

**UNIVERSIDAD NACIONAL AGRARIA
LA MOLINA**

**ESCUELA DE POSGRADO
MAESTRÍA EN RECURSOS HÍDRICOS**



**“FRECUENCIA DE PRECIPITACIONES MÁXIMAS SEGÚN
LOS L-MOMENTOS EN LA CUENCA HIDROGRÁFICA DEL
LAGO TITICACA”**

Presentada por:

CARLOS ANTONIO FERNÁNDEZ PALOMINO

**TESIS PARA OPTAR EL GRADO DE MAGISTER SCIENTIAE EN
RECURSOS HÍDRICOS**

Lima - Perú

2015

AGRADECIMIENTOS

Deseo expresar mi más sincero agradecimiento al Dr. Waldo Sven Lavado Casimiro por su supervisión constante, consejos cruciales, inestimables sugerencias y entera cooperación desde el principio hasta la conclusión de esta tesis.

A todos mis profesores de la maestría en Recursos Hídricos por sus aportes académicos y opiniones constructivas en mi formación profesional. A mis amigos y compañeros de estudio, así como a mis colegas del Servicio Nacional de Meteorología e Hidrología del Perú. Especialmente a Samuelson, Guido, Jennifer, Teófila, Fiorella, Yelko, Víctor, Boris, Jaime y otros, sus discusiones y la amistad fueron valiosos académicamente, espiritualmente y socialmente a lo largo de mis estudios y esto será por siempre apreciado e inolvidable.

Por último, agradezco a toda mi familia por la motivación interminable, en especial a mi madre por su constante e incondicional apoyo, y por depositar su entera confianza hacia mi persona.

ÍNDICE GENERAL

	Pág.
I. INTRODUCCIÓN	01
1.1 Inundaciones en la vertiente del Lago Titicaca	01
1.2 Análisis de frecuencias de precipitaciones máximas	02
1.3 L-momentos	03
1.4 Objetivos del estudio	04
1.5 Importancia del Estudio	04
II. REVISIÓN DE LITERATURA	05
2.1 Análisis de frecuencia de eventos extremos	05
2.2 Análisis exploratorio de datos	06
2.2.1 Diagrama de caja (Boxplot).	06
2.2.2 El Test de Grubbs de detección de outliers	07
2.3 Supuestos del análisis regional	08
2.3.1 Prueba de Mann Kendall (MK) de detección de tendencias	08
2.3.2 Prueba de Pettitt de detección de ruptura de series	09
2.3.3 Prueba de Ljung-Box (Q-estadístico) de detección de autocorrelación	10
2.4 Análisis regional de frecuencias	11
2.5 L-momentos	13
2.6 El método del índice de avenida	14
2.7 Procedimiento del análisis regional	16
2.7.1 Filtrado de datos usando medidas de discordancia	16
2.7.2 Identificación de regiones homogéneas	17
2.7.3 Elección de la distribución de frecuencia	21
2.7.4 Estimación de los cuantiles de frecuencia	23
2.8 Mapeo de índice de avenida y eventos extremos	29
2.8.1 Método de ponderación inverso a la distancia (DIP)	29
2.8.2 Kriging ordinario (KO)	30

III. MATERIALES Y MÉTODOS	34
3.1 Descripción del área de estudio	34
3.1.1 Climatología	35
3.2 Materiales	37
3.2.1 Información básica de sistema de información geográfica	37
3.2.2 Información de datos observados	38
3.2.3 Programas	41
3.3. Metodología	42
3.3.1 Análisis exploratorio de datos	44
3.3.2 Supuestos del análisis regional de frecuencias	44
3.3.3 Filtrado de las observaciones empleando la medida de discordancia	45
3.3.4 Identificación de regiones homogéneas	46
3.3.5 Selección de una función de distribución de frecuencia para la región	46
3.3.6 Estimación de los cuantiles regionales de frecuencia y locales	46
3.3.7 Comparación con el método de análisis local	47
3.3.8 Mapeo de índice de avenida y eventos extremos	47
IV. RESULTADOS Y DISCUSIÓN	49
4.1 Análisis exploratorio de datos	49
4.1.1 Diagramas de Boxplots	49
4.1.2 Prueba de Grubbs para valores atípicos	50
4.2 Supuestos del análisis regional de frecuencias	52
4.2.1 Análisis de tendencia	52
4.2.2 Análisis de homogeneidad	54
4.2.3 Análisis de independencia serial	55
4.3 Filtrado de datos usando medidas de discordancia	56
4.4 Heterogeneidad regional y la formación de regiones homogéneas	59
4.5 Selección de la distribución regional de mejor ajuste	61
4.6 Estimación de los cuantiles de frecuencia	63
4.6.1 Estimación de la curva de crecimiento regional	63
4.6.2 Decisión acerca de la mejor curva de crecimiento regional	64
4.6.3 Cuantiles de frecuencia regional y local	67

4.7	Comparación con el método de análisis local	71
4.8	Mapeo de índice de avenida y eventos extremos	76
4.8.1	Mapeo de índice de avenida por Kriging ordinario (KO)	76
4.8.2	Mapeo de índice de avenida por distancia inversa ponderado (DIP)	79
4.8.3	Mapeo de eventos extremos	82
V.	CONCLUSIONES	85
VI.	RECOMENDACIONES	86
VII.	REFERENCIAS BIBLIOGRÁFICAS	87
VIII	ANEXO	93
8.1	Distribuciones teóricas y sus relaciones	93
8.2	Cuantiles de precipitación máxima en 24 horas estimadas por ARF	96
8.3	Comparación de los cuantiles estimados por el análisis local y ARF	105

ÍNDICE DE CUADROS

Cuadro 1:	Valores críticos de la discordancia D_i	17
Cuadro 2:	Aproximaciones polinómicas de $L - C_k$ en función de $L - C_s$	22
Cuadro 3:	Algoritmo para la simulación del Algoritmo de L-momento regional	28
Cuadro 4:	Red de estaciones Meteorológicas	39
Cuadro 5:	Estadísticas descriptivas de la estación Ananea	50
Cuadro 6:	Estadístico Z de test de Grubbs para la estación Ananea	50
Cuadro 7:	Datos atípicos detectados por el test de Grubbs	51
Cuadro 8:	Resumen del test de Mann Kendall para el análisis de tendencias	53
Cuadro 9:	Resumen del test de Pettit para el análisis de homogeneidad	54
Cuadro 10:	Valores de los ratios de L – momentos de las estaciones	57
Cuadro 11:	Valores de la medida de discordancia de las estaciones	58
Cuadro 12:	Estadísticas de medias ponderadas de L-momentos regionales	60
Cuadro 13:	Parámetros de la distribución Kappa para la región	60
Cuadro 14:	Medidas de heterogeneidad de la región sin la estación discordante	60
Cuadro 15:	Estadísticas de heterogeneidad de la región con y sin estación discordante	61
Cuadro 16:	Estadístico Z^{DIST} para varias distribuciones para la región	62
Cuadro 17:	Parámetros y cuantiles regionales para las dos distribuciones de mejor ajuste	63
Cuadro 18:	Resultados de la simulación para la curva de crecimiento regional	66
Cuadro 19:	Resultado de cuantiles para la curva de crecimiento regional	67
Cuadro 20:	Índice de avenida [mm] de las estaciones de la región	68
Cuadro 21:	Parámetros del semivariograma empírico del modelo exponencial	77
Cuadro 22:	Validación cruzada para Kriging ordinario	79
Cuadro 23:	Validación cruzada para distancia inversa ponderado	80

Cuadro 24: Parámetros del semivariograma empírico en la interpolación de eventos extremos	82
Cuadro 25: Fórmulas para estimación de parámetros por el método de máxima verosimilitud para la distribución Gumbel	93
Cuadro 26: Fórmulas por distribución, L-momentos y su ratios para las distribuciones analizadas.	94
Cuadro 27: Coeficientes de las aproximaciones para GNO	96
Cuadro 28: Coeficientes de las aproximaciones para PE3	96
Cuadro 29: Cuantiles de precipitación máxima en 24 horas estimadas por ARF para cada estación en mm y sus medidas de precisión	96
Cuadro 30: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 2 y 5 años.	105
Cuadro 31: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 10 y 20 años	106
Cuadro 32: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 50 y 100 años	106
Cuadro 33: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 200 y 500 años	107
Cuadro 34: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 1000 años	108

ÍNDICE DE FIGURAS

Figura 1: Comportamiento típico de un semivariograma acotado con una representación de los parámetros básicos. SEMEXP corresponde al semivariograma experimental y MODELO al ajuste de un modelo teórico	32
Figura 2: Mapa de ubicación de la cuenca hidrográfica del Lago Titicaca	34
Figura 3: Climatología del Altiplano Peruano (periodo 1961-1990). a) Climatología de la precipitación observada y b) Climatología de la temperatura observada.	37
Figura 4: Estaciones con registro de precipitación máxima diaria en el tiempo.	40
Figura 5: Ubicación de las estaciones meteorológicas.	40
Figura 6: Diagrama esquemático de la metodología.	43
Figura 7: Gráficos boxplots para detección de outliers.	49
Figura 8: Autocorrelaciones detectadas por el test de Ljun Box.	55
Figura 9: Ratios de los L – momentos de las estaciones.	57
Figura 10: Medidas de discordancia de las estaciones.	59
Figura 11: Diagrama de L-momentos ratio con el L-momentos ratio regional para las distribuciones candidatas.	62
Figura 12: Curva de crecimiento regional para distribuciones candidatas	64
Figura 13: Correlación entre pares de estaciones significantes a un nivel de 0.05.	65
Figura 14: Representación gráfica de la curva de crecimiento regional	68
Figura 15: a) Distribución espacial de estaciones definitivas analizadas en ARF y Cuantiles de PMA estimadas por el ARF para las estaciones de b) Munañi, c) Taraco, d) Laraqueri, e) Desaguadero f) Capazo.	70
Figura 16: Gráfica de los cuantiles estimados por ARF versus AL para varios periodos de retorno.	73
Figura 17: Comparación gráfica de los cuantiles estimados por ARF y AL para periodos de retorno de 2, 5, 10, 20, 50, 100, 200, 500 y 1000 años	74

Figura 18: Desviación relativa (%) entre los cuantiles de PMA estimados por el método local y regional para 1000 años de periodo de retorno.	75
Figura 19: Diferencias entre los cuantiles estimados por Gumbel respecto a GNO	75
Figura 20: a) Distribución espacial de índice de avenida. b) Frecuencias de índice de avenida. c) Gráfico normal de cuantil – cuantil de índice de avenida.	77
Figura 21: Semivariograma experimental direccional (suroeste – noreste) ajustado por el modelo exponencial	78
Figura 22: a) Mapa de índice de avenida para vertiente del Lago Titicaca en mm. b) Mapa de incertidumbre del índice de avenida	81
Figura 23: Mapa de precipitación máxima de 24 horas para tiempos de retorno de 2, 5, 10 y 20 años en mm.	83
Figura 24: Mapa de precipitación máxima de 24 horas para tiempos de retorno de 50, 100, 200 y 500 años en mm.	84
Figura 25: Cuantiles para las estaciones de Arapa, Cabanillas, Capachica, y Capazo.	100
Figura 26: Cuantiles para las estaciones de Chilligua, Chuquibambilla, Cojata, Cuyo Cuyo, Desaguadero y Huancané.	101
Figura 27: Cuantiles para las estaciones de Huaraya Moho, Ichuña, Ilave , Juli, Lagunillas y Lampa.	102
Figura 28: Cuantiles para las estaciones de Laraqueri, Llally, Macusani, Mazo Cruz, Muñani, y Pampahuta	103
Figura 29: Cuantiles para las estaciones de Pizacoma, Progreso, Pucará, Puno, Putina y Santa Rosa.	104
Figura 30: Cuantiles para las estaciones de Taraco.	105

“FRECUENCIA DE PRECIPITACIONES MÁXIMAS SEGÚN LOS L-MOMENTOS EN LA CUENCA HIDROGRÁFICA DEL LAGO TITICACA”

RESUMEN

Las precipitaciones extremas suelen dar lugar a la ocurrencia de eventos de inundación con las consiguientes pérdidas de vidas, de producción agrícola y de infraestructura en la Vertiente del Lago Titicaca (VT). Sin embargo, una comprensión de la frecuencia de ocurrencia de estos eventos extremos, ya sea para fines de diseño o planificación de desastres, es a menudo limitada por la escasa disponibilidad y calidad de los datos en la escala temporal y espacial deseada. Por ello, en esta tesis se desarrolla el análisis regional de frecuencia (ARF) de los extremos de precipitación en 24 horas, empleando L-momentos y el procedimiento del índice de avenida para la estimación de los cuantiles de precipitación en la VT (33 estaciones pluviométricas seleccionadas). Inicialmente se realizaron los análisis detallados de la calidad y verificación de los supuestos del análisis de frecuencias de las series de precipitaciones máximas, aplicándose para ello diferentes pruebas de verificación de datos atípicos, tendencia, homogeneidad e independencia de la serie de tiempo. El método de L-momentos a través de la medida de heterogeneidad, permite definir objetivamente que la vertiente del Lago Titicaca conforma una sola región hidrológicamente homogénea en cuanto a la frecuencia de las precipitaciones máximas. Según el diagrama de L-momento ratios y Z-estadístico, las distribuciones de valor extremo generalizado (GEV) y normal generalizado (GNO) son las mejores opciones para las estimaciones de los cuantiles regionales. Sin embargo, un criterio de análisis de robustez basado en simulaciones de Monte Carlo revela que GNO es ligeramente más robusto que la distribución GEV para estimar los cuantiles de diseño en grandes periodos de retorno (50 años a más). Se constató que la aplicación en la VT de la función de distribución Gumbel en el análisis local de frecuencia de las precipitaciones máximas conlleva a una sobrestimación de los cuantiles estimados. Finalmente se generó el mapa de índice avenida, válido para las estimaciones del índice de avenida en sitios sin medición de precipitación en la VT.

PALABRAS CLAVES: *Precipitaciones máximas, análisis regional de frecuencias, L-momentos, Titicaca.*

“FREQUENCY OF MAXIMUM PRECIPITATIONS ACCORDING TO THE L-MOMENTS IN THE WATERSHED OF TITICACA LAKE”

ABSTRACT

Extreme precipitation often leads to the occurrence of flood events with consequent loss of life, agricultural production and infrastructure in the Vertiente of Titicaca Lake (VT). However, an understanding of the frequency of occurrence of extreme events, either for design purposes or disaster planning, is often limited by poor availability and quality of data in the desired temporal and spatial scale. Therefore, this thesis develops the regional frequency analysis (ARF) of extreme precipitation in 24 hours, using L-moments and the index flood proceeding for estimating quantiles of precipitation in VT (33 raingauge stations selected). Initially detailed quality analysis and verification of the assumptions of frequency analysis of the series of maximum precipitation were performed, applying for this different verification tests for outliers data, trend, homogeneity and independence of the time series. The method L-moments through the measure of heterogeneity, allows to define objectively that Vertiente of Titicaca Lake forms one hydrologically homogeneous region in terms of the frequency of maximum precipitation. According to diagram L-moment ratios and Z-statistic, generalized extreme value (GEV) and generalized normal (GNO) distributions are the best choices for estimates of regional quantiles. However, a robustness analysis approach based on Monte Carlo simulations reveals that GNO is slightly more robust than GEV distribution to estimate quantiles in large design return periods (50 years to more). It is found that the application at VT of the Gumbel distribution function in the local frequency analysis of maximum precipitation leads to an overestimation of the estimated quantiles. Finally, the index flood map was generated, valid for index flood estimates in sites without measuring precipitation in VT.

KEYWORDS: *Maximum precipitation, regional frequency analysis, L-moments, Titicaca.*

I. INTRODUCCIÓN

1.1 Inundaciones en la vertiente del Lago Titicaca

La sociedad humana se enfrenta a grandes problemas debido a los fenómenos ambientales extremos. Por ejemplo, inundaciones, tormentas, sequías y otros; que destruyen casi todo lo que está en sus inmediaciones en el momento de las apariciones. Según el primer reporte de World Water Assessment Programme (WWAP, 2003) los eventos más extremos en la cuenca de la vertiente del Lago Titicaca están relacionados con las condiciones de riesgo de inundación alrededor del Lago Titicaca, la sequía al sur e incidencia de granizo y las heladas durante todo año. También mencionan que la región en estudio tiene como principal actividad económica la agricultura y dada la pobreza general en el área, es particularmente vulnerable a los fenómenos extremos como sequías e inundaciones; donde las inundaciones tienen un impacto significativo en la economía de la zona aunque el riesgo de pérdidas humanas es mínimo debido a la lenta crecida de las aguas de inundación en las llanuras y planicies, los daños económicos para la producción agrícola, pecuaria y la infraestructura durante un período de diez años se han estimado en 890 000 dólares y además las pérdidas por inundaciones en un período de veinticinco años se estiman en 1 506 000 dólares. Estos montos son enormes dada la extrema pobreza de la región.

Es frecuente los sucesos de inundaciones en el área de estudio, por ejemplo en la página web de la Autoridad Nacional del Agua (ANA, 2012) se encuentra el resumen de eventos y daños ocurridos para la temporada de precipitación en el año 2012, donde en la Administración Local del Agua de Juliaca se identifica que ocurrió dos eventos de inundaciones, siendo los días tres y nueve de enero, la misma que produjo daños afectando 24 viviendas y 94 hectáreas de cultivos de pan llevar. De lo anterior no existe una valoración de la pérdida económica, sin embargo se refleja los daños que ocasiona la inundación, siendo la causa fundamental la incidencia de las fuertes precipitaciones y la gran concentración resultante de la escorrentía, lo que excede la capacidad del río. Por tanto conocer con qué frecuencia se produce la ocurrencia de una inundación de una

magnitud dada es de gran importancia, para planificar y prevenir desastres por inundación.

1.2 Análisis de frecuencias de precipitaciones máximas

La estimación de la magnitud y frecuencia de los eventos extremos de variables hidrometeorológicas, tales como la precipitación máxima diaria es fundamental en el diseño de estructuras de los recursos hídricos, zonificación de inundación y estimación económica de los proyectos de protección contra inundaciones en llanuras y planicies.

A menudo, el interés se centra en los raros eventos con períodos de retorno (T), de más de 50 años. Esto se debe principalmente a su naturaleza destructiva de la actividad socioeconómica y de las infraestructuras. Sin embargo, la estimación fiable de tales eventos extremos requiere largos registros de observaciones en las estaciones, para ser caracterizados por una función de distribución.

Tradicionalmente, la mayoría de los métodos de análisis de frecuencias extremas se basan en el uso local de las series de observación, siendo el uso cotidiano de la función de distribución biparamétrica de valores extremos de Gumbel, en los diferentes ámbitos relacionados con la ingeniería (Álvarez *et al.*, 1999). Incluso con enfoques regionales como en recientes estudios realizados. Por ejemplo, Luna y Domínguez (2013), probaron diversas distribuciones de 2 parámetros para el análisis regional de frecuencias aplicando la técnica de estaciones-año; los mejores resultados se obtuvieron con la función de distribución Gumbel, para los Andes bolivianos.

La escasa disponibilidad en el tiempo y la calidad de datos son a menudo un desafío en muchas partes del mundo, sin excepción del Perú, por lo que la estimación de los cuantiles de frecuencia de los eventos extremos puede no siempre justificarse a partir del uso de los datos locales. Para superar este problema, diversos enfoques que utilizan fuentes de datos alternativos o adicionales, se han ideado. En esta investigación se desarrolla uno de ellos el denominado análisis regional de frecuencias (de aquí en adelante ARF), que según Hosking y Wallis (1997), este problema puede resolverse “paliando la carencia de datos en el tiempo con su abundancia en el espacio”, pues son muchas las situaciones en las que se dispone de observaciones en diferentes estaciones cercanas entre sí. En esta idea se cimienta el ARF; que constituye una alternativa útil y toma en consideración las deficiencias de los métodos de análisis local, aumentando la longitud de los registros

disponibles mediante la transferencia de información entre las diferentes estaciones que componen una supuesta región hidrológicamente homogénea.

1.3 L-momentos

Recientemente, los hidrólogos están aplicando el método más popularizado de ARF que compromete el empleo del método de los momentos lineales (L-momentos) conjuntamente con el uso del procedimiento de *índice de avenida* (Hosking y Wallis, 1997; Saf, 2008; citados por Alem, 2011). Se trata de una metodología robusta que aún en presencia de datos atípicos (*outliers*), y utilizando correctamente puede conducir a estimaciones suficientemente precisas y de gran utilidad en diversos problemas reales en el amplio rango de la ingeniería (Hosking y Wallis, 1997).

El papel principal de los L-momentos es para la estimación de los parámetros de las distribuciones de probabilidad. Los L-momentos son análogos a los momentos convencionales. Sin embargo, Hosking y Wallis (1997) establecen que los L-momentos tienen capacidades superiores a momentos convencionales en la discriminación entre diferentes distribuciones. Asimismo, constatan mayor robustez en presencia de valores atípicos y menor sensibilidad a la asimetría u observaciones extremas que los momentos convencionales.

Hosking *et al.*, (1985) y Hosking y Wallis (1987), encontraron que con muestras pequeñas y moderada, el método de los L-momentos con frecuencia es más eficiente que de máxima verosimilitud. Estos resultados son para los estimadores basados en una sola muestra de datos y que no están directamente relacionados con el análisis de frecuencia regional. Sin embargo, demuestran que el método de L-momentos produce estimaciones eficientes y computacionalmente convenientes de los parámetros y cuantiles, por tanto se espera razonablemente que estas propiedades continúen si se aplica en los procedimientos de índices de avenida para el análisis de frecuencia regional.

1.4 Objetivos del estudio

El principal objetivo de este estudio consiste en analizar la distribución de frecuencias de las precipitaciones máximas para las regiones hidrológicamente homogéneas en la vertiente del Lago Titicaca, empleando L-momentos en base al enfoque del índice de avenida.

Los objetivos específicos que permiten lograr el objetivo general del estudio son:

- Identificar las regiones hidrológicamente homogéneas en la vertiente del Lago Titicaca.
- Identificar una apropiada función de distribución que relacione la magnitud de las precipitaciones máximas con su frecuencia de ocurrencia en la vertiente del Lago Titicaca.
- Comparar los resultados del análisis regional; con los resultados que se obtendría a partir de un análisis local.
- Determinar el mapa de índice de avenida y el mapa de eventos extremos de precipitación para distintos periodos de retorno.

1.5 Importancia del Estudio

Los resultados de este estudio son de beneficio para los estudios hidrológicos. Los beneficiarios directos del estudio son los ingenieros e hidrólogos que trabajan en las áreas de la ingeniería e investigación. La estimación bastante precisa de la ocurrencia de los eventos extremos de precipitación, permitirá una mejor estimación de crecidas o inundaciones, que, no solo va dirigido a la prevención de posibles riesgos debido al desbordamiento de los ríos (catástrofes), sino también ayuda al estado a planificar de manera adecuada considerando un análisis estadístico robusto.

II. REVISIÓN DE LITERATURA

2.1 Análisis de frecuencia de eventos extremos

Álvarez *et al.*, (1999) menciona que el análisis de frecuencias es la estimación de los sucesos extremos (precipitaciones máximas, avenidas, etc.) correspondientes a diferentes períodos de retorno mediante el uso de funciones de distribución de probabilidad. La relación resultante entre las magnitudes de los eventos extremos y sus correspondientes períodos de retorno, se le conoce como curva de frecuencia y es de gran utilidad en el diseño en ingeniería.

Chow (1964) menciona que el análisis de frecuencia de datos hidrológicos comienza con el tratamiento de datos brutos y finalmente determina la frecuencia o probabilidad de un valor de diseño.

Para efectuar el análisis de frecuencias con enfoque local o regional, se debe cumplir tres suposiciones:

- Los procesos naturales son estacionarios o libres de tendencia con respecto al tiempo.
- Las observaciones en cualquier estación dado están idénticamente distribuidas o son homogéneas.
- Las observaciones en cualquier estación dado son serialmente independiente.

En este estudio para verificar la estacionariedad de la serie se prueba el test de tendencias y para corroborar si todos los valores de la muestra, provienen estadísticamente de una población se emplea el test de homogeneidad.

La prueba de independencia es utilizada para demostrar que los valores que conforman la serie son aleatorios. Esta afirmación implica que la probabilidad de ocurrencia de cualquier valor de la serie no depende de la ocurrencia del o de los valores precedentes, y no afecta de ninguna manera a la probabilidad de ocurrencia de los datos posteriores.

2.2 Análisis exploratorio de datos

La primera condición esencial de cualquier análisis de datos estadísticos es comprobar que las observaciones de la serie de datos sean adecuadas para el análisis de frecuencia. Para el caso específico de observaciones de variables hidrológicas para las que se ha de realizar un análisis de frecuencia, dos tipos de errores son habituales en las muestras:

Primero, que los valores de los datos puedan ser incorrectos, debido a posibles fallos en la grabación o transcripción, que conducen a valores incorrectos y pone en duda a la muestra para cualquier análisis de frecuencia posterior. En segundo lugar, las circunstancias en que se recogieron los datos pueden haber cambiado con el tiempo, como puede ser: el reemplazo de los aparatos de medición o su traslado a una ubicación diferente, manifestación de tendencias a través del tiempo que pueden haber surgido a partir de los cambios ambientales y climáticos en el entorno donde se ubica el dispositivo de medición. Esto significa que la muestra de observaciones no es homogénea o constante en el tiempo, y el análisis de frecuencia de los datos no será una base válida para estimaciones de las mediciones futuras en el sitio.

En el contexto de este estudio el análisis exploratorio consiste en el análisis de datos atípicos (*outliers*) mediante técnicas visuales utilizando el diagrama de caja (*boxplot*). Los supuestos datos *outliers* visualmente detectados son verificados estadísticamente con el test de Grubbs.

2.2.1 Diagrama de caja (*Boxplot*).

La regla de boxplot es un test visual para reconocer outliers. El rango intercuartil es incluido dentro de una caja y los intervalos de confianza de 5% y 95% son indicados con barras de error fuera de la caja. Aquellos valores que caen fuera de los intervalos de confianza son candidatos a ser outliers (Iglewicz y Hoaglin, 1993).

Límite para el intervalo de confianza al 95%:

$$\frac{x - Q_3}{Q_3 - Q_1} > L_5 \quad (1)$$

Límite para el intervalo de confianza al 5%:

$$\frac{Q_1 - x}{Q_3 - Q_1} > L_l \quad (2)$$

$$L_l = Q_1 - (Q_3 - Q_1) \times 1.5 \quad (3)$$

$$L_s = Q_3 + (Q_3 - Q_1) \times 1.5 \quad (4)$$

Donde L_l, L_s es el valor crítico en la regla de boxplot, x es el punto fuera del límite más alto o más bajo en la regla de boxplot y Q_i = Cuartil i-ésimo.

2.2.2 El Test de Grubbs de detección de outliers

El Test de Grubbs es recomendado por la Agencia de Protección Medioambiental de los Estados Unidos (*Environmental Protection Agency - EPA*), como un test estadístico para outliers (EPA, 1992). La EPA recomienda tomar los datos medioambientales en logaritmos, que son a menudo distribuidos log-normal. Los datos son clasificados en orden ascendente y se calcula la media y la desviación típica. Aquellos puntos más altos o más bajos pueden ser declarados estadísticamente como un outlier. Este test se encuentra en Barnett y Lewis (1984).

El estadístico Z para el valor más pequeño es:

$$Z = \frac{\bar{X} - X_1}{s} \quad (5)$$

El estadístico Z para el valor más grande es:

$$Z = \frac{X_n - \bar{X}}{s} \quad (6)$$

Donde s es la desviación típica.

El estadístico Z , También denominado G es comparado con un valor crítico de Z o $G_{crítico}$ según el tamaño de la muestra y el nivel de significación α seleccionado. Si el estadístico Z es más grande que el Z crítico, la hipótesis nula es rechazada y la conclusión es que el dato bajo consideración es un outlier.

2.3 Supuestos del análisis regional

Verificar los supuestos del análisis de frecuencia regional, es una práctica habitual para poner a prueba la hipótesis de que las observaciones en diversos sitios son estacionarios (libres de tendencias), independientes y homogéneos y/o idénticamente distribuidos.

2.3.1 Prueba de Mann Kendall (MK) de detección de tendencias

La prueba de MK, también llamada prueba de Kendall tau por Mann (1945) y Kendall (1975), es la prueba no paramétrica basada en el rango para evaluar la importancia de una tendencia, y ha sido ampliamente utilizado en detección de tendencia en estudios hidrológicos. La hipótesis nula H_0 : No existe una tendencia en la serie $\{X_i, i = 1, 2, \dots, n\}$. La hipótesis alternativa es H_a : Hay una tendencia en la serie.

El estadístico S de Tau de Kendall se define como sigue (Yue *et al.*, 2002).

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(X_j - X_i) \quad (7)$$

Donde el X_j son los valores de datos secuenciales, n es la longitud del conjunto de datos, y:

$$\text{sgn}(\theta) = \begin{cases} 1 & \text{si } \theta > 0 \\ 0 & \text{si } \theta = 0 \\ -1 & \text{si } \theta < 0 \end{cases} \quad (8)$$

Mann (1945) y Kendall (1975) han documentado que cuando $n \geq 8$, la estadística S es aproximadamente normal distribuido con la media y la varianza de la siguiente manera:

$$E(S) = 0 \quad (9)$$

$$V(S) = \frac{n(n-1)(2n+5) - \sum_{m=1}^n t_m m(m-1)(2m+5)}{18} \quad (10)$$

Donde t_m es el número de vínculos de grado m . La prueba estadística estandarizada Z se calcula:

$$Z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & S > 0 \\ 0 & S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & S < 0 \end{cases} \quad (11)$$

El valor de probabilidad P de la estadística S de MK para datos de la muestra se puede estimar usando la función de distribución acumulativa normal como:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z e^{-t^2/2} dt \quad (12)$$

Si los datos de la muestra están correlacionados serialmente, será necesario blanquear previamente los datos y aplicar una corrección para calcular la varianza. Una descripción a detalle del pre blanqueo para la detección de tendencias en series hidrológicas significativamente correlacionados se encuentra en el estudio de Yue *et al.*, (2002).

2.3.2 Prueba de Pettitt de detección de ruptura de series

La prueba de Pettitt se basa en la prueba de Mann-Whitney, donde una ruptura puede ser definida de manera general por un cambio en la ley de probabilidad $f(x)$ de una serie cronológica en un instante dado (normalmente desconocido). Según CERESTA (1986) el fundamento de esta prueba es el siguiente, la serie estudiada es dividida en dos sub-muestras de tamaño m y n respectivamente. Los valores de las dos muestras son reagrupados y ordenados en forma creciente.

Se calcula entonces, la suma de los rangos de los elementos de cada sub-muestra respecto a la muestra total. Un estadístico es definido a partir de estas dos sumas y se prueba una hipótesis nula en donde se asume que las dos sub-muestras pertenecen a la misma población. Esta prueba fue modificada por Pettitt (Pettitt, 1979), de esta forma, la hipótesis nula de la prueba denota la ausencia de ruptura en la serie. Esto es:

$$D_{ij} = \text{sgn}(x_i - x_j) \quad (13)$$

$$\text{sgn}(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases} \quad (14)$$

Se considera la variable $U_{t,N}$ tal que:

$$U_{t,N} = \sum_{i=1}^t \sum_{j=t+1}^N D_{ij} \quad (15)$$

Sea K_N la variable definida por el máximo valor absoluto del estadístico $U_{t,N}$ para t que varía de 1 a $N-1$. Si k designa el valor de K_N tomado de la serie estudiada, con la hipótesis nula, la probabilidad de excedencia del valor k estará dada aproximadamente por:

$$\text{Pr ob}(K_N > k) \approx 2 \exp\left[\frac{-6k^2}{N^3 + N^2}\right] \quad (16)$$

Para un riesgo α de primer orden, si $\text{Pr ob}(K_N > k)$ es inferior a α , la hipótesis nula es rechazada.

2.3.3 Prueba de Ljung-Box (Q-estadístico) de detección de autocorrelación

La función de autocorrelación (FAC) y la función de autocorrelación parcial (FAP) son herramientas cualitativas útiles para evaluar la presencia de autocorrelación en retardos individuales. El Q-Ljung-Box es una forma más cuantitativa para probar la autocorrelación en múltiples retardos conjuntamente Ljung y Box (1978).

Para comprobar la correlación serial de la serie de precipitaciones máximas diarias para todos los sitios, una autocorrelación basado en Q-estadístico de Ljung y Box se aplica. La hipótesis nula para esta prueba es que los primeros m autocorrelaciones son conjuntamente a cero. $H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0$

La estadística de prueba Ljung-Box está dada por:

$$Q(m) = N(N+2) \sum_{h=1}^m \frac{\rho_h^2}{N-h} \quad (17)$$

Donde N es el tamaño de la muestra, ρ_h es la autocorrelación de la muestra en el retraso h y m es el número de retardos en prueba no debe ser más que $N/4$ (Box *et al.*,

1994). Por nivel de significación α , la región crítica para el rechazo de la hipótesis de aleatoriedad es:

$$Q(m) > X_{1-\alpha, m}^2 \quad (18)$$

Donde $X_{1-\alpha, m}^2$ es la cuantil α de la distribución chi-cuadrado con m grados de libertad.

2.4 Análisis regional de frecuencias

La teoría del análisis regional de frecuencias basado en los L-momentos que se describen en la presente tesis se fundamenta en Hosking y Wallis (1997).

El análisis probabilístico de eventos extremos de carácter regional permite estimar estas en cualquier sitio dentro de la región o zona estudiada, con base en todos los datos registrados en las diferentes estaciones de medición, incluyendo aquellos de un sitio específico, cuando tales datos están disponibles. Algunos métodos del análisis regional, requieren que la zona donde se aplican, esté integrada por estaciones (pluviométricas, hidrológicas, etc.) cuyo comportamiento estadístico – hidrológico es homogéneo en alguna manera cuantificable a nivel interanual. El análisis regional explota esta homogeneidad para producir estimaciones las cuales son, en la mayoría de los casos, más confiables que aquéllas obtenidas al hacer uso únicamente de los datos de una sola estación. Los principios de análisis regional de frecuencia, sin embargo, se aplican cuando varias muestras de datos similares están disponibles.

Supóngase que las observaciones se realizan a intervalos regulares en algún sitio de interés. Sea Q la magnitud del evento que ocurre en un momento dado en un sitio dado, entonces se considera a Q como una cantidad aleatoria (una variable aleatoria), potencialmente puede tomar cualquier valor entre cero y el infinito. La cantidad fundamental del análisis de frecuencia estadística es la distribución de frecuencias, que especifica cómo ocurren frecuentemente los valores de Q . Denotemos por $F(x)$ la probabilidad de que el valor actual de Q es a lo más x :

$$F(x) = \Pr[Q \leq x] \quad (19)$$

$F(x)$ es la función de distribución acumulada de la distribución de frecuencia. Su función inversa $x(F)$ es la función cuantil de la distribución de frecuencia, expresa la magnitud de un evento en términos de su probabilidad de no excedencia F . El cuantil Q_T del período de retorno T , es un evento de magnitud tan extrema que tiene una probabilidad $1/T$ de ser excedido por cualquier evento único. Para un evento extremo alto, en la cola superior de la distribución de frecuencias, Q_T está dada por:

$$Q_T = x(1-1/T) \quad (20)$$

$$F(Q_T) = 1-1/T \quad (21)$$

El objetivo del análisis de frecuencia es obtener una estimación útil del cuantil Q_T para uno o para una serie de períodos de retorno. A su vez para ser útil una estimación no sólo debe estar cerca del verdadero cuantil sino también debe de venir con una evaluación de precisión.

Si los datos están disponibles en el sitio de interés, entonces los datos observados brindan una muestra de las realizaciones de Q . En muchas aplicaciones medioambientales el tamaño de la muestra es escasamente suficiente para permitir que la estimación de los cuantiles sea fiable. Se suele considerar que el cuantil de período de retorno T puede ser estimado con fiabilidad de un registro de datos de longitud n sólo si $T \leq n$; sin embargo, en muchas aplicaciones de ingeniería basados en datos anuales (por ejemplo, la precipitación máxima diaria) esta condición rara vez se satisface, donde típicamente $n < 50$ y T entre 50 a 1000 años. Para superar este problema, diversos enfoques se han ideado que utilizan fuentes de datos alternativos o adicionales. En esta investigación se desarrolla uno de ellos; denominado el análisis regional de frecuencias.

El análisis regional de frecuencia aumenta los datos para el sitio de interés mediante el uso de datos de otros sitios que se considera que tienen distribuciones de frecuencia similares a la del sitio de interés; excepto de un factor de escala denominado *índice de avenida* o inundación que particulariza a cada sitio de medición. Si un conjunto de N sitios cada uno con n años de registro se pueden encontrar, entonces uno podría ingenuamente esperar que los valores de datos Nn proporcionen estimaciones precisas de los cuantiles tan extremos como el cuantil Q_{Nn} del año Nn . En la práctica esto no es razonable; ya que surgen problemas porque las distribuciones de frecuencia en diferentes

sitios no son exactamente idénticas y porque las magnitudes de eventos en diferentes sitios pueden no ser estadísticamente independientes.

2.5 L-momentos

L-momentos o momentos lineales (L significa linealidad), es un sistema alternativo a los métodos tradicionales de los momentos convencionales para describir las formas de las distribuciones de probabilidad. Hosking (1986) citado por Rahman *et al.*, (2013), describe los L-momentos como combinaciones lineales de los momentos ponderados probabilísticamente (MPP) definidos por Greenwood *et al.*, (1979).

El estimador de MPP a partir de una muestra con variable aleatoria X de tamaño n , con elementos en orden ascendente: $X_{1:n} < X_{2:n} < \dots < X_{n:n}$ es:

$$b_r = n^{-1} \binom{n-1}{r}^{-1} \sum_{j=r+1}^n \binom{j-1}{r} x_{j:n} \quad (22)$$

Donde $x_{j:n}$ es el j -ésimo elemento en orden ascendente.

Desarrollando los primeros MPP se tiene:

$$b_0 = \frac{1}{n} \sum_{j=1}^n x_{j:n} \quad (23)$$

$$b_1 = \frac{1}{n} \sum_{j=2}^n \frac{(j-1)}{(n-1)} x_{j:n} \quad (24)$$

$$b_2 = \frac{1}{n} \sum_{j=3}^n \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{j:n} \quad (25)$$

$$b_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n} \quad (26)$$

A partir de los MPP; los primeros cuatro L-momentos l_r , están dados por:

$$l_1 = b_0 \quad (27)$$

$$l_2 = 2b_1 - b_0 \quad (28)$$

$$l_3 = 6b_2 - 6b_1 + b_0 \quad (29)$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \quad (30)$$

El momento lineal de primer orden l_1 es el parámetro de localización o media de la muestra; el de segundo orden l_2 mide la escala o la variación, indicando el grado de dispersión de los datos; el momento de tercer orden l_3 hace referencia a su asimetría; y el de cuarto orden l_4 indica la kurtosis de la muestra.

Los L-momentos de la muestra que son independientes de las unidades de medición, llamados *L-momentos ratios* o cocientes (Hosking, 1990), se definen dividiendo aquellos de mayor orden por la medida de escala l_2 :

$$t = l_2 / l_1 = L - C_v, \quad \text{donde} \quad 0 \leq t \leq 1 \quad (31)$$

$$t_3 = l_3 / l_2 = L - C_s, \quad \text{donde} \quad -1 \leq t_3 \leq 1 \quad (32)$$

$$t_4 = l_4 / l_2 = L - C_k, \quad \text{donde} \quad -1 \leq t_4 \leq 1 \quad (33)$$

Donde t es L-coeficiente de variación ($L - C_v$), t_3 es L-coeficiente de asimetría ($L - C_s$) y t_4 es L-coeficiente de kurtosis ($L - C_k$).

Los momentos lineales de la muestra (t, t_3 y t_4) se relacionan con los momentos lineales de la función de distribución (τ, τ_3 y τ_4) para determinar los parámetros de las distribuciones (véase Anexo 8.1).

2.6 El método del índice de avenida

El término “índice de avenida”, aplicado por vez primera en el estudio de las avenidas por Dalrymple (1960) y de ahí su nombre, se ha utilizado desde entonces en el análisis de frecuencia de otras variables diferentes a la original (caudales), entre ellas las precipitaciones máximas.

Supongamos que los datos son disponibles en una región con N estaciones, cada estación i con un tamaño muestral n_i y los datos en cada estación i es $Q_{ij}, j = 1, \dots, n_i$. Si $Q_i(F), 0 < F < 1$; es la función cuantil de la distribución de frecuencia en la estación i . La asunción clave del procedimiento del índice de avenida es que las estaciones formen una región homogénea, es decir que la distribuciones de frecuencia de las N estaciones son idénticas excepto por un factor de escala específico de cada estación denominado *índice de avenida*. Entonces se define:

$$Q_i(F) = \mu_i q(F), \quad i = 1, \dots, N. \quad (34)$$

Donde μ_i es el *índice de avenida*, que toma el valor de la media de la distribución de frecuencia en la estación i considerada; y $q(F)$, es la curva de crecimiento regional o la función cuantil adimensional común para todas las estaciones. La función cuantil de la distribución de frecuencia regional es la distribución común para Q_{ij} / μ_i .

El *índice de avenida* se estima naturalmente por $\hat{\mu}_i = \bar{Q}_i$, que es la media muestral de las observaciones en la estación i .

Los datos adimensionales se reajustan a $q_{ij} = Q_{ij} / \hat{\mu}_i, j = 1, \dots, n_i; i = 1, \dots, N$, que son la base para estimar la curva de crecimiento regional $q(F), 0 < F < 1$.

Es usualmente asumir que la forma de $q(F)$ se conoce, excepto de parámetros indeterminados $p : \theta_1, \dots, \theta_p$, por lo que se escribe $q(F)$ como $q(F; \theta_1, \dots, \theta_p)$. Por ejemplo estos parámetros pueden ser el coeficiente de variación y el sesgo de la distribución, o los ratios de L-momentos t, t_3, t_4, \dots , definidos en la sección 2.5. El enfoque planteado por Hosking y Wallis (1997), estima los parámetros por separado para cada estación, la estimación en la estación i de θ_k se denota por $\hat{\theta}_k^{(i)}$. Las estimaciones en cada estación se combinan para obtener la estimación regional de cada parámetro, de modo que se obtiene una media ponderada, donde cada estación tiene un peso proporcional correspondiente al tamaño de su muestra n_i . Es decir:

$$\hat{\theta}_k^R = \frac{\sum_{i=1}^N n_i \hat{\theta}_k^{(i)}}{\sum_{i=1}^N n_i} \quad (35)$$

Sustituyendo los parámetros regionales estimados mediante la ecuación 35 en $q(F)$, se obtiene la curva de crecimiento regional $\hat{q}(F) = q(F; \hat{\theta}_1^R, \dots, \hat{\theta}_p^R)$. Este método de obtener las estimaciones regionales es esencialmente la de (Wallis, 1981; citado por Hosking y Wallis, 1997), excepto la ponderación proporcional por n_i que es una adición posterior; sugerida por (Wallis, 1982; citado por Hosking y Wallis, 1997).

Obtenidos el estimador de *índice de avenida* $\hat{\mu}_i$ y el estimador de la curva de crecimiento regional $\hat{q}(F)$, la estimación de los cuantiles de interés en una estación i se reduce a la forma de la ecuación de partida:

$$\hat{Q}_i(F) = \hat{\mu}_i \hat{q}(F) \quad (36)$$

2.7 Procedimiento del análisis regional

La literatura sobre el ARF fue revisada sobre los siguientes subtítulos que constituyen el procedimiento general del análisis:

- i. Filtrado de datos usando medidas de discordancia.
- ii. Identificación de regiones homogéneas
- iii. Elección de la distribución de frecuencia
- iv. Estimación de los cuantiles de frecuencia

2.7.1 Filtrado de datos usando medidas de discordancia

Dado un grupo de sitios, el objetivo consiste en identificar las estaciones que son groseramente discordantes con el grupo como un todo. Discordancia se mide en términos de los L-momentos de los datos de las estaciones. Hosking y Wallis (1997), definen la medida discordancia para el sitio i como:

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}). \quad (37)$$

Donde

$$u_i = \left[t^{(i)} \quad t_3^{(i)} \quad t_4^{(i)} \right]^T \quad (38)$$

Vector que contiene los valores de los ratios L-momentos t, t_3 y t_4 para la estación i : el superíndice T indica la transposición de un vector. La media del grupo (sin ponderar) es:

$$\bar{u} = N^{-1} \sum_{i=1}^N u_i \quad (39)$$

Y la matriz de sumas de cuadrados y productos cruzados se define como:

$$A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (40)$$

La estación i se declara discordante si el valor de D_i es elevado, y ello depende del número de estaciones que forman el grupo. Hosking y Wallis (1997) establecen que una estación ha de ser considerada discordante si su valor D_i supera el valor crítico establecido en función del número de estaciones en la región, dado en el Cuadro 1:

Cuadro 1: Valores críticos de la discordancia D_i .

Número de estaciones en la región	Valor crítico	Número de estaciones en la región	Valor crítico
5	1.333	10	2.491
6	1.648	11	2.632
7	1.917	12	2.757
8	2.140	13	2.869
9	2.329	14	2.971
		≥ 15	3

FUENTE: Hosking y Wallis (1997).

Hosking y Wallis (1997), recomienda el empleo de D_i en dos etapas. Inicialmente, se puede aplicar a un gran grupo de sitios, procurando abarcar la totalidad de las estaciones o regiones suficientemente amplias. Las estaciones que difieren significativamente de la mayoría, en sus datos serán señalados como discordante y serán excluidos, pudiéndose considerar la posibilidad de desplazarlo a otra región.

Posteriormente, una vez identificadas las diferentes regiones homogéneas, se debe recalcular la medida de la discordancia para cada sitio en su región propuesta.

2.7.2 Identificación de regiones homogéneas

De todas las etapas en el análisis regional de frecuencias, la identificación de regiones homogéneas es generalmente la de mayor dificultad dada la necesidad de una elevada cantidad de juicios subjetivos. El objetivo principal de formar regiones homogéneas es formar grupos de estaciones que satisfagan la condición de homogeneidad, que consiste en asumir que las funciones de distribución de probabilidades son idénticas en toda la región, excepto por un factor de escala local (índice de avenida).

Para definir las regiones hidrológicamente homogéneas, dos pasos básicos deben llevarse a cabo. Primero, formación de las regiones mediante diferentes métodos como la cuenca, el medio ambiente y la información sobre el clima. Segundo, aplicación de las pruebas de discordancia y heterogeneidad (es decir, evaluar si las regiones contienen sitios estadísticamente similares o no).

a. Formación de regiones homogéneas

Debido a la complejidad en la comprensión de los factores que tienen un efecto directo e indirecto sobre la generación de precipitaciones, no hay guías metodológicas simples para identificar regiones homogéneas. Siendo la experiencia, la información previa y juicios personales que pueden proporcionar posibles directrices para formar regiones con características hidrológicas similares.

Hubo varios intentos realizados por diferentes autores para identificar regiones hidrológicamente homogéneas y su énfasis eran o bien en consideraciones geográficas o en características hidrológicas o una combinación de ambos (Kachroo et al., 2000). Por ejemplo, Hosking y Wallis (1997) discuten algunos de los métodos tales como: la *conveniencia geográfica* basada en las áreas administrativas, *demarcación subjetiva* que define subjetivamente la región mediante la inspección de las características del sitio, *particionamiento objetivo* son regiones formadas mediante la asignación de sitios a uno de dos grupos dependiendo de algunos valores de umbral, y el *método de agrupación* que es el método estándar de análisis estadístico multivariante para dividir un conjunto de datos en regiones. Es obvio que las cuencas posiblemente no tengan exactamente el mismo comportamiento, dado el tamaño limitado de la muestra, la dinámica y los infinitos factores que influyen en la generación de inundaciones.

Consiguientemente, no es necesariamente para los sitios que conforman una región satisfacer exactamente las pruebas de homogeneidad (Hosking y Wallis, 1997). Es decir, una homogeneidad aproximada podría ser suficiente para asegurar que se utilice el análisis de frecuencia regional.

En este estudio con la medida de heterogeneidad se verifica que la zona de estudio conforma una sola región, por lo que no es necesario la formación en 2 o más regiones. Sin embargo el autor sugiere la agrupación jerárquica (*Cluster*) de las características locales, como el método más práctico para conformar regiones en presencia de una gran cantidad

de datos, empleando el análisis de agrupación jerárquica en base al algoritmo Ward's, considerando variables predictores como los vectores característicos de la precipitación promedio multianual, longitud, latitud y la elevación de cada estación; donde estas variables deben ser estandarizados para evitar el predominio de vectores característicos con grandes valores absolutos (por ejemplo, altitud).

b. Medida de Heterogeneidad

Hosking y Wallis (1997) presenta la medida de heterogeneidad para estimar el grado de heterogeneidad en un grupo de estaciones y evaluar cuando pueden ser tratadas en su conjunto como una región homogénea.

La medida de la heterogeneidad compara la dispersión observada y simulada de L-momentos para N estaciones considerados. Para este propósito la simulación de Monte Carlo se realiza mediante la distribución Kappa de cuatro parámetros definidos por:

$$f(x) = \alpha^{-1} [1 - k(x - \xi) / \alpha]^{1/k-1} [F(x)]^{1-h} \quad (41)$$

ξ es un parámetro de localización, α es parámetro de escala, k y h son parámetros de forma.

La razón para utilizar la distribución de Kappa es que se trata de una distribución generalizada que produce muchas distribuciones como casos especiales se tiene la logística generalizada, general de valores extremos, y general de Pareto, por lo que es capaz de abarcar la mayor parte de las distribuciones empleadas en estudios medioambientales.

Los cuatro parámetros de la distribución Kappa son ajustada a los ratios medios regionales $1, t^R, t_3^R$ y t_4^R , para simular un número N_{sim} de realizaciones de una región con N estaciones como la original. Para comparar los valores de dispersión observados con los simulados, el estadístico apropiado es la medida de heterogeneidad, H_j ($j = 1, 2, 3$), definido como:

$$H_j = \frac{(V_j - \mu_{V_j})}{\sigma_{V_j}} \quad (42)$$

Donde para cada simulación se calcula V_j . Se determina además la media μ_{V_j} , y la desviación estándar σ_{V_j} , de los N_{sim} valores de V_j . Siendo:

V_1 , es la desviación estándar ponderada observada de los t valores:

$$V_1 = \left\{ \frac{\sum_{i=1}^N n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^N n_i} \right\}^{1/2} \quad (43)$$

V_2 , es la distancia media observada de t/t_3 :

$$V_2 = \sum_{i=1}^N n_i \left[(t^{(i)} - t^R)^2 + (t_3^{(i)} - t_3^R)^2 \right]^{1/2} / \sum_{i=1}^N n_i \quad (44)$$

V_3 , es la distancia media observada de t_3/t_4 :

$$V_3 = \sum_{i=1}^N n_i \left[(t_3^{(i)} - t_3^R)^2 + (t_4^{(i)} - t_4^R)^2 \right]^{1/2} / \sum_{i=1}^N n_i \quad (45)$$

Donde la región propuesta tiene N estaciones, cada estación i tiene la longitud de registro n_i y los ratios de L -momentos $t^{(i)}$, $t_3^{(i)}$ y $t_4^{(i)}$. t^R , t_3^R y t_4^R son los ratios medios regionales de L -CV, L -asimetría y L -kurtosis, ponderados proporcionalmente en función de la longitud de registro; por ejemplo:

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i} \quad (46)$$

La región se declara heterogénea si el valor de H_j es suficientemente elevado. Hosking y Wallis (1997) sugieren considerarla como “aceptablemente homogénea” si $H_j < 1$, “posiblemente heterogénea” si $1 \leq H_j < 2$, y “definitivamente heterogénea” si $H_j \geq 2$.

El uso del estadístico H con un número mayor de repeticiones constituye una alternativa más fiable y menos subjetiva. Se considera adecuado un valor M de 500 simulaciones, requiriéndose de valores mayores en caso de que H_j resulte cercano a los límites 1 y 2.

2.7.3 Elección de la distribución de frecuencia

Definidas las regiones, se procede a determinar la función de distribución más apropiada para cada una de ellas.

Muchas distribuciones se pueden utilizar para la estimación de cuantiles para los datos regionales. Las distribuciones de tres parámetros, logística generalizada (GLO), valor extremo generalizado (GEV), lognormal (LN3) o normal generalizado (GNO), Pareto generalizada (GPA) y Pearson tipo III (PE3) se han considerado en este análisis regional.

Hosking y Wallis (1997), justifican la elección entre estas cinco distribuciones por el hecho de poseer tres parámetros lo cual, en un análisis regional, facilita el ajuste al disponerse de observaciones de varias estaciones y de incluir entre sus casos particulares las distribuciones más empleadas en el estudio de variables ambientales.

De acuerdo a Hosking y Wallis (1997), distribuciones de dos parámetros pueden causar sesgos en la cola de los cuantiles estimados si la forma de la cola de la distribución de frecuencia verdadera no está bien aproximada por la distribución ajustada. La mejor distribución de ajuste es uno que da estimaciones robustas para la curva de crecimiento regional, así como para los cuantiles en las estaciones.

Las distribuciones y sus parámetros resumidos se presentan en anexo 8.1 y para mayor detalle véase en Hosking y Wallis (1997), página 191-209.

El diagrama de la relación L-momento y Z-estadístico se han utilizado como los mejores criterios de ajuste para identificar la distribución regional subyacente para regiones homogéneas.

a. Los diagramas de las relaciones de los L-momentos

En Hosking (1990), aparecen por vez primera los diagramas de las relaciones de los L-momentos, que resulta una herramienta útil en la selección adecuada de la función de distribución regional de frecuencia de los eventos hidrológicos extremos. Recomendado por Stendiger *et al.*, (1993), Vogel y Fennessey (1993) y Hosking y Wallis (1995).

En Hosking y Wallis (1997) se han desarrollado expresiones polinómicas que permiten elaborar los diagramas de los L-momentos para cada una de las funciones de distribución más frecuentemente utilizadas, en las que L-coeficiente de kurtosis ($L-C_k$) se

estima a partir de L-coeficiente de asimetría ($L - C_s$); mediante la aproximación del siguiente polinomio (ecuación (47)), los coeficientes A_k , se dan en el Cuadro 2.

$$L - C_k = \sum_{k=0}^8 A_k (L - C_s)^k \quad (47)$$

Cuadro 2: Aproximaciones polinómicas de $L - C_k$ en función de $L - C_s$.

A_k	GPA	GEV	GLO	LN3	PE3
A_0	0	0.10701	0.16667	0.12282	0.1224
A_1	0.20196	0.1109	-	-	-
A_2	0.95924	0.84838	0.83333	0.77518	0.30115
A_3	-0.20096	-0.06669	-	-	-
A_4	0.04061	0.00567	-	0.12279	0.95812
A_5	-	-0.04208	-	-	-
A_6	-	0.03763	-	-0.13668	-0.57488
A_7	-	-	-	-	-
A_8	-	-	-	0.11368	0.9383

FUENTE: Hosking y Wallis (1997).

Cuando se dispone de los diagramas en el plano $(L - C_s, L - C_k)$, de acuerdo a Vogel *et al.*, (1993) y Hosking y Wallis (1995), la proximidad de la media regional a un candidato particular de la curva de distribución teórica, se interpreta como una indicación de la distribución de mejor ajuste para describir los datos regionales.

Como una alternativa complementaria a los diagramas analizados, Hosking y Wallis (1993) propusieron el empleo de una medida cuantitativa, o lo que es lo mismo, una prueba de medida de bondad de ajuste denominado *estadístico* z^{DIST} , con el objetivo de reducir el posible grado de subjetividad en la selección de la función de distribución.

b. Medida de bondad de ajuste - estadístico Z^{DIST}

El criterio para seleccionar la distribución de mejor ajuste es el estadístico z^{DIST} definidas por Hosking y Wallis (1993), cuyo objetivo principal es comparar $L - C_s$ y $L - C_k$ simulado de una distribución ajustada con los valores de la media regional $L - C_s$ y $L - C_k$ obtenidos de datos observados.

La bondad de la medida de ajuste para una distribución z^{DIST} se define por:

$$Z^{DIST} = (\tau_4^{DIST} - t_4^R + B_4) / \sigma_4. \quad (48)$$

Donde

τ_4^{DIST} es el coeficiente L -kurtosis teórico de la distribución ajustada, donde $DIST$ hace referencia a GLO, GEV, GPA, GNO y PE3 se determina directamente de los diagramas $L-C_s$ y $L-C_k$ o mediante la ecuación (47).

t_4^R es la media regional de L -kurtosis del conjunto de datos observados en la región homogénea.

B_4 y σ_4 bias y desviación estándar de t_4^R , respectivamente, definido como:

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} (t_4^{[m]} - t_4^R), \quad (49)$$

$$\sigma_4 = \left\{ (N_{sim} - 1)^{-1} \left[\sum_{m=1}^{N_{sim}} (t_4^{[m]} - t_4^R)^2 - N_{sim} (B_4)^2 \right] \right\}^{1/2}, \quad (50)$$

N_{sim} , es el número de conjunto de datos regionales simulados, generados usando una distribución de Kappa. Se insiste en la posibilidad de emplear las mismas simulaciones mediante técnicas de simulación de Monte Carlo; llevadas a cabo para el cálculo de la heterogeneidad. m es la m -ésima región simulada.

Finalmente se considera que el ajuste de una determinada distribución es adecuado si Z^{DIST} es suficientemente cercano a cero, siendo un valor razonable para este criterio $|Z^{DIST}| \leq 1.64$ lo cual corresponde a la aceptación de la distribución, planteado la hipótesis en un nivel de confianza de 90%.

2.7.4 Estimación de los cuantiles de frecuencia

a. Algoritmo regional de L - momentos

En esta etapa el objetivo es ajustar los datos de las estaciones de la región homogénea a una distribución regional de frecuencias que se particulariza en cada una de ellas mediante un factor de escala (*índice de avenida*). Después, la estimación de cuantiles se lleva a cabo por medio de la distribución local.

El algoritmo regional de L-momentos descrito por Hosking y Wallis (1997) se ha aplicado para estimar la distribución de frecuencias regionales. Que se define:

Dada la región de N estaciones, cada estación i con longitud muestral n_i , media muestral $l_1^{(i)}$, y ratios de L-momentos $t^{(i)}, t_3^{(i)}, t_4^{(i)}, \dots$, donde t^R, t_3^R, t_4^R, \dots , son los ratios medios regionales de L-momentos, ponderados proporcionalmente en función de la longitud de la muestra:

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i}, \quad (51)$$

$$t_r^R = \frac{\sum_{i=1}^N n_i t_r^{(i)}}{\sum_{i=1}^N n_i}, \quad (52)$$

El valor de la media regional l_1^R es 1, es decir $l_1^R = 1$.

El algoritmo regional de L-momentos consiste en ajustar la distribución mediante sus ecuaciones de L-momentos ratios $\lambda_1, \tau, \tau_3, \tau_4, \dots$, a las medias ponderadas de los ratios regionales de L-momentos muestrales $l_1^R, t^R, t_3^R, t_4^R, \dots$, calculados con la ecuación (51) y (52). La función cuantil de la distribución regional de frecuencia ajustada, se denota como $\hat{q}(\cdot)$.

Las estimaciones de los cuantiles en la estación i se determinan por combinación de las estimaciones de μ_i y $q(\cdot)$. Así, el cuantil para la probabilidad de no excedencia F es:

$$\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F). \quad (53)$$

Pese a que Hosking y Wallis (1997) recomiendan el empleo del método propuesto, de él se derivan multitud de variantes. Por ejemplo, es posible asumir como *índice de avenida* un valor diferente a la media local de las observaciones. La mediana o un cuantil de frecuencia concreto constituyen otras opciones.

b. Evaluación de la precisión de los cuantiles estimados

Los resultados obtenidos en los análisis estadísticos se enmarcan dentro de un rango de incertidumbre, es necesario determinar si se quiere maximizar su utilidad.

Cuando más de una distribución es adecuada para modelar la región, la elección de la distribución se realiza en función a aquello que realiza estimaciones más robustas para periodos de retorno más extremos. Hosking y Wallis (1997) recomienda una herramienta

eficaz para establecer las propiedades de los procedimientos estadísticos complejos, como el *algoritmo de L-momento regional* a través de la simulación de Monte Carlo, es una forma de evaluar la precisión de las estimaciones, teniendo en cuenta la esperable dependencia interestacional de las observaciones, y la posibilidad de que la región sea moderadamente heterogénea.

En el curso del proceso de simulación, las estimaciones de los cuantiles se calculan para diversas probabilidades de no excedencia. En la repetición m -ésima de la estación i , denótese la estimación del cuantil de probabilidad de no excedencia F como $\hat{Q}_i^{[m]}(F)$. Su error relativo es $[\hat{Q}_i^{[m]}(F) - Q_i(F)] / Q_i(F)$, y esta cantidad puede ser cuadrada y promediada sobre las M repeticiones para obtener, bias relativo $B_i(F)$ y el error cuadrático medio relativo $R_i(F)$ de los estimadores. Para valores elevados de M , dicho error se expresa como:

$$B_i(F) = M^{-1} \sum_{m=1}^M \frac{\{\hat{Q}_i^{[m]}(F) - Q_i(F)\}}{Q_i(F)} \quad (54)$$

$$R_i(F) = \left[M^{-1} \sum_{m=1}^M \left\{ \frac{\hat{Q}_i^{[m]}(F) - Q_i(F)}{Q_i(F)} \right\}^2 \right]^{1/2} \quad (55)$$

El resumen de la precisión de las estimaciones de los cuantiles a lo largo de todas las estaciones de la región viene dado por el valor medio regional de: bias relativo $B^R(F)$, bias relativo absoluto $A^R(F)$ y el error cuadrático medio $R^R(F)$; las mismas se expresan:

$$B^R(F) = N^{-1} \sum_{i=1}^N B_i(F). \quad (56)$$

$$A^R(F) = N^{-1} \sum_{i=1}^N |B_i(F)|. \quad (57)$$

$$R^R(F) = N^{-1} \sum_{i=1}^N R_i(F). \quad (58)$$

Adicionalmente, se pueden calcular valores análogos para la estimación de la curva de frecuencia. Sea $q_i(F)$ la curva del sitio i , definida por:

$$Q_i(F) = \mu_i q_i(F). \quad (59)$$

Las cantidades $q_i(F)$ se necesitan para la simulación de regiones heterogéneas, mientras que en las homogéneas cada $q_i(F)$ es igual a la curva regional de frecuencia $q(F)$. En la repetición m -ésima, denótese la curva regional estimada como $\hat{q}^{[m]}(F)$. Las medidas de precisión de las estimaciones de las curvas se definen también con las ecuaciones (55) a (58), reemplazando $Q_i(F)$ y $\hat{Q}_i^{[m]}(F)$ por $q_i(F)$ y $\hat{q}^{[m]}(F)$, respectivamente. Estas medidas son de especial interés para aquellas situaciones en las que únicamente interesa la estimación de las curvas de frecuencia, como ocurre cuando el *índice de avenida* se estima mediante métodos que no implican el empleo de datos locales.

Otros medidores útiles, particularmente cuando la distribución de las estimaciones es asimétrica, son los cuantiles empíricos de dicha distribución. Estos pueden ser obtenidos calculando el ratio entre los valores estimados y los verdaderos - $\hat{Q}_i(F)/Q_i(F)$ en el caso de los cuantiles, y $\hat{q}_i(F)/q_i(F)$ en las curvas; promediando dichos valores para todas las estaciones de la región y acumulando sobre las diferentes realizaciones un histograma de los valores que toma el ratio. Por ejemplo, para una probabilidad de no excedencia F puede determinarse que el 5% de los valores simulados de $\hat{Q}_i(F)/Q_i(F)$ se encuentra por debajo del valor $L_{.05}(F)$, mientras que otro 5% queda por encima de $U_{.05}(F)$; así, el 90% de la distribución de $\hat{Q}(F)/Q(F)$ se incluye en el intervalo:

$$L_{.05}(F) \leq \frac{\hat{Q}(F)}{Q(F)} \leq U_{.05}(F), \quad (60)$$

e invirtiéndolo para expresar Q en términos de \hat{Q} :

$$\frac{\hat{Q}(F)}{U_{.05}(F)} \leq Q(F) \leq \frac{\hat{Q}(F)}{L_{.05}(F)}. \quad (61)$$

La expresión (61) adquiere forma de intervalo de confianza, pero sólo puede ser interpretada como tal si la distribución de $\hat{Q}(F)/Q(F)$ es independiente de los parámetros implicados en la especificación del modelo que subyace en el método del *índice de avenida*; para el algoritmo regional de L-momentos, las medias locales de las observaciones y los ratios medios regionales de L-momentos. Pese a que en la práctica esta

independencia no se cumple, el intervalo constituye una buena indicación de la variación entre las cantidades estimadas y las verdaderas.

Los límites $\hat{Q}(F)/U_{.05}(F)$ y $\hat{Q}(F)/L_{.05}(F)$ reciben el nombre de “límites de error al 90%” para $\hat{Q}(F)$, y pueden ser de gran utilidad en la cola inferior de la distribución. Si ésta toma valores negativos, puede ocurrir que $L_{.05}(F)$ sea muy pequeño o negativo, conduciendo a un límite superior demasiado elevado o incluso infinito. En estos casos, el valor medio regional del error cuadrático medio $R^R(F)$ de las estimaciones (ecuación (58)) constituye una medida de precisión más eficiente.

Hay que destacar, no obstante, que esta forma de simulación basada en los límites de la ecuación (61) es menos exacta que la construcción formal de intervalos de confianza, pero aun así aporta estimaciones razonables de los errores que se pueden esperar en los cuantiles y las curvas de frecuencia. En este sentido, la precisión viene condicionada por el número de repeticiones M de la simulación. Valores de $M = 100$ pueden resultar suficientes, aunque Hosking y Wallis (1997) recomiendan $M = 1000$ o incluso $M = 10000$ para un mayor rigor.

En el Cuadro 3, se describe el procedimiento para evaluar la elección de la distribución que realiza estimaciones más robustas para periodos de retorno más extremos; con y sin la consideración de la dependencia (correlación cruzada) entre estaciones en la simulación.

Cuadro 3: Algoritmo para la simulación del Algoritmo de L-momento regional

1. Especificar N y para cada uno de los N estaciones su longitud de registro n_i y los L-momentos de su distribución de frecuencias.
 2. Calcular los parámetros de la distribución frecuencia para cada estación, dado sus ratios de L-momentos.
 3. Para cada una de las M repeticiones del procedimiento de simulación, se lleva a cabo los siguientes pasos.
 - 3.1. Generar muestra de datos para cada estación. Si no hay dependencia entre estaciones, esto simplemente requiere la generación de una muestra aleatoria de tamaño n_i a partir de la distribución de frecuencias para el sitio i , $i = 1, \dots, n$. Si la dependencia entre sitios se incluye en la simulación, el siguiente procedimiento puede ser utilizado.
 - 3.1.1. Si $n_0 = \max n_i$, es la mayor de las longitudes de registro de las estaciones. Para cada punto de tiempo $k = 1, \dots, n_0$, se genera una realización de un vector aleatorio y_k con elementos y_{ik} , $i = 1, \dots, N$, que tiene una distribución normal multivariada con el vector de media cero y matriz de covarianza R .
 - 3.1.2. Transformar cada y_{ik} , $k = 1, \dots, n_i$, $i = 1, \dots, N$, para la distribución marginal requerida, es decir, el cálculo de los valores de los datos $Q_{ik} = Q_i(\Phi(y_{ik}))$, donde Q_i es el cuantil para la estación i y Φ es la función de distribución acumulada de la distribución normal estándar.
 - 3.2. Aplicar el algoritmo de L-momento regional para la muestra de los datos regionales. Esto implica los siguientes pasos:
 - 3.2.1. Calcular ratios de L-momento en cada estación y el promedio regional de los ratios de L-momento.
 - 3.2.2. Ajustar a la distribución elegida
 - 3.2.3. Cálculos de las estimaciones de la curva de crecimiento de la región y cuantiles en cada estación.
 - 3.3. Calcular el error relativo de la curva de crecimiento regional estimado y cuantiles en las estaciones, y acumular las sumas necesarias para el cálculo de las medidas generales de precisión.
 4. Calcular las medidas generales de la exactitud de los cuantiles estimados y curva de crecimiento regional.
-

2.8 Mapeo de índice de avenida y eventos extremos

Szolgay *et al.*, (2008) señala que recientemente, se ha informado de intensos esfuerzos para desarrollar métodos de interpolación estadísticos y espaciales complejos para estimar precipitaciones máximas de diseño en la literatura y también por varias autoridades nacionales y las oficinas meteorológicas del mundo. Por ejemplo Wallis *et al.* (2007) ha actualizado el atlas de frecuencias de precipitación publicadas por el Servicio Meteorológico Nacional de EE.UU. en el año 1973; utilizando el sistema de mapeo PRISM y el algoritmo regional L-momentos para las estimaciones de frecuencias y magnitud de precipitación para la estimación de frecuencias de precipitación de 2 horas y 24 horas.

Siendo la cartografía de las magnitudes de precipitación para una frecuencia dada, complemento a los análisis estadísticos de frecuencias. En este estudio se propone comparar dos técnicas de interpolación para el mapeo de frecuencias de precipitación máxima de 24 horas, que se obtienen en el análisis regional de frecuencia.

2.8.1 Método de ponderación inverso a la distancia (DIP)

Este algoritmo calcula los valores de las celdas por medio de una composición ponderada de los datos de la muestra, el modelo sugiere que el resultado predicho reduce su incidencia en la medida en que aumenta la separación entre el punto a evaluar y los puntos de su entorno. Es así como se acepta que los puntos más cercanos al centroide poseen un mayor peso en el cálculo del valor definitivo de la celda, su modelo matemático (Slocum, 1999):

$$Z = \left(\sum \left(\frac{Z_i}{d_i^n} \right) \right) / \sum \left(\frac{1}{d_i^n} \right) \quad (62)$$

Donde Z es el valor a interpolar, Z_i son los valores de los puntos vecinos, d_i Distancia entre los puntos vecinos y el punto de la celda a predecir, n Es la potencia a la cual se eleva la distancia, oscila entre 0.5 a 4. Para este estudio, el parámetro de potencia se fija para que sea igual a dos.

Este método se utiliza a menudo en aplicaciones de ingeniería, ya que es simple y eficiente. Por ejemplo, Rahman *et al.*, (2013) en Bangladesh mapearon los cuantiles extremos determinados mediante el análisis regional de frecuencias con el enfoque de L-

momentos, para ello emplearon la técnica de interpolación de Ponderación Inverso a la Distancia.

2.8.2 Kriging ordinario (KO)

Kriging ordinario es el método de interpolación definitiva investigado en este análisis. Tveito *et al.*, (2008), señala que la familia Kriging, también conocido como métodos geoestadísticos, se basa en la idea de que los valores medidos en lugares cercanos tienden a estar más relacionada que los valores medidos en otros lugares.

Similar a DIP, KO también utiliza una ponderación, sin embargo, como un método probabilístico, depende de las relaciones espaciales y estadísticos para predecir puntos no medidos (Johnston *et al.*, 2001). Las funciones de semivariogramas empíricos proporcionan medios para cuantificar la autocorrelación espacial de los conjuntos de datos que luego se incluyen en el modelo espacial (Goovaerts, 1997).

Este modelo busca minimizar la varianza del error, y busca llevar la media del error de los valores predichos a cero, para que no se presenten sobre, ni subestimaciones, su expresión matemática es:

$$Z_{(x_0)} = \lambda_1 Z_{(x_1)} + \lambda_2 Z_{(x_2)} + \lambda_3 Z_{(x_3)} + \dots + \lambda_n Z_{(x_n)} = \sum_{i=1}^n \lambda_i Z_{(x_i)} \quad (63)$$

Donde $Z_{(x_0)}$ es el valor estimado de $Z_{(x)}$ en el punto x_0 , λ_i son los pesos asignados a la i -ésima medición para realizar la predicción del valor $Z_{(x_0)}$, $Z_{(x_i)}$ valores de los puntos vecinos que intervienen en el procedimiento.

Los factores de ponderación se pueden hallar resolviendo el siguiente sistema de ecuaciones que se obtiene mediante la minimización de la varianza del error e imponiendo la condición de ausencia de sesgo:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma_{(x_i-x_j)} - u_{(x_0)} = \gamma_{(x_i-x_0)}, & i = 1, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (64)$$

Siendo $\gamma(\cdot)$ el valor correspondiente del modelo del semivariograma experimental

La semivarianza: llamado también semivariograma, caracteriza las propiedades de dependencia espacial del proceso. Dada una realización del fenómeno, la función de semivarianza es estimada, por el método de momentos, a través del semivariograma experimental, que se calcula mediante (Wackernagel, 1995):

$$\gamma(h) = \sum_{i=1}^{n(h)} \frac{(Z(x+h) - Z(x))^2}{2n} \quad (65)$$

Donde $Z(x)$ es el valor de la variable en un sitio x , $Z(x+h)$ es otro valor muestral separado del anterior por una distancia h y n es el número de parejas que se encuentran separadas por dicha distancia.

Para poder realizar la predicción del valor es necesario ajustar el semivariograma empírico a un semivariograma experimental para cuantificar el grado y escala de variación espacial. Por ello se hace necesario el ajuste de modelos que generalicen lo observado en el semivariograma experimental a cualquier distancia.

Existen diversos modelos teóricos de semivarianza que pueden ajustarse al semivariograma experimental. En Samper y Carrera (1990) se presenta una discusión respecto a las características y condiciones que éstos deben cumplir. Warrick *et al.*, (1986) menciona que en general dichos modelos pueden dividirse en no acotados (lineal, logarítmico, potencial) y acotados (esférico, exponencial, gaussiano). Los del segundo grupo garantizan que la covarianza de los incrementos es finita, por lo cual son ampliamente usados cuando hay evidencia de que presentan buen ajuste. Todos estos modelos tienen tres parámetros comunes (Figura 1) que son descritos a continuación:

Efecto Pepita: Se denota por C_0 y representa una discontinuidad puntual del semivariograma en el origen (Figura 1). Puede ser debido a errores de medición en la variable o a la escala de la misma.

Rango: muestra la distancia donde la separación entre puntos indica una autocorrelación significativa. Indica la amplitud del radio de búsqueda de los vecinos locales.

Meseta: máxima semivarianza hallada entre un par de puntos, que es encargada de encontrar la incertidumbre en el momento de la interpolación.

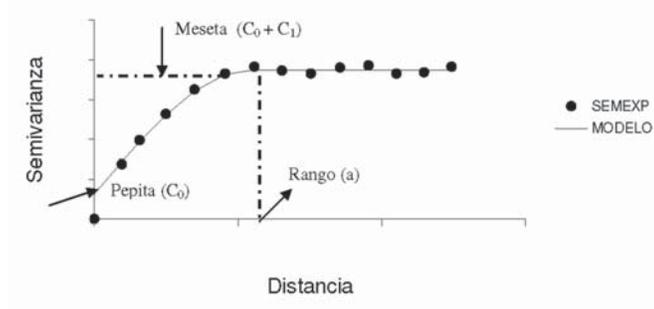


Figura 1: Comportamiento típico de un semivariograma acotado con una representación de los parámetros básicos. SEMEXP corresponde al semivariograma experimental y MODELO al ajuste de un modelo teórico.

FUENTE: Giraldo (2002).

En este estudio se evalúa el ajuste del variograma experimental por los modelos teóricos: esférico y exponencial por ser de mayor aplicación con variables de precipitación.

Modelo Esférico: En este modelo se muestra una disminución progresiva de la autocorrelación espacial (así como un aumento en la semivarianza) hasta cierta distancia, después de la cual la autocorrelación es cero. Su expresión matemática es la siguiente:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left(\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right) & h \leq a \\ C_0 + C_1 & h > a \end{cases} \quad (66)$$

En donde C_1 representa la meseta, a el rango y h la distancia.

Modelo Exponencial: Este modelo se aplica cuando la dependencia espacial tiene un crecimiento exponencial respecto a la distancia entre las observaciones. Su expresión matemática es la siguiente:

$$\gamma(h) = C_0 + C_1 \left(1 - \exp \left(\frac{-3h}{a} \right) \right) \quad (67)$$

Validación cruzada: Para evaluar la bondad de ajuste, el más empleado es la validación cruzada, que consiste en excluir la observación de uno de los n puntos muestrales y con los $n-1$ valores restantes se predice el valor de la variable en estudio en la ubicación del punto que se excluyó; vía Kriging ordinario o ponderación de la distancia inversa. Este

procedimiento se realiza en forma secuencial con cada uno de los puntos muestrales y así se obtiene un conjunto de n “errores de predicción”; con el cual se determina medidas de precisión como la raíz del error cuadrático medio (RECM) y el coeficiente de determinación R^2 .

III. MATERIALES Y MÉTODOS

3.1 Descripción del área de estudio

El área de estudio corresponde a la cuenca del Lago Titicaca parte peruana, que forma parte del Altiplano del Sur de América y está ubicado al sur del Perú, entre la latitud (13°S - 17°S) y longitud (69°W – 71°W). En la Figura 2 se muestra la cuenca de la vertiente del Lago Titicaca, las sub cuencas y los principales ríos aportantes al Lago (Ramis, Ilave, Coata, Huancané y Zapatilla). Política y administrativamente la zona de estudio se encuentra en el departamento de Puno.

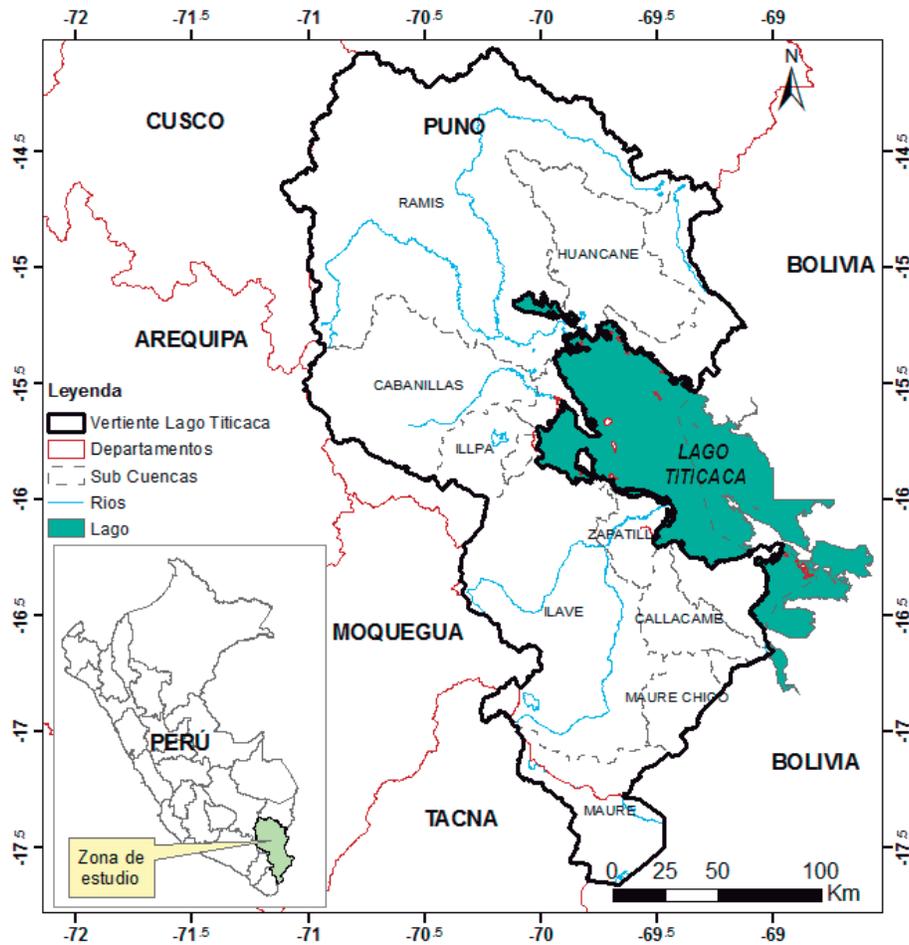


Figura 2: Mapa de ubicación de la cuenca hidrográfica del Lago Titicaca

La cuenca del Lago Titicaca es uno de los sistemas hidrológicos activos y separados que se distinguen en la cuenca endorreica del Altiplano: el Lago Titicaca (3809.5 m.s.n.m.), vierte al Lago Poopó (3686 m.s.n.m.) por intermedio del río Desaguadero. El río Desaguadero que comienza en el Lago Titicaca (parte Boliviana), después de un recorrido de 398 km desemboca en el lago Poopó.

La superficie total de la hondonada del lago Titicaca, hasta el comienzo del Desaguadero, es de 56300 km² (incluidas las aguas superficiales); de esta superficie de la cuenca, las tres cuartas partes están en el Perú (subregión de Puno) y el resto en Bolivia (departamento de La Paz). La superficie de la cuenca de drenaje del lago Titicaca es de 47900 km², o sea, el 85% de la cuenca total, 1/5 de la cual situada en Bolivia y 4/5 en el Perú.

3.1.1 Climatología

Sanabria *et al.*, (2009) describe la climatología en el Altiplano (Vertiente del Lago Titicaca) y menciona que la precipitación está restringida a los meses de noviembre a marzo acumulando valores entre 500 a 900 mm/año (véase Figura 3a), así mismo menciona que las condiciones térmicas están caracterizadas por temperaturas máximas absolutas que oscilan entre 14 a 18 °C y mínimas entre -7 a 3 °C y las temperaturas medias que oscilan entre 4 a 9 °C (véase Figura 3b).

La mayor parte de la precipitación (>70%) se produce durante el verano austral [diciembre a febrero (DEF)]; cuando los vientos del este de nivel medio (600 hPa) y superior (200 hPa), trae aire húmedo y alimentan las tormentas convectivas sobre la meseta del altiplano (Garreaud, 1999; Falvey y Garreaud, 2005). La precipitación de verano exhibe variabilidad sinóptica significativa, explicado en gran medida por la humedad de la capa límite local y el viento zonal en alto nivel de 200 hPa (cerca de 12000 m.s.n.m.) (Garreaud, 2001). El resto del año el Altiplano está influenciado por el flujo de viento del oeste de nivel medio (600 hPa) y alto (200 hPa), que trae el aire muy seco desde el Pacífico y las precipitaciones son casi inexistentes (Vuille y Ammann, 1997).

En escalas de tiempo interanuales, hay una tendencia a una mayor (menor) precipitación en toda la cuenca durante los años de La Niña (El Niño) que también se explica en gran medida por la intensidad de la circulación zonal en alto nivel (Garreaud y Aceituno, 2001; Vuille y Keimig, 2004). En una reciente investigación Lavado y Espinoza

(2014) evalúan los impactos de El niño (EN) y la Niña (LN) en las lluvias del Perú utilizando datos de lluvias mensuales (1965-2007) donde utilizaron 21 estaciones distribuidas sobre la Vertiente hidrográfica del Lago Titicaca, para ello primero sobre la base de años hidrológicos (septiembre a agosto) clasifican los eventos del El niño y La Niña en El Niño fuerte (ENF), El Niño moderado (ENM), La Niña moderada (LNM) y La Niña fuerte (LNF) utilizando el Índice Troup de Oscilación del Sur (IOS) y encuentran que la Vertiente hidrográfica del Lago Titicaca presenta un importante déficit de lluvias durante El Niño fuerte (ENF). Además analizan los principales modos de variabilidad interanual de las lluvias en el Perú utilizando la técnica de funciones ortogonales empíricas (EOF) y los resultados de los modos relacionan con la variabilidad de la temperatura superficial del mar de las diferentes regiones del Pacífico ecuatorial, usando índices que describen de manera independiente la variabilidad en la región Pacífico central (índice C) y en la región del Pacífico este (índice E) definidos por Takahashi et al., (2011) y encuentran que el principal modo de variabilidad de las lluvias (primer componente principal-CP1, que explica el 37% de la varianza total) está asociado a los dos eventos El Niño extraordinarios (1983 y 1998), lo cual generó abundantes lluvias en el norte de la Vertiente del Pacífico y sequías en la Vertiente del Titicaca y la Vertiente del Amazonas, mientras el segundo modo de variabilidad de las lluvias (segundo componente principal-CP2, que explica el 25% de la varianza total) se correlaciona con las anomalías de la temperatura superficial del mar en el Pacífico ecuatorial central, donde condiciones frías en esta región (LN) causan más lluvias de lo normal en la región Andina de la Vertiente del Pacífico, la Vertiente del Titicaca y en la Vertiente del Amazonas.

En cuanto a las tendencias hidrológicas a escala interanual en la Vertiente del Titicaca Lavado et al., (2012) encuentran que no hay cambios y tendencias significativas en cuanto a la precipitación, escorrentía media, escorrentía máxima, pero si en la escorrentía mínima en la cuenca del río Ramis; donde el incremento de esta puede ser asociado con el derretimiento acelerado de los glaciares debido al cambio climático.

Con respecto a las inundaciones Gálvez *et al.*, (2006) mencionan que durante la historia reciente, se ha observado inundaciones debido a las condiciones húmedas prolongadas. Donde condiciones de lluvia persistente en el altiplano superior hacen que los niveles del Lago Titicaca se eleven y el lago se desborde, produciendo: (1) inundaciones en el sector poblado del Lago Titicaca (Bourges *et al.*, 1992) y (2) una migración hacia el sur

del desbordamiento, que a veces llenan el Salar de Uyuni con varios centímetros de agua (Sylvestre *et al.*, 2001).

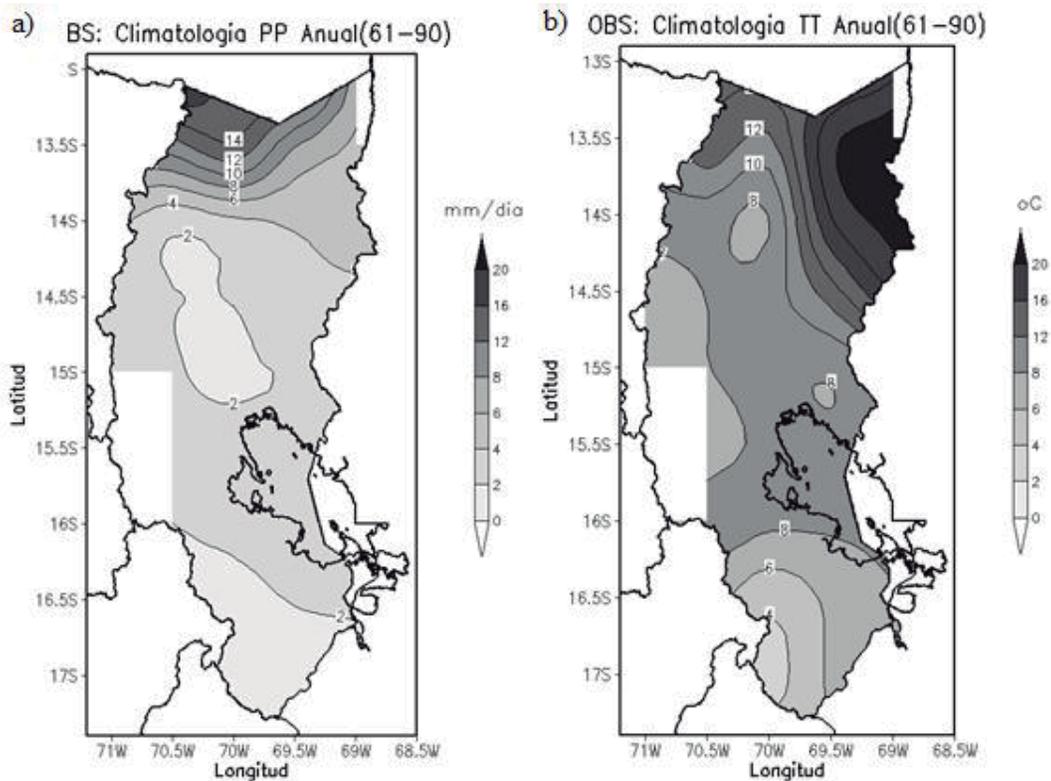


Figura 3: Climatología del Altiplano Peruano (periodo 1961-1990). a) Climatología de la precipitación observada y b) Climatología de la temperatura observada.

FUENTE: Sanabria *et al.*, (2009).

3.2 Materiales

En el desarrollo del análisis regional de frecuencia de las precipitaciones máximas anuales en la cuenca hidrográfica del Titicaca, empleando el método del índice avenida conjuntamente con el uso de los L-momentos, ha de emplearse las siguientes informaciones:

3.2.1 Información básica de sistema de información geográfica

Para identificar el área de influencia de la vertiente del Lago Titicaca (parte Peruana) se ha recopilado inicialmente la información de sistemas hidrográficos de la cuenca del vertiente del Lago Titicaca en formatos de SIG, para ello se recurrió al banco de datos de la Autoridad Nacional del Agua (ANA-PERU).

Para comprender la distribución espacial y altitudinalmente de las estaciones en el área de estudio se ha recopilado el modelo de elevación digital de terreno (DEM) de 90 m de resolución a partir del Shuttle Radar Topography Mission (SRTM) disponible en: <http://srtm.csi.cgiar.org/>

3.2.2 Información de datos observados

Para conformar series de datos anuales de precipitación máxima diaria y precipitación promedio multianual, correspondiente a las estaciones pluviométricas que se encuentran administradas por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), en la cuenca hidrográfica del Lago Titicaca y estaciones vecinas; se han recopilado información de las mismas en el portal web de la Autoridad Nacional del Agua (ANA-PERU) disponible en: <http://www.ana.gob.pe:8080/snirh2/consPluviometria.aspx>

Es importante conocer la longitud mínima de datos para obtener valores adecuados. En tal sentido, algunos autores han realizado recomendaciones sobre la base longitud mínima de registro de variables hidrológicas, por ejemplo la IACWD (Interagency Advisory Committee on Water Data), recomienda utilizar como mínimo 25 años de longitud de registro (IACWD, 1982).

En este estudio se tomó como referencia las recomendaciones de IACWD (1982), por lo que fueron seleccionadas solamente aquellas estaciones con longitud de registro mayor a 25 años, identificadas en el Cuadro 4.

En la Figura 4 se describe la cantidad de estaciones que presentan registro de precipitación máxima diaria en el tiempo. Este gráfico muestra que existe solamente ocho años donde todas las estaciones cuentan con registro. La distribución espacial de las estaciones empleadas se muestra en la Figura 5.

Cuadro 4: Red de estaciones Meteorológicas

Nº	Id estación	Nombre	Cuenca	Latitud	Longitud	Altitud	Longitud Registro
1	826	Ananea	Ramis	-14.6787	-69.5345	4660	42
2	783	Arapa	Ramis	-15.1356	-70.1186	3830	44
3	776	Ayaviri	Ramis	-14.8811	-70.5928	3928	39
4	781	Azangaro	Ramis	-14.9147	-70.1864	3863	34
5	780	Cabanillas	Coata	-15.6439	-70.3386	3890	46
6	788	Capachica	Coata	-15.6097	-69.8469	3819	32
7	158326	Capazo	Mauri	-17.1864	-69.7456	4419	45
8	158329	Chilligua	Ilave	-16.5422	-69.6778	4164	41
9	764	Chuquibambilla	Ramis	-14.7964	-70.7283	3910	36
10	827	Cojata	Suches	-15.0167	-69.3556	4344	37
11	7415	Crucero	Ramis	-14.3556	-70.0169	4130	39
12	157418	Cuyo Cuyo	Inambari	-14.4892	-69.5499	3494	39
13	883	Desaguadero	Alto desaguadero	-16.5761	-69.0339	3860	39
14	786	Huancane	Ramis	-15.2033	-69.7625	3860	44
15	787	Huaraya Moho	Ramis	-15.3897	-69.4914	3890	37
16	158310	Ichuña	Tambo*	-16.1408	-70.5372	3792	43
17	879	Ilave	Ilave	-16.0847	-69.6436	3850	38
18	880	Juli	Titicaca	-16.2033	-69.4572	3825	32
19	763	Lagunillas	Coata	-15.7794	-70.6606	4250	35
20	779	Lampa	Coata	-15.3558	-70.3728	3900	45
21	889	Laraqueri	Ilave	-16.1525	-70.0678	3900	37
22	761	Llally	Ramis	-14.9489	-70.8981	4111	31
23	777	Macusani	Ramis	-14.0678	-70.4236	4331	30
24	878	Mazo Cruz	Ilave	-16.7456	-69.7117	3970	42
25	785	Muñani	Huancane	-14.7794	-69.9658	4119	43
26	762	Pampahuta	Coata	-15.4914	-70.6775	4320	39
27	881	Pizacoma	Alto desaguadero	-16.915	-69.3728	3940	39
28	778	Progreso	Alto desaguadero	-14.6947	-70.3556	3905	42
29	815	Pucará	Ramis	-15.0336	-70.3728	3885	45
30	708	Puno	Titicaca	-15.8303	-70.0167	3840	43
31	157414	Putina	Huancane	-14.9149	-69.868	3878	37
32	823	Santa Rosa	Ramis	-14.6269	-70.7964	3940	27
33	816	Taraco	Ramis	-15.305	-69.9825	3820	45

* Estación que se encuentra en la vertiente del Amazonas,

** Estación que se encuentra en la vertiente del pacífico, el resto pertenece a la vertiente del Lago Titicaca.

Todas las estaciones son del tipo de estaciones meteorológicas convencionales.

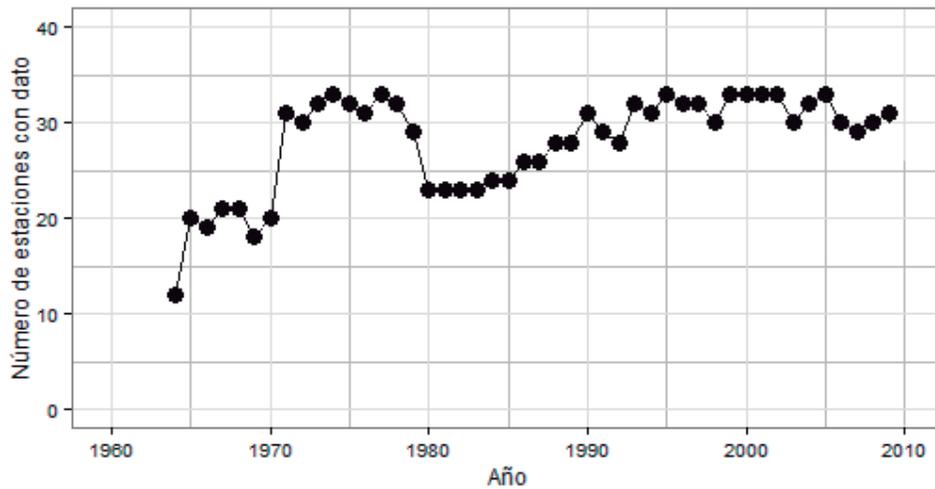


Figura 4: Estaciones con registro de precipitación máxima diaria en el tiempo.

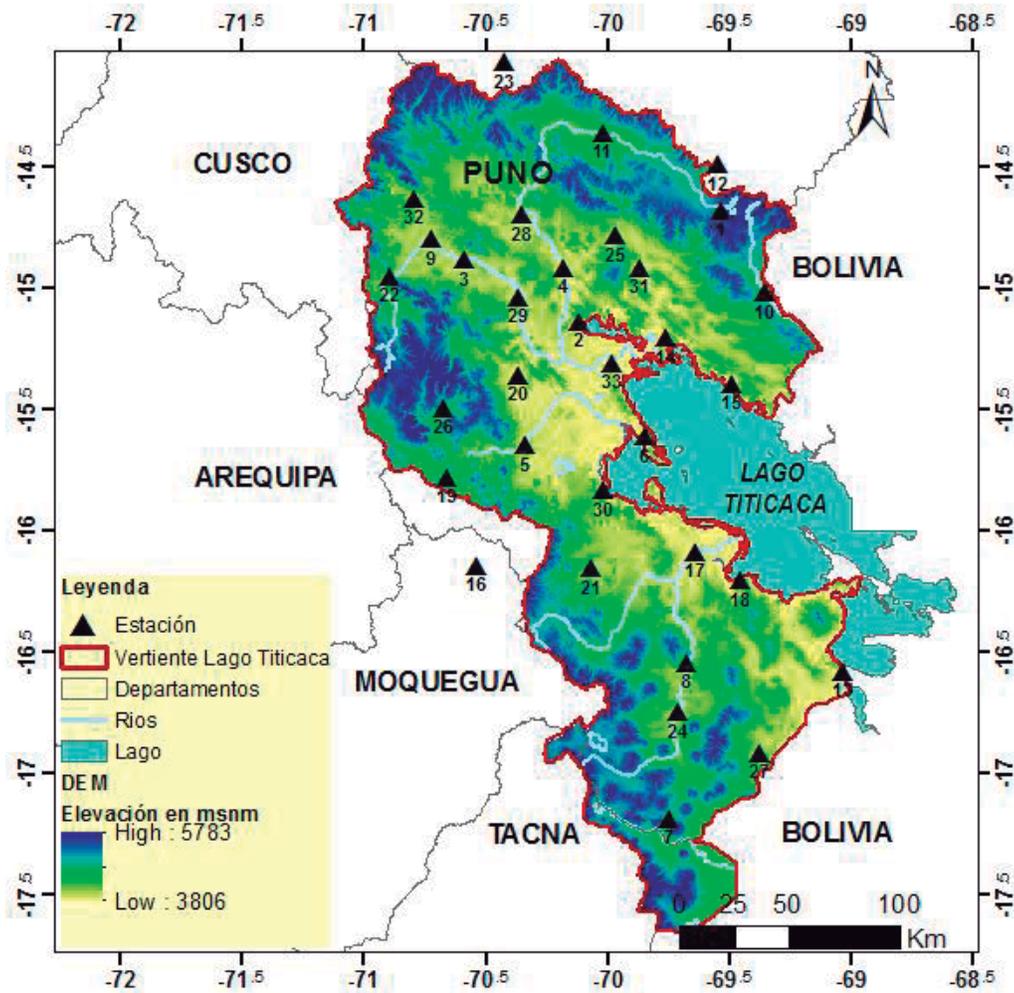


Figura 5: Ubicación de las estaciones meteorológicas.

3.2.3 Programas

Los programas utilizados en la ejecución de la metodología fueron las siguientes:

- **Hydraccess 4.5**, permitió que las observaciones de precipitación diaria sean organizadas en una base de datos en Microsoft Access para seleccionar las observación máxima diaria por año en todas las estaciones, así mismo sirvió para la crítica de datos de precipitación anual, está disponible en: <http://www.ore-hybam.org/>
- **XLSTAT**, es un complemento de estadística para Microsoft Excel, se ha empleado en el análisis de datos outliers mediante la prueba de Grubbs, en la verificación de tendencias por la prueba de Tau de Mann Kendall y en el análisis de ruptura de series por el test de Pettitt. La versión libre es disponible en forma gratuita durante 30 días en el sitio web <http://www.xlstat.com/es/descargar.html>
- **EViews 8**, es un paquete ideal para trabajar con series de tiempo y estadística en general. Se ha empleado para probar el test de Ljun Box en el análisis de autocorrelaciones de series de tiempo. La versión gratuita durante 30 días, está disponible en el sitio web: <http://www.eviews.com/home.html>
- **R (lenguaje de programación)**, es un lenguaje y entorno de programación para análisis estadístico y gráfico.

Se trata de un proyecto de software libre < <http://www.r-project.org/>>, resultado de la implementación GNU del premiado lenguaje S. R y S-Plus -versión comercial de S- son, probablemente, los dos lenguajes más utilizados en investigación por la comunidad estadística.

Fue desarrollado inicialmente por Robert Gentleman y Ross Ihaka del Departamento de Estadística de la Universidad de Auckland en 1993. Su desarrollo actual es responsabilidad del *R Development Core Team*.

En el contexto de esta monografía para el análisis exploratorio y todo el procedimiento del análisis regional de frecuencias basado en L-momentos se han codificado en el programa R, haciendo uso principalmente de los siguientes paquetes:

*lmomRFA*¹, funciones para el análisis regional de frecuencia utilizando los métodos de Hosking y Wallis (1997).

*lmom*², contiene funciones para evaluar la función de distribución acumulativa y la función cuantil de la distribución, para calcular los L-momentos dados los parámetros y para calcular los parámetros dados los L-momentos.

*Fitdistrplus*³, se ha empleado para determinar los parámetros de la distribución Gumbel, por el método de máxima verosimilitud. Luego los cuantiles determinados para distintos periodos de retorno por el método local se comparan con los resultados del análisis regional.

*Gstat*⁴, es un paquete para el modelado geoestadístico espacial y espacio-temporal, la predicción y simulación (Pebesma, 2004). Se ha empleado para el mapeo mediante la interpolación de ponderado de distancia inversa y Kriging ordinario.

3.3. Metodología

Una vez conformada las series de datos anuales de precipitación máxima diaria y la precipitación promedio multianual para todas las estaciones, para cumplir con los objetivos se sigue la siguiente estructura del trabajo de tesis que se muestra en la Figura 6. Este esquema en resumen consiste en realizar el análisis exploratorio para detectar los datos atípicos, verificar los supuestos del análisis regional principalmente para determinar la estacionariedad, homogeneidad e independencia serial de los datos, realizar el ARF aplicando la técnica de L-momento y realizar la interpolación espacial para obtener mapas de variaciones espaciales de índice de avenida y cuantiles de precipitación extrema para diferentes periodos de retorno. Adicionalmente los cuantiles estimados por el ARF serán comparados con un análisis local donde los cuantiles son estimados a partir de una distribución biparamétrica (Gumbel).

¹ <http://cran.r-project.org/web/packages/lmomRFA/index.html>

² <http://cran.r-project.org/web/packages/lmom/index.html>

³ <http://cran.r-project.org/web/packages/fitdistrplus/index.html>

⁴ <http://cran.r-project.org/web/packages/gstat/index.html>

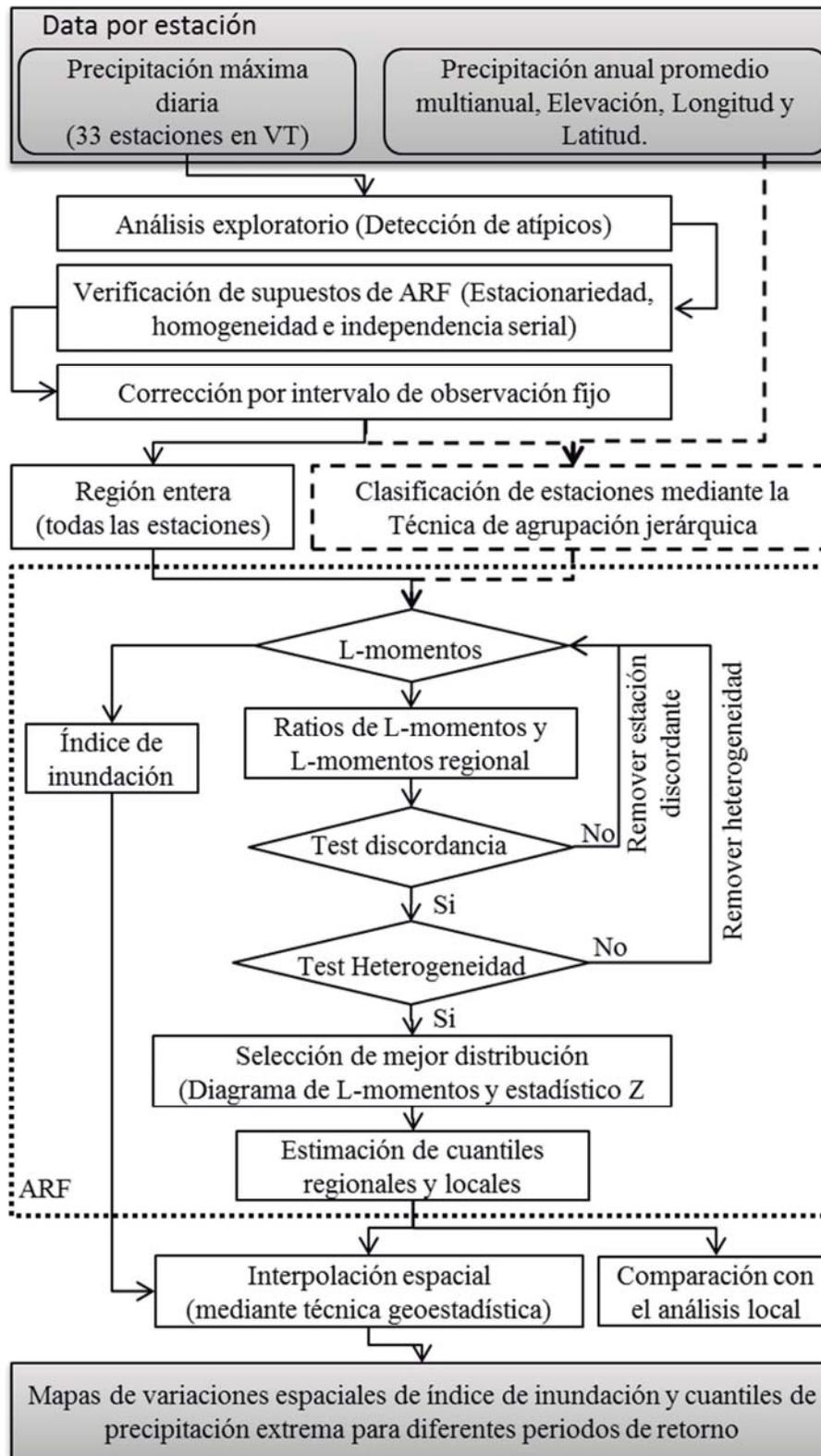


Figura 6: Diagrama esquemático de la metodología.

3.3.1 Análisis exploratorio de datos

Esta etapa consiste en la detección de datos atípicos (outliers) de manera cualitativa y cuantitativa, mediante los gráficos de Boxplots y el test de Grubbs respectivamente. Los posibles outliers que serán detectados con los gráficos de Boxplots, son verificados con la prueba estadística de Grubbs, y en caso de ser probado estadísticamente, se contrasta conociendo su fecha de ocurrencia con el comportamiento de eventos de estaciones contiguas; de aquí se pone en juicio los datos atípicos detectados para ser eliminados o incluidos en la siguiente etapa de análisis.

La serie de datos de precipitación máxima diaria previamente a ser sometido al análisis de detección de outliers mediante el test de Grubbs, será normalizada tomando para ello los logaritmos a las observaciones.

Cabe señalar que en las series anuales de precipitación máxima diaria, los datos faltantes no serán completados; puesto que esta variable es sumamente aleatoria en comparación con las series mensuales o anuales, donde se tiene diferentes técnicas de completación de datos faltantes.

3.3.2 Supuestos del análisis regional de frecuencias

Debido a que la variable analizada (series de precipitación máxima diaria) no sigue necesariamente una distribución normal, en esta etapa las series libres de datos outliers se someterá a las pruebas estadísticas no paramétricas, mediante el uso del programa XLSTAT y EViews 8, para verificar los supuestos del ARF.

El primer supuesto es que las series no presenten tendencias a disminuir o aumentar en el tiempo, para ello se aplica el test no paramétrico de Mann Kendall. La homogeneidad de las series (libre de cambios en la media) se corrobora con el test no paramétrico de Pettit. Mientras la independencia serial y la aleatoriedad que se debe a procesos naturales se verifica con el test de Ljun Box. Las series que cumplan los tres supuestos pasan a la siguiente etapa, y los que no cumplen serán excluidos del ARF basado en L - momentos.

Cabe señalar que a los datos de precipitación máxima diaria al final de esta etapa se aplicará un factor de corrección (C_{OITF}), debido que la medida y reporte de las observaciones analizadas son realizados en intervalos de tiempo fijo. El factor de corrección C_{OITF} varía con la longitud del periodo de observación (24 horas para medición

diaria). Un factor de 1.13 ha sido estimada desde consideraciones teóricas (Weiss, 1964) y también se han encontrado en numerosos estudios para corregir los datos de precipitación máxima diaria (Miller *et al.*, 1973; citada por Wallis *et al.*, 2007).

En este estudio se aplicará el factor de corrección $C_{OFF} = 1.13$; según lo recomendado por Weiss, para ajustar las observaciones de las precipitaciones máximas diarias y denominar de aquí en adelante más bien como los valores de precipitación máxima anual en 24 horas de duración (PMA). Los análisis posteriores serán realizados con series de datos de precipitación máxima de 24 horas de duración.

3.3.3 Filtrado de las observaciones empleando la medida de discordancia

Específicamente es la primera etapa del análisis regional de frecuencias basado en L-momentos, que consiste en identificar las estaciones que son groseramente discordantes con el grupo como un todo. Se determina la medida de discordancia D_i para cada estación lo cual mide la similitud entre las distribuciones de frecuencia de la estaciones, visualizada en términos de sus L-momentos, detectando las estaciones que difieren significativamente del resto en función a sus L-CV, L-asimetría y L-curtosis. Los valores de D_i calculados será comparado con un valor crítico que depende del número de estaciones que conforman la región. Para su cálculo se utiliza el paquete *lmomRFA* en el programa R.

La medida de discordancia señalada en la sección 2.7.1, se evalúa en primera instancia asumiendo que todas las estaciones forman una sola región homogénea, de aquí:

- (a) Cuando se contraste que todo o la gran mayoría de estaciones no son discordantes se procede a evaluar si todas las estaciones forman una sola región homogénea con el estadístico de heterogeneidad (H) según lo señalado en la sección 2.7.2 en caso se cumpla las medidas de región homogénea, se procede con la etapa de selección de una función de distribución de frecuencia para la región, descrita en la sección 3.3.5.
- (b) En caso se encuentre varias estaciones discordantes se continuará con la formación de regiones homogéneas (sección 3.3.4), sin excluir las estaciones discordantes. Una vez conformada las regiones según la sección 3.3.4; nuevamente se realiza el análisis de discordancia.

3.3.4 Identificación de regiones homogéneas

De acuerdo al procedimiento anterior en caso de requerir, se procede a formar regiones homogéneas o formar grupos de estaciones que satisfagan la medida de condición de homogeneidad ($H_j < 1$), para formar regiones homogéneas, se realizará la agrupación preliminar por el análisis clúster u otros métodos, considerando a su vez formar regiones coherentes con la concepción física de la cuenca.

En el contexto de la presente tesis, después de un proceso de verificación con la medida de heterogeneidad se ha encontrado que todas las estaciones conforman una sola región homogénea por lo que no es necesaria la conformación de varias regiones.

3.3.5 Selección de una función de distribución de frecuencia para la región

Para la selección de una o varias funciones de distribución de frecuencia que modelizen la región homogénea, se evalúa la performance de cinco distribuciones de probabilidad más empleadas en el estudio de variables ambientales y recomendados por Hosking y Wallis (1997).

Se evaluará las funciones de distribución siguientes: logística generalizada (GLO), general de valores extremos (GEV), Pareto generalizado (GPA), log normal de tres parámetros (LN3) o normal generalizado (GNO) y Pearson tipo III (PE3). De éstas se selecciona las distribuciones candidatas de mejor ajuste en función a los diagramas de los ratios de L-momentos y el estadístico $|Z^{DIST}| \leq 1.64$ a un nivel de confianza del 90 por ciento. Los cálculos han sido realizados con el paquete lmomRFA en el programa R.

3.3.6 Estimación de los cuantiles regionales de frecuencia y locales

Seleccionada la función de distribución adecuada para la región homogénea definida, se estima los cuantiles regionales y locales según el *algoritmo regional de L – momentos*, señalado en la sección 2.7.4 e implementada en el paquete lmomRFA.

Los cuantiles se definen como los montos de precipitación asociados a una determinada probabilidad. Es así como los cuantiles dependen directamente de la función de probabilidad. En este contexto, determinada los parámetros de las distribuciones de probabilidad candidatas se estiman los cuantiles regionales o la curva de crecimiento regional para periodos de retorno de 5, 10, 20, 50, 100, 200, 500 y 1000 años.

Cuando más de una distribución dada se ajusta a los datos regionales, se elegirá la distribución más robusta, que podría dar estimaciones razonablemente precisas para periodos de retorno más extremos. La robustez de la distribución es medida mediante la comparación de las medidas de precisión como bias relativo $B^R(F)$, bias relativo absoluto $A^R(F)$ y el error cuadrático medio relativo $R^R(F)$ de los cuantiles extremos estimados. Para este propósito se realiza una evaluación estadística de la precisión de la curva de crecimiento regional empleando el paquete *lmomRFA* que también implementa la simulación para probar la robustez de las distribuciones candidatas el cual se fundamenta en el algoritmo regional de L-momentos de Hosking y Wallis (1997); descritas en el Cuadro 3. Como es de conocimiento que la precisión viene por el número de simulaciones, por lo que en esta fase se realizará 10 000 repeticiones de cara a un mayor vigor.

Finalmente se elegirá una sola distribución adecuada según las medidas de precisión, para estimar los cuantiles locales y graficar los mismos con sus respectivos intervalos de confianza al 90% para el rango de periodos de retorno de 2 a 1000 años y en la misma gráfica con propósitos de comparación en este estudio se incluye la posición de los datos reales de PMA según la posición de frecuencia empírica de Weibull, por ser de mayor aplicación en el Perú.

3.3.7 Comparación con el método de análisis local

Para cada estación, se ajusta la distribución Gumbel a la serie de datos de PMA, sus parámetros son determinados mediante el método de máxima verosimilitud empleando el paquete *Fitdistrplus*, luego se estiman los cuantiles y éstas se comparan con los resultados que se obtendrían mediante el enfoque del análisis regional.

3.3.8 Mapeo de índice de avenida y eventos extremos

Se obtendrán mapas de regionalización espacial del índice de avenida y PMA para distintos periodos de retorno en la cuenca hidrográfica del Lago Titicaca empleando las técnicas de interpolación de ponderado de distancia inversa y Kriging ordinario. La performance de las dos técnicas para el mapeo final se verifica con una validación cruzada y medidas de precisión.

El paquete gstat se ha empleado para el mapeo por ambas técnicas, tanto como para el modelado de variogramas, predicción y validación cruzada. Y la visualización de datos de alta resolución del mapeo se realizó con el paquete lattice⁵. Cabe señalar que todos los paquetes mencionados para su uso fueron codificados en el programa R.

⁵ <http://cran.r-project.org/web/packages/lattice/index.html>

IV. RESULTADOS Y DISCUSIÓN

4.1 Análisis exploratorio de datos

En esta etapa se ha realizado los gráficos de Boxplots, para detectar los datos atípicos (outliers) de manera cualitativa, la misma que se corrobora mediante el test estadístico de Grubbs.

4.1.1 Diagramas de Boxplots

En la Figura 7, se muestra la distribución de los datos para cada estación donde la mayoría de las estaciones tiene la concentración de sus datos dentro de los límites de confianza al 90%. Con excepción de algunas estaciones donde se presentan extremos (probables outliers).

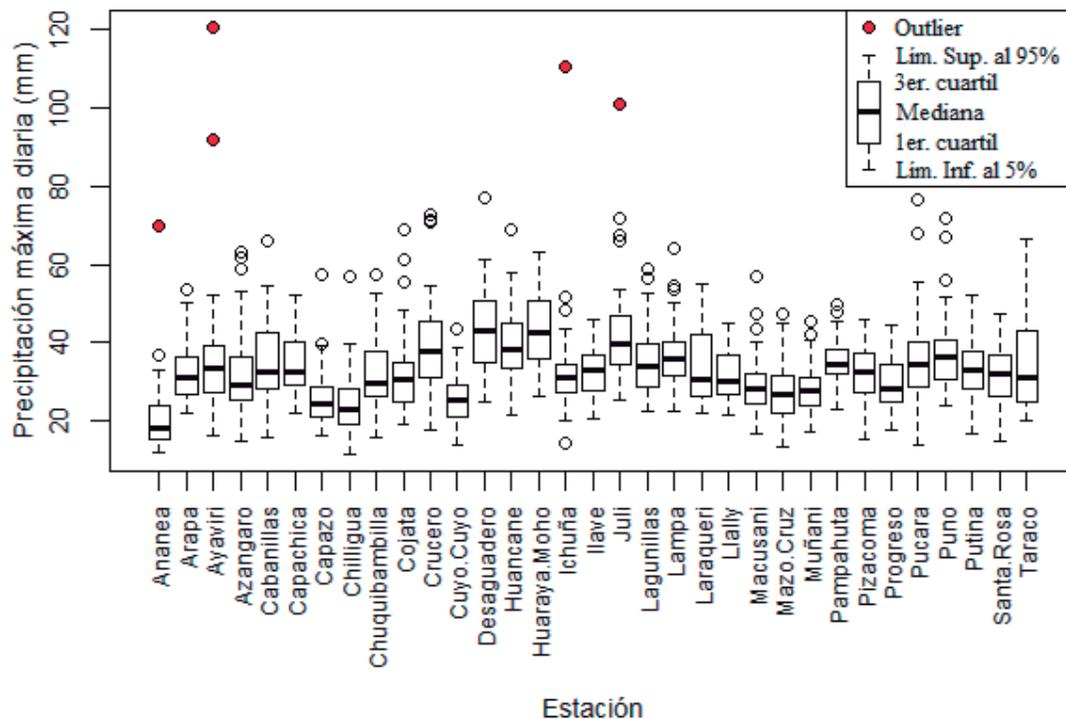


Figura 7: Gráficos boxplots para detección de outliers.

De la Figura 7, se puede catalogar que las estaciones de Ananea (01 outliers), Ayavirí (02), Ichuña (01) y Juli (01); parecen presentar datos atípicos por superar el rango intercuartil según los diagramas de boxplots. También se visualiza que las observaciones en todas las estaciones no tienen una distribución simétrica con respecto a la mediana por lo que no siguen una distribución normal, de aquí, para someter al análisis de outliers con el test estadístico de Grubbs; primero se normaliza las series tomando los logaritmos, ya que el test de Grubbs se basa en la normalidad de la serie.

4.1.2 Prueba de Grubbs para valores atípicos

Se muestra a manera de ejemplo el procedimiento completo de detección de outliers para la estación Ananea.

Cuadro 5: Estadísticas descriptivas de la estación Ananea

Variable	Observaciones	Mínimo	Máximo	Media	Desviación típica
Ananea (reales)	42	11.6	70	19.57	12.12
Ananea (normalizado)	42	2.451	4.248	2.974	0.351

Los valores reales de las observaciones de precipitación máxima diaria (mm), primeramente se normaliza tomando los logaritmos, luego estas fueron sometidas a la prueba o test de Grubbs. Aplicando la ecuación 5 y 6 se obtiene el estadístico Z (Cuadro 6) para cada observación.

Cuadro 6: Estadístico Z de test de Grubbs para la estación Ananea

Nº	Dato real	Dato normalizado	Estadístico Z	Nº	Dato real	Dato normalizado	Estadístico Z
1	17.80	2.879	-0.270	22	14.00	2.639	-0.955
2	15.00	2.708	-0.758	23	11.60	2.451	-1.491
3	20.00	2.996	0.063	24	16.00	2.773	-0.574
4	29.40	3.381	1.161	25	15.20	2.721	-0.720
5	70.00	4.248	3.636	26	17.30	2.851	-0.351
6	16.10	2.779	-0.556	27	13.50	2.603	-1.058
7	16.20	2.785	-0.538	28	12.10	2.493	-1.371
8	22.60	3.118	0.411	29	18.20	2.901	-0.206
9	14.00	2.639	-0.955	30	22.40	3.109	0.386
10	15.40	2.734	-0.683	31	16.20	2.785	-0.538
11	14.60	2.681	-0.835	32	19.20	2.955	-0.054
12	18.30	2.907	-0.191	33	25.20	3.227	0.722
13	15.60	2.747	-0.646	34	23.60	3.161	0.535

14	16.10	2.779	-0.556	35	23.80	3.170	0.559
15	16.50	2.803	-0.486	36	36.90	3.608	1.809
16	14.60	2.681	-0.835	37	31.00	3.434	1.313
17	24.50	3.199	0.641	38	21.40	3.063	0.256
18	12.20	2.501	-1.347	39	31.50	3.450	1.358
19	26.70	3.285	0.887	40	33.10	3.500	1.499
20	19.10	2.950	-0.069	41	22.40	3.109	0.386
21	19.30	2.960	-0.039	42	23.00	3.135	0.461

El estadístico Z, se verifica con el Z crítico para un nivel de significación de $\alpha = 0.05$ y se prueba la siguiente hipótesis:

Hipótesis nula: No hay valores atípicos en los datos,

Hipótesis alternativa: El valor mínimo o máximo es un valor atípico

Para las 42 observaciones y $\alpha = 0.05$ se tiene un Z crítico de 3.057 y un p-valor de 0.003, de aquí se desprende que los valores de Z mayores que Z crítico hacen que se rechace la hipótesis nula y aceptar la hipótesis alternativa.

A partir del Cuadro 6 se tiene que la estación Ananea presenta un $Z=3.636$ que corresponde al dato real de 70 mm; es mayor al Z crítico (3.057), por tanto se rechaza la hipótesis nula y se acepta la hipótesis alternativa. El riesgo de rechazar la hipótesis nula H_0 cuando es verdadera es de 0.003%, menor que 5% para que sea significativa.

En el Cuadro 7, se muestra el resumen del test de Grubbs para las estaciones que presentan datos atípicos, en la cual se observa que en total se detectaron 5 outliers (0.4% del total de datos analizados), confirmando lo señalado con los diagramas de Boxplots, las mismas fueron corroboradas en la base de datos conociendo su fecha de ocurrencia con estaciones contiguas, al no encontrar relación alguna se procedió con su exclusión para el análisis de supuestos del análisis regional.

Cuadro 7: Datos atípicos detectados por el test de Grubbs

Estación	Dato real	Normalizado	Z	Z(Valor crítico)	p-valor	Paso
Ananea	70.0	4.248	3.636	3.057	0.003	1
Ayavirí	120.5	4.792	3.542	3.025	0.004	1
	92.0	4.522	3.493	3.014	0.004	1
Ichuña	110.4	4.704	4.115	3.067	0.000	1
Juli	100.8	4.613	3.065	2.938	0.028	1

4.2 Supuestos del análisis regional de frecuencias

Se analiza los tres supuestos del análisis regional de frecuencias: estacionariedad (libre de tendencias), independencia serial y distribución idéntica de los datos o sin cambios en la media. Debido que la variable analizada no tiene necesariamente una distribución normal de acuerdo a los diagramas de Boxplots, las series libres de datos atípicos se sometió a las pruebas estadísticas no paramétricas.

4.2.1 Análisis de tendencia

La prueba no-paramétrica más utilizada en la identificación de tendencia en series de variables hidrometeorológicas ha sido la de Mann Kendall. La autocorrelación presente en algunas de estas series, afecta los resultados de la prueba, causando problemas en la detección y evaluación del nivel de significación. La presencia de autocorrelación positiva tiende a aumentar la probabilidad de detección de tendencia, cuando en efecto, ésta no existe, mientras que la autocorrelación negativa tiende a disminuirla (Yue *et al.*, 2002).

Para corregir este efecto, Yue *et al.*, (2002) presentan una técnica de pre-blanqueo (trend-free prewhitening, TFPW) que remueve la correlación serial sin alterar la tendencia presente en la serie. Esta última técnica fue aplicada a las series de precipitación máxima diaria puesto que ésta y las variables hidrológicas casi siempre presentan autocorrelaciones seriales.

En el Cuadro 8 se describe los resultados del análisis de tendencias por el test estadístico de Mann Kendall mediante la técnica del pre blanqueo de Yue y Wang, para un nivel de significación $\alpha = 0.05$. Dónde:

H₀ (Hipótesis Nula): No existe una tendencia en la serie

H_a (Hipótesis Alternativa): Hay una tendencia en la serie

Cuadro 8: Resumen del test de Mann Kendall para el análisis de tendencias

Estación	Tau de Kendall	S	Var(S)	p	Hipótesis Nula
Ananea	0.29	233	20842.86	0.11	aceptado
Arapa	0.11	107	21731.83	0.47	aceptado
Ayavirí	0.34	225	12804.35	0.05	rechazado
Azangaro	0.11	64	9599.00	0.51	aceptado
Cabanillas	0.18	184	24549.55	0.24	aceptado
Capachica	-0.04	-21	9849.16	0.83	aceptado
Capazo	-0.14	-139	21954.67	0.35	aceptado
Chilligua	0.32	261	19340.33	0.06	aceptado
Chuquibambilla	-0.01	-7	11253.98	0.95	aceptado
Cojata	-0.15	-98	12143.35	0.37	aceptado
Crucero	-0.37	-277	15069.41	0.02	rechazado
Cuyo Cuyo	0.10	79	15961.69	0.53	aceptado
Desaguadero	0.16	122	14547.44	0.31	aceptado
Huancané	-0.07	-69	23170.76	0.65	aceptado
Huaraya Moho	-0.04	-29	11850.98	0.79	aceptado
Ichuña	-0.17	-146	17464.91	0.27	aceptado
Ilave	-0.06	-44	13882.79	0.71	aceptado
Juli	-0.16	-72	7406.04	0.40	aceptado
Lagunillas	-0.07	-40	10775.16	0.70	aceptado
Lampa	0.08	79	23515.21	0.61	aceptado
Laraqueri	0.04	26	12270.36	0.81	aceptado
Llally	0.12	57	7367.63	0.51	aceptado
Macusani	0.17	74	6719.02	0.37	aceptado
Mazo Cruz	-0.11	-92	17524.28	0.49	aceptado
Muñani	-0.06	-58	18756.95	0.67	aceptado
Pampahuta	0.26	189	14522.12	0.12	aceptado
Pizacoma	0.27	200	14223.96	0.09	aceptado
Progreso	0.12	105	17396.61	0.43	aceptado
Pucará	0.11	110	23322.45	0.47	aceptado
Puno	0.07	60	19420.44	0.67	aceptado
Putina	-0.06	-37	12646.73	0.74	aceptado
Santa Rosa	-0.33	-117	5056.89	0.10	aceptado
Taraco	0.20	196	21791.93	0.18	aceptado

De acuerdo al Cuadro 8, la hipótesis nula: No existe una tendencia en la serie fue rechazada en las estaciones de Ayavirí y Crucero, que representan el 6.06% de las series analizadas.

4.2.2 Análisis de homogeneidad

Se aplicó el test de Pettitt para el análisis de rotura de series o cambios en la media, para verificar la homogeneidad de la serie. La hipótesis nula y alternativa probada para un nivel de significación $\alpha = 0.05$ es:

H0 (Hipótesis Nula): Los datos son homogéneos

Ha (Hipótesis Alternativa): Hay una fecha en la que hay un cambio en los datos

Cuadro 9: Resumen del test de Pettit para el análisis de homogeneidad

Estación	K	Año de cambio (t)	P	Hipótesis Nula
Ananea	281.0	1993	0.001	rechazado
Arapa	152.0	1979	0.306	aceptado
Ayavirí	254.0	1978	0.000	rechazado
Azángaro	77.0	1969	0.627	aceptado
Cabanillas	259.0	1992	0.020	rechazado
Capachica	93.0	1999	0.302	aceptado
Capazo	167.0	1984	0.241	aceptado
Chilligua	266.0	1992	0.001	rechazado
Chuquibambilla	61.0	2000	0.904	aceptado
Cojata	125.0	1988	0.228	aceptado
Crucero	245.0	1994	0.001	rechazado
Cuyo Cuyo	128.0	1996	0.333	aceptado
Desaguadero	140.0	1998	0.201	aceptado
Huancané	188.0	1999	0.112	aceptado
Huaraya Moho	59.0	1979	0.943	aceptado
Ichuña	132.0	1987	0.374	aceptado
Ilave	113.0	1986	0.375	aceptado
Juli	98.0	1988	0.192	aceptado
Lagunillas	88.0	1987	0.529	aceptado
Lampa	128.0	1971	0.528	aceptado
Laraqueri	76.0	1992	0.782	aceptado
Llally	74.0	1966	0.497	aceptado
Macusani	122.0	1997	0.041	rechazado
Mazo Cruz	117.0	1986	0.515	aceptado
Muñani	105.0	1971	0.696	aceptado
Pampahuta	205.0	1991	0.016	rechazado
Pizacoma	179.0	1979	0.045	rechazado
Progreso	127.0	1991	0.414	aceptado
Pucará	166.0	2000	0.244	aceptado
Puno	101.0	1981	0.732	aceptado
Putina	88.0	2004	0.622	aceptado
Santa Rosa	112.0	1978	0.021	rechazado
Taraco	213.0	1983	0.063	aceptado

Con la aplicación estadística de la prueba de Pettitt se detectaron rupturas importantes en las series históricas anuales de precipitación máxima diaria de 9 estaciones (Cuadro 9), que representa el 27.27% de las estaciones analizadas.

Se observa (Cuadro 9), cuando el valor p calculado es mayor que el nivel de significación $\alpha=0.05$, no se puede rechazar la hipótesis nula H_0 . Debido que el riesgo de rechazar la hipótesis nula H_0 cuando es verdadera es de $p\%$. Así mismo, se observa que los cambios en su mayoría ocurrió entre los años de 1990 y 1998, pudiendo atribuirse a efectos locales como posible variabilidad o en mayores casos a un cambio climático, que podrían ser propios de la zona; los cuales son comportamientos propios en la naturaleza, o simplemente son posibles errores de medición.

4.2.3 Análisis de independencia serial

Se verifica la independencia serial aplicando la función de autocorrelación (FAC) y el estadístico Q de Ljung-Box el cual es una forma más cuantitativa para probar la autocorrelación en múltiples retardos conjuntamente Ljung y Box (1978).

La hipótesis nula para un nivel de significación $\alpha = 0.05$ es que los primeros m autocorrelaciones son conjuntamente a cero. El número de retardos se considera 10, puesto que la longitud promedio de las series es aproximadamente 39 años.

En la Figura 8, se aprecia los resultados de la prueba para aquellas estaciones en las cuales se detectó la autocorrelación significativa.

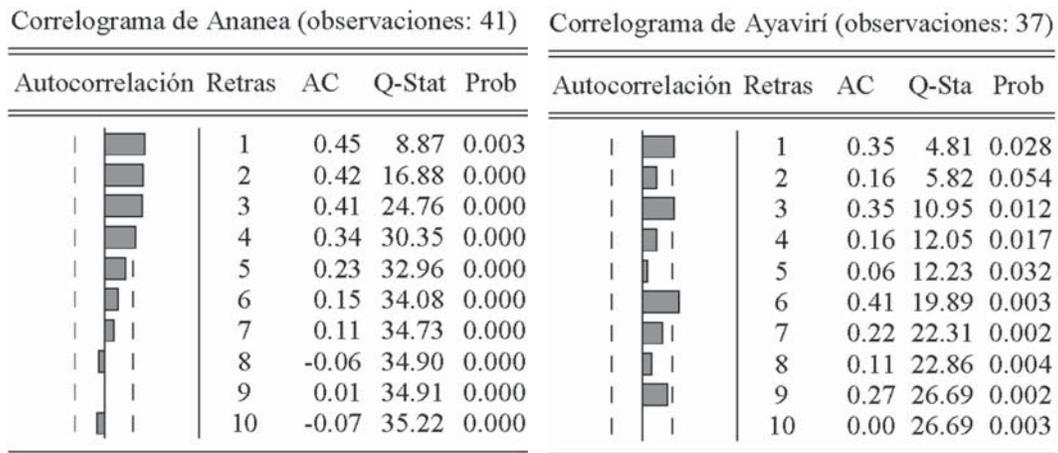


Figura 8: Autocorrelaciones detectadas por el test de Ljung Box.

Cabe señalar que la estación Ananea y Ayavirí que representan el 6.06 % de las estaciones; presentan autocorrelación significativa con varios retrasos. Por tanto las dos estaciones no presentan independencia serial; siendo la ocurrencia de los eventos extremos posiblemente gobernada por procesos físicos no aleatorios en estas dos estaciones.

Realizada el análisis de los supuestos del análisis regional de frecuencias se observa que las estaciones que presentan tendencia son: Ayavirí y Crucero. Con datos no homogéneos o con saltos en la media: Ananea, Ayavirí, Cabanillas, Chilligua, Crucero, Macusani, Pampahuta, Pizacoma y Santa rosa. Y con datos sin independencia serial, altamente autocorrelacionados: Ananea y Ayavirí.

El enunciado para el análisis de frecuencias es que se cumpla los tres supuestos, de aquí, las estaciones que no cumplen son Ananea y Ayavirí; siendo excluidos para el análisis de frecuencias. La estación Crucero no cumple el supuesto de libre de tendencia y homogeneidad de manera que también fue excluida. Las estaciones de Cabanillas, Chilligua, Macusani, Pampahuta, Pizacoma y Santa rosa; fueron consideradas para el análisis de frecuencias a pesar de no cumplir con el supuesto de homogeneidad, pudiendo ser atribuida probablemente a la variabilidad propia de la zona; puesto que la mayoría de los cambios ocurre en el periodo de 1990 y 1998. El comportamiento de estas estaciones fue sometida por un análisis más riguroso con la medida de discordancia, en caso persistieran ser discordantes con las estaciones que conforman una región homogénea son eliminadas o excluidas del análisis regional de frecuencias.

4.3 Filtrado de datos usando medidas de discordancia

Los resultados y análisis que se muestra en esta etapa son para las series de datos de precipitación máxima en 24 horas de duración de las estaciones que pasaron el análisis de supuestos (30 estaciones).

La medida D_i indica cuán lejos se encuentran los ratios de L – momentos de cada estación con respecto a los ratios de L – momentos regionales, Las mismas se muestran en el Cuadro 10 y Figura 9.

Se aprecia que las estaciones de Ilave, Azángaro, Cojata, Capazo, Taraco, Llally, Ichuña y Laraqueri (Figura 9); parecen ser discordantes con el resto de las estaciones ya que sus valores se encuentran alejados del promedio regional (punto en azul).

Cuadro 10: Valores de los ratios de L – momentos de las estaciones

Estación N°	Estación nombre	Longitud n	Media l_1	L-CV t	L-Asimetría t_3	L-Curtosis t_4
1	Arapa	44	37.27	0.1388	0.2221	0.1386
2	Azángaro	34	37.74	0.2113	0.2895	0.1895
3	Cabanillas	46	39.18	0.1694	0.1162	0.1535
4	Capachica	32	39.29	0.1353	0.1825	0.1049
5	Capazo	45	29.67	0.1456	0.2978	0.2024
6	Chilligua	41	27.24	0.1735	0.1529	0.2078
7	Chuquibambilla	36	37.13	0.1557	0.2114	0.1139
8	Cojata	37	36.48	0.1758	0.2976	0.2258
9	Cuyo Cuyo	40	29.18	0.1356	0.1713	0.1363
10	Desaguadero	39	49.66	0.1503	0.0981	0.0978
11	Huancané	44	44.39	0.1422	0.1228	0.1650
12	Huaraya Moho	37	48.86	0.1293	0.0813	0.0870
13	Ichuña	42	35.34	0.1261	0.0863	0.2415
14	Ilave	38	36.62	0.1149	0.0011	0.0727
15	Juli	31	47.26	0.1409	0.2529	0.1930
16	Lagunillas	35	39.77	0.1454	0.2036	0.1422
17	Lampa	45	41.36	0.1373	0.1366	0.1839
18	Laraqueri	37	38.80	0.1559	0.1841	0.0373
19	Llally	31	36.13	0.1211	0.0662	0.0287
20	Macusani	30	33.05	0.1591	0.2271	0.2534
21	Mazo Cruz	42	30.80	0.1483	0.1396	0.1410
22	Muñani	43	31.73	0.1254	0.0731	0.1631
23	Pampahuta	39	39.49	0.1010	0.0518	0.1742
24	Pizacoma	39	36.53	0.1398	0.0224	0.0884
25	Progreso	42	33.12	0.1409	0.1192	0.1139
26	Pucará	45	40.97	0.1658	0.2004	0.2430
27	Puno	43	42.29	0.1446	0.2233	0.2242
28	Putina	37	37.90	0.1376	0.0614	0.1145
29	Santa Rosa	27	36.07	0.1372	0.0279	0.1169
30	Taraco	45	38.64	0.1870	0.2637	0.0856

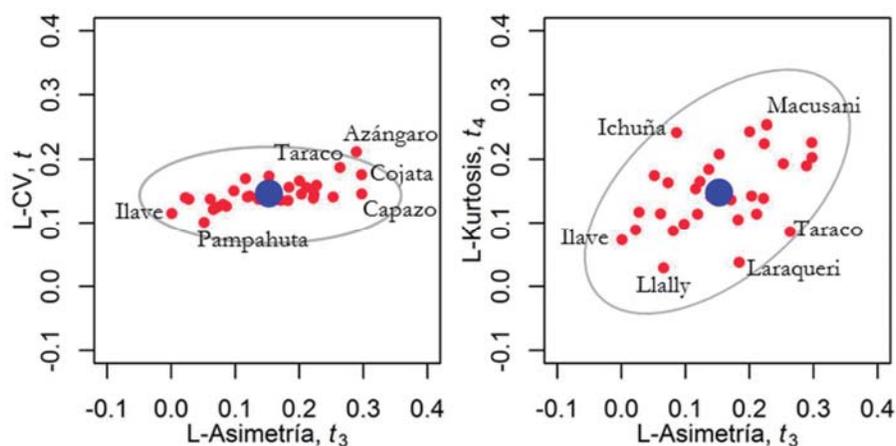


Figura 9: Ratios de los L – momentos de las estaciones.

Hosking y Wallis (1997) sugirieron que una estación es considerada como discordante si su valor D_i excede el valor crítico. De acuerdo a la metodología propuesta primero se considera el análisis considerando todo el conjunto de 30 estaciones como una sola región. Para el filtrado de la estación discordante, la medida de discordancia, D_i se ha calculado para cada estación (ver Cuadro 11). Para 30 estaciones de la región, el valor crítico de D_i es 3. La estación 2, Azángaro, tiene $D_i = 3.04$ ligeramente mayor que el valor crítico de 3. Azángaro cuenta con alta L-asimetría y moderada L-kurtosis pero muy alta L-CV que puede causar el alto valor D_i .

Cuadro 11: Valores de la medida de discordancia de las estaciones

Estación N°	Estación nombre	Discordancia	
		D*	D
1	Arapa	0.83	0.81
2	Azángaro	3.04	-----
3	Cabanillas	1.19	1.53
4	Capachica	0.72	0.71
5	Capazo	1.80	1.73
6	Chilligua	1.41	1.82
7	Chuquibambilla	0.48	0.51
8	Cojata	1.15	1.44
9	Cuyo Cuyo	0.31	0.30
10	Desaguadero	0.47	0.51
11	Huancane	0.12	0.12
12	Huaraya Moho	0.45	0.45
13	Ichuña	1.79	1.75
14	Ilave	1.22	1.30
15	Juli	1.12	1.08
16	Lagunillas	0.30	0.28
17	Lampa	0.25	0.23
18	Laraqueri	1.74	1.74
19	Llally	1.53	1.57
20	Macusani	1.05	1.12
21	Mazo Cruz	0.03	0.05
22	Muñani	0.54	0.55
23	Pampahuta	1.79	2.16
24	Pizacoma	1.14	1.10
25	Progreso	0.12	0.11
26	Pucará	1.06	1.25
27	Puno	0.79	0.75
28	Putina	0.45	0.43
29	Santa Rosa	0.93	0.89
30	Taraco	2.17	2.73

En el Cuadro 11, D^* es la medida de discordancia para todas las estaciones que forman una sola región. D es la medida de discordancia para las estaciones que forman una región entera menos la estación discordante.

Puesto que la estación Azángaro es discordante con el resto del total de las estaciones, fue removida del análisis y la medida de discordancia D (Cuadro 11) es determinada por segunda vez. En la Figura 10 se muestra que removida la estación de Azángaro, las medidas de discordancia D son menores que el D crítico. De aquí se desprende evaluar la posible región homogénea considerando todo el conjunto de las estaciones (29 estaciones), evaluando con la medida heterogeneidad.

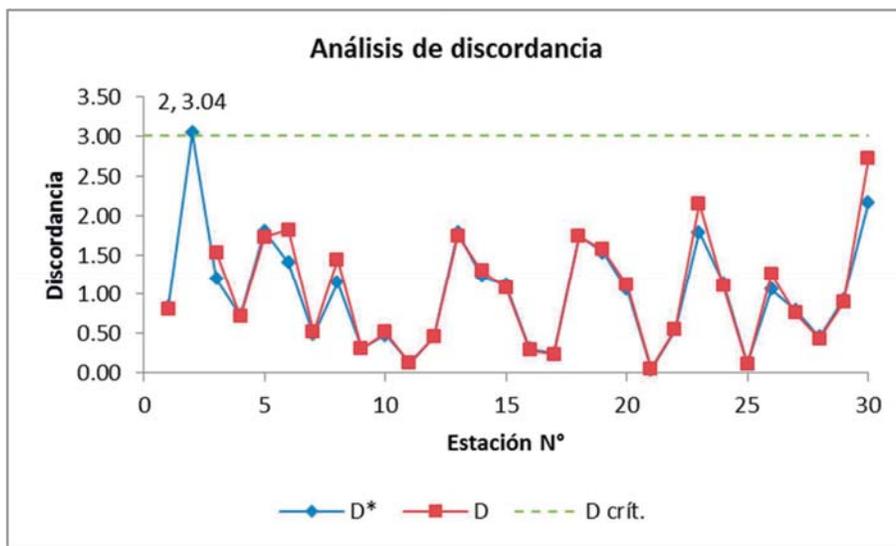


Figura 10: Medidas de discordancia de las estaciones.

4.4 Heterogeneidad regional y la formación de regiones homogéneas

Previa la formación de 2 o más regiones homogéneas, fue corroborado la posible región hidrológicamente homogénea considerando todas las estaciones (29), calculando la medida de heterogeneidad H de Hosking y Wallis con y sin estación discordante.

Según lo señalado en la sección 2.7.2, para generar regiones artificiales homogéneas se ha empleado la simulación de Monte Carlo mediante la distribución Kappa, donde las regiones simuladas tienen el mismo número de estaciones y longitud de registros. Los parámetros de la distribución Kappa (Cuadro 13): ξ (localización), α (escala), k y h (forma); se obtienen ajustando a los ratios regionales medios de L-momentos (Cuadro 12).

La distribución Kappa de cuatro parámetros definidos en la ecuación (41) tiene la forma: $f(x) = \alpha^{-1} [1 - k(x - \xi) / \alpha]^{l/k-1} [F(x)]^{1-h}$. Para el presente estudio se ha realizado 10 000 simulaciones para obtener medidas de heterogeneidad más precisas, puesto que el tiempo de simulación no es significativo.

Para la región sin la estación discordante (Azángaro), se tiene la media ponderada de los L-momentos regionales en el Cuadro 12.

Cuadro 12: Estadísticas de medias ponderadas de L-momentos regionales

N° de estaciones	L-momentos regional			
	L ^R	T ^R	T ₃ ^R	T ₄ ^R
29 estaciones sin estación discordante	1.0000	0.1446	0.1495	0.1485

Cuadro 13: Parámetros de la distribución Kappa para la región

N° de estaciones de la región	Kappa parámetros			
	xi	alpha	k	h
29 estaciones sin estación discordante	0.8916	0.2049	0.0102	-0.0752

Hecha la simulación de Monte Carlo para la región sin el sitio discordante se obtiene: V_1 que es la desviación estándar ponderada observada de los valores t (L-CV), μ_{V_1} y σ_{V_1} son la media y la desviación estándar de los valores de V_1 de las regiones simuladas; V_2 es la distancia media observada de t/t_3 , μ_{V_2} y σ_{V_2} son la media y la desviación estándar de los valores de V_2 de las regiones simuladas; V_3 es la distancia media observada de t_3/t_4 , μ_{V_3} y σ_{V_3} son la media y la desviación estándar de los valores de V_3 de las regiones simuladas. A partir de los cuales mediante la ecuación 41, se obtiene los estadísticos de la medida de heterogeneidad H_1 , H_2 y H_3 (Cuadro 14).

Cuadro 14: Medidas de heterogeneidad de la región sin la estación discordante

Estadísticos para H ₁				Estadísticos para H ₂				Estadísticos para H ₃			
V ₁	Mv1	D1	H ₁	V ₂	U1	D1	H ₂	V ₃	U1	D1	H ₃
0.0187	0.0172	0.0023	0.64	0.0718	0.0657	0.0086	0.71	0.0898	0.0857	0.0093	0.44

De manera similar se ha realizado 10,000 simulaciones para obtener medidas de heterogeneidad más precisas, considerando todas las estaciones, incluyendo la estación discordante (Azángaro). Las medidas de heterogeneidad para todas las estaciones, considerando que forman una sola región con y sin estación discordante se muestra el resumen en el Cuadro 15.

Cuadro 15: Estadísticas de heterogeneidad de la región con y sin estación discordante

N° de estaciones	Heterogeneidad		
	H ₁	H ₂	H ₃
30 estaciones con estación discordante	1.71	0.95	0.56
29 estaciones sin estación discordante	0.64	0.71	0.44

Del Cuadro 15, la región se declara heterogénea si el valor de H_j es suficientemente elevado. Hosking y Wallis (1997) sugieren considerarla como “aceptablemente homogénea” si $H_j < 1$, “posiblemente heterogénea” si $1 \leq H_j < 2$, y “definitivamente heterogénea” si $H_j \geq 2$. Según Hosking y Wallis (1997), H_1 tiene mejor poder discriminatorio para detectar la homogeneidad de la región que H_2 y H_3 . De aquí se desprende que la región formada con 30 estaciones incluida el sitio discordante (Azángaro) parece ser heterogénea puesto que $H_1(1.707)$ es mayor que 1, mientras la región de 29 estaciones sin el sitio discordante es una región homogénea con $H_1(0.64)$ menor que 1.

Por tanto, no es necesaria la posible formación de 2 o más regiones en base a descriptores físicos de la cuenca mediante técnicas de agrupamientos en función a su similitud. Siendo definitivamente considerada la cuenca de la vertiente del Lago Titicaca como una sola región hidrológicamente homogénea con respecto a la frecuencia del régimen de las precipitaciones máximas diarias.

4.5 Selección de la distribución regional de mejor ajuste

Las distribuciones consideradas en el análisis regional para estimar los cuantiles son cinco y todas las distribuciones son de tres parámetros, siendo, la logística generalizada (GLO), generalizada del valor extremo (GEV), generalizado normal (GNO), generalizada de Pareto (GPA) y Pearson tipo III (PE3). De acuerdo a Hosking y Wallis

(1997), distribuciones de dos parámetros pueden causar sesgos en la cola de los cuantiles estimados si la forma de la cola de la distribución de frecuencia verdadera no está bien aproximada por la distribución ajustada. La mejor distribución de ajuste es uno que da estimaciones robustas para la curva de crecimiento regional, así como para los cuantiles en cada estación.

El diagrama de la relación L-momentos (Figura 11) se ha empleado para identificar cualitativamente las distribuciones de mejor ajuste, complementariamente aprovechando la simulación de Monte Carlo realizada para la medida de heterogeneidad se determinó el Z-estadístico para las cinco distribuciones en análisis (Cuadro 16), para identificar la distribución regional de mejor ajuste.

La media del ratio L-momentos regionales (punto en negrita) se representan en la curva teórica de la relación de L-momentos de las distribuciones candidatas en la Figura 11 para la región entera. El punto para la región se encuentra cerca de la curva de distribución GEV y GNO. Siendo estas las distribuciones candidatas de mejor ajuste.

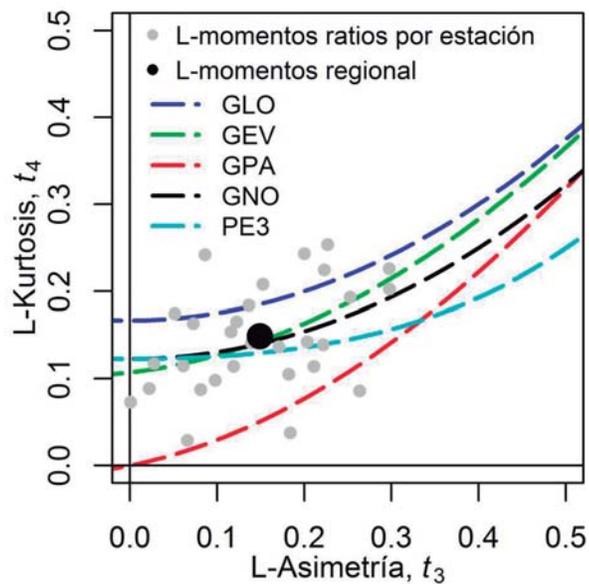


Figura 11: Diagrama de L-momentos ratio con el L-momentos ratio regional para las distribuciones candidatas.

Cuadro 16: Estadístico Z^{DIST} para varias distribuciones para la región

Región	Z^{GLO}	Z^{GEV^*}	Z^{GNO^*}	Z^{PE3}	Z^{GPA}
1	3.080	-0.565	-0.779	-1.681	-8.432

* se resalta las distribuciones con aceptación a un nivel de significancia de 10%

Considerando que el ajuste es bueno si $|Z^{DIST}| \leq 1.64$ lo cual corresponde a la aceptación de la distribución, planteado la hipótesis en un nivel de confianza de 90%. Un número de distribuciones se puede calificar con este criterio, la más adecuada es una que tiene el valor $|Z^{DIST}|$ cerca a cero.

De acuerdo al estadístico Z^{DIST} calculado para las distribuciones propuestas (Cuadro 16), lo cual muestra que el GPA, GLO y PE3 pueden ser eliminados por ser opciones de distribuciones inadecuadas para la región. Arreglando el valor de Z^{DIST} en orden ascendente para seleccionar la distribución, GEV (-0.565) da un mejor ajuste para la Región y GNO (-0.779) es la opción alternativa.

4.6 Estimación de los cuantiles de frecuencia

4.6.1 Estimación de la curva de crecimiento regional

La estimación de los cuantiles regionales $\hat{q}(F)$, para varias probabilidades de no excedencia, así como los parámetros de las distribuciones candidatas se presentan en el Cuadro 17. Los parámetros de las distribuciones se obtiene ajustando sus L-momentos a los L-momentos promedios regionales de la muestra de datos de todas las estaciones que forman la región.

Las curvas de crecimiento regional para la región se presentan en la Figura 12. Estos pueden ser interpretados como, por ejemplo $\hat{q}_{GEV}(0.99)$ la cantidad de lluvia que se producirá una vez cada 100 años y es 1.801 veces más grande que el promedio para todos los sitios en la región. Las curvas de crecimiento regional para la región son aproximadamente igual para periodo de retorno de hasta 500 años.

Cuadro 17: Parámetros y cuantiles regionales para las dos distribuciones de mejor ajuste

Distri- bución	Parámetros			Cuantiles regionales para probabilidades de no excedencia F								
	ξ	α	κ	F = 0.5 TR = 2	0.8 5	0.9 10	0.95 20	0.98 50	0.99 100	0.995 200	0.998 500	0.999 1000
GEV	0.883	0.215	0.032	0.961	1.197	1.349	1.491	1.670	1.801	1.929	2.093	2.213
GNO	0.961	0.246	-0.308	0.961	1.198	1.348	1.489	1.667	1.799	1.929	2.102	2.233

F es la probabilidad de no excedencia, y TR es el tiempo de retorno

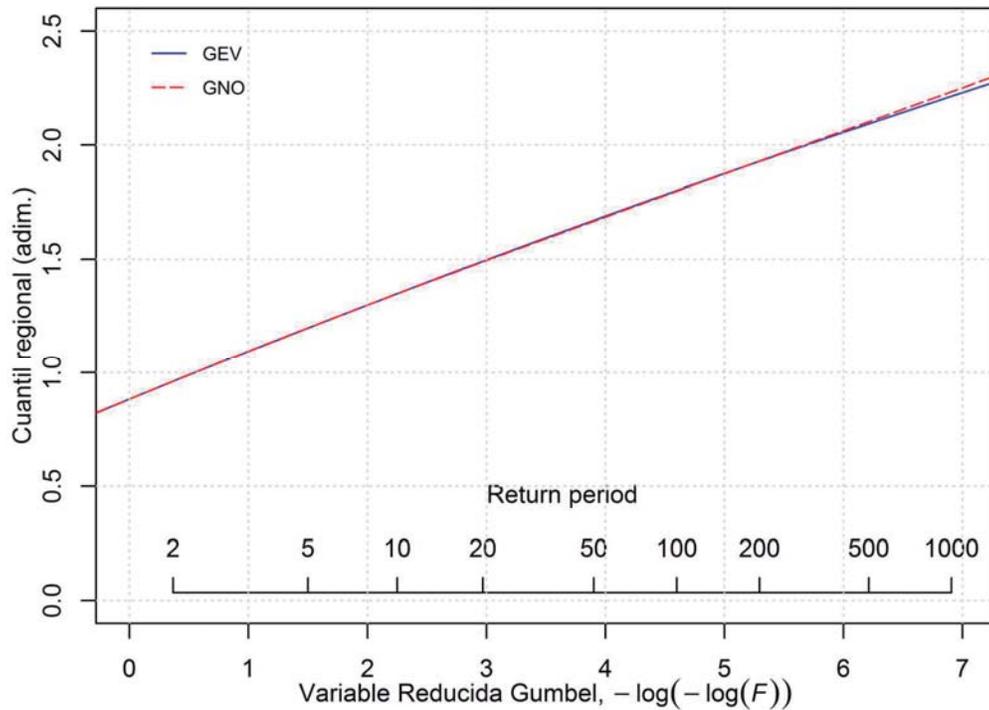


Figura 12: Curva de crecimiento regional para distribuciones candidatas

Las frecuencias de precipitaciones máximas en la región de la cuenca de la vertiente de Lago Titicaca fácilmente podrían ser modeladas por ambas distribuciones GEV y GNO. Sin embargo con fines científicos para seleccionar las estimaciones robustas de cuantiles regionales, siempre hay una necesidad de algún análisis de evaluación.

4.6.2 Decisión acerca de la mejor curva de crecimiento regional

Para medir la robustez de la precisión de las curvas de crecimiento regional estimados, se han llevado a cabo simulaciones sobre la base de las dos distribuciones (GEV y GNO), de acuerdo con el algoritmo de L-momento regional propuesto por Hosking y Wallis (1997), descrito en el Cuadro 3, donde se tiene dos posibilidades a seguir, con o sin la consideración de la matriz de dependencia entre pares de estaciones en la simulación. Por ello en este estudio primero se verifica la significancia de la correlación entre los pares de estaciones a un nivel de significancia 0.05, mostrada en la (Figura 13), en la parte superior de esta figura se aprecia que pocos pares de estaciones presentan correlaciones positivas significantes, de manera similar pocos pares de estaciones ubicados en la parte inferior de esta figura presentan correlaciones negativas significantes. Sin embargo hay gran cantidad de pares de estaciones que no presentan correlaciones

significativas; por lo que para la simulación de acuerdo con el Cuadro 3, no se considera la dependencia entre estaciones.

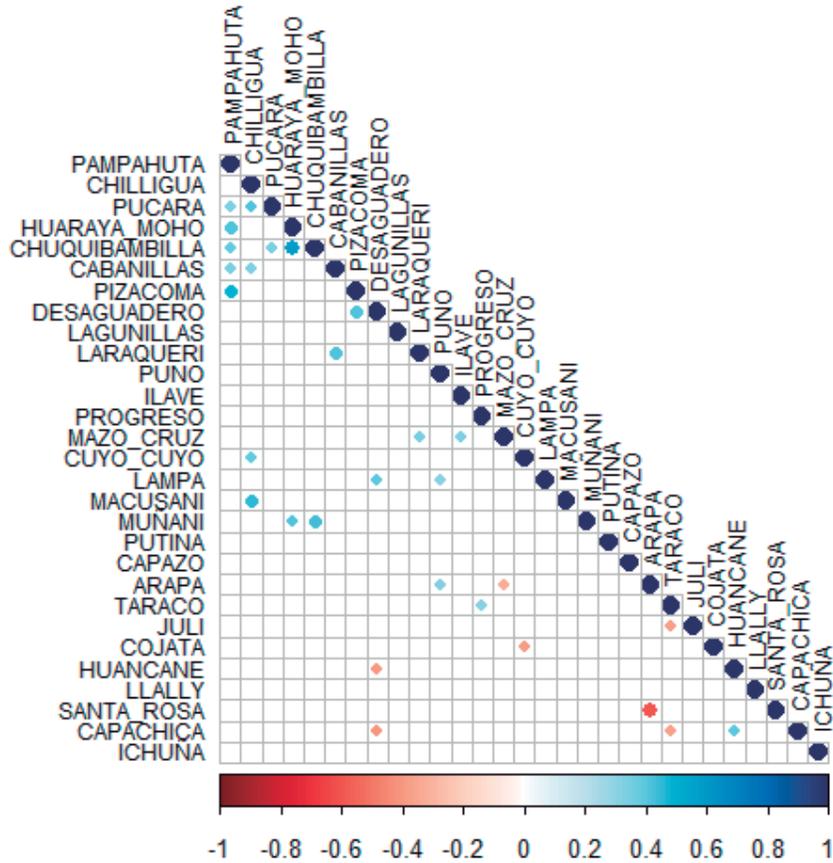


Figura 13: Correlación entre pares de estaciones significantes a un nivel de 0.05.

Para las dos distribuciones candidatas GEV y GNO en la simulación se ha establecido que las regiones artificiales simuladas a partir de ellas presenten las siguientes características:

- Mismo número de estaciones que la región real, es decir, 29 estaciones.
- Mismas longitudes de registro como las estaciones originales.
- Las series de las estaciones generadas deben tener la misma media que su correspondiente serie real.
- L-CV que varía en el rango de 0.1010 a 0.1870 para la posible heterogeneidad, estos valores son el mínimo y máximo respectivamente de L-CV calculados en el Cuadro 10.

- L-Cs = 0.1446 es lo mismo en cada estación simulado. Este valor es la media ponderada regional de L-asimetría (Cuadro 12).

La región se ha simulado $M = 10\ 000$ veces, esto significa que se tendrían M-ésimos valores de cuantiles para cada estación y para cada probabilidad de no excedencia, comparando estas con los cuantiles reales se obtienen las medidas de precisión, Bias relativo ($B^R(F)$), Bias absoluto relativo ($A^R(F)$) y la raíz de error cuadrático medio (RECM) relativo ($R^R(F)$).

Las medidas de precisión se muestran en el Cuadro 18. Los resultados de la simulación para la región, muestran que en general para periodos de retorno de 2 – 20 años el $B^R(F)$, $A^R(F)$ y $R^R(F)$; son similares para ambas distribuciones. Los indicadores de medida de precisión no afectos a los signos como el $A^R(F)$ y $R^R(F)$ muestran que a partir del periodo de retorno de 50 a 1000 años son relativamente bajos para la distribución GNO y altos para GEV. Por otra parte los límites de error al 90%; que significa que 90% de los datos simulados se encuentran dentro de los límites de error superior (LES) y límites de error inferior (LEI), son más estrechos para la distribución GNO en periodos de retorno de 50 a 1000 años. En consecuencia, se puede concluir que GNO es la mejor distribución para la estimación de cuantiles para grandes periodos de retorno (> 50 años) y para menores periodos de retorno GEV o GNO podrían emplearse.

Cuadro 18: Resultados de la simulación para la curva de crecimiento regional

Distribución	F =	0.5	0.8	0.9	0.95	0.98	0.99	0.995	0.998	0.999
		TR =	2	5	10	20	50	100	200	500
GEV	$B^R(F)$	0.0010	0.0013	0.0016	0.0019	0.0022	0.0022	0.0022	0.0020	0.0018
	$A^R(F)$	0.0070	0.0255	0.0401	0.0514	0.0631	0.0705	0.0770	0.0845	0.0896
	$R^R(F)$	0.0084	0.0296	0.0466	0.0597	0.0737	0.0826	0.0905	0.0999	0.1064
	LEI	0.9877	0.9566	0.9324	0.9140	0.8944	0.8817	0.8703	0.8564	0.8467
	LES	1.0145	1.0484	1.0769	1.0995	1.1240	1.1400	1.1543	1.1715	1.1836
GNO	$B^R(F)$	0.0011	0.0011	0.0015	0.0019	0.0023	0.0026	0.0028	0.0031	0.0033
	$A^R(F)$	0.0070	0.0256	0.0401	0.0512	0.0628	0.0701	0.0765	0.0838	0.0888
	$R^R(F)$	0.0084	0.0296	0.0465	0.0596	0.0733	0.0820	0.0897	0.0986	0.1047
	LEI	0.9876	0.9563	0.9320	0.9137	0.8946	0.8826	0.8720	0.8596	0.8512
	LES	1.0148	1.0483	1.0768	1.0993	1.1238	1.1396	1.1537	1.1704	1.1818

F es la probabilidad de no excedencia, y TR es el tiempo de retorno

$B^R(F)$ =Bias relativo, $A^R(F)$ =Bias absoluto relativo, $R^R(F)$ =RECM relativo

LEI=límite de error inferior, LES=límite de error superior

En este trabajo de tesis se definió definitivamente la distribución GNO como la mejor para modelar la región hidrológicamente homogénea de la vertiente del Lago Titicaca; puesto que las medidas de precisión para periodos de retorno menores a 50 años es aproximadamente similar que para el GEV, sin embargo para mayores de 50 años en general presenta bajas medidas de errores.

La distribución GNO para el caso de este estudio está definida por la función de distribución acumulada $F(x) = \Phi\left(-k^{-1}, \ln\left[1 - k(x - \xi)/\alpha\right]\right)$ y sus tres parámetros estimados: localización ($\varepsilon = 0.961$), escala ($\alpha = 0.246$) y forma ($k = -0.308$). Donde Φ , es la función de distribución acumulada de la distribución normal estándar.

4.6.3 Cuantiles de frecuencia regional y local

Seleccionada la distribución más robusta (GNO) y determinada sus parámetros (Cuadro 17), se computan los cuantiles regionales adimensionales de precipitación máxima en 24 horas para los periodos de retorno T de 2, 5, 10, 20, 50, 100, 200, 500 y 1000 años, también se evalúa la incertidumbre de su estimación mediante la medida de precisión como el RECM absoluto y los límites de error al 90% (Límite de error inferior y límite de error superior).

Cuadro 19: Resultado de cuantiles para la curva de crecimiento regional

F =	0.5	0.8	0.9	0.95	0.98	0.99	0.995	0.998	0.999
TR =	2	5	10	20	50	100	200	500	1000
Cuantil	0.9612	1.1979	1.3484	1.4889	1.6670	1.7988	1.9295	2.1019	2.2328
RECM	0.0082	0.0356	0.0629	0.0889	0.1226	0.1480	0.1736	0.2081	0.2347
LEI	0.9472	1.1427	1.2522	1.3543	1.4833	1.5784	1.6724	1.7959	1.8892
LES	0.9733	1.2527	1.4467	1.6295	1.8633	2.0380	2.2127	2.4452	2.6232

F es la probabilidad de no excedencia, y TR es el tiempo de retorno

El Cuadro 19 y la Figura 14, describen que la incertidumbre aumenta para grandes periodos de retorno.

Establecida la curva de crecimiento regional, los cuantiles para cada estación se determinan mediante la ecuación 53:

$$\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F).$$

Donde los valores adimensionales $\hat{q}(F)$ estimados para la curva de crecimiento regional (Cuadro 19), se multiplica por el índice de avenida $I_1^{(i)}$, considerado como la media de las observaciones de cada estación, la misma se muestra en el Cuadro 20.

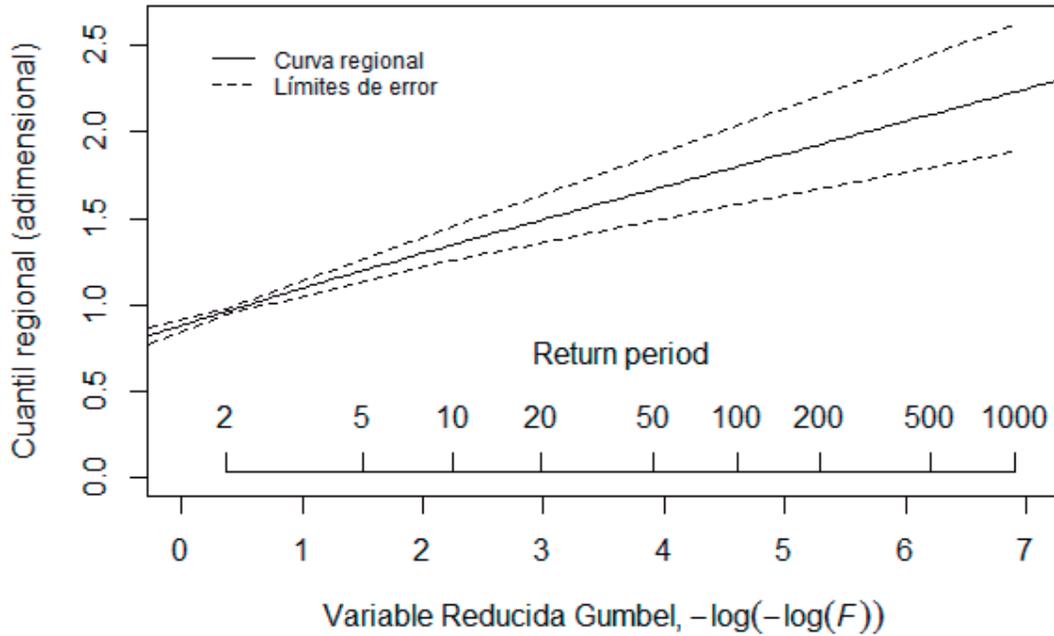


Figura 14: Representación gráfica de la curva de crecimiento regional

Cuadro 20: Índice de avenida de las estaciones de la región

Arapa	Cabanillas	Capachica	Capazo	Chilligua	Chuquibambilla	Cojata	Cuyo Cuyo
37.27	39.18	39.29	29.67	27.24	37.13	36.48	29.18
Desaguadero	Huancané	Huaraya Moho	Ichuña	Ilave	Juli	Lagunillas	Lampa
49.66	44.39	48.86	35.34	36.62	47.26	39.77	41.36
Laraqueri	Llally	Macusani	Mazo Cruz	Muñani	Pampahuta	Pizacoma	Progreso
38.80	36.13	33.05	30.80	31.73	39.49	36.53	33.12
Pucara	Puno	Putina	Santa Rosa	Taraco			
40.97	42.29	37.90	36.07	38.64			

A continuación se presenta los resultados de los cuantiles estimados por el ARF para cinco estaciones representativas espacialmente de la cuenca de VT (Figura 15), para

el norte de la cuenca se muestra para la estación de Muñani, para el centro para las estaciones de Taraco y Laraqueri, y para el sur se presenta para Desaguadero y Capazo.

En la Figura 15a las estaciones en azul son aquellos cuyos resultados de los cuantiles estimados por el ARF; muestran buenos ajustes en comparación con la posición de datos según la frecuencia empírica de Weibull, como se observa por ejemplo en la Figura 15b para la estación Muñani. Mientras tanto la estación Taraco (Figura 15c) es la única estación que no presenta buen ajuste, aparentemente en ésta estación se subestima. Al contrastar los resultados para las estaciones vecinas de Taraco, como en Arapa, Huancané y Capachica (ver Anexo 8.2) se encuentra que los resultados son buenos; siendo válido el ARF. Así mismo cabe señalar que comparando los cuantiles estimados por ARF y mediante el análisis local para la estación no se encuentra diferencias considerables, por lo que se podría considerar aun buena las estimaciones.

Los resultados tabulares de los cuantiles estimados por ARF para todas las estaciones, así como los gráficos de los mismos se presentan en el Anexo 8.2.

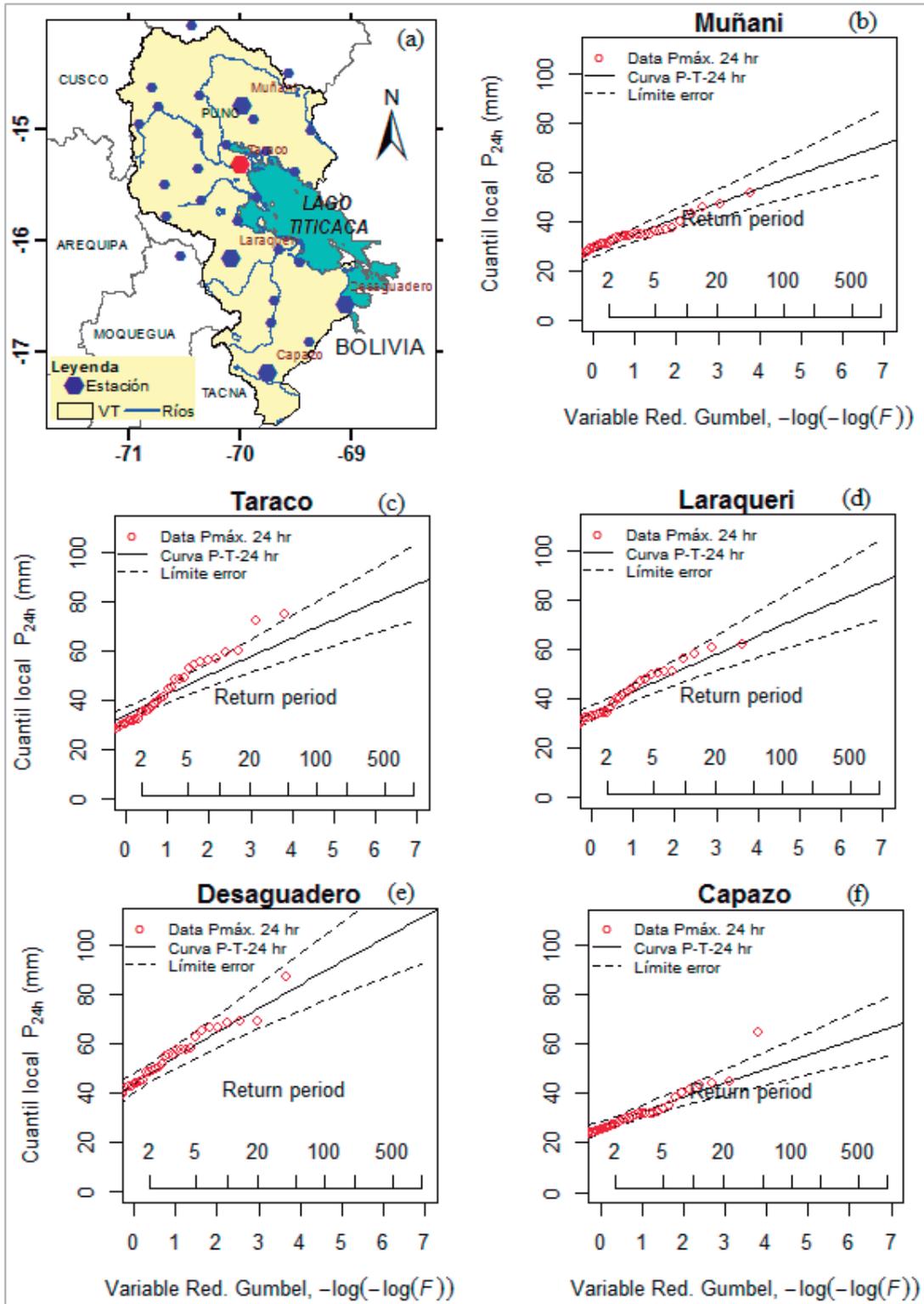


Figura 15: a) Distribución espacial de estaciones definitivas analizadas en ARF y Cuantiles de PMA estimadas por el ARF para las estaciones de b) Munañi, c) Taraco, d) Laraqueri, e) Desaguadero f) Capazo.

4.7 Comparación con el método de análisis local

La comparación entre ARF, en la cual se selecciona la función de distribución GNO con sus parámetros estimados regionalmente por el enfoque de la metodología de L-momentos y el método de análisis local (AL) que consiste en el ajuste a la serie de observaciones de PMA de cada estación por la función de distribución Gumbel con sus parámetros estimados por máxima verosimilitud (MV) resulta de gran interés práctico, ya que la distribución Gumbel se usa con frecuencia en muchos estudios en el campo de la hidrología. Así, Luna y Domínguez (2013) utiliza la función de distribución Gumbel para estimar los cuantiles de PMA en los Andes Bolivianos.

En la Figura 17 se presenta los valores de los cuantiles de PMA para los períodos de retorno de 2, 5, 10, 20, 100, 200, 500 y 1000 años respectivamente donde se han ordenado las estaciones en el sentido de sur a norte. Del Cuadro 30 al Cuadro 34 mostrados en Anexo 8.3 se describen los resultados de esta comparación.

Considerando mayor precisión en la estimación por el ARF al comparar los cuantiles del período de retorno de 2, 5 y 10 años, estimados para la distribución GNO (ARF) y Gumbel (AL) (Cuadro 30 y Cuadro 31, Figura 17), observamos que no existen diferencias marcadas donde el promedio en la que subestima el AL es en el orden de 0.82, 1.67 y 2.21 por ciento respectivamente, mientras sobrestima en el orden de 0.54, 1.67 y 2.80 por ciento respectivamente. En la Figura 16 se aprecia que la relación entre los cuantiles para estos periodos de retorno tiene una pendiente prácticamente igual a 1. Estos resultados indican las buenas estimaciones por ambos métodos para bajos periodos de retorno. Para el periodo de retorno de 20 años, las diferencias promedio de los cuantiles estimados por Gumbel se incrementa subestimando en (-2.63%) y sobrestimando en (+3.67%).

En los cuantiles para el tiempo de retorno de 50 años (Figura 17), AL tiene una tendencia a sobrestimar en varias estaciones, siendo marcadamente la sobrestimación en las estaciones de Chilligua (+11.04%), Cabanillas (+10.79%), pucará (+9.38%), Taraco (+7.05%), Santa Rosa (+6.77%) y otros en pequeña escala. Mientras la subestimación marcada se presencia en la estación Pampahuta (-7.45%) y otros en menor escala. Estas diferencias entre los dos enfoques tienen tendencia al incremento para los periodos de retorno de 100, 200, 500 y 1000 años; los mismos que observamos en la Figura 17. En la Figura 16 también observamos como el análisis local tiende a sobrestimar en la mayoría de

las estaciones para periodos de 50 años a más. Generalmente es más significativa las sobrestimaciones que se obtiene con el análisis local en la mayoría de las estaciones, en promedio estas diferencias alcanzan para los tiempos de retorno de 50 (+4.92%), 100 (+5.62%), 200 (+6.50%), 500 (+7.49) y 1000 (+7.70%) años.

La Figura 18 muestra espacialmente para un periodo de retorno de 1000 años que la desviación relativa de los cuantiles de PMA estimadas por el AL respecto al ARF en 66 % (34 %) de las estaciones analizadas el AL tiende a sobrestimar (subestimar) los cuantiles, alcanzando una variación de hasta 17.2% (-9.21%). En este estudio las diferencias encontradas para un periodo de retorno de 1000 años entre los cuantiles estimados por los dos enfoques no son tan grandes como en otros estudios. Por ejemplo, Kondratieva y Amarchi (2015) encuentra que el AL con respecto al enfoque regional subestima en 62% de las 50 estaciones que analizaron, obteniendo diferencias máximas de subestimación hasta de -46.5% para 1000 años de periodo de retorno, asimismo también encontraron una sola región homogénea en el noreste de Argelia.

Estas diferencias entre ambos enfoques están relacionadas fundamentalmente por la metodología de obtención de parámetros de las distribuciones y con el número de parámetros de una y otra distribución. La distribución Gumbel al disponer tan sólo de dos parámetros (su parámetro de forma k es igual a cero) con el cual en un gráfico de probabilidades se presenta como una recta, mientras que la GNO, al presentar el parámetro regional de forma ($k=-0.308$) es ligeramente cóncava hacia abajo. Esta circunstancia explica que en los períodos de retorno altos ($T>50$ años), la función de distribución Gumbel tienda a sobrestimar los cuantiles obtenidos.

Sin embargo, los mejores resultados encontrados en este estudio con respecto al de Kondratieva y Amarchi (2015) se deben al tamaño de la muestra. Así, Kondratieva y Amarchi (2015) utilizan para su análisis estaciones incluso de 14 años de longitud de registro mientras que nosotros utilizamos entre 27 a 46 años; es decir a menor longitud de registro se presenta mayor incertidumbre en el AL.

La subestimación significativa presenciada en la estación Pampahuta es probablemente asociada al cambio en la media detectada según el test de Pettit, en la serie de observaciones de esta estación, en la cual el análisis local (Gumbel) no realiza buena estimación. Mientras el ARF al considerar los datos de todas las estaciones que conforman

una región, es menos sensible al salto presentado en esta estación y puede ser considerado mejor la estimación.

A manera de resumen, la Figura 19 recoge todo lo visto hasta el momento con respecto a las diferencias promedio de sobrestimación y subestimación en función del período de retorno, existentes entre la distribución Gumbel y GNO.

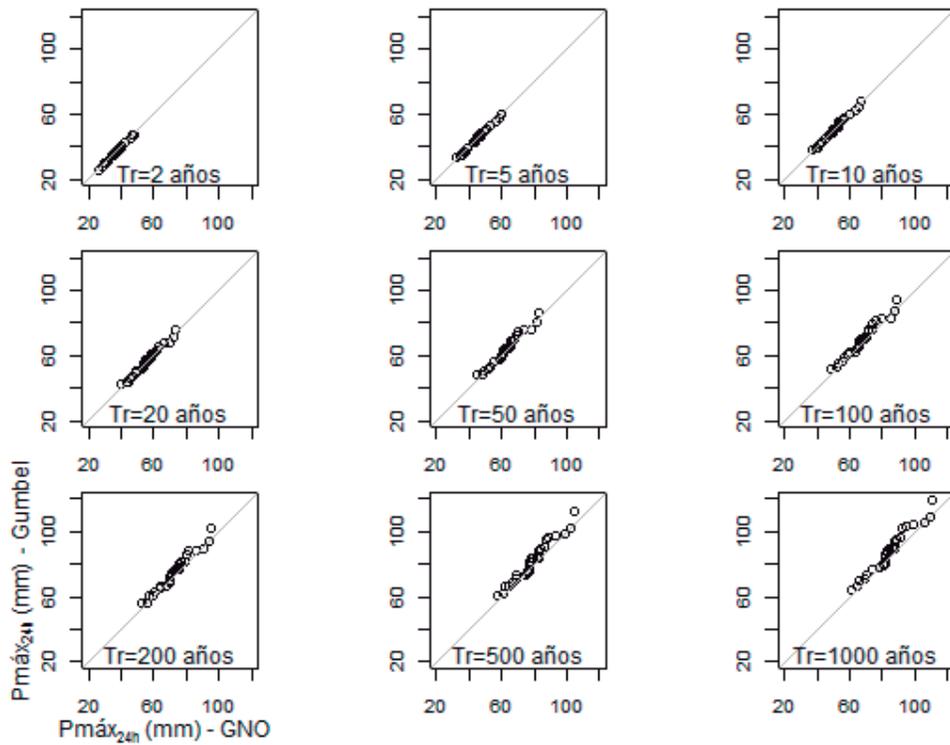


Figura 16: Gráfica de los cuantiles estimados por ARF versus AL para varios periodos de retorno.

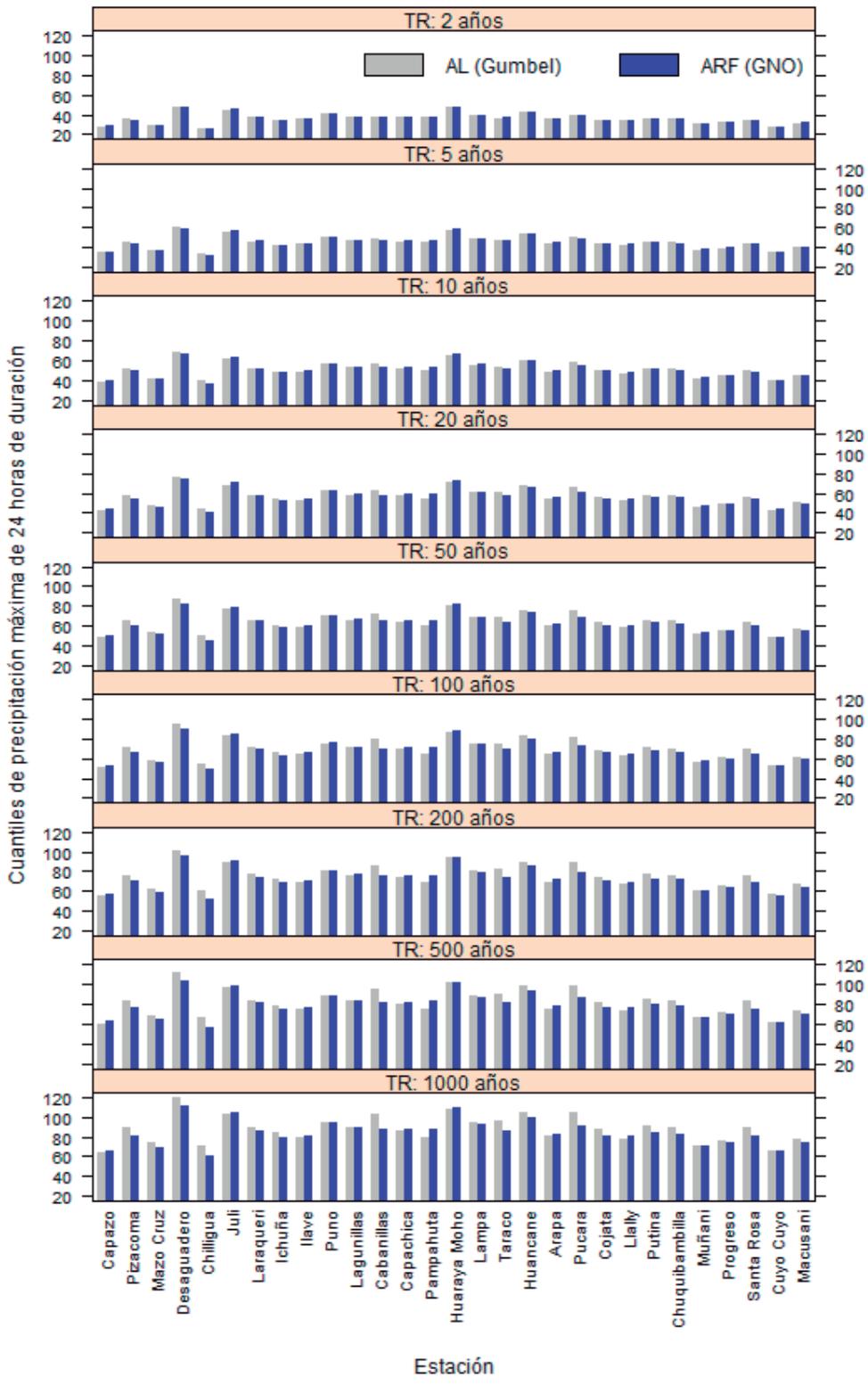


Figura 17: Comparación gráfica de los cuantiles estimados por ARF y AL para periodos de retorno de 2, 5, 10, 20, 50, 100, 200, 500 y 1000 años.

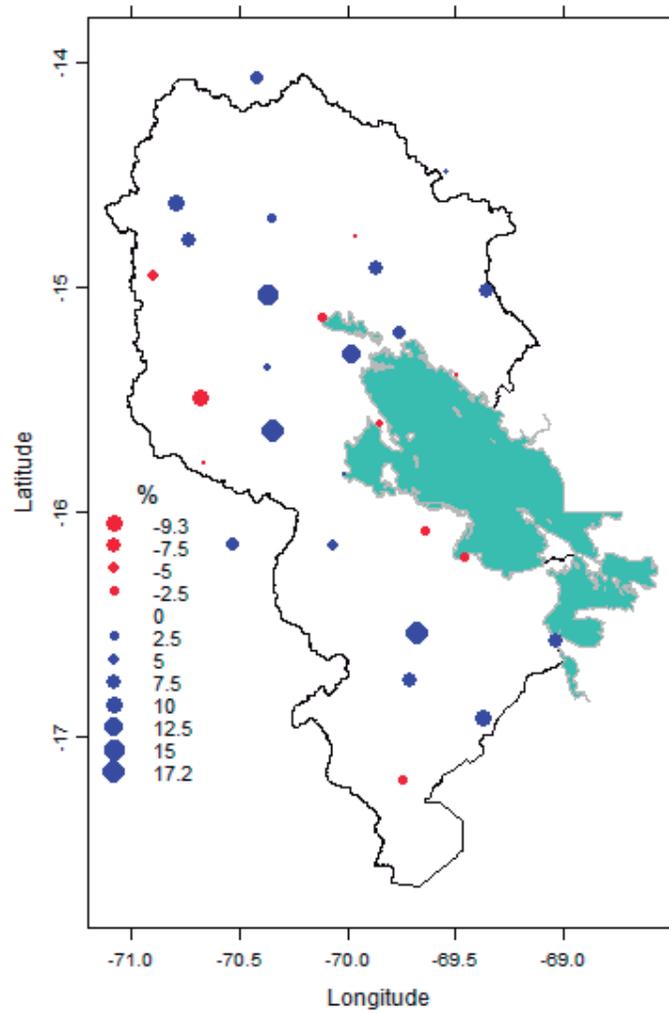


Figura 18: Desviación relativa (%) entre los cuantiles de PMA estimados por el método local y regional para 1000 años de periodo de retorno.

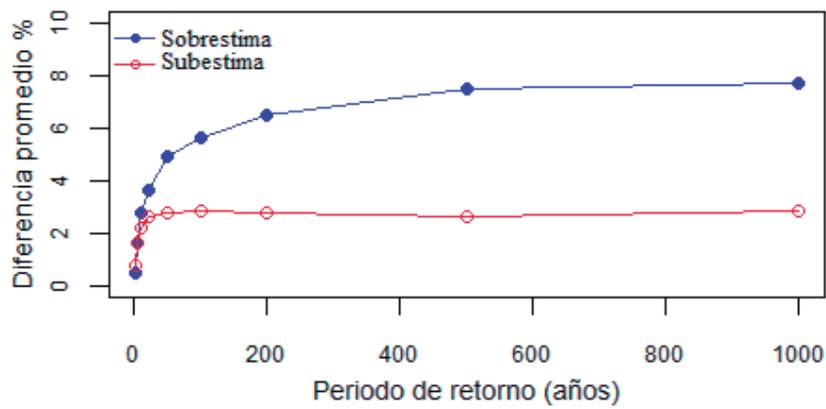


Figura 19: Diferencias entre los cuantiles estimados por Gumbel respecto a GNO

4.8 Mapeo de índice de avenida y eventos extremos

Para estimar cuantiles de precipitación máxima de 24 horas de duración asociado a diferentes periodos de retorno, esencialmente en sitios sin medición a nivel de la cuenca de la vertiente del Lago Titicaca, se carece de un patrón espacial de índices de avenida para los eventos extremos de precipitación, y ésta es sumamente importante para estimar los cuantiles empleando la metodología del ARF. Debido a ello para la utilidad práctica de este trabajo de tesis se evaluaron dos técnicas de mapeo geoestadísticos, entre ellas la técnica de interpolación de ponderado de distancia inversa (PDI) y el Kriging ordinario (KO). La validación cruzada fue realizada para determinar medidas de precisión y evaluar la técnica de mapeo más eficiente.

4.8.1 Mapeo de índice de avenida por Kriging ordinario (KO)

Para el mapeo por KO, como en todo análisis geoestadístico, es necesario realizar un análisis exploratorio del comportamiento espacial del índice de avenida como lo que observamos en la Figura 20a, que muestra claramente como los índices son mayores alrededor del Lago Titicaca y en la dirección suroeste a noreste, detectándose la posible influencia local (anisotropía local) del lago en el comportamiento espacial de los índices, en la Figura 20b observamos que la frecuencia de los índices de avenida se aproximan a una distribución normal, confirmándose el buen ajuste visualmente en el gráfico de cuantiles en la Figura 20c. Cumpliéndose el supuesto de normalidad para todo análisis geoestadístico.

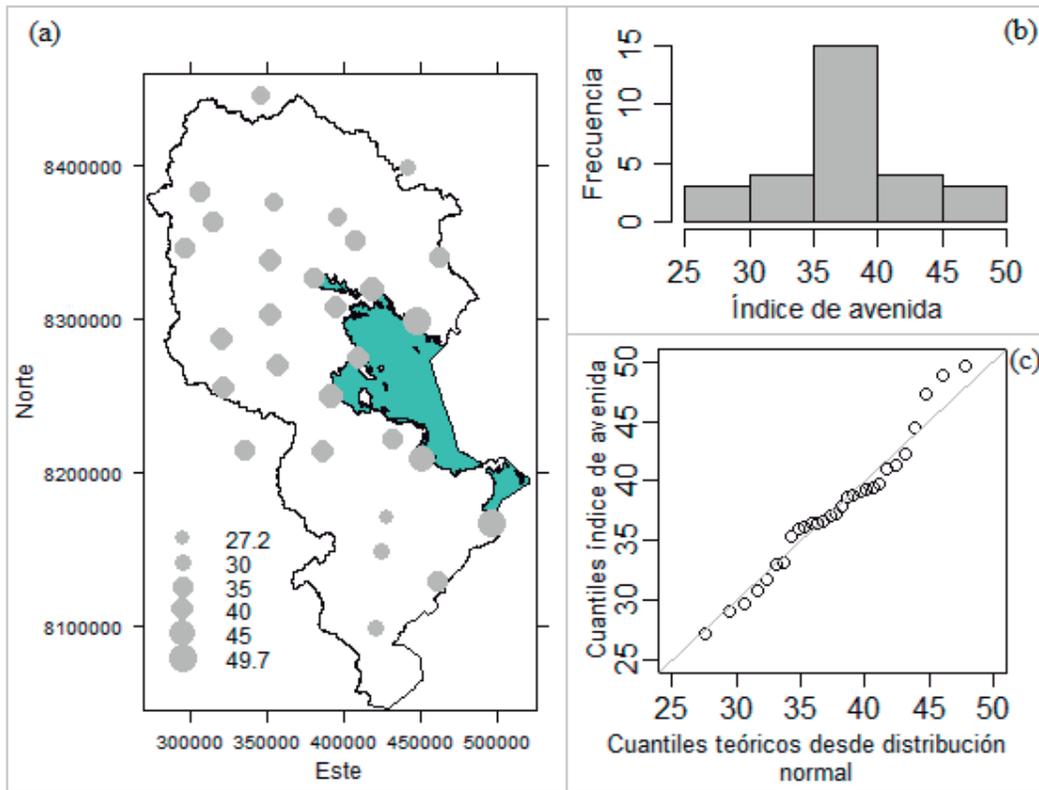


Figura 20: a) Distribución espacial de índice de avenida. b) Frecuencias de índice de avenida. c) Gráfico normal de cuantil – cuantil de índice de avenida.

En este trabajo de tesis únicamente se describe los resultados para el mejor semivariograma experimental direccional (sureste a noroeste), para el cual se ajustó el mejor semivariograma empírico (modelo exponencial) para cuantificar el grado y escala de variación espacial; como se observa en la Figura 21. Los parámetros del modelo exponencial (efecto pepita, rango y meseta) se estimaron mediante la técnica de optimización de mínimos cuadrados generalizados que se ejecuta por defecto en el paquete Gstat; las mismas se observa en el Cuadro 21.

Cuadro 21: Parámetros del semivariograma empírico del modelo exponencial

Modelo	Pepita [mm ²]	Meseta [mm ²]	Rango [km]	Dirección	Anisotropía
Exponencial	4.31	28.79	95972.59	SE-NO (135°)	0.5

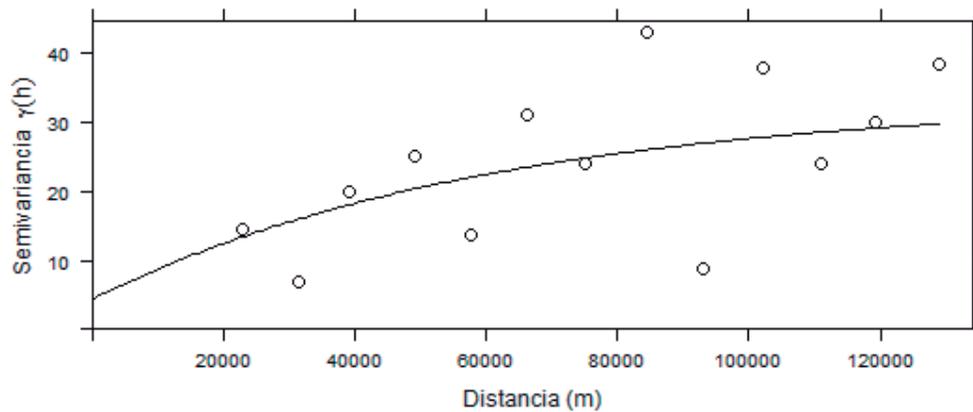


Figura 21: Semivariograma experimental direccional (suroeste – noreste) ajustado por el modelo exponencial.

De la Figura 21 y Cuadro 21, se observa que la semivarianza se incrementa con la distancia entre estaciones, hasta un valor casi constante (meseta = 28.79), a una distancia conocida como rango (aproximadamente 96 km) de dependencia espacial. Las estaciones separadas por unas distancias menores de 96 km están relacionadas espacialmente y aquellos separados por distancias mayores por el rango no están relacionados espacialmente, el rango de búsqueda de 96 km sería para un proceso isotrópico, sin embargo la variable índice de avenida presenta dependencia de la interacción espacial del Lago Titicaca en la dirección de sureste a noroeste; por lo que el proceso es anisotrópico. El factor de anisotropía (0.5), indica la relación entre el rango direccional de SO-NE y el rango en la dirección SE-NO. A partir del cual se tendría en la dirección de SO-NE un rango de 48 km aproximadamente. Esto significa que en la interpolación realizada en una celda, intervienen todas las estaciones vecinas que se encuentran dentro de una área de forma elíptica con un rango de 96 km en la dirección SE-NO y un rango de 48 km de la dirección de SO-NE.

Definida el semivariograma fue realizada la validación cruzada (Cuadro 22) mediante la técnica de interpolación de Kriging ordinario utilizando el paquete Gstat en el programa R. Cabe señalar que la interpolación se realizó en celdas o grillado con resolución de 1000 x 1000 m.

En el Cuadro 22, KO tiende a sobrestimar el índice de avenida como máximo en la estación Chilligua (+34.23 %) y subestima más en la estación Desaguadero (-19.69 %). La medida de precisión R^2 (coeficiente de determinación=0.39) es relativamente bajo; puesto

que el comportamiento espacial del índice de avenida tan solamente es explicada en un 39% por la técnica de mapeo KO.

Cuadro 22: Validación cruzada para Kriging ordinario

N°	Estación	Punto de control		Validación			
		Este	Norte	Predicción	Observado	Residual	Error %
1	Arapa	379810.1	8326375.0	39.20	37.27	1.93	5.19
2	Cabanillas	356518.1	8270003.0	40.49	39.18	1.30	3.33
3	Capachica	409207.2	8274054.0	41.49	39.29	2.20	5.59
4	Capazo	420718.0	8099673.0	35.15	29.67	5.48	18.48
5	Chilligua	427680.6	8170964.0	36.57	27.24	9.33	34.23
6	Chuquibambilla	313992.0	8363477.0	37.12	37.13	-0.01	-0.03
7	Cojata	461778.2	8339800.0	37.38	36.48	0.91	2.49
8	Cuyo Cuyo	440747.4	8398102.0	35.75	29.18	6.57	22.52
9	Desaguadero	496384.6	8167336.0	39.88	49.66	-9.78	-19.69
10	Huancané	418101.8	8319041.0	41.22	44.39	-3.17	-7.14
11	Huaraya Moho	447268.6	8298508.0	41.11	48.86	-7.75	-15.85
12	Ichuña	335626.4	8214882.0	36.85	35.34	1.51	4.28
13	Ilave	431167.2	8221586.0	43.36	36.62	6.74	18.39
14	Juli	451130.4	8208518.0	39.77	47.26	-7.49	-15.84
15	Lagunillas	322119.9	8254760.0	37.54	39.77	-2.23	-5.61
16	Lampa	352650.6	8301849.0	39.25	41.36	-2.11	-5.10
17	Laraqueri	385838.0	8213900.0	35.42	38.80	-3.38	-8.71
18	Llally	295863.8	8346455.0	38.25	36.13	2.11	5.85
19	Macusani	346293.9	8444314.0	33.68	33.05	0.64	1.92
20	Mazo Cruz	424144.4	8148457.0	32.08	30.80	1.28	4.17
21	Muñani	396055.5	8365845.0	36.37	31.73	4.64	14.61
22	Pampahuta	320052.7	8286619.0	38.86	39.49	-0.63	-1.60
23	Pizacoma	460302.0	8129810.0	33.56	36.53	-2.97	-8.12
24	Progreso	354051.3	8375001.0	35.90	33.12	2.78	8.40
25	Pucará	352426.7	8337498.0	37.32	40.97	-3.65	-8.91
26	Puno	391128.6	8249575.0	38.64	42.29	-3.65	-8.63
27	Putina	406640.7	8350906.0	36.08	37.90	-1.82	-4.80
28	Santa Rosa	306516.4	8382169.0	36.27	36.07	0.19	0.53
29	Taraco	394520.8	8307700.0	40.08	38.64	1.44	3.73
						RECM	4.31
						R2	0.39

4.8.2 Mapeo de índice de avenida por distancia inversa ponderado (DIP)

Similar al KO, la interpolación por DIP se ejecuta con el paquete Gstat, para celdas con resolución de 1000 x 1000 m. Con esta técnica no se requiere el modelado de semivariograma, la interpolación se realiza únicamente con el modelo matemático de la ecuación (62). Las medidas de precisión mediante la validación cruzada para esta técnica se presenta en el Cuadro 23.

Cuadro 23: Validación cruzada para distancia inversa ponderado

N°	Estación	Punto de control		Validación			
		Este	Norte	Predicción	Observado	Residual	Error %
1	Arapa	379810.1	8326375.0	38.91	37.27	1.64	4.41
2	Cabanillas	356518.1	8270003.0	39.39	39.18	0.20	0.51
3	Capachica	409207.2	8274054.0	40.18	39.29	0.89	2.26
4	Capazo	420718.0	8099673.0	36.02	29.67	6.35	21.41
5	Chilligua	427680.6	8170964.0	36.35	27.24	9.11	33.44
6	Chuquibambilla	313992.0	8363477.0	36.66	37.13	-0.47	-1.26
7	Cojata	461778.2	8339800.0	39.69	36.48	3.22	8.81
8	Cuyo Cuyo	440747.4	8398102.0	37.58	29.18	8.40	28.78
9	Desaguadero	496384.6	8167336.0	37.03	49.66	-12.63	-25.43
10	Huancané	418101.8	8319041.0	38.93	44.39	-5.46	-12.31
11	Huaraya Moho	447268.6	8298508.0	39.04	48.86	-9.82	-20.09
12	Ichuña	335626.4	8214882.0	38.82	35.34	3.48	9.86
13	Ilave	431167.2	8221586.0	41.14	36.62	4.51	12.32
14	Juli	451130.4	8208518.0	36.94	47.26	-10.32	-21.84
15	Lagunillas	322119.9	8254760.0	38.56	39.77	-1.20	-3.02
16	Lampa	352650.6	8301849.0	38.72	41.36	-2.64	-6.38
17	Laraqueri	385838.0	8213900.0	38.42	38.80	-0.37	-0.96
18	Llally	295863.8	8346455.0	37.48	36.13	1.35	3.74
19	Macusani	346293.9	8444314.0	36.72	33.05	3.67	11.12
20	Mazo Cruz	424144.4	8148457.0	33.46	30.80	2.66	8.65
21	Muñani	396055.5	8365845.0	37.82	31.73	6.09	19.20
22	Pampahuta	320052.7	8286619.0	38.97	39.49	-0.52	-1.32
23	Pizacoma	460302.0	8129810.0	35.92	36.53	-0.60	-1.64
24	Progreso	354051.3	8375001.0	37.44	33.12	4.32	13.05
25	Pucará	352426.7	8337498.0	37.56	40.97	-3.41	-8.32
26	Puno	391128.6	8249575.0	38.83	42.29	-3.46	-8.18
27	Putina	406640.7	8350906.0	36.27	37.90	-1.63	-4.31
28	Santa Rosa	306516.4	8382169.0	36.94	36.07	0.87	2.42
29	Taraco	394520.8	8307700.0	39.59	38.64	0.96	2.48
						RECM	5.08
						R2	0.13

Cuadro 23, DIP tiende a sobrestimar el índice de avenida como máximo en la estación Chilligua (+33.44 %) que es relativamente menor a la sobrestimación máxima de KO (+34.23 %), y subestima más en la estación Desaguadero (-25.43 %) que es mayor a KO (-19.69 %). La medida de precisión R^2 (coeficiente de determinación=0.13) es relativamente muy bajo; puesto que el comportamiento espacial del índice de avenida tan solamente es explicada en un 13% por la técnica de mapeo DIP.

La decisión de la técnica más eficiente para el mapeo del índice de avenida, así como para los eventos extremos, es de acuerdo a la proximidad a cero del RECM. Bajo

esta consideración del Cuadro 22 y Cuadro 23, observamos que RECM para KO (4.31) es más próximo a cero que la DIP (5.08). Consecuentemente la técnica de mapeo seleccionada es KO. A continuación (Figura 22a) se presenta el mapa de índice de avenida para la cuenca de la vertiente del Lago Titicaca, válidos para estimar índices de avenida en sitios sin medición con la precaución que en algunos sitios la incertidumbre es alto donde se podría subestimar o sobrestimar. La incertidumbre es relativamente baja en los sitios cercanos a las estaciones de medición, mientras en sitios alejados se tiene incertidumbres altos como se observa en el mapa de incertidumbre en la Figura 22b.

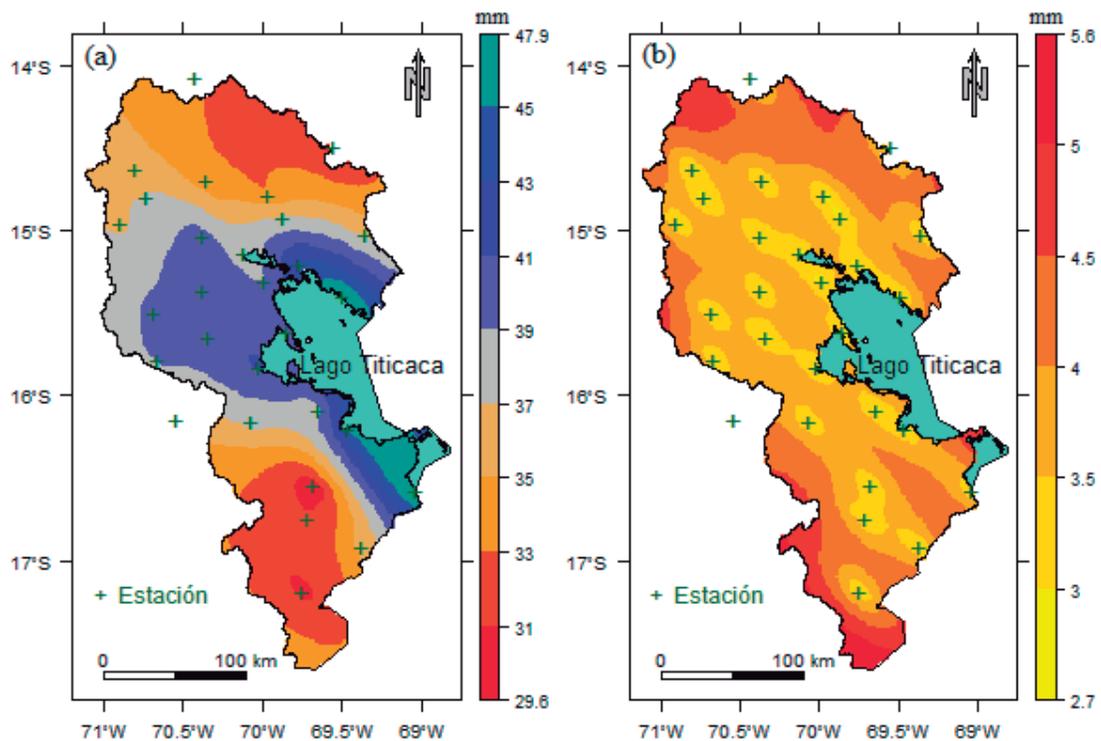


Figura 22: a) Mapa de índice de avenida para vertiente del Lago Titicaca en mm. b) Mapa de incertidumbre del índice de avenida.

Estimada el índice de avenida, multiplicando ésta por los cuantiles adimensionales asociados a diferentes periodos de retorno de la curva de crecimiento regional se obtendría el cuantil de diseño. De acuerdo a esta mapa de índice de avenida se observa claramente que el índice es mayor alrededor del Lago Titicaca, alcanzando valores de 47.9 mm y ésta disminuye en la dirección de SE a NO hasta alcanzar valores de 38 mm, mientras tanto en la parte norte y sur de la cuenca se presenta tasas de índice de avenida menores, presentándose valores próximos a 32 mm.

4.8.3 Mapeo de eventos extremos

Después de estimar los cuantiles extremos de precipitación en base a la distribución GNO en el análisis regional de frecuencias, los mapas de los mismos correspondientes a diferentes periodos de retorno fueron desarrollados usando la técnica más eficiente del mapeo de índice de avenida (Kriging ordinario). Entre ellas, las Figura 23 y 25 muestran los mapas de precipitación extrema que corresponden a periodos de retorno de 2, 5, 10, 20, 50, 100, 200 y 500 años, que proporcionan las precipitaciones máximas de diseño, necesarias para el diseño de drenaje y estructuras hidráulicas en la Cuenca del Lago Titicaca. En todas las mapas la magnitud de los cuantiles extremos presentan la misma caracterización espacial que el índice de avenida, siendo mayores alrededor del Lago Titicaca, disminuyendo éstas en la dirección de SO a NE, mientras tanto en la parte norte y sur de la cuenca se presenta las menores magnitudes de los cuantiles extremos.

En el Cuadro 24 se observa los parámetros del semivariograma teórico, donde mediante el modelo exponencial se ajustó el semivariograma experimental para realizar la interpolación en celdas con resolución de 1000 m x 1000 m en todos los casos. Así mismo se observa las medidas de precisión como RECM y R^2 productos de la validación cruzada.

Cuadro 24: Parámetros del semivariograma empírico en la interpolación de eventos extremos.

TR	Modelo	Pepita [mm ²]	Meseta [mm ²]	Rango [m]	Dirección	Anisotropía	RECM [mm]	R^2
2	Exponencial	3.99	26.59	95972.74	SO-NE (135°)	0.5	4.14	0.39
5	Exponencial	6.19	41.31	95972.97	SO-NE (135°)	0.5	5.16	0.39
10	Exponencial	7.84	52.33	95972.75	SO-NE (135°)	0.5	5.81	0.39
20	Exponencial	9.56	63.81	95972.64	SO-NE (135°)	0.5	6.42	0.39
50	Exponencial	11.99	79.99	95972.66	SO-NE (135°)	0.5	7.18	0.39
100	Exponencial	13.96	93.14	95972.79	SO-NE (135°)	0.5	7.75	0.39
200	Exponencial	16.06	107.16	95972.57	SO-NE (135°)	0.5	8.31	0.39
500	Exponencial	19.06	127.17	95972.60	SO-NE (135°)	0.5	9.06	0.39

Donde se puede apreciar que la incertidumbre de los mapas según RECM aumenta para grandes periodos de retorno, mientras el R^2 nos indica que los mapas realizados solamente explican en 39 % la variabilidad de los cuantiles extremos verdaderos obtenidos en el ARF. Cabe señalar que en la validación cruzada las correlaciones entre los cuantiles producto de la interpolación y los cuantiles determinados por el ARF, en todos los casos presentan una correlación de 0.62 donde con 27 grados de libertad resultan ser significantes.

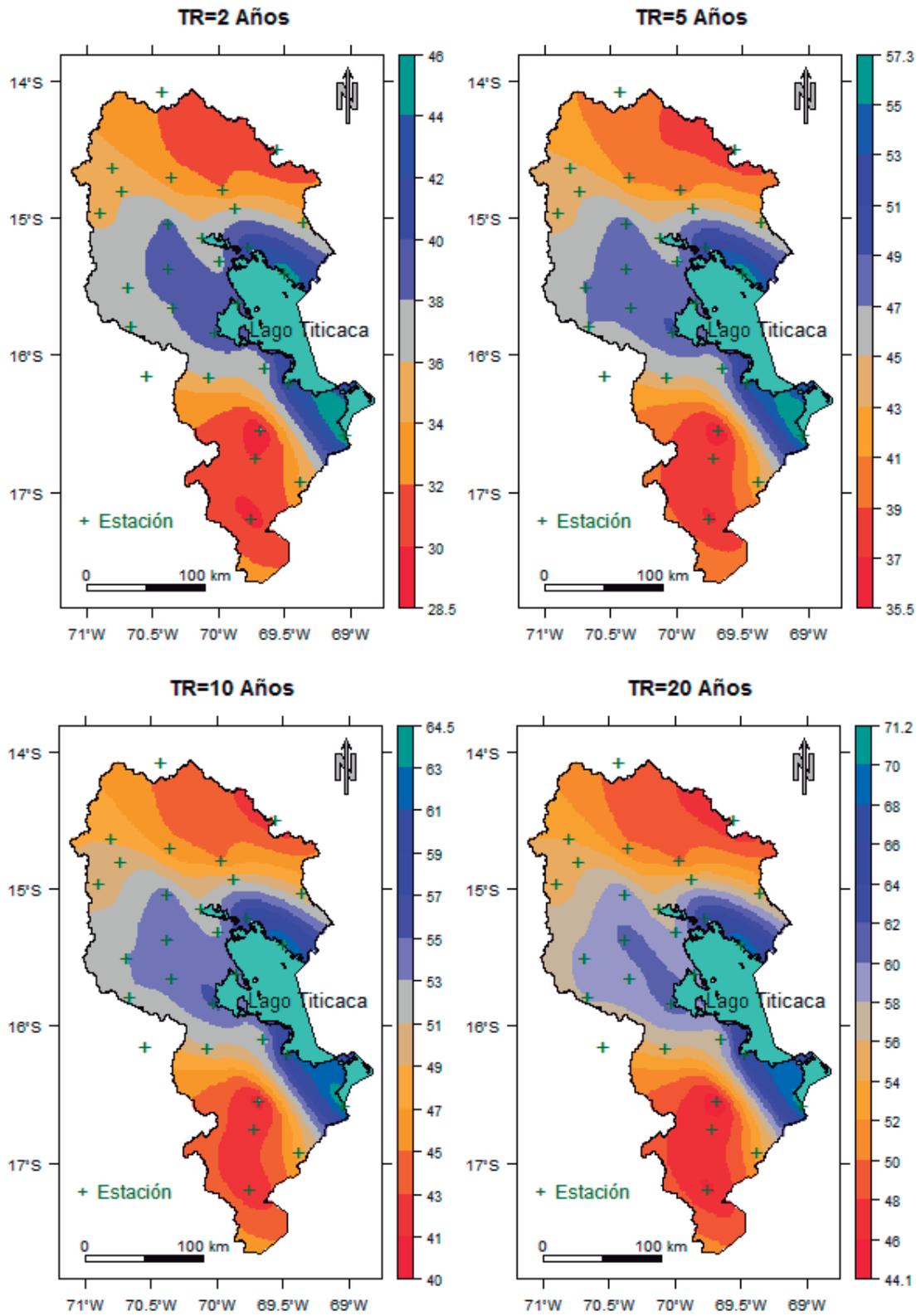


Figura 23: Mapa de precipitación máxima de 24 horas para tiempos de retorno de 2, 5, 10 y 20 años en mm.

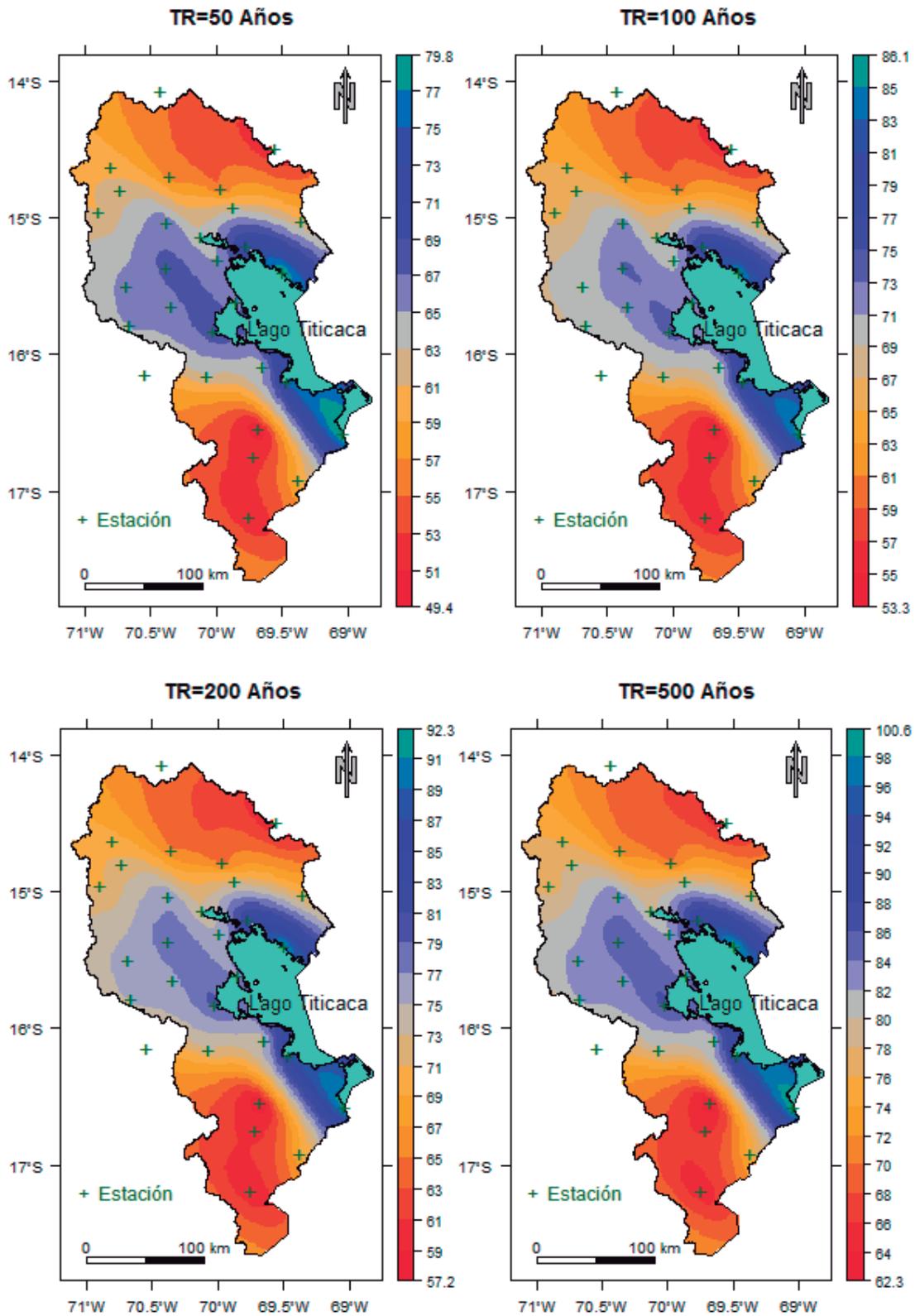


Figura 24: Mapa de precipitación máxima de 24 horas para tiempos de retorno de 50, 100, 200 y 500 años en mm.

V. CONCLUSIONES

El análisis de eventos extremos de precipitación realizada en este estudio, mediante una correcta aplicación del análisis regional de frecuencias empleando L-momentos y su variante del índice de avenida, se constituye como una alternativa útil en la estimación de las frecuencias de precipitación máximas en la vertiente del Lago Titicaca, válido para sitios con y sin registro pluviométrico. A partir de esta aplicación se extrae las siguientes conclusiones:

- El método de L-momentos a través de la medida de heterogeneidad, permite definir objetivamente que la vertiente del Lago Titicaca conforma una sola región hidrológicamente homogénea.
- El diagrama convencional de ratios de L-momentos y el test de bondad de ajuste denominado Z estadístico permiten identificar que existen dos distribuciones de probabilidad de mejor ajuste a los eventos de precipitaciones máximas regionales, siendo mejor el valor extremo generalizado (GEV) que el normal generalizado (GNO). Sin embargo un criterio de análisis de robustez basado en la simulación de Monte Carlo revela que GNO es ligeramente más robusto que la distribución GEV, para estimar los cuantiles de diseño para grandes periodos de retorno.
- Se constató que, en la Vertiente del Lago Titicaca, la aplicación de la función de distribución Gumbel, en el análisis local de frecuencias de las precipitaciones máximas conlleva a una sobrestimación de los cuantiles estimados en comparación a los cuantiles estimados por el análisis regional de frecuencias.
- Finalmente se generó el mapa de estimación espacial del índice de avenida (factor de escala específico de cada estación que es la media muestral de las observaciones) con la mejor técnica de interpolación geoestadística para la zona de estudio (Kriging ordinario), para estimar el valor del índice de avenida en sitios sin medición en la vertiente del Lago Titicaca. Así mismo utilizando la misma técnica de interpolación se estima espacialmente los eventos extremos para distintos periodos de retorno.

VI. RECOMENDACIONES

Se dan algunas recomendaciones y sugerencias para futuras investigaciones y desarrollo de estudios afines. La siguiente es una lista de algunas sugerencias a tener en cuenta.

- Se recomienda que previa a la aplicación de la metodología del análisis regional de frecuencias, realizar indispensablemente un exhaustivo tratamiento de la calidad de datos y verificar los supuestos del análisis de frecuencias, puesto que el análisis regional de frecuencias es aplicable para datos estacionarios e independientes.
- Debido a su robustez estadística y su uso popularizado en los últimos años del análisis regional de frecuencias de precipitaciones máximas empleando L-momentos, se recomienda extender su aplicación en las cuencas o regiones propensas a las inundaciones en el país.
- Investigar el enfoque de análisis regional de frecuencias empleando L-momentos en otros fenómenos extremos como los caudales máximos, eventos de índices de sequías extremos y otros.
- Investigar los procesos atmosféricos que explican el suceso de los eventos extremos de precipitación en la vertiente del Lago Titicaca.
- Se recomienda investigar la aplicación de *distribuciones mezcladas*, para eventos de precipitación que se originan como resultado de múltiples factores influyentes en el clima, donde se considere que el comportamiento de la variable aleatoria es una combinación de dos o más distribuciones univariadas de probabilidad.
- Investigar adicionalmente algunas técnicas de interpolación multivariada, no utilizadas en esta investigación, para mejorar la estimación regional del índice de avenida.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Alem Tadesse, Haile., 2011. Regional Flood Frequency Analysis for Southern Africa. Masteroppgave, Master Thesis, Department of Geosciences, University of Oslo.
- Álvarez, M., Puertas, J., Soto, B. y Díaz-Fierros, F., 1999. Análisis regional de las precipitaciones máximas en Galicia mediante el método del índice de avenida. *Ingeniería del Agua* 6(4), 283-290.
- Autoridad Nacional del Agua, 2012. ANA trabaja en el Monitoreo, Reporte y Coordinación de apoyo en zonas afectadas por inundaciones. Recuperado 12/07/2014 desde <http://www.ana.gob.pe/servicios-al-usuario/eventos-extremos-inundaciones-2012/presentaciones.aspx>
- Barnett, V. y Lewis, T., 1984. *Outliers in statistical data*. John Wiley y Sons, Nueva York.
- Bílková, D., 2014. Robust parameter estimations using L-moments, TL-moments and the order statistics. *American Journal of Applied Mathematics*. 2(2), 36-53.
- Bourges, J., Cortes, J. y Salas, E., 1992. Hydrological Potential. In “Lake Titicaca a synthesis of limnological knowledge” (C. Dejoux and A. Iltis, eds). Kluwer Academic Publishers, Dordrecht, Boston, London. 523-538.
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., 1994. *Time Series Analysis: Forecasting and Control*. Pearson Education, Delhi.
- CERESTA (Centre d'Enseignement et de Recherche de Statistique Appliquée), 1986. Aide-mémoire pratique des techniques statistiques pour ingénieurs et techniciens supérieurs. *Revue de statistique appliquée*, vol. XXXIV numéro spécial.
- Chow, V., 1964. Statistical and probability analysis of hydrologic data, part 1: frequency analysis. In *Handbook of applied hydrology; a compendium of water resources technology*. New York. McGraw-Hill. pp: 8.1-8.42.

- Dalrymple, T., 1960. Flood frequency analyses. Water Supply Paper 1543-A, U.S. Geological Survey, Reston, Va.
- Environmental Protection Agency, 1992. Statistical training course for ground-water monitoring data analysis. US EPA/530-R-93-003, Office of Solid Waste, Washington, DC.
- Falvey, M., y Garreaud, R., 2005. Moisture variability over the South American Altiplano during the South American Low Level Jet Experiment (SALLJEX) observing season. *J. Geophys. Res.* (110), D22105, doi:10.1029/2005JD006152.
- Gálvez, J. M., Orozco, R. K., Reyes, C. R. y Douglas, M. W., 2006. Observed diurnal circulations and rainfall over the altiplano during the salljex. Proceedings of 8 ICSHMO, Foz do Iguaçu, Brazil, April 24-28, INPE, 1041-1047.
- Garreaud, R. D., 1999. Multiscale analysis of the summertime precipitation over the central Andes. *Mon. Wea. Rev.* (127), 901–921.
- Garreaud, R. D., 2001. Subtropical cold surges: Regional aspects and global signatures. *Int. J. Climatol* (21), 1181–1197.
- Garreaud, R. D., y Aceituno, P., 2001. Interannual rainfall variability over the South American Altiplano. *J. Climate* (14), 2779–2789.
- Giraldo, R., J. 2002. Introducción a la geoestadística: Teoría y aplicación. Departamento de Estadística, Universidad Nacional de Colombia.
- Goovaerts, P. G., 1997. Geostatistics for natural resources evaluation. Oxford University Press, New York.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., y Wallis, J. R., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research* 15 (5), 1049–1054
- Hosking, J. R. M, y Wallis, J. R., 1993. Some statistics useful in regional frequency analysis. *Water Resources Research* 29(2), 271-281.
- Hosking, J. R. M., 1986. The theory of probability weighted moments. Research Report RC12210, IBM Research Division, Yorktown Heights, N.Y.

- Hosking, J. R. M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52(2), 105-124.
- Hosking, J. R. M., Wallis, J. R., 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29(3), 339-349.
- Hosking, J. R. M., Wallis, J. R., 1997. *Regional Frequency Analysis: An approach Based on L-moments*. Cambridge University Press, New York.
- Hosking, J. R. M., Wallis, J. R., y Wood, E. F., 1988. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27(3), 351-261.
- Hosking, J. R. M., y Wallis, J. R., 1995. A comparison of unbiased and plotting-position estimators of L-moments. *Water Resources Research* 31(8), 2019-2025.
- Iglewicz, P. J. y Hoaglin, D. C., 1993. *How to Detect and Handle Outliers*. American Society for Quality Control, Milwaukee, WI.
- Interagency Advisory Committee on Water Data, 1982. Guidelines for determining flood flow frequency, Bull 17 B of the hydrology sub committee, office of water data coordination, geological survey, U.S. Department of the interior, Washington, D.C.
- Johnston, K., Ver Hoef, J. M., Krivoruchko, K. y Lucas, N., 2001. *Using ArcGIS Geostatistical Analyst*. ESRI, Redlands.
- Kachroo, R. K., Mkhani, S. H., y Parida, B. P., 2000. Flood frequency analysis of southern Africa: I. Delineation of homogeneous regions. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques* 45(3), 437-447.
- Kendall, M. G., 1975. *Rank Correlation Methods*, 4th ed., Charles Griffin: London.
- Lavado, C. W. S., Ronchail, J., Labat, D., Espinoza, J. C. y Guyot, J. L. A., 2012 basin-scale analysis of rainfall and runoff in Peru (1969-2004) Pacific, Titicaca and Amazonas drainages. *Hydrological Sciences Journal* 57(4), 1-18, DOI:10.1080/02626667.2012.672985.

- Lavado, C.W.S. y Espinoza, J. C., 2014. Impactos de El Niño y La Niña en las lluvias del Perú (1965-2007). *Revista Brasileira de Meteorologia* 29(2), 171-182.
- Ljung, G. M. y Box, G. E. P., 1978. On a measure of lack of fit in time series models. *Biometrika* 65(1), 297–303.
- Luna, J. A. y Domínguez, R., 2013. Un método para el análisis de frecuencia regional de lluvias máximas diarias: aplicación en los Andes bolivianos. *Revista chilena de ingeniería* 21 (1), 111-124.
- Mann, H. B., 1945. Nonparametric tests against trend. *Econométrica* 13(1): 245–259.
- Miller, J. F., Frederick, R. H. y Tracey, R. S., 1973. NOAA ATLAS 2, Precipitation - frequency Atlas of the Western United States. Publication U.S. Dept. of Commerce, NOAA, National Weather Service, Washington DC, USA.
- Pebesma, E. J., 2004. Multivariable geostatistics in S: the gstat package. *Computers y Geosciences* 30(7), 683-691
- Pettitt, A. N., 1979. A non-parametric approach to the change-point problem. *Applied Statistics* 28(2), 126-135.
- Rahman, M., Sarkar, S., Najafi, M., y Rai, R., 2013. Regional Extreme Rainfall Mapping for Bangladesh Using L-Moment Technique. *J. Hydrol. Eng.* 18(5), 603–615.
- Saf, B., 2008. Application of index procedures to flood frequency analysis in Turkey. *Journal of the American Water Resources Association* 44(1), 37-47.
- Samper, F. J. y Carrera, J., 1990. *Geoestadística. Aplicaciones a la hidrogeología subterránea*. Centro Internacional de Métodos Numéricos en Ingeniería. Universitat Politècnica de Catalunya. Barcelona.
- Sanabria, J., Marengo, J. y Valverde, M., 2009. Escenarios de cambio climático con modelos regionales sobre el Altiplano Peruano (Departamento de Puno). *Revista Peruana Geo-Atmosférica* (1), 134-149.
- Slocum, T. A., 1999. Chapter 8: Interpolation methods for smooth continuous phenomena. In *Thematic Cartography and Visualization*. Prentice Hall, Upper Saddle River, New Jersey, 136-152.

- Smail Mahdi, S. y Cenac, M., 2005. Estimating parameters of gumbel distribution using the methods of moments, probability weighted moments and maximum likelihood. *Revista de Matemática: Teoría y Aplicaciones*. 12(1&2), 151–156.
- Stendinger, J. R., Vogel, R.M., y Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events, In: D. R. Maidment (ed.) *Handbook of Hydrology*, McGraw Hill, New York.
- Sylvestre, F., Servant-Vildary, S. y Roux, M., 2001. Diatombased ionic concentration and salinity models from the south Bolivian Altiplano (15-23°S). *Journal of Paleolimnology* (25), 279-295.
- Szolgay, J., Parajka, J., Kohnová, S., Hlavčová, K., 2009. Comparison of mapping approaches of design annual maximum daily precipitation. *Atmospheric Research* 92(3), 289–307.
- Takahashi, K., Montecinos, A., Goubanova, K. y Dewitte, B., 2011 ENSO regimes: Reinterpreting the canonical and Modoki El Niño. *Geophysical Research Letters* (38), L10704, DOI:10.1029/2011GL047364.
- Tamara Kondratieva y Hocine Amarchi (2015). Régionalisation des précipitations journalières extrêmes: cas de la région située au Nord-Est de l'Algérie, *Hydrological Sciences Journal*, 60:3, 498-507, DOI: 10.1080/02626667.2014.988154
- Tveito, O., Wegehenkel, M., Van der Wel, F., y Dobesch, H., 2008. *The Use of Geographic Information Systems in Climatology and Meteorology: COST 719*. Luxemburg: COST Action719.
- United Nations (UN) & World Water Assessment Programme, 2003. *UN World Water Development Report: Water for People, Water for Life*. Paris, New York and Oxford, UNESCO and Berghahn Books.
- Vogel, R. M., y Fennessey, N. M., 1993. L-moment diagrams should replace product moment diagrams, *Water resources Research* 29(6), 1745-1754.

- Vogel, R. M., Thomas, W. O., y McMahon, T. A., 1993. Flood flow frequency model selection in southwestern U.S.A. *Journal of Water Resources Planning and Management ASCE* 119(3), 353–366.
- Vuille, M., y Ammann, C., 1997. Regional snowfall patterns in the high, arid Andes. *Climatic Change* (36), 413–423.
- Vuille, M., y Keimig, F., 2004. Interannual variability of summertime convective cloudiness and precipitation in the central Andes derived from ISCCP-B3 data. *J. Climate* (17), 3334–3348.
- Wackernagel, H., 1995. *Multivariate geostatistics. An introduction with applications.* Springer-Verlag, Berlin.
- Wallis, J. R., 1981. Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures. In *Piense e Siccita*, edited by E. Guggino, G. Rossi, and E. Todini, pp. 3-36. Fondazione Politecnica del Mediterraneo. Catania, Italy.
- Wallis, J. R., 1982. Hydrologic problems associated with oilshale development. In *Environmental Systems and Management*, edited by S.Rinaldi, pp. 85-102. North-Holland, Amsterdam.
- Wallis, J. R., Schaefer, M. G., Barker, B. L., y Taylor, G. H., 2007. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrology and Earth System Sciences* 11(1), 415-442.
- Warrick, A. W., Myers, D. E. y Nielsen, D. R., 1986. *Geostatistical Methods Applied to Soil Science. Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods- Agronomy Monograph 9: 53 - 81.*
- Weiss, L. L., 1964. Ratio of true to fixed interval maximum rainfall. *journal hydraulics*, ASCE 90(1), 77-82.
- Yue, S., Pilon, P., Phinney, B. y Cavadias, G., 2002. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes* 16 (9), 1807–1829. doi: 10.1002/hyp.1095

VIII ANEXO

8.1 Distribuciones teóricas y sus relaciones

Cuadro 25: Fórmulas para estimación de parámetros por el método de máxima verosimilitud para la distribución Gumbel.

Función de distribución $F(x)$ o Función cuantil $x(F)$	Estimación de parámetros por máxima verosimilitud
$F(x) = e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}$ $x(F) = \mu - \sigma \cdot \ln[-\ln(F)]$	<p>La función de verosimilitud en base a una muestra aleatoria x_1, \dots, x_n está dada por:</p> $L(\sigma, \mu) = -\sum_{i=1}^n \frac{x_i - \mu}{\sigma} - n \cdot \ln(\sigma) - \sum_{i=1}^n \exp\left[-\frac{x_i - \mu}{\sigma}\right]$ $\frac{\partial \ln L(\sigma, \mu)}{\partial \mu} = \frac{1}{\sigma} \left[n - \sum_{i=1}^n \exp\left[-\frac{x_i - \mu}{\sigma}\right] \right]$ $\frac{\partial \ln L(\sigma, \mu)}{\partial \sigma} = \sum_{i=1}^n \left[\frac{x_i - \mu}{\sigma^2} \right] - \frac{n}{\sigma} - \sum_{i=1}^n \left[\frac{x_i - \mu}{\sigma^2} \right] \exp\left[-\frac{x_i - \mu}{\sigma}\right]$ <p>para $\sigma \neq 0$. derivando :</p> $\frac{\partial \ln L(\sigma, \mu)}{\partial \mu} = \frac{\partial \ln L(\sigma, \mu)}{\partial \sigma} = 0.$ <p>Se produce la máxima verosimilitud y resolviendo simultáneamente las siguientes ecuaciones se obtienen los parámetros.</p> $\mu = \sigma \left\{ \ln n - \ln \sum_{i=1}^n \exp\left[-\frac{x_i}{\sigma}\right] \right\}, \bar{x} = \sigma + \frac{\sum_{i=1}^n x_i \exp\left[-\frac{x_i}{\sigma}\right]}{\sum_{i=1}^n \exp\left[-\frac{x_i}{\sigma}\right]}$

Fuente: Mahdi y Cenac (2005).

Cuadro 26: Fórmulas por distribución, L-momentos y su ratios para las distribuciones analizadas.
 Parámetros 3: localización (ε, u^*), escala (α, σ^*) y forma (k, γ^*).

Distribución	Función de distribución $F(x)$ o Función cuantil $x(F)$	L-momentos y ratios de L- momentos	Estimación de parámetros.
Pareto Generalizado (GPA)	$x(F) = \varepsilon + \alpha \cdot \frac{1 - \left[\frac{1 - F(x)}{F(x)} \right]^k}{k}$	$\lambda_1 = \xi + \frac{\alpha}{1+k}$ $\lambda_2 = \frac{\alpha}{(1+k) \cdot (2+k)}$ $\tau_3 = \frac{1-k}{3+k}, \tau_4 = \frac{(1-k) \cdot (2-k)}{(3+k) \cdot (4+k)}$	$k = \frac{(1-3\tau_3)}{1+\tau_3}$ $\alpha = (1+k)(2+k)\lambda_2$ $\xi = \lambda_1 - (2+k)\lambda_2$
Valor extremo Generalizado (GEV)	$x(F) = \varepsilon + \alpha \cdot \frac{1 - \left[-\ln F(x) \right]^k}{k}$	$\lambda_1 = \xi + \alpha \cdot \frac{1 - \Gamma(1+k)}{k}$ $\lambda_2 = \alpha \cdot \frac{(1-2^{-k}) \cdot \Gamma(1+k)}{k}$ $\tau_3 = \frac{2 \cdot (1-3^{-k})}{1-2^{-k}} - 3$ $\tau_4 = \frac{1 - 6 \cdot 2^{-k} + 10 \cdot 3^{-k} - 5 \cdot 4^{-k}}{1-2^{-k}}$	$c = \frac{2}{3+\tau_3} - \frac{\ln 2}{\ln 3}$ $k = 7.8590c + 2.9554c^2$ $\alpha = \frac{\lambda_2 k}{(1-2^{-k}) \Gamma(1+k)}$ $\xi = \lambda_1 - \frac{\alpha [1 - \Gamma(1+k)]}{k}$
Logístico Generalizado (GLO)	$x(F) = \varepsilon + \alpha \cdot \frac{1 - \left[\frac{1 - F(x)}{F(x)} \right]^k}{k}$	$\lambda_1 = \xi + \alpha \cdot \frac{1 - \Gamma(1+k) \cdot \Gamma(1-k)}{k}$ $\lambda_2 = \alpha \cdot \Gamma(1+k) \cdot \Gamma(1-k)$ $\tau_3 = -k$ $\tau_4 = \frac{1+5k^2}{6}$	$k = -\tau_3$ $\alpha = \frac{\lambda_2 \sin k\pi}{k\pi}$ $\xi = \lambda_1 - \alpha \left(\frac{1}{k} - \frac{\pi}{\sin k\pi} \right)$

<p>Normal Generalizado (GNO)</p>	$f(x) = \frac{e^{-k \cdot y - y^2/2}}{\alpha \sqrt{2\pi}}$ $y = \begin{cases} -k^{-1} \ln[1 - k(x - \xi)/\alpha], & k \neq 0 \\ (x - \xi)/\alpha, & k = 0 \end{cases}$ <p>$F(x) = \Phi(y)$</p> <p>$x(F)$ No tiene forma analítica explícita.</p> <p>Φ, es la función de distribución acumulada de la distribución normal estándar.</p>	$\lambda_1 = \xi + \alpha \cdot \frac{1 - \exp(k^2/2)}{k}$ $\lambda_2 = \frac{\alpha e^2}{k} \left[1 - 2\Phi\left(-\frac{k}{\sqrt{2}}\right) \right]$ $\tau_3 \approx -k \frac{A_0 + A_1 k^2 + A_2 k^4 + A_3 k^6}{1 + B_1 k^2 + B_2 k^4 + B_3 k^6}$ $\tau_4 \approx \tau_4^0 - k \frac{C_0 + C_1 k^2 + C_2 k^4 + C_3 k^6}{1 + D_1 k^2 + D_2 k^4 + D_3 k^6}$ <p>τ_3, τ_4; Estos no tienen simples expresiones, por lo que se utiliza aproximaciones con precisión de 2×10^{-7} y 5×10^{-7} respectivamente.</p> <p>$\lambda_1 = \xi + \alpha \cdot \beta$</p> $\lambda_2 = \pi^{-1/2} \beta \Gamma\left(\alpha + \frac{1}{2}\right) / \Gamma(\alpha)$ $\tau_3 \approx -k \frac{A_0 + A_1 k^2 + A_2 k^4 + A_3 k^6}{1 + B_1 k^2 + B_2 k^4 + B_3 k^6}$ $\tau_4 \approx \tau_4^0 - k \frac{C_0 + C_1 k^2 + C_2 k^4 + C_3 k^6}{1 + D_1 k^2 + D_2 k^4 + D_3 k^6}$ <p>Las aproximaciones de τ_3 y τ_4 tienen precisión de 10^{-6}.</p>	$k \approx -\tau_3 \frac{E_0 + E_1 \tau_3^2 + E_2 \tau_3^4 + E_3 \tau_3^6}{1 + F_1 \tau_3^2 + F_2 \tau_3^4 + F_3 \tau_3^6}$ $\alpha = \frac{\lambda_2 k e^{-k^2/2}}{1 - 2\Phi(-k/\sqrt{2})}$ $\xi = \lambda_1 - \frac{\alpha}{k} (1 - e^{k^2/2})$
<p>Pearson Tipo III (PE3)</p>	<p>Si: $\gamma \neq 0, \alpha = 4/\gamma^2, \beta = \frac{1}{2} \sigma \gamma$,</p> $y \xi = \mu - 2 \frac{\sigma}{\gamma}$ <p>si: $\gamma > 0, \rightarrow \xi \leq x < \infty, y$:</p> $F(x) = \frac{G\left(\alpha, \frac{x - \xi}{\beta}\right)}{\Gamma \alpha}$ $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, G(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t} dt$	<p>si: $0 < \tau_3 < \frac{1}{3} : z = 3\pi \cdot \tau_3^2$</p> $\alpha = \frac{1 + 0.2906z}{z + 0.1882z^2 + 0.0442z^3}$ <p>si: $\frac{1}{3} < \tau_3 < 1 : z = 1 - \tau_3$</p> $\alpha = \frac{0.36067z - 0.59567z^2 + 0.25361z^3}{1 - 2.78861z + 2.56096z^2 - 0.77045z^3}$ <p>$\gamma = 2\alpha^{-1/2} \text{sign}(\tau_3)$</p> <p>$\sigma = \lambda_2 \pi^{1/2} \alpha^{1/2} \Gamma(\alpha) / \Gamma(\alpha + 1/2)$</p> <p>$\mu = \lambda_1$</p>	

Fuente: Hosking y Wallis (1997) y Bílková (2014).

Cuadro 27: Coeficientes de las aproximaciones para GNO

A0	4.8860251×10^{-1}	C1	1.8756590×10^{-1}	E1	2.0466534
A1	4.4493076×10^{-3}	C2	$-2.5352147 \times 10^{-3}$	E2	-3.6544371
A2	8.8027039×10^{-4}	C3	2.6995102×10^{-4}	E3	1.8396733
A3	1.1507084×10^{-6}	C4	$-1.8446680 \times 10^{-6}$	E4	-0.20360244
B1	6.4662924×10^{-2}	D1	8.2325617×10^{-2}	F1	-2.0182173
B2	3.3090406×10^{-3}	D2	4.2681448×10^{-3}	F2	1.2420401
B3	7.4290680×10^{-5}	D3	1.1653690×10^{-4}	F3	-0.21741801

Cuadro 28: Coeficientes de las aproximaciones para PE3

A0	3.2573501×10^{-1}	C1	1.2260172×10^{-1}	B1	4.6697102×10^{-1}
A1	1.6869150×10^{-1}	C2	5.3730130×10^{-1}	B2	2.4255406×10^{-1}
A2	7.8327243×10^{-2}	C3	4.3384378×10^{-2}	D1	1.8324466×10^{-1}
A3	2.9120539×10^{-3}	C4	1.1101277×10^{-3}	D2	2.0166036×10^{-1}

8.2 Cuantiles de precipitación máxima en 24 horas estimadas por ARF

Cuadro 29: Cuantiles de precipitación máxima en 24 horas estimadas por ARF para cada estación en mm y sus medidas de precisión.

Estación	F =	0.5	0.8	0.9	0.95	0.98	0.99	0.995	0.998	0.999
	TR =	2	5	10	20	50	100	200	500	1000
Arapa	Cuantil	35.82	44.65	50.25	55.49	62.13	67.04	71.91	78.34	83.22
	RECM	1.49	2.24	3.08	4.00	5.18	6.13	7.08	8.32	9.36
	LEI	33.46	41.13	45.50	49.38	54.32	57.80	61.34	65.97	69.45
	LES	38.30	48.39	55.59	62.35	70.97	77.44	84.08	92.53	99.52
Cabanillas	Cuantil	37.66	46.94	52.83	58.34	65.32	70.48	75.60	82.36	87.49
	RECM	1.51	2.29	3.22	4.16	5.43	6.43	7.41	8.72	9.79
	LEI	35.26	43.33	47.83	51.97	57.06	60.71	64.60	69.51	73.08
	LES	40.19	50.82	58.28	65.36	74.34	81.19	88.03	97.09	104.24
Capachica	Cuantil	37.77	47.07	52.98	58.50	65.50	70.68	75.81	82.59	87.73
	RECM	1.83	2.65	3.52	4.48	5.73	6.78	7.81	9.07	10.14
	LEI	34.86	42.90	47.63	51.75	56.90	60.58	64.30	69.28	72.93
	LES	40.84	51.48	59.07	66.26	75.34	82.17	89.22	98.59	105.42

Continuación

Estación	F =	0.5	0.8	0.9	0.95	0.98	0.99	0.995	0.998	0.999
	TR =	2	5	10	20	50	100	200	500	1000
Capazo	Cuantil	28.52	35.54	40.01	44.18	49.46	53.37	57.25	62.37	66.25
	RECM	1.16	1.76	2.45	3.20	4.14	4.90	5.64	6.65	7.43
	LEI	26.67	32.77	36.24	39.36	43.24	46.01	48.88	52.66	55.46
	LES	30.52	38.55	44.23	49.69	56.65	61.73	66.92	73.88	79.14
Chilligua	Cuantil	26.18	32.63	36.73	40.56	45.41	49.00	52.56	57.26	60.82
	RECM	1.12	1.68	2.32	2.96	3.88	4.55	5.26	6.11	6.88
	LEI	24.38	30.05	33.19	36.07	39.64	42.29	44.85	48.26	50.79
	LES	28.08	35.57	40.71	45.62	52.05	56.94	61.59	67.87	72.80
Chuquibambilla	Cuantil	35.69	44.48	50.06	55.28	61.89	66.79	71.64	78.04	82.90
	RECM	1.62	2.39	3.24	4.18	5.30	6.26	7.20	8.51	9.49
	LEI	33.15	40.64	45.03	48.85	53.85	57.33	60.94	65.46	68.99
	LES	38.39	48.47	55.54	62.19	70.75	77.37	83.94	92.56	99.05
Cojata	Cuantil	35.06	43.70	49.18	54.31	60.81	65.61	70.38	76.67	81.45
	RECM	1.58	2.33	3.17	4.04	5.18	6.20	7.07	8.42	9.33
	LEI	32.58	40.05	44.21	48.08	53.12	56.33	59.96	64.38	67.81
	LES	37.74	47.70	54.65	61.19	69.75	76.22	82.44	91.29	97.81
Cuyo Cuyo	Cuantil	28.05	34.96	39.34	43.44	48.64	52.49	56.30	61.33	65.15
	RECM	1.22	1.81	2.48	3.17	4.10	4.82	5.56	6.57	7.34
	LEI	26.13	32.16	35.51	38.63	42.54	45.32	48.13	51.77	54.54
	LES	30.13	38.04	43.64	48.89	55.72	60.85	66.01	72.92	78.09
Desaguadero	Cuantil	47.73	59.49	66.95	73.93	82.77	89.32	95.81	104.37	110.87
	RECM	2.08	3.11	4.26	5.46	7.07	8.28	9.52	11.15	12.57
	LEI	44.43	54.65	60.38	65.55	72.21	76.87	81.59	87.90	92.49
	LES	51.27	64.86	74.27	83.28	94.95	103.45	112.17	123.76	132.77
Huancane	Cuantil	42.67	53.18	59.86	66.10	74.00	79.85	85.66	93.31	99.12
	RECM	1.79	2.69	3.74	4.73	6.20	7.33	8.37	9.94	11.13
	LEI	39.82	48.95	54.11	58.89	64.69	68.92	73.34	78.72	82.91
	LES	45.68	57.73	66.21	74.15	84.65	92.41	100.18	110.70	118.42
Huaraya Moho	Cuantil	46.96	58.53	65.88	72.74	81.45	87.89	94.27	102.70	109.09
	RECM	2.10	3.11	4.19	5.40	6.96	8.18	9.45	11.09	12.34
	LEI	43.62	53.58	59.40	64.52	70.91	75.80	80.42	86.18	91.02
	LES	50.51	63.76	72.87	82.03	93.23	101.79	110.29	121.73	130.45

Continuación

Estación	F =	0.5	0.8	0.9	0.95	0.98	0.99	0.995	0.998	0.999
	TR =	2	5	10	20	50	100	200	500	1000
Ichuña	Cuantil	33.97	42.33	47.65	52.62	58.91	63.57	68.19	74.28	78.91
	RECM	1.44	2.15	2.95	3.78	4.95	5.81	6.70	7.87	8.82
	LEI	31.65	38.99	43.08	46.85	51.57	54.85	58.24	62.71	66.02
	LES	36.37	46.02	52.78	59.19	67.50	73.55	79.88	87.89	94.33
Ilave	Cuantil	35.20	43.87	49.38	54.53	61.05	65.88	70.66	76.98	81.77
	RECM	1.57	2.34	3.18	4.06	5.23	6.25	7.13	8.41	9.33
	LEI	32.70	40.20	44.48	48.35	53.07	56.59	60.14	64.47	68.14
	LES	37.84	47.80	54.78	61.34	69.95	76.55	82.97	91.53	97.83
Juli	Cuantil	45.42	56.61	63.72	70.36	78.77	85.00	91.18	99.33	105.51
	RECM	2.22	3.22	4.28	5.42	7.01	8.14	9.41	11.05	12.28
	LEI	41.90	51.56	57.14	62.15	68.24	72.90	77.31	83.24	87.54
	LES	49.18	62.07	71.09	79.56	90.84	98.94	107.33	118.58	127.01
Lagunillas	Cuantil	38.22	47.64	53.62	59.21	66.29	71.53	76.73	83.59	88.79
	RECM	1.76	2.59	3.50	4.43	5.74	6.77	7.78	9.16	10.17
	LEI	35.42	43.58	48.19	52.47	57.63	61.36	65.11	70.01	73.90
	LES	41.17	52.06	59.64	66.75	76.07	83.04	89.98	99.49	106.51
Lampa	Cuantil	39.76	49.55	55.77	61.58	68.95	74.40	79.80	86.94	92.35
	RECM	1.63	2.47	3.43	4.41	5.71	6.78	7.82	9.22	10.34
	LEI	37.21	45.73	50.57	54.86	60.43	64.36	68.24	73.49	77.15
	LES	42.48	53.78	61.72	69.22	78.84	86.05	93.14	102.97	110.64
Laraqueri	Cuantil	37.29	46.47	52.31	57.76	64.67	69.78	74.85	81.54	86.62
	RECM	1.69	2.52	3.40	4.33	5.53	6.58	7.59	8.94	9.88
	LEI	34.59	42.48	47.01	51.17	56.37	59.96	63.65	68.32	72.23
	LES	40.10	50.66	58.18	65.12	74.20	81.02	87.76	96.86	103.77
Llally	Cuantil	34.73	43.29	48.72	53.80	60.23	65.00	69.72	75.95	80.68
	RECM	1.71	2.48	3.31	4.18	5.33	6.29	7.21	8.50	9.44
	LEI	32.03	39.42	43.61	47.42	52.25	55.72	59.16	63.56	67.17
	LES	37.62	47.50	54.47	61.04	69.39	75.90	82.14	90.71	97.36
Macusani	Cuantil	31.77	39.59	44.56	49.20	55.09	59.45	63.77	69.47	73.79
	RECM	1.58	2.28	3.00	3.82	4.88	5.72	6.54	7.69	8.56
	LEI	29.28	36.03	39.94	43.39	47.81	50.96	54.12	58.14	61.05
	LES	34.45	43.51	49.73	55.80	63.56	69.33	75.20	82.94	88.69

Continuación

Estación	F =	0.5	0.8	0.9	0.95	0.98	0.99	0.995	0.998	0.999
	TR =	2	5	10	20	50	100	200	500	1000
Mazo Cruz	Cuantil	29.60	36.89	41.53	45.85	51.34	55.40	59.42	64.73	68.77
	RECM	1.24	1.88	2.59	3.36	4.36	5.14	5.91	7.00	7.88
	LEI	27.64	33.93	37.55	40.70	44.78	47.73	50.64	54.46	57.14
	LES	31.70	40.12	45.90	51.57	58.67	64.08	69.40	76.59	82.09
Muñani	Cuantil	30.50	38.01	42.79	47.24	52.90	57.08	61.23	66.70	70.85
	RECM	1.26	1.91	2.64	3.42	4.45	5.29	6.09	7.14	8.00
	LEI	28.50	34.95	38.68	42.05	46.20	49.15	51.99	56.18	59.11
	LES	32.66	41.26	47.23	53.04	60.37	66.05	71.35	78.84	84.52
Pampahuta	Cuantil	37.96	47.31	53.25	58.80	65.83	71.04	76.20	83.01	88.18
	RECM	1.66	2.48	3.38	4.34	5.61	6.67	7.63	9.05	10.09
	LEI	35.34	43.44	47.94	52.23	57.42	61.29	64.98	69.82	73.70
	LES	40.81	51.54	59.00	66.30	75.44	82.37	89.24	98.58	105.80
Pizacoma	Cuantil	35.11	43.75	49.25	54.38	60.89	65.70	70.47	76.77	81.55
	RECM	1.52	2.26	3.12	4.01	5.20	6.12	7.09	8.29	9.30
	LEI	32.69	40.20	44.43	48.26	53.07	56.52	60.04	64.66	67.92
	LES	37.69	47.63	54.58	61.13	69.76	76.15	82.49	90.99	97.55
Progreso	Cuantil	31.83	39.67	44.65	49.31	55.20	59.57	63.90	69.61	73.94
	RECM	1.36	2.03	2.79	3.57	4.65	5.50	6.34	7.46	8.33
	LEI	29.68	36.51	40.38	43.93	48.22	51.39	54.45	58.52	61.84
	LES	34.10	43.14	49.40	55.48	63.15	69.00	74.71	82.48	88.32
Pucará	Cuantil	39.38	49.08	55.24	61.00	68.29	73.69	79.05	86.11	91.47
	RECM	1.61	2.46	3.42	4.39	5.77	6.76	7.80	9.23	10.25
	LEI	36.75	45.17	49.93	54.32	59.58	63.58	67.43	72.62	76.50
	LES	42.06	53.22	61.08	68.41	78.11	85.20	92.33	101.86	109.23
Puno	Cuantil	40.65	50.66	57.02	62.97	70.50	76.07	81.60	88.89	94.43
	RECM	1.69	2.54	3.53	4.57	5.94	6.94	8.01	9.51	10.63
	LEI	37.98	46.72	51.51	56.05	61.55	65.70	69.82	74.93	79.01
	LES	43.48	54.98	63.09	70.75	80.62	87.96	95.42	105.47	112.89
Putina	Cuantil	36.43	45.40	51.10	56.42	63.17	68.17	73.12	79.66	84.62
	RECM	1.64	2.41	3.28	4.19	5.42	6.36	7.38	8.63	9.66
	LEI	33.79	41.60	45.96	50.07	55.06	58.70	62.22	66.87	70.38
	LES	39.14	49.51	56.75	63.61	72.40	79.09	85.57	94.29	101.35

Continuación

Estación	F =	0.5	0.8	0.9	0.95	0.98	0.99	0.995	0.998	0.999
		TR =	2	5	10	20	50	100	200	500
Santa Rosa	Cuantil	34.67	43.21	48.64	53.71	60.13	64.89	69.60	75.82	80.54
	RECM	1.83	2.59	3.42	4.26	5.44	6.36	7.29	8.53	9.54
	LEI	31.82	39.20	43.37	47.26	51.89	55.48	58.77	63.39	66.60
	LES	37.77	47.71	54.59	61.13	69.50	75.89	82.43	90.74	97.27
Taraco	Cuantil	37.14	46.28	52.09	57.52	64.40	69.50	74.55	81.21	86.27
	RECM	1.53	2.32	3.19	4.09	5.41	6.29	7.31	8.62	9.66
	LEI	34.66	42.66	47.14	51.24	56.13	59.98	63.64	68.38	72.02
	LES	39.69	50.15	57.37	64.45	73.41	80.11	86.88	95.82	102.85

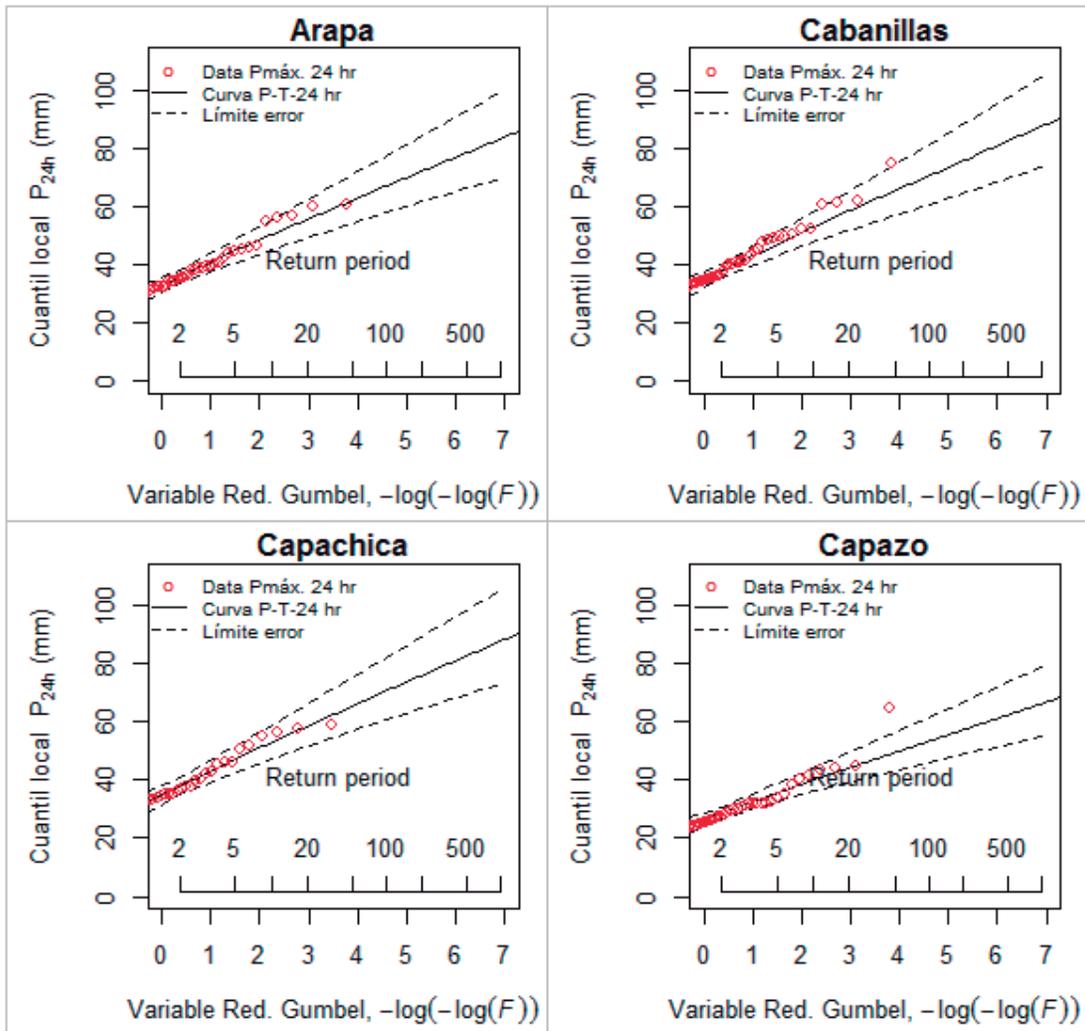


Figura 25: Cuantiles para las estaciones de Arapa, Cabanillas, Capachica, y Capazo.

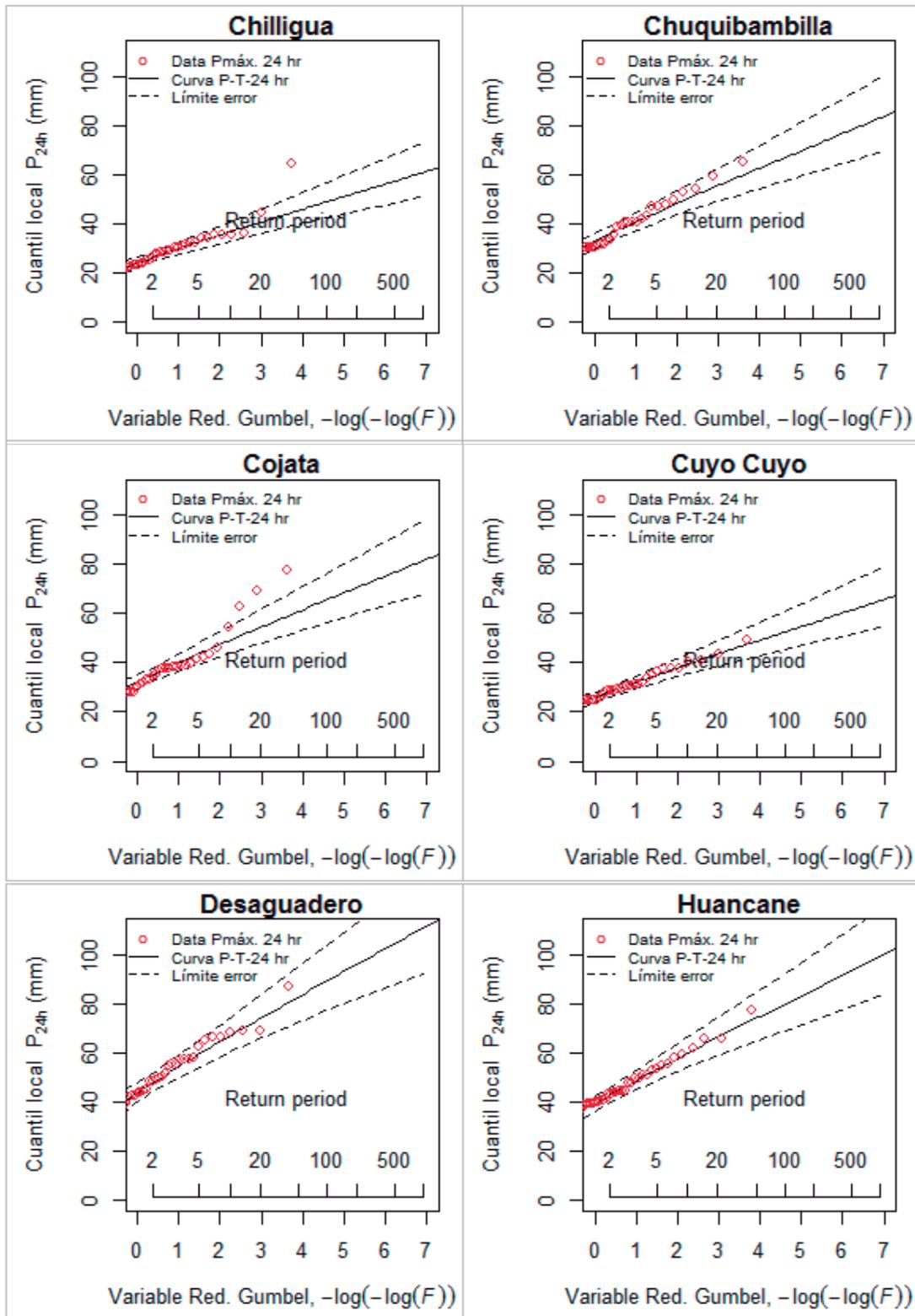


Figura 26: Cuantiles para las estaciones de Chilligua, Chuquibambilla, Cojata, Cuyo Cuyo, Desaguadero y Huancané.

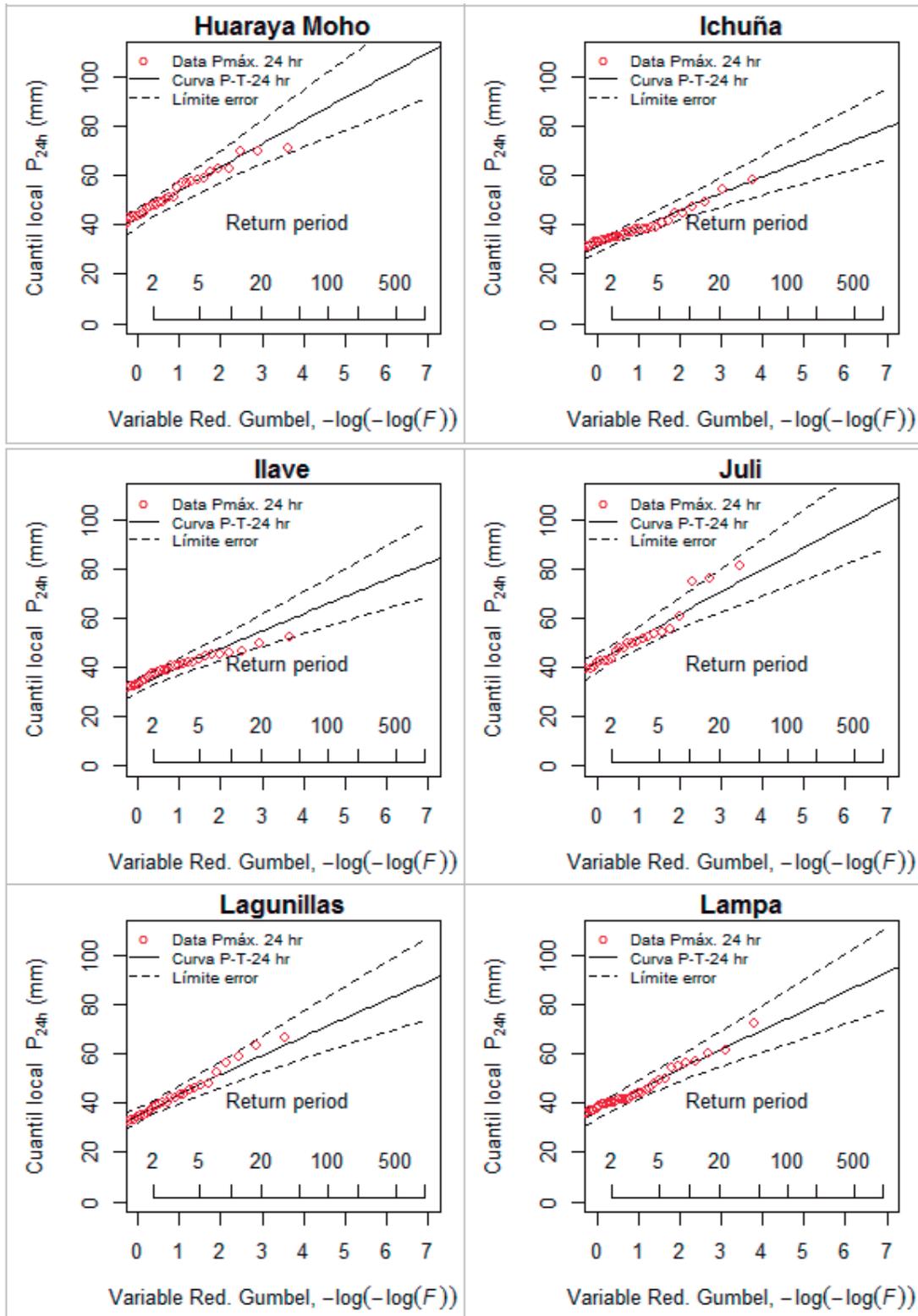


Figura 27: Cuantiles para las estaciones de Huaraya Moho, Ichuña, Ilave, Juli, Lagunillas y Lampa.

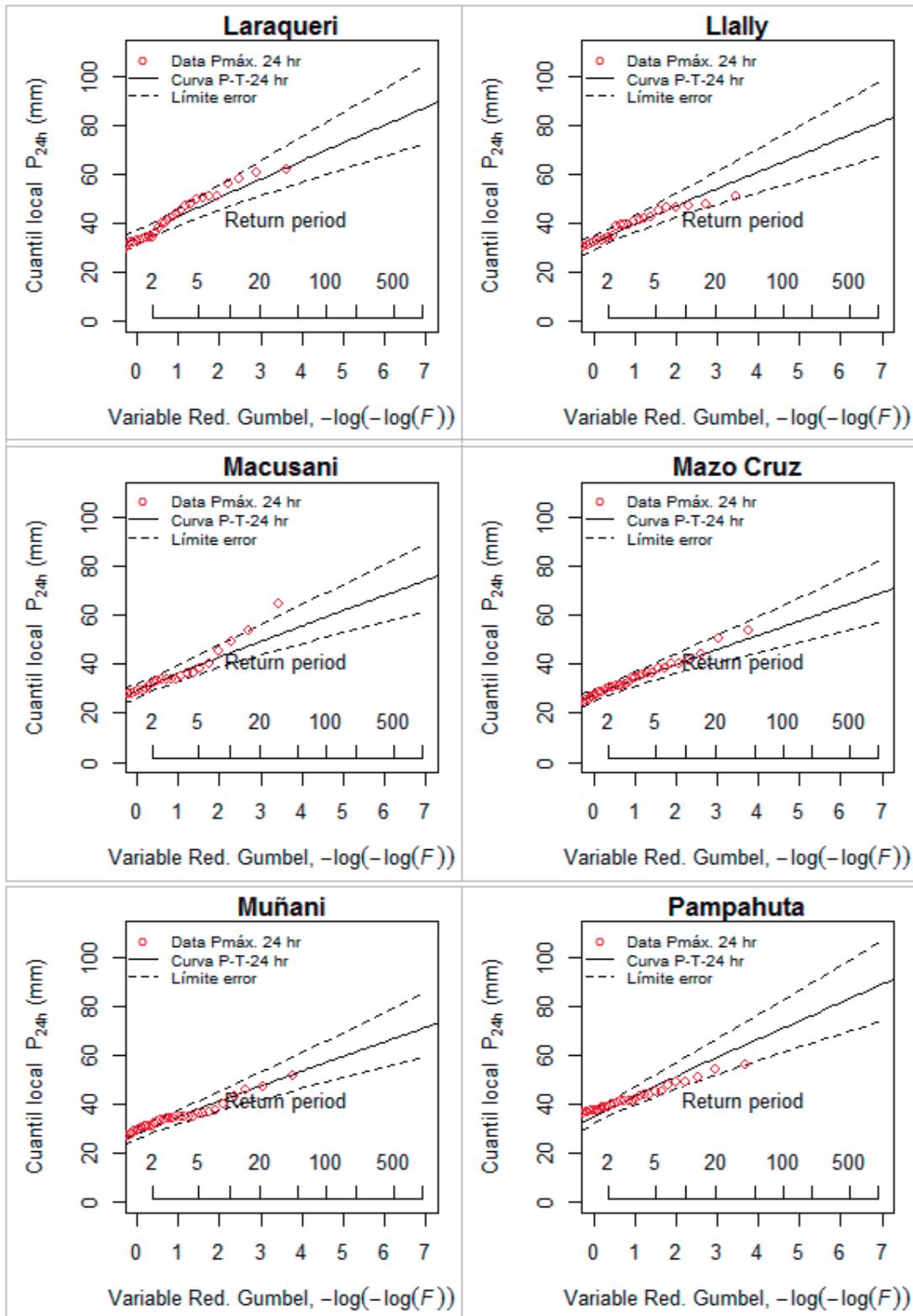


Figura 28: Cuantiles para las estaciones de Laraqueri, Llally, Macusani, Mazo Cruz, Muñani, y Pampahuta

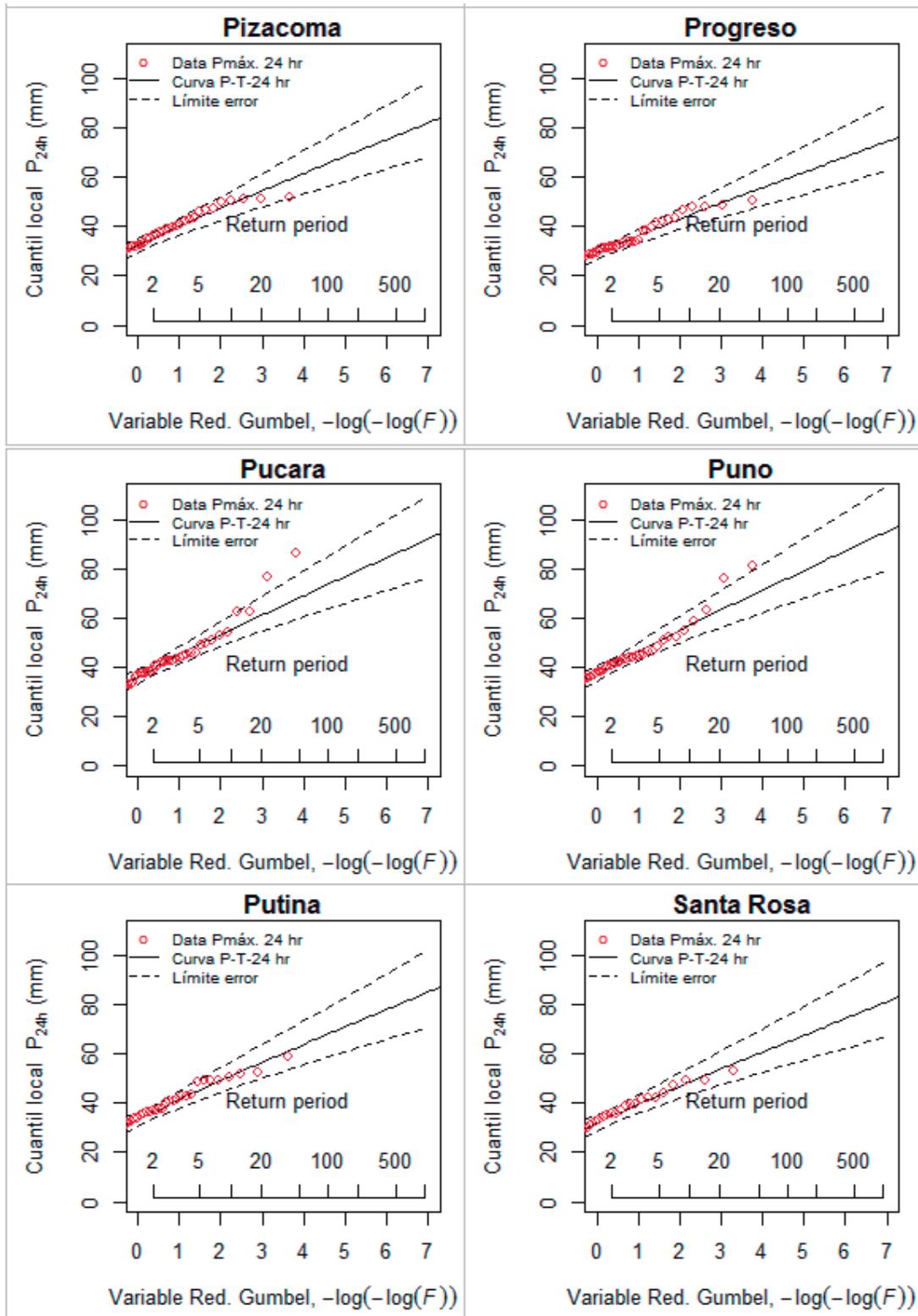


Figura 29: Cuantiles para las estaciones de Pizacoma, Progreso, Pucará, Puno, Putina y Santa Rosa.

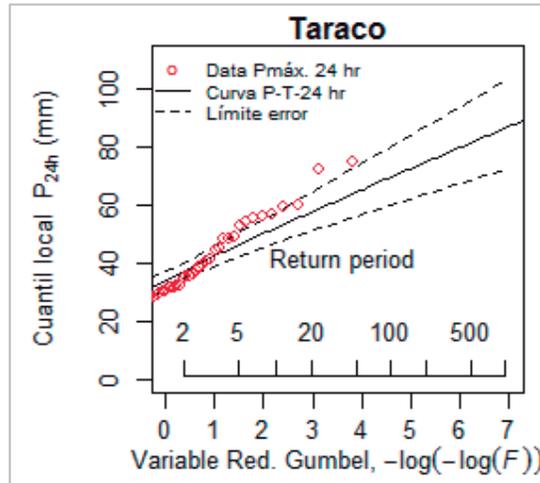


Figura 30: Cuantiles para las estaciones de Taraco.

8.3 Comparación de los cuantiles estimados por el análisis local y ARF

Cuadro 30: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 2 y 5 años.

N°	Estación nombre	P _{máx} 24 hr, TR=2 años				P _{máx} 24 hr, TR=5 años			
		GNO (ARF)	Gumbel (AL)	Dif.	%	GNO (ARF)	Gumbel (AL)	Dif.	%
1	Arapa	35.82	35.62	-0.20	(0.56)	44.65	43.40	-1.25	(2.80)
2	Cabanillas	37.66	37.30	-0.36	(0.95)	46.94	48.55	1.61	3.42
3	Capachica	37.77	37.69	-0.08	(0.21)	47.07	46.04	-1.03	(2.20)
4	Capazo	28.52	28.24	-0.28	(1.00)	35.54	34.50	-1.05	(2.94)
5	Chilligua	26.18	25.87	-0.32	(1.21)	32.63	33.74	1.11	3.39
6	Chuquibambilla	35.69	35.43	-0.26	(0.72)	44.48	44.82	0.34	0.76
7	Cojata	35.06	34.32	-0.75	(2.13)	43.70	43.48	-0.22	(0.50)
8	Cuyo Cuyo	28.05	28.05	0.00	0.01	34.96	34.60	-0.36	(1.02)
9	Desaguadero	47.73	47.53	-0.20	(0.42)	59.49	60.02	0.54	0.90
10	Huancané	42.67	42.61	-0.06	(0.13)	53.18	53.33	0.15	0.28
11	Huaraya Moho	46.96	47.09	0.13	0.27	58.53	57.79	-0.74	(1.26)
12	Ichuña	33.97	34.25	0.29	0.84	42.33	42.83	0.49	1.17
13	Ilave	35.20	35.50	0.30	0.85	43.87	43.06	-0.81	(1.85)
14	Juli	45.42	45.15	-0.27	(0.60)	56.61	55.14	-1.47	(2.60)
15	Lagunillas	38.22	37.97	-0.25	(0.65)	47.64	46.77	-0.87	(1.82)
16	Lampa	39.76	39.70	-0.06	(0.14)	49.55	49.11	-0.43	(0.88)
17	Laraqueri	37.29	36.88	-0.41	(1.10)	46.47	46.05	-0.43	(0.93)
18	Llally	34.73	34.89	0.16	0.45	43.29	42.22	-1.07	(2.46)
19	Macusani	31.77	31.43	-0.34	(1.06)	39.59	39.57	-0.02	(0.05)
20	Mazo Cruz	29.60	29.50	-0.10	(0.34)	36.89	37.13	0.23	0.63
21	Muñani	30.50	30.62	0.12	0.39	38.01	37.51	-0.51	(1.34)
22	Pampahuta	37.96	38.42	0.46	1.22	47.31	45.64	-1.67	(3.53)
23	Pizacoma	35.11	35.24	0.13	0.37	43.75	44.69	0.94	2.15
24	Progreso	31.83	31.76	-0.07	(0.23)	39.67	39.47	-0.20	(0.52)
25	Pucará	39.38	39.03	-0.35	(0.88)	49.08	50.46	1.39	2.83
26	Puno	40.65	40.37	-0.28	(0.68)	50.66	49.77	-0.90	(1.77)
27	Putina	36.43	36.53	0.11	0.29	45.40	45.91	0.51	1.13
28	Santa Rosa	34.67	34.90	0.23	0.67	43.21	44.30	1.09	2.51
29	Taraco	37.14	36.22	-0.92	(2.47)	46.28	46.71	0.43	0.92
		Sobrestima (Promedio)		0.54		Sobrestima (Promedio)		1.67	
		Subestima (Promedio)		(0.82)		Subestima (Promedio)		(1.67)	

Cuadro 31: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 10 y 20 años

N°	Estación nombre	Pmáx 24 hr, TR=10 años				Pmáx 24 hr, TR=20 años			
		GNO (ARF)	Gumbel (AL)	Dif.	%	GNO (ARF)	Gumbel (AL)	Dif.	%
1	Arapa	50.25	48.54	-1.71	(3.40)	55.49	53.48	-2.01	(3.62)
2	Cabanillas	52.83	55.99	3.15	5.97	58.34	63.13	4.79	8.21
3	Capachica	52.98	51.56	-1.42	(2.68)	58.50	56.86	-1.64	(2.80)
4	Capazo	40.01	38.65	-1.36	(3.40)	44.18	42.62	-1.55	(3.51)
5	Chilligua	36.73	38.95	2.22	6.05	40.56	43.95	3.39	8.37
6	Chuquibambilla	50.06	51.03	0.97	1.93	55.28	56.99	1.71	3.09
7	Cojata	49.18	49.55	0.36	0.74	54.31	55.37	1.06	1.94
8	Cuyo Cuyo	39.34	38.93	-0.41	(1.04)	43.44	43.09	-0.35	(0.81)
9	Desaguadero	66.95	68.29	1.34	2.00	73.93	76.22	2.29	3.10
10	Huancané	59.86	60.42	0.56	0.94	66.10	67.22	1.13	1.71
11	Huaraya Moho	65.88	64.88	-1.00	(1.52)	72.74	71.68	-1.07	(1.47)
12	Ichuña	47.65	48.50	0.85	1.79	52.62	53.95	1.33	2.54
13	Ilave	49.38	48.07	-1.31	(2.66)	54.53	52.87	-1.66	(3.04)
14	Juli	63.72	61.76	-1.96	(3.08)	70.36	68.10	-2.26	(3.21)
15	Lagunillas	53.62	52.60	-1.02	(1.90)	59.21	58.19	-1.02	(1.72)
16	Lampa	55.77	55.35	-0.42	(0.76)	61.58	61.33	-0.25	(0.41)
17	Laraqueri	52.31	52.11	-0.20	(0.38)	57.76	57.93	0.17	0.30
18	Llally	48.72	47.08	-1.64	(3.38)	53.80	51.73	-2.06	(3.84)
19	Macusani	44.56	44.96	0.40	0.89	49.20	50.13	0.92	1.87
20	Mazo Cruz	41.53	42.18	0.65	1.56	45.85	47.02	1.16	2.54
21	Muñani	42.79	42.07	-0.72	(1.68)	47.24	46.44	-0.81	(1.70)
22	Pampahuta	53.25	50.41	-2.84	(5.33)	58.80	54.99	-3.80	(6.47)
23	Pizacoma	49.25	50.95	1.71	3.46	54.38	56.96	2.58	4.74
24	Progreso	44.65	44.57	-0.08	(0.18)	49.31	49.47	0.16	0.33
25	Pucará	55.24	58.04	2.79	5.06	61.00	65.30	4.30	7.05
26	Puno	57.02	55.99	-1.04	(1.82)	62.97	61.95	-1.01	(1.61)
27	Putina	51.10	52.12	1.02	1.99	56.42	58.07	1.65	2.92
28	Santa Rosa	48.64	50.52	1.88	3.86	53.71	56.48	2.78	5.17
29	Taraco	52.09	53.66	1.56	3.00	57.52	60.32	2.79	4.86
		Sobrestima (Promedio)			2.80	Sobrestima (Promedio)			3.67
		Subestima (Promedio)			(2.21)	Subestima (Promedio)			(2.63)

Cuadro 32: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 50 y 100 años

N°	Estación nombre	Pmáx 24 hr, TR=50 años				Pmáx 24 hr, TR=100 años			
		GNO (ARF)	Gumbel (AL)	Dif.	%	GNO (ARF)	Gumbel (AL)	Dif.	%
1	Arapa	62.13	59.87	-2.25	(3.63)	67.04	64.66	-2.38	(3.54)
2	Cabanillas	65.32	72.37	7.05	10.79	70.48	79.29	8.81	12.50
3	Capachica	65.50	63.72	-1.77	(2.71)	70.68	68.87	-1.81	(2.56)
4	Capazo	49.46	47.77	-1.69	(3.41)	53.37	51.63	-1.74	(3.26)
5	Chilligua	45.41	50.42	5.01	11.04	49.00	55.27	6.27	12.80
6	Chuquibambilla	61.89	64.71	2.81	4.55	66.79	70.49	3.70	5.54
7	Cojata	60.81	62.90	2.09	3.44	65.61	68.54	2.93	4.46
8	Cuyo Cuyo	48.64	48.48	-0.17	(0.34)	52.49	52.51	0.02	0.04
9	Desaguadero	82.77	86.49	3.72	4.49	89.32	94.19	4.87	5.45
10	Huancané	74.00	76.03	2.03	2.74	79.85	82.63	2.78	3.48
11	Huaraya Moho	81.45	80.47	-0.97	(1.19)	87.89	87.07	-0.82	(0.93)
12	Ichuña	58.91	61.00	2.09	3.55	63.57	66.28	2.71	4.27
13	Ilave	61.05	59.09	-1.96	(3.21)	65.88	63.75	-2.13	(3.24)
14	Juli	78.77	76.31	-2.46	(3.12)	85.00	82.47	-2.53	(2.98)
15	Lagunillas	66.29	65.42	-0.87	(1.31)	71.53	70.84	-0.69	(0.97)
16	Lampa	68.95	69.07	0.12	0.17	74.40	74.86	0.47	0.63
17	Laraqueri	64.67	65.47	0.80	1.23	69.78	71.12	1.33	1.91

Continuación

18	Llally	60.23	57.76	-2.47	(4.10)	65.00	62.28	-2.72	(4.18)
19	Macusani	55.09	56.82	1.73	3.14	59.45	61.83	2.38	4.01
20	Mazo Cruz	51.34	53.29	1.95	3.79	55.40	57.98	2.58	4.67
21	Muñani	52.90	52.10	-0.80	(1.50)	57.08	56.34	-0.73	(1.29)
22	Pampahuta	65.83	60.93	-4.91	(7.45)	71.04	65.37	-5.67	(7.98)
23	Pizacoma	60.89	64.73	3.85	6.32	65.70	70.56	4.86	7.40
24	Progreso	55.20	55.81	0.60	1.09	59.57	60.56	0.99	1.66
25	Pucará	68.29	74.70	6.40	9.38	73.69	81.74	8.05	10.92
26	Puno	70.50	69.68	-0.82	(1.16)	76.07	75.46	-0.61	(0.80)
27	Putina	63.17	65.78	2.61	4.13	68.17	71.56	3.39	4.97
28	Santa Rosa	60.13	64.20	4.07	6.77	64.89	69.99	5.11	7.87
29	Taraco	64.40	68.94	4.54	7.05	69.50	75.40	5.91	8.50
				Sobrestima (Promedio)		Sobrestima (Promedio)		5.62	
				Subestima (Promedio)		Subestima (Promedio)		(2.88)	

Cuadro 33: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 200 y 500 años

N°	Estación nombre	Pmáx 24 hr, TR=200 años				Pmáx 24 hr, TR=500 años			
		GNO (ARF)	Gumbel (AL)	Dif.	%	GNO (ARF)	Gumbel (AL)	Dif.	%
1	Arapa	71.91	69.44	-2.47	(3.44)	78.34	75.73	-2.61	(3.33)
2	Cabanillas	75.60	86.19	10.59	14.01	82.36	95.30	12.94	15.71
3	Capachica	75.81	73.99	-1.82	(2.41)	82.59	80.75	-1.84	(2.23)
4	Capazo	57.25	55.48	-1.77	(3.10)	62.37	60.55	-1.82	(2.91)
5	Chilligua	52.56	60.11	7.55	14.36	57.26	66.48	9.22	16.11
6	Chuquibambilla	71.64	76.25	4.61	6.43	78.04	83.85	5.81	7.44
7	Cojata	70.38	74.17	3.78	5.38	76.67	81.59	4.91	6.41
8	Cuyo Cuyo	56.30	56.53	0.23	0.40	61.33	61.83	0.50	0.81
9	Desaguadero	95.81	101.85	6.04	6.31	104.37	111.97	7.60	7.28
10	Huancané	85.66	89.21	3.55	4.15	93.31	97.88	4.57	4.90
11	Huaraya Moho	94.27	93.64	-0.64	(0.67)	102.70	102.30	-0.39	(0.38)
12	Ichuña	68.19	71.54	3.36	4.92	74.28	78.48	4.20	5.66
13	Ilave	70.66	68.39	-2.28	(3.22)	76.98	74.51	-2.47	(3.21)
14	Juli	91.18	88.60	-2.58	(2.83)	99.33	96.69	-2.63	(2.65)
15	Lagunillas	76.73	76.24	-0.49	(0.64)	83.59	83.36	-0.22	(0.26)
16	Lampa	79.80	80.64	0.84	1.05	86.94	88.27	1.33	1.53
17	Laraqueri	74.85	76.74	1.89	2.52	81.54	84.16	2.62	3.21
18	Llally	69.72	66.78	-2.94	(4.21)	75.95	72.72	-3.23	(4.25)
19	Macusani	63.77	66.83	3.06	4.80	69.47	73.42	3.95	5.69
20	Mazo Cruz	59.42	62.66	3.24	5.45	64.73	68.84	4.10	6.34
21	Muñani	61.23	60.57	-0.66	(1.07)	66.70	66.15	-0.55	(0.83)
22	Pampahuta	76.20	69.80	-6.40	(8.40)	83.01	75.64	-7.37	(8.88)
23	Pizacoma	70.47	76.36	5.89	8.36	76.77	84.02	7.25	9.44
24	Progreso	63.90	65.29	1.39	2.18	69.61	71.53	1.92	2.76
25	Pucará	79.05	88.76	9.71	12.28	86.11	98.02	11.90	13.82
26	Puno	81.60	81.23	-0.37	(0.45)	88.89	88.84	-0.05	(0.06)
27	Putina	73.12	77.31	4.19	5.73	79.66	84.90	5.25	6.59
28	Santa Rosa	69.60	75.76	6.16	8.85	75.82	83.36	7.54	9.95
29	Taraco	74.55	81.84	7.30	9.79	81.21	90.34	9.13	11.24
				Sobrestima (Promedio)		Sobrestima (Promedio)		7.49	
				Subestima (Promedio)		Subestima (Promedio)		(2.64)	

Cuadro 34: Comparación de los cuantiles estimados por el análisis local y regional para el periodo de retorno de 1000 años

Estación		P _{máx} 24 hr, TR=1000 años				Estación		P _{máx} 24 hr, TR=1000 años			
N°	nombre	GNO	Gumbel	Dif.	%	N°	nombre	GNO	Gumbel	Dif.	%
1	Arapa	83.22	80.49	-2.72	(3.27)	Continuación					
2	Cabanillas	87.49	102.18	14.69	16.79	17	Laraqueri	86.62	89.77	3.15	3.64
3	Capachica	87.73	85.85	-1.88	(2.14)	18	Llally	80.68	77.21	-3.47	(4.30)
4	Capazo	66.25	64.38	-1.87	(2.82)	19	Macusani	73.79	78.40	4.61	6.24
5	Chilligua	60.82	71.30	10.48	17.22	20	Mazo Cruz	68.77	73.50	4.74	6.89
6	Chuquibambilla	82.90	89.59	6.69	8.07	21	Muñani	70.85	70.36	-0.49	(0.69)
7	Cojata	81.45	87.19	5.75	7.06	22	Pampahuta	88.18	80.05	-8.12	(9.21)
8	Cuyo Cuyo	65.15	65.84	0.69	1.05	23	Pizacoma	81.55	89.81	8.25	10.12
9	Desaguadero	110.87	119.61	8.74	7.88	24	Progreso	73.94	76.25	2.31	3.12
10	Huancané	99.12	104.44	5.32	5.36	25	Pucará	91.47	105.01	13.54	14.80
11	Huaraya Moho	109.09	108.85	-0.24	(0.22)	26	Puno	94.43	94.59	0.16	0.17
12	Ichuña	78.91	83.73	4.83	6.12	27	Putina	84.62	90.64	6.02	7.12
13	Ilave	81.77	79.14	-2.64	(3.22)	28	Santa Rosa	80.54	89.11	8.57	10.64
14	Juli	105.51	102.81	-2.71	(2.56)	29	Taraco	86.27	96.76	10.49	12.16
15	Lagunillas	88.79	88.75	-0.04	(0.05)			Sobrestima (Promedio)		7.70	
16	Lampa	92.35	94.03	1.68	1.82			Subestima (Promedio)		(2.85)	

Observación: Los cuantiles estimados por el ajuste exclusivo a las series de las precipitaciones máximas de cada estación, mediante la distribución de dos parámetros Gumbel, cumplen con el test de bondad de ajuste de Kolmogórov-Smirnov.