sarricolea et al., 2017)

3.2 Medidas de tendencia central y de dispersión

En la Tabla 2 se presenta un resumen de las medidas de tendencia central y dispersión mencionadas en los puntos 2.6 y 2.7. Este se realizó por cada método y estación. En la tabla 2 se aprecia que los valores más altos de medias para la precipitación diaria máxima anual (PDMA) se encuentran en las estaciones de Potrero Grande, Colorado y Bullileo Embalse, valores que se explicarían debido a la cercanía a la precordillera que estas estaciones comparten, a lo que se suma que Bullileo E. se encuentra cercana al embalse del mismo nombre (Pizarro et al., 2013). Al contrario, las estaciones ubicadas dentro de la depresión intermedia son las que presentan valores

promedio de menor envergadura. Con respecto a la mediana, se puede afirmar que el comportamiento es similar a lo expuesto para la media en el párrafo anterior. Cabe destacar que esta medida de tendencia central es un mejor indicador que la media, debido a la naturaleza no paramétrica de los datos. El coeficiente de variación por su parte presenta los valores menores en las estaciones de Colorado y Bullileo Embalse, con valores que varían entre 31% y 34%. En contraste, las estaciones que presentaron coeficientes de variación mayores son las de Potrero Grande y Melozal, con valores que rondan el 50% y el 60%.

Tabla 1. Medidas de tendencia central y de dispersión para cada estación

Estación	Método de medición	Número de datos	Media	Mediana	Desviación Estándar	Coefficiente de Variación
Potrero Grande	PDMAC	38	82.13	82.09	41.12	50%
	PDMAD	38	70.46	71.55	33.35	47%
Pencahue	PDMAC	36	54.38	53.31	23.86	44%
	PDMAD	36	47.07	43.51	20.31	43%
Talca	PDMAC	26	53.29	56.59	22.36	42%
	PDMAD	26	45.68	47.57	17.93	39%
Colorado	PDMAC	38	103.69	105.06	32.43	31%
	PDMAD	38	92.99	89.98	31.18	34%
Melozal	PDMAC	37	58.42	59.25	34.09	58%
	PDMAD	37	50.18	47.41	31.72	63%
Parral	PDMAC	35	71.50	70.82	28.67	40%
	PDMAD	35	60.69	60.57	22.16	37%
Bullileo Embalse	PDMAC	39	144.07	147.20	47.98	33%
	PDMAD	39	126.20	123.51	40.22	32%

PDMAC: Precipitación diaria máxima anual en horario continuo, PDMAD: Precipitación diaria máxima anual en horario discreto.

Para realizar una comparación de las medidas de tendencia central y dispersión entre ambos métodos de medición (medición continua y discreta), se estableció un análisis que estima las diferencias

porcentuales entre ambos métodos, lo que se expresa en la Tabla 3. Para las tres medidas se estableció un indicador definido como sigue:

$$\text{Indicador porcentual} = \frac{(PDMAC - PDMAD) \cdot 100}{PDMAD}$$

Donde PDMAC: Precipitación diaria máxima anual medida en horario continuo; PDMAD: Precipitación diaria máxima anual medida en horario discreto.

Tabla 3. Indicador porcentual (%) para las 3 medidas

Estación	Media	Mediana	Coef. Var.
Potrero Grande	16.56	14.73	5.76
Pencahue	15.51	22.52	1.67
Talca	16.66	18.96	6.88
Colorado	11.51	16.77	-6.73
Melozal	16.43	24.97	-7.68
Parral	17.81	16.92	9.83
Bullileo Embalse	14.16	19.18	4.51

Con respecto a la diferencia porcentual de la media, esta fluctúa entre 11.5% y 17.8% entre todas las estaciones. La mediana por otro lado presenta diferencias porcentuales mayores entre ambos métodos de medición, variando desde 14.7% hasta un 25%. Sin embargo, es importante destacar que las diferencias porcentuales entre ambos métodos de medición entregan siempre valores positivos para las medidas de tendencia central, lo que demuestra que tanto las medias como las medianas obtenidas por mediciones continuas son más altas que aquellas obtenidas con medición discreta de 8AM a 8AM. Al contrario, las diferencias porcentuales presentadas en el coeficiente de variación presentan tanto signos positivos como negativos, lo que indicaría a primera

vista que el método de medición escogido no influye positiva o negativamente en los resultados entregados por el coeficiente de variación. Con el objetivo de profundizar acerca de las diferencias de los métodos de medición (PDMAC y PDMAD), la Tabla 4 indica el promedio de las medidas de tendencia central y dispersión por cada método. En el cuadro se pueden notar las diferencias existentes entre los métodos de medición a modo global, las cuales son 10.6 mm para la media y 12.9 mm para la mediana, y siempre al favor del método continuo, lo que ratifica lo expuesto en el párrafo precedente. En cuanto al coeficiente de variación, este prácticamente no presenta diferencias al momento de elegir diferentes métodos de medición.

Tabla 4. Promedio de las medidas de tendencia central y dispersión por cada método de medición

Estación	Media		Mediana		Coef. Var.	
	PDMAC	PDMAD	PDMAC	PDMAD	PDMAC	PDMAD
Promedio	81.07	70.47	82.05	69.16	43%	42%

3.3 Diagrama de Violín

En términos generales, se aprecia que la distribución de los datos de precipitación diaria, estimados de forma continua o discreta, muestran similitudes para todas las estaciones, a excepción de la estación Bullileo Embalse. Esto porque en esta última estación se aprecian diferencias entre ambos métodos (Figura 9). Así mismo, la distribución de las precipitaciones se encuentra más acotada en el método discreto en comparación con el continuo para todas las estaciones analizadas, lo que se verifica en las gráficas correspondientes. Por otra parte, las

distribuciones de ambos métodos son multimodales, es decir, existen varias modas al interior de la distribución.

3.4 Análisis de concordancia de Bland y Altman (ACBA)

Como ya se mencionó previamente en el punto 2.9, la prueba de Bland y Altman compara los datos de ambas técnicas de medición, en donde una diferencia promedio (dp) más pequeña indicaría que no existen grandes diferencias entre métodos. De acuerdo con Doğan (2018), el análisis de Bland y Altman es un método simple y preciso para cuantificar la

concordancia entre dos variables y puede servir de ayuda a los investigadores al comparar un nuevo método de medición contra otro tradicional. Así

mismo, usar este enfoque facilita hacer un juicio objetivo de las propiedades del método (Cao et al., 2020).

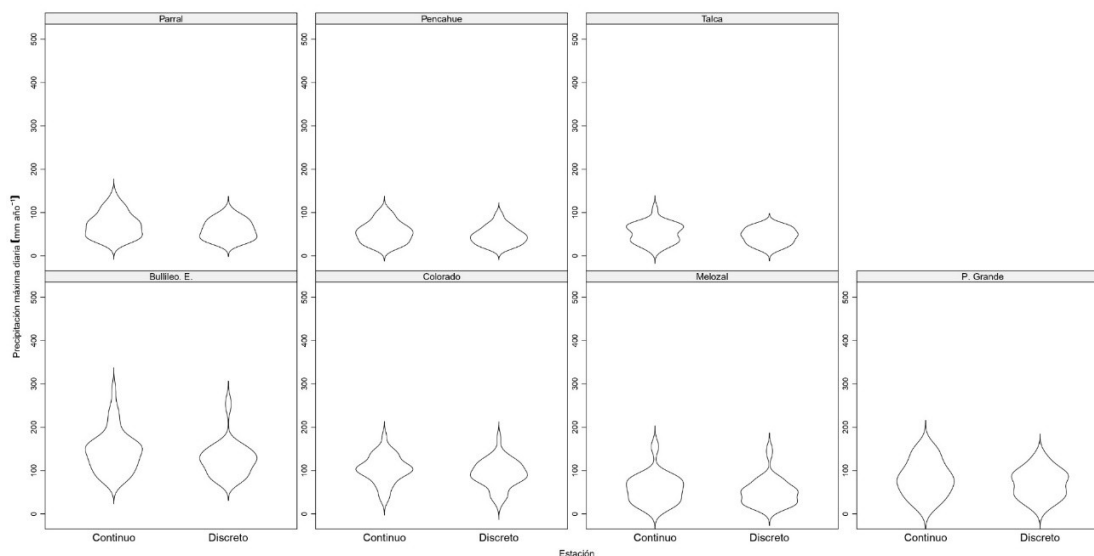


Figura 9. Diagramas de violín para las estaciones

Tabla 5. Prueba de Bland y Altman, entre ambos planteamientos (PDMAC y PDMAD) por cada estación.

Estación	dp	DS	LS	LI
Potrero Grande	11.67	15.24	41.55	-18.20
Pencoahue	7.30	8.67	24.30	-9.69
Talca	7.61	8.40	24.07	-8.85
Colorado	10.70	9.33	29.00	-7.59
Melozal	8.25	8.51	24.92	-8.43
Parral	10.81	11.49	33.33	-11.71
Bullileo Embalse	17.87	19.33	55.74	-20.01

Diferencias promedio (dp), desviación estándar de las diferencias (DS) y límites de concordancia (LS y LI)

En la Tabla 5 se muestran los estadísticos de Bland y Altman. Las mayores diferencias promedio (dp) se manifiestan en las estaciones ubicadas en la zona precordillerana andina, a saber, Potrero Grande, Colorado y Bullileo Embalse. En estas, las dp fluctúan entre los 10.7 mm y 17.87 mm. Por otra parte, las estaciones de Pencoahue, Talca y Melozal presentan valores menores, con diferencias promedios que fluctúan entre los 7.3 mm y 8.3 mm,

ubicándose estas en la depresión intermedia. La estación de Parral denota una dp alta, aunque se ubica en la depresión intermedia, siendo la única de este grupo que manifiesta esta particularidad. De la Figura 10 se puede afirmar en términos generales que, a medida que aumenta el promedio de los métodos comparados, existe una mayor probabilidad de encontrar diferencias mayores entre los métodos de medición.

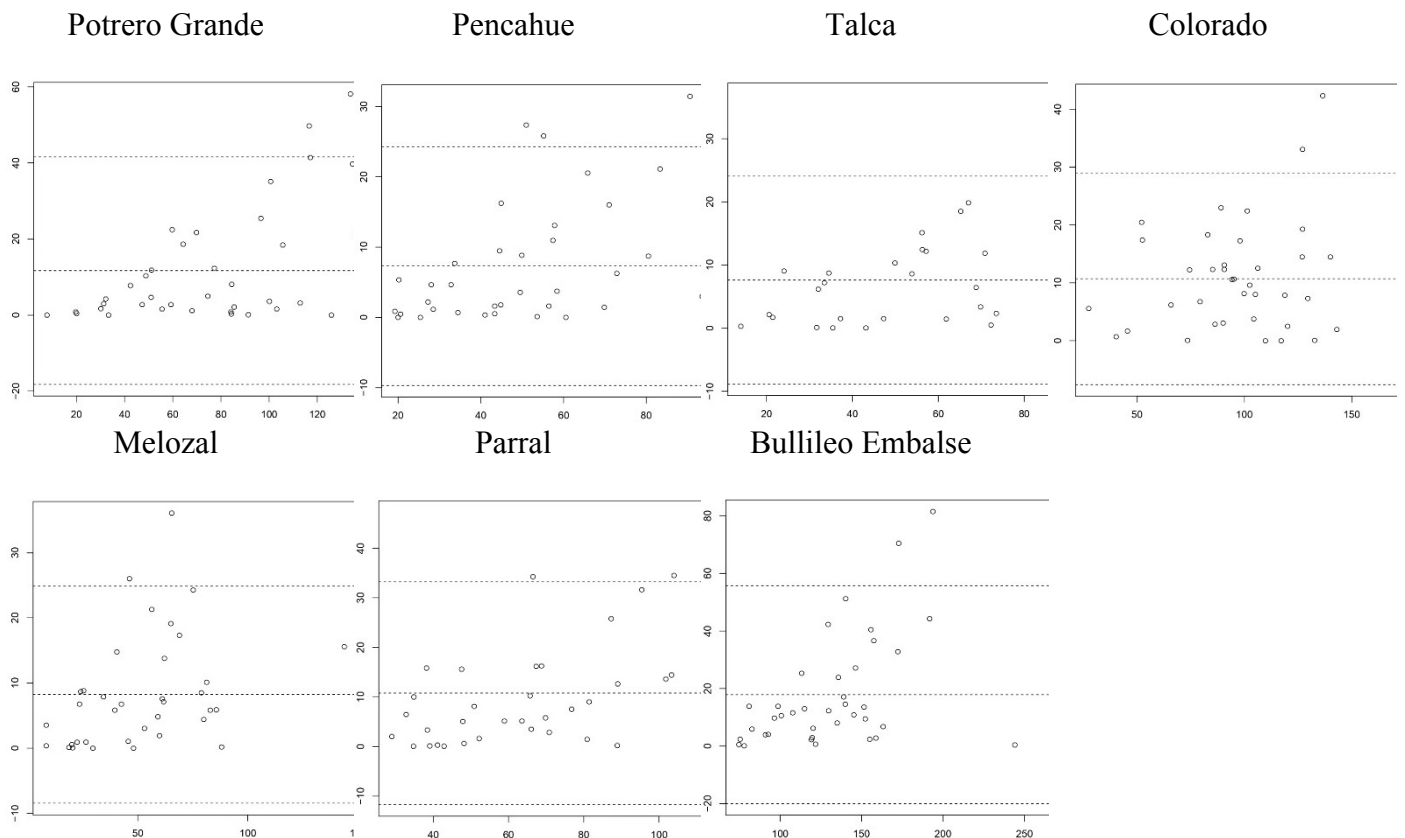


Figura 10. Gráficos de la prueba Bland-Altman por estación; En las ordenadas, la diferencia entre ambos métodos de medición, y en las abscisas, el promedio entre los métodos.

3.5 Prueba U de Mann-Whitney

Mauget (2010) señaló que el método de Mann-Whitney se presenta como una robusta opción en el análisis de series de tiempo, pues es capaz de detectar una extensa gama de variaciones de datos, derivado que establece pocas suposiciones limitantes sobre cómo varían los datos a lo largo del tiempo. Los resultados encontrados al aplicar el estadístico de Mann-Whitney muestran que a nivel macro existen

diferencias estadísticamente significativas ($P=0.0098$) al comparar ambos métodos (Tabla 6). Sin embargo, y a pesar de los prominentes resultados descritos en el punto 3.2 de este estudio, solo Bullileo Embalse presenta diferencias estadísticamente significativas entre los datos continuos y discretos ($p=0.0281$), presentando todas las demás un valor p mayor a 0.05.

Tabla 6. Valor P de U de Mann Whitney al 95% de confiabilidad, para las estaciones en estudio.

Estación	Valor P	Estación	Valor P
Global	0.0098*	Colorado	0.0614
Potrero Grande	0.1091	Melozal	0.0832
Pencahue	0.0919	Parral	0.0656
Talca	0.0893	Bullileo. E.	0.0281*

Ahora bien, si el umbral se cambiara al 10% de significancia ($\alpha=0.1$) la totalidad de las estaciones presentaría diferencias estadísticamente significativas, a excepción de Potrero Grande, como lo expresa la Tabla 6. De lo anterior se desprende que, con un 99% de confiabilidad no existen diferencias significativas entre ambos métodos; que al 95% de confiabilidad, solo Bullileo Embalse presenta diferencias; y que al 90% de confiabilidad todas las estaciones presentan diferencias significativas entre ambos métodos, a excepción de Potrero Grande. Esto determina que dependiendo del tipo de confiabilidad que se quiera aceptar y del nivel de precisión, los dos métodos en análisis podrán ser cuestionados para su uso o no. No obstante lo anterior, en hidrología es altamente utilizado un $\alpha=0.05$, lo que determinaría que en general no hay diferencias entre ambos métodos, a excepción de una de las 7 estaciones. En el presente estudio se utilizaron datos de la misma estación para comparar datos continuos y datos discretos de precipitación en 24 horas, es decir se tomó igual longitud de serie para ambos tipos de datos. Los resultados definieron que el método continuo obtendrá valores más altos de precipitación que el método discreto para una misma serie de tiempo. Sin embargo, esto no ocurre en la realidad para la aplicación del método K o Índice de Tormenta, porque dicho método utiliza información de una estación pluviográfica cercana, para su aplicación en otra estación pluviométrica. Aún más, sucede que generalmente los datos tomados en forma tradicional desde estaciones pluviométricas cuentan con una mayor longevidad que las estaciones pluviográficas, debido a su facilidad y simpleza al momento de registrar datos. Por ende, una estación pluviométrica (registros de 8AM a 8AM), probablemente tendrá una mayor longitud de la serie, y por tanto una mayor probabilidad de encontrar valores mayores de precipitación, lo que podría compensar las carencias del método discreto frente al método continuo. Cabe mencionar que la estaciones pluviográficas utilizadas no cuentan con un pluviómetro a su lado, por lo que fue necesario "simular" las mediciones de un pluviómetro

obteniendo las precipitaciones diarias en un horario fijo y después filtrando las máximas anuales desde los datos del pluviógrafo. Si existiese dicha información sería posible efectuar el mismo análisis, puesto que esas son las diferencias que busca encontrar este estudio, y en parte sería beneficioso pues estaría incluido el factor humano dentro del análisis, lo que sería más apegado a la realidad.

4. CONCLUSIONES

De ambas series de datos obtenidas para las 7 estaciones en análisis, la información de precipitación en 24 horas, generada por el método continuo presenta valores superiores respecto a la información obtenida con el método discreto. Lo anterior es corroborado por las medidas de tendencia central (media y mediana) y el análisis de concordancia de Bland y Altman. Asimismo, el método de Bland y Altman, identifica que las diferencias entre el método continuo y el método discreto para la estimación de la precipitación en 24 horas se acrecientan en aquellas estaciones ubicadas en la precordillera andina, a lo que se suma la estación Parral, que es la más austral de la zona de estudio. Ello permitiría concluir que a mayor altitud hay mayor diferencia entre ambos métodos. Finalmente, mediante la prueba U de Mann-Whitney, se logra determinar que solo la estación Bullileo Embalse presenta diferencias estadísticamente significativas, entre el método discreto y el método continuo, con un 95% de confiabilidad. De acuerdo a los resultados, el método índice de tormenta estaría subestimando la información que este genera (curvas IDF), no obstante, esta diferencia se "minimiza" al tener una mayor cantidad de años lo que aumenta la probabilidad de encontrar tormentas de mayor magnitud. Por otra parte, si el estudio se expande a una mayor cantidad de estaciones sería posible encontrar cierto patrón en las diferencias lo que eventualmente podría llevar a una suerte "corrección" del método índice de tormenta el que proyecte el incremento de las precipitaciones diarias máximas anuales de ciertas estaciones.

REFERENCIAS

Bland, J., & Altman, D. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8(2), 135-160.

- Burn, D. H. (2014). A framework for regional estimation of intensity-duration-frequency (IDF) curves. *Hydrological Processes*, 4209-4218.
- Campos-Aranda, D. F. (2010). Intensidades máximas de lluvia para el diseño hidrológico urbano en la república mexicana. *Ingeniería Investigación y Tecnología*, XI(2), 179-188.
- Canavos, G. (1995). *Probabilidad y Estadística. Aplicaciones y métodos*. México: Mc Graw-Hill.
- Cao, X., Bai, Z., Sun, C., Wen, J., Lin, X., & Sun, W. (2020). Bland-Altman Analysis of Different Radiographic Measurements of the Hallux Valgus Angle and the Intermetatarsal Angle After Distal Osteotomy. *Orthopaedic Surgery*(12), 1621-1626.
- Carrasjo, J., & Jover, L. (2004). *Métodos estadísticos para evaluar la Concordancia*. Barcelona: Med Clin.
- Chow, V. T., Maidment, D. R., & Mays, L. W. (1998). *Hidrología aplicada*. Bogotá: McGraw-Hill.
- Doğan, N. Ö. (2018). Bland-Altman analysis: A paradigm to understand correlation and agreement. *Turkish journal of emergency medicine*(18), 139-141.
- Hintze, J. L., & Nelson, R. D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *American Statistician*, 52(2), 181-184.
- Köppen, W., & Geiger, R. (1936). *Handbuch der Klimatologie*. Berlin: Gebrüder Borntraeger.
- Lind, D. A., Mason, R. D., & Marshall, W. G. (2001). *Estadística para la administración y economía*. México: McGraw-Hill Interamericana.
- Mauget, S. (2010). Time series analysis based on running Mann-Whitney Z Statistics. *Journal of Time Series Analysis*(32), 47-53.
- Mendenhall, W., & Sincich, T. (1997). *Probabilidad y estadística para la ingeniería y ciencias*. México: Prentice-Hall.
- Moncho, J. (2014). *Estadística Aplicada a Las Ciencias de la Salud*. Barcelona: Elsevier.
- Pizarro, R., Ben Ingram, F. G.-L.-P., Sangüesa, C., Delgado, N., García-Chevesich, P., & Valdés, J. (2018). WEBSEidf: A web-based system for the Estimation of IDF Curves in Central Chile. *Hidrology*, 5,40.
- Pizarro, R., Garcia Chevesich, P., Valdes, R., Dominguez, F., Hossain, F., Ffolliott, P., . . . Bro, P. (2013). Inland water bodies in Chile can locally increase rainfall intensity. *Journal of Hidrology*, 481(2), 56-63.
- Pizarro, R., Valdés, R., Abarza, A., & Garcia-Chevesich, P. (2015). A simplified storm index method to extrapolate intensity-duration-frequency (IDF) curves for ungauged stations in central Chile. *Hydrological Processes*, 641-652.
- Sarricolea, P., Herrera-Ossandón, M., & Meseguer-Ruiz, Ó. (2017). Climatic regionalisation of continental Chile. *Journal of maps*, 13(2), 66-73.
- Seoáñez Calvo, M. (2001). *Tratado de climatología aplicada a la ingeniería ambiental*. Madrid: Mundi-Prensa.
- UNESCO. (2013). *Curvas Intensidad Duración Frecuencia para las regiones Metropolitana, Meule y Biobío- Intensidades de 14 minutos a 24 horas*. Montevideo, Uruguay: PHI-LAC, N°29.