



PERÚ

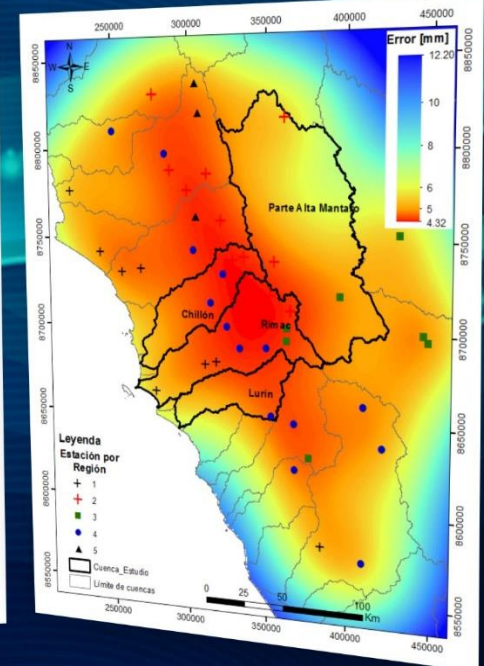
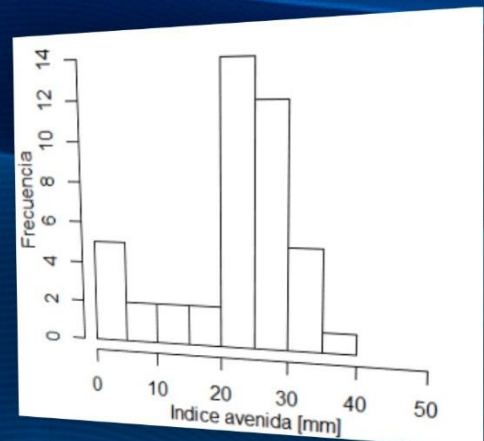
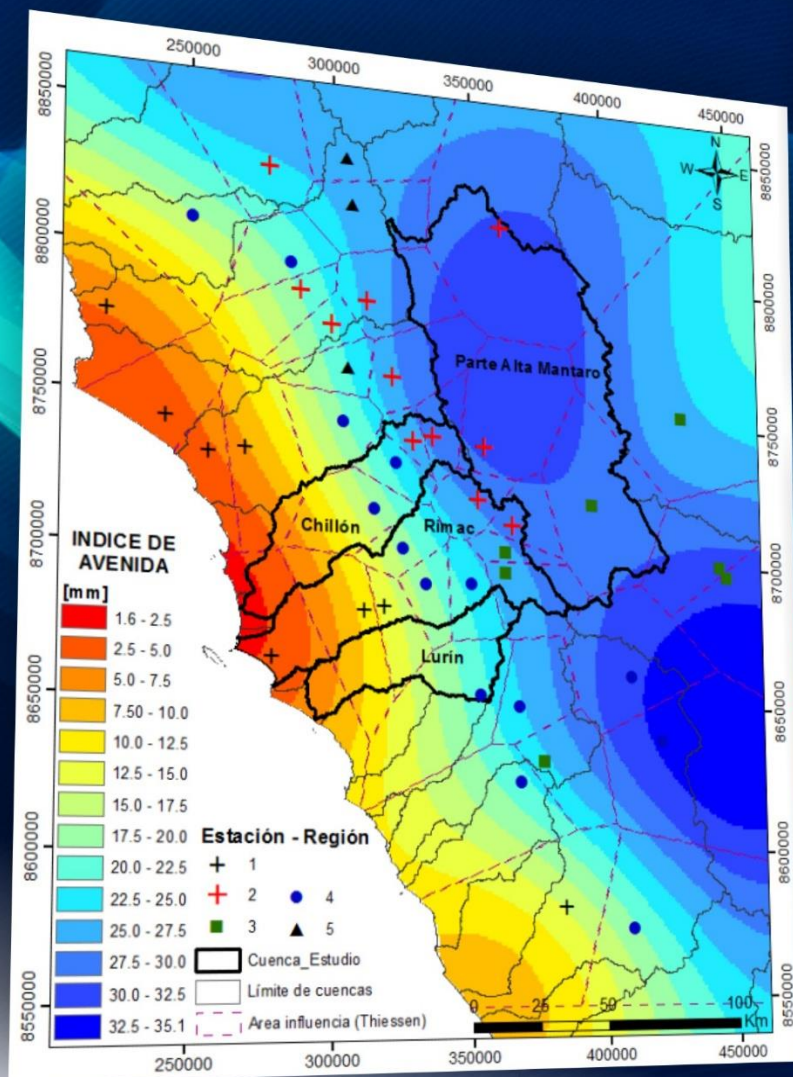
Ministerio del Ambiente

Servicio Nacional de Meteorología e Hidrología del Perú - SENAMHI

META: ESTUDIOS E INVESTIGACIÓN EN HIDROLOGÍA

ESTUDIO: ANÁLISIS DE INFORMACIÓN PARA GENERACIÓN DE ESTUDIOS HIDROMETEOROLÓGICOS

ANÁLISIS REGIONAL DE PRECIPITACIONES MÁXIMAS EN LAS CUENCAS CHILLÓN, RÍMAC, LURÍN Y PARTE ALTA DEL MANTARO





DIRECTORIO

Presidenta Ejecutiva del SENAMHI

Ing. Amelia Díaz Pabló

Director Científico

Ing. Esequiel Villegas Paredes

Director General de Hidrología y Recursos Hídricos

Ing. Oscar G. Felipe Obando

RESPONSABLES DE LA ELABORACION

Especialista en Hidrología

Ing. Carlos Fernández Palomino

Director de Hidrología Aplicada

Dr. Waldo Sven Lavado Casimiro

Diciembre – 2014

LIMA - PERÚ

ÍNDICE

	Pág.
I. INTRODUCCIÓN.....	2
1.1 Análisis de frecuencias de precipitaciones máximas	2
1.2 L-momentos	3
1.3 Objetivos del estudio	4
1.4 Importancia del Estudio	4
II. REVISIÓN BIBLIOGRÁFICA	5
2.1 Análisis de frecuencia de eventos extremos	5
2.2 Análisis exploratorio de datos	6
2.2.1 Diagrama de caja (<i>Boxplot</i>).	6
2.3 Supuestos del análisis regional	7
2.3.1 Prueba de Mann Kendall (MK) de detección de tendencias	7
2.3.2 Prueba de Ljung-Box (Q-estadístico) de detección de autocorrelación	8
2.4 Análisis regional de frecuencias	9
2.5 L-momentos	11
2.6 El método del <i>índice de avenida</i>	13
2.7 Procedimiento del análisis regional	14
2.7.1 Filtrado de datos usando medidas de discordancia	14
2.7.2 Identificación de regiones homogéneas	16
2.7.3 Elección de la distribución de frecuencia	19
2.7.4 Estimación de los cuantiles de frecuencia	22
2.8 Mapeo de índice de avenida	27
2.8.1 Método de Co-kriging (Co-Ko).....	27
III. MATERIALES Y METODOLOGÍA.....	32
3.1 Descripción del área de estudio	32
3.2 Materiales	33

3.2.1 Información básica de sistema de información geográfica.....	33
3.2.2 Información de datos observados	33
3.2.3 Programas	38
3.3. Metodología.....	39
3.3.1 Análisis exploratorio de datos	41
3.3.2 Supuestos del análisis regional de frecuencias	41
3.3.3 Filtrado de las observaciones empleando la medida de discordancia.....	42
3.3.4 Identificación de regiones homogéneas	42
3.3.5 Selección de una función de distribución de frecuencia para la región.....	43
3.3.6 Estimación de los cuantiles regionales de frecuencia y locales.....	43
3.3.7 Mapeo de índice de avenida	44
IV. RESULTADOS Y DISCUSIÓN	45
4.1 Análisis exploratorio de datos	45
4.1.1 Diagramas de Boxplots	45
4.2 Supuestos del análisis regional de frecuencias	47
4.2.1 Análisis de tendencia	48
4.2.2 Análisis de independencia serial.....	51
4.3 Filtrado de datos usando medidas de discordancia	55
4.4 Heterogeneidad regional y la formación de regiones homogéneas	57
4.5 Selección de la distribución regional de mejor ajuste	60
4.6 Estimación de los cuantiles de frecuencia	62
4.6.1 Estimación de la curva de crecimiento regional.....	62
4.6.2 Decisión acerca de la mejor curva de crecimiento regional	64
4.6.3 Cuantiles de frecuencia regional y local.....	67
4.8 Mapeo del índice de avenida	69
4.8.1 Mapeo de índice de avenida por Co - Kriging (Co-Ko).....	70
V. CONCLUSIONES	76

VI. RECOMENDACIONES	77
VII. REFERENCIAS BIBLIOGRÁFICAS	78
VIII ANEXO	83
8.1 Distribuciones teóricas y sus relaciones	83
8.2 Cuantiles de precipitación máxima en 24 horas estimadas por ARF	87

RELACIÓN DE CUADROS

Cuadro 1: Valores críticos de la discordancia D_i	15
Cuadro 2: Aproximaciones polinómicas de $L - C_k$ en función de $L - C_s$	20
Cuadro 3: Algoritmo para la simulación del Algoritmo de L-momento regional.....	26
Cuadro 4: Red de estaciones Meteorológicas	34
Cuadro 5: Resumen del test de Mann Kendall para el análisis de tendencias	48
Cuadro 6: Lista de estaciones aptas para el ARF.	54
Cuadro 7: Valores de los ratios de L – momentos de las estaciones y medida de discordancia considerando región entera.	55
Cuadro 8: Medida de discordancia para las estaciones de cada región.....	58
Cuadro 9: Estadísticas de medias ponderadas de L-momentos regionales	59
Cuadro 10: Parámetros de la distribución Kappa para la región.....	59
Cuadro 11: Medidas de heterogeneidad	59
Cuadro 12: Estadístico Z^{DIST} de varias distribuciones para cada región.....	61
Cuadro 13: Parámetros y cuantiles regionales para las distribuciones candidatas por región.	63
Cuadro 14: Resultados de la simulación para la curva de crecimiento regional.....	65
Cuadro 15: Resumen de las distribuciones más robustas para cada región	66
Cuadro 16: Resultado de cuantiles para la curva de crecimiento regional de cada región.	67
Cuadro 17: Índice de avenida de las estaciones de la región	69
Cuadro 18: Parámetros del semivariograma empírico del modelo Gaussiano	72
Cuadro 19: Fórmulas para estimación de parámetros por el método de máxima verosimilitud para la distribución Gumbel.	83
Cuadro 20: Fórmulas por distribución, L-momentos y su ratios para las distribuciones analizadas.....	84

Cuadro 21: Coeficientes de las aproximaciones para GNO 86
Cuadro 22: Coeficientes de las aproximaciones para PE3 86

RELACIÓN DE FIGURAS

Figura 1: Mapa de ubicación de la zona de estudios (cuencas del río Chillón, Rímac, Lurín y parte alta de Mantaro) 32
Figura 2: Ubicación de las estaciones meteorológicas. 36
Figura 3: Estaciones con registro de precipitación máxima diaria en el tiempo. 37
Figura 4: Cantidad de registro de precipitación máxima diaria por estación. 37
Figura 5: Diagrama esquemático de la metodología. 40
Figura 6: Gráficos box plots para detección de outliers. 45
Figura 7: Superior: Gráficos box plots con datos atípicos, Inferior: Gráficos box plots de series libre de atípicos. 46
Figura 8: Superior: Bar plots de estaciones con longitud de registro mayor a 15 años. Inferior: Gráficos box plots de series libre de atípicos con longitud de registro mayor a 15 años. 47
Figura 9: Gráficos de series de tiempo para las estaciones con tendencia significativa a un nivel de confianza de 5%. 51
Figura 10: Resumen de autocorrelaciones significativas detectadas para varios retardos a un nivel de 5% por el test de Ljun Box. 52
Figura 11: Autocorrelaciones detectadas por el test de Ljun Box. 53
Figura 12: Ratios de los L – momentos de las estaciones. 56
Figura 13: Izquierda: Clasificación de estaciones mediante agrupación jerárquica y enlace de Ward. Centro: Clasificación de estaciones mediante K-means. Derecha: Regiones definitivos. 57
Figura 14: Diagrama de L-momentos ratio con el L-momentos ratio regional para las distribuciones candidatas. 61
Figura 15: Curva de crecimiento regional para las distribuciones candidatas 64
Figura 16: Representación gráfica de la curva de crecimiento regional 68
Figura 17: Histograma de frecuencias del índice de avenida. 70
Figura 18: Gráfica de nube de semivariancia de índice de avenida. 71
Figura 19: Gráfica de relación de índice de avenida versus elevación. 71

Figura 20: Gráfica de análisis estructural mediante semivariogramas: Ajuste de semivariograma cruzado por el modelo teórico Gaussiano.	72
Figura 21: Gráfico de cuantiles de índice de avenida observada y estimada en mm	73
Figura 22: Mapa de índice de avenida.	74
Figura 23: Mapa de incertidumbre del índice de avenida.	75
Figura 24: Diagrama de uso práctico del estudio.	77
Figura 25: Cuantiles locales para las estaciones de la región 1.	87
Figura 26: Cuantiles locales para las estaciones de la región 2.	88
Figura 27: Cuantiles locales para las estaciones de la región 3.	89
Figura 28: Cuantiles locales para las estaciones de la región 4.	91
Figura 29: Cuantiles locales para las estaciones de la región 5.	91

RESUMEN

La Dirección General de Hidrología y Recursos Hídricos del SENAMHI viene desarrollando investigación en el análisis y caracterización de Eventos Extremos, vista desde una perspectiva regional basados en Análisis Regional de Frecuencias Hidrológicas con L-Moments. En ese contexto se ha programado en el Plan Operativo del presente año POI-2014 el trabajo de investigación denominado “Análisis regional de Precipitaciones Máximas en las cuencas de los ríos Chillón, Rímac, Lurín y Alto Mantaro.

Las precipitaciones extremas suelen dar lugar a la ocurrencia de eventos de inundación con las consiguientes pérdidas de vidas, la producción agrícola y la infraestructura en las Cuencas de los ríos Chillón, Rímac, Lurín, parte Alta de Mantaro y cuencas vecinas. 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 este trabajo, 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. Inicialmente se realizó 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 atípicos, tendencia e independencia serial. El método de L-momentos a través de la medida de heterogeneidad, permite definir objetivamente que en la región de estudio se identifica 5 regiones hidrológicamente homogéneas. Para cada región homogénea se ha determinado la función de distribución de probabilidad más robusta para estimar los cuantiles de diseño en grandes periodos de retorno (>50 años), siendo para la región 1 y 2 la distribución normal generalizado (GNO), región 3 (Logístico generalizado - GLO), región 4 y 5 (Pearson tipo 3). 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 técnica de interpolación geoestadística Co-Kriging. Válido para estimar el valor del índice de avenida en sitios sin medición.

I. INTRODUCCIÓN

1.1 Análisis de frecuencias de precipitaciones máximas

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.

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, 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.

Sin embargo 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 este estudio se desarrolla uno de ellos; denominado el 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.

1.2 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.

1.3 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 las Cuencas Chillón, Rímac, Lurín y parte Alta del Mantaro, 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.
- Identificar una apropiada función de distribución que relacione la magnitud de las precipitaciones máximas con su frecuencia de ocurrencia para cada región homogénea.
- Determinar el mapa de índice de avenida y su incertidumbre.

1.4 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 BIBLIOGRÁFICA

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%:

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

Límite para el intervalo de confianza al 5%:

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

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

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

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.

2.3 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).

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(X_j - X_i) \quad (5)$$

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

$$\text{sgn}(\theta) = \begin{cases} 1 & \text{si } \theta > 0 \\ 0 & \text{si } \theta = 0 \\ -1 & \text{si } \theta < 0 \end{cases} \quad (6)$$

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:

$$E(S) = 0 \quad (7)$$

$$V(S) = \frac{n(n-1)(2n+5) - \sum_{m=1}^n t_m m(m-1)(2m+5)}{18} \quad (8)$$

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

$$Z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & S > 0 \\ 0 & S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & S < 0 \end{cases} \quad (9)$$

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:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z e^{-t^2/2} dt \quad (10)$$

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 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:

$$Q(m) = N(N+2) \sum_{h=1}^m \frac{\rho_h^2}{N-h} \quad (11)$$

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:

$$Q(m) > X_{1-\alpha, m}^2 \quad (12)$$

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 el presente estudio 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. El análisis regional explota esta homogeneidad para producir estimaciones las cuales son, en la mayoría de los casos, más confiables que aquellas 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 :

$$F(x) = \Pr[Q \leq x] \quad (13)$$

$F(x)$ 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:

$$Q_T = x(1 - 1/T) \quad (14)$$

$$F(Q_T) = 1 - 1/T \quad (15)$$

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 este estudio 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:

$$b_r = n^{-1} \binom{n-1}{r}^{-1} \sum_{j=r+1}^n \binom{j-1}{r} x_{j:n} \quad (16)$$

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

Desarrollando los primeros MPP se tiene:

$$b_0 = \frac{1}{n} \sum_{j=1}^n x_{j:n} \quad (17)$$

$$b_1 = \frac{1}{n} \sum_{j=2}^n \frac{(j-1)}{(n-1)} x_{j:n} \quad (18)$$

$$b_2 = \frac{1}{n} \sum_{j=2}^n \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{j:n} \quad (19)$$

$$b_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n} \quad (20)$$

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

$$l_1 = b_0 \quad (21)$$

$$l_2 = 2b_1 - b_0 \quad (22)$$

$$l_3 = 6b_2 - 6b_1 + b_0 \quad (23)$$

$$l_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \quad (24)$$

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_1 :

$$t = l_2 / l_1 = L - C_v, \quad \text{donde} \quad 0 \leq t \leq 1 \quad (25)$$

$$t_3 = l_3 / l_2 = L - C_s, \quad \text{donde} \quad -1 \leq t_3 \leq 1 \quad (26)$$

$$t_4 = l_4 / l_2 = L - C_k, \quad \text{donde} \quad -1 \leq t_4 \leq 1 \quad (27)$$

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 (t , t_3 y t_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:

$$Q_i(F) = m_i q(F), \quad i=1, \dots, N. \quad (28)$$

Donde m_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} / m_i .

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

Los datos adimensionales se reajustan a $q_{ij} = Q_{ij} / \hat{m}_i, j=1, \dots, n_i; i=1, \dots, N$, 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 $p: q_1, \dots, q_p$ por lo que se escribe $q(F)$ como $q(F; q_1, \dots, q_p)$. 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 q_k se denota por $\hat{q}_k^{(i)}$. 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:

$$\hat{q}_k^R = \frac{\sum_{i=1}^N n_i \hat{q}_k^{(i)}}{\sum_{i=1}^N n_i} \quad (29)$$

Sustituyendo los parámetros regionales estimados mediante la ecuación 35 en $q(F)$, se obtiene la curva de crecimiento regional $\hat{q}(F) = q(F; \hat{q}_1^R, \dots, \hat{q}_p^R)$. 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{m}_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:

$$\hat{Q}_i(F) = \hat{m}_i \hat{q}(F) \quad (30)$$

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:

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}). \quad (31)$$

Donde

$$u_i = \begin{pmatrix} t_1^{(i)} \\ t_2^{(i)} \\ t_3^{(i)} \\ t_4^{(i)} \end{pmatrix} \mathbf{u}^T \quad (32)$$

Vector que contiene los valores de los ratios L-momentos t_1, t_2, 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:

$$\bar{u} = N^{-1} \sum_{i=1}^N u_i \quad (33)$$

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

$$A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (34)$$

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 geográficos 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 se utiliza 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 se estandarizan 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:

$$f(x) = a^{-1} \frac{1 - k(x - x) / a}{u}^{1/k-1} \frac{1}{u} F(x) \frac{1-h}{u} \quad (35)$$

x es un parámetro de localización, a 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:

$$H_j = \frac{(V_j - m_{V_j})}{s_{V_j}} \quad (36)$$

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

v_1 , es la desviación estándar ponderada observada de los t valores:

$$V_1 = \left\{ \frac{\sum_{i=1}^N n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^N n_i} \right\}^{1/2} \quad (37)$$

v_2 , es la distancia media observada de t / t_3 :

$$V_2 = \frac{\sum_{i=1}^N n_i \left(\frac{t^{(i)} - t^R}{t_3} \right)^2 + \left(\frac{t_3^{(i)} - t_3^R}{t_3} \right)^2}{\sum_{i=1}^N n_i}^{1/2} \quad (38)$$

v_3 , es la distancia media observada de t_3 / t_4 :

$$V_3 = \frac{\sum_{i=1}^N n_i \left(\frac{t_3^{(i)} - t_3^R}{t_4} \right)^2 + \left(\frac{t_4^{(i)} - t_4^R}{t_4} \right)^2}{\sum_{i=1}^N n_i}^{1/2} \quad (39)$$

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:

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i} \quad (40)$$

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 (41)), los coeficientes A_k , se dan en el Cuadro 2.

$$L - C_k = \sum_{k=0}^8 A_k (L - C_s)^k \tag{41}$$

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:

$$Z^{DIST} = (t_4^{DIST} - t_4^R + B_4) / s_4. \tag{42}$$

Donde

t_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 (41).

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

B_4 y s_4 bias y desviación estándar de t_4^R , respectivamente, definido como:

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} (t_4^{[m]} - t_4^R), \tag{43}$$

$$s_4 = \left[\sum_{m=1}^{N_{sim}} (N_{sim} - 1)^{-1} \sum_{m=1}^{N_{sim}} (t_4^{[m]} - t_4^R)^2 - N_{sim} (B_4)^2 \right]^{1/2}, \tag{44}$$

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:

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i}, \quad (45)$$

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

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 t_1, t_3, t_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 (45) y

(46). 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 m_i y $q(\cdot)$. Así, el cuantil para la probabilidad de no excedencia F es:

$$\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F). \quad (47)$$

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 $\frac{\hat{Q}_i^{[m]}(F) - Q_i(F)}{Q_i(F)}$, 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:

$$B_i(F) = M^{-1} \sum_{m=1}^M \frac{\{\hat{Q}_i^{[m]}(F) - Q_i(F)\}}{Q_i(F)} \quad (48)$$

$$R_i(F) = \left(\sum_{m=1}^M \frac{\{\hat{Q}_i^{[m]}(F) - Q_i(F)\}^2}{Q_i(F)} \right)^{1/2} \quad (49)$$

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:

$$B^R(F) = N^{-1} \sum_{i=1}^N B_i(F). \quad (50)$$

$$A^R(F) = N^{-1} \sum_{i=1}^N |B_i(F)|. \quad (51)$$

$$R^R(F) = N^{-1} \sum_{i=1}^N R_i(F). \quad (52)$$

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:

$$Q_i(F) = m_i q_i(F). \quad (53)$$

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 (49) a (52), 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:

$$L_{.05}(F) \leq \frac{\hat{Q}(F)}{Q(F)} \leq U_{.05}(F), \quad (54)$$

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

$$\frac{\hat{Q}(F)}{U_{.05}(F)} \leq Q(F) \leq \frac{\hat{Q}(F)}{L_{.05}(F)}. \quad (55)$$

La expresión (55) 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 (52)) constituye una medida de precisión más eficiente.

Hay que destacar, no obstante, que esta forma de simulación basada en los límites de la ecuación (55) 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} , $k=1, \dots, n_0$, $i=1, \dots, N$, 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

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 genera el mapa de índice de avenida utilizando la técnica de interpolación Co-Kriging para determinar de los cuantiles de diseño en sitios sin medición.

2.8.1 Método de Co-kriging (Co-Ko)

Co-kriging es una extensión de Kriging donde se pueden agregar más de una variable auxiliar para la predicción. En un modelo de Kriging convencional, se asume una respuesta a ser un proceso aleatorio espacial con función de covarianza estacionaria, lo que implica que la suavidad de una respuesta es bastante uniforme en cada región de la zona de dominio (Paciorek, 2003).

Sin embargo, los casos son comunes donde el nivel de suavidad de una respuesta podría cambiar considerablemente debido a las características geofísicas (Xiong et al., 2007). En tales situaciones, co-Kriging es una técnica utilizada regularmente donde interpolaciones se mejoran mediante la adición de atributos secundarios como por ejemplo longitud, latitud y elevación; que pueden conducir a mejorar la distribución espacial de la variable analizada, (Stein et al., 1991). Co-Kriging es más eficaz cuando las covariables están altamente correlacionados (Nalder y Wein, 1998).

Algoritmos de estimación Kriging y co-kriging

Si, $z^\alpha(s_i)$ y $z^\beta(s_i)$ representan realizaciones de las variables aleatorias $Z^\alpha(s_i)$ y $Z^\beta(s_i)$ en los puntos particulares s_i dentro de un campo de S . Es obvio que el co-kriging estimador es el mejor estimador lineal insesgado, así: El co-kriging estimador de $Z^*(s_0)$ es una combinación lineal de los valores de la muestra $Z^\alpha(s_i)$ y $Z^\beta(s_i)$:

$$Z_\alpha^*(s_0) = \sum_{i=1}^{n_\alpha} \lambda_{\alpha i} Z_\alpha(s_i) + \sum_{j=1}^{n_\beta} \lambda_{\beta j} Z_\beta(s_j) \quad (56)$$

Donde $Z_\alpha^*(s_0)$ es la estimación de Z_α en el punto s_0 , n_α y n_β son los números de puntos de datos de Z_α y Z_β utilizados en la estimación, y $\lambda_{\alpha i}$ y $\lambda_{\beta j}$ son las ponderaciones asociadas. Para kriging ordinario $Z_\alpha(s_i)$ representa los valores de la precipitación media mensual histórica en el punto de muestra s_i y las ponderaciones $\lambda_{\beta j}$ con ceros, ya que solo la precipitación media mensual contribuye al proceso de estimación.

Para co-kriging, $Z_\alpha(s_i)$ y $Z_\beta(s_j)$ representan los valores de la precipitación media mensual histórica y la elevación en los puntos de muestra s_i y s_j respectivamente.

Los pesos en la ecuación (2) son determinados por la minimización de la estimación de la varianza.

$$\text{var} [Z_\alpha^*(s_0) - Z_\alpha(s_0)] \quad (57)$$

Sujeto a la restricción que la estimación debe ser insesgada:

$$E[Z_{\alpha}^*(s_0) - Z_{\alpha}(s_0)] = 0 \quad (58)$$

Esto produce el sistema Kriging de ecuaciones:

$$\sum_{i=1}^{n_{\alpha}} \lambda_{\alpha i} \gamma_{\alpha\alpha}(s_m, s_i) + \mu = \gamma_{\alpha\alpha}(s_m, s_0) \quad m = 1, \dots, n_{\alpha} \quad (59)$$

$$\sum_{i=1}^{n_{\alpha}} \lambda_{\alpha i} = 1 \quad (60)$$

Y el sistema co-kriging de ecuaciones:

$$\sum_{i=1}^{n_{\alpha}} \lambda_{\alpha i} \gamma_{\alpha\alpha}(s_m, s_i) + \sum_{j=1}^{n_{\beta}} \lambda_{\beta j} \gamma_{\alpha\beta}(s_m, s_j) + \mu_{\alpha} = \gamma_{\alpha\alpha}(s_m, s_0) \quad m = 1, \dots, n_{\alpha} \quad (61)$$

$$\sum_{i=1}^{n_{\alpha}} \lambda_{\alpha i} \gamma_{\alpha\beta}(s_m, s_i) + \sum_{j=1}^{n_{\beta}} \lambda_{\beta j} \gamma_{\beta\beta}(s_m, s_j) + \mu_{\beta} = \gamma_{\alpha\beta}(s_m, s_0) \quad m = 1, \dots, n_{\beta} \quad (62)$$

$$\sum_{i=1}^{n_{\alpha}} \lambda_{\alpha i} = 1 \quad (63)$$

$$\sum_{j=1}^{n_{\beta}} \lambda_{\beta j} = 1 \quad (64)$$

Donde $\gamma_{\alpha\beta}$ es la semivarianza de Z_{α} (cuando $\alpha = \beta$) o semivarianza cruzada de Z_{α} y Z_{β} (cuando $\alpha \neq \beta$) en una distancia de separación h . El μ , μ_{α} y μ_{β} son los valores multiplicadores de Lagrange.

Al derivar el sistema de co-kriging, hay una restricción adicional de insesgamiento, por lo tanto, se requiere la última ecuación en el sistema. Resolviendo este sistema de ecuaciones para los pesos λ y los multiplicadores de Lagrange μ permite calcular el valor de los puntos estimado $Z_{\alpha}^*(s_0)$ por la ecuación (2), y la estimación de la varianza:

$$\sigma^2(s_0) = \mu_{\alpha} + \sum_{i=1}^{n_{\alpha}} \lambda_{\alpha i} \gamma_{\alpha\alpha}(s_i, s_0) + \sum_{j=1}^{n_{\beta}} \lambda_{\beta j} \gamma_{\alpha\beta}(s_j, s_0) \quad (65)$$

a. Análisis Estructural

El análisis estructural es una de las etapas fundamentales de todo estudio geoestadístico. Tiene como objetivo la caracterización de la estructura espacial del fenómeno estudiado.

El enfoque geoestadístico se basa en la teoría de las variables regionalizadas (Matheron 1970). Se supone que las muestras espaciales se consideran como la realización de un proceso espacial aleatorio. Esto permite el uso de un poderoso instrumento estadístico para la estimación espacial: la semivariograma (Feki y Slimani 2006).

Si, $z^\alpha(s_i)$ y $z^\beta(s_i)$ representan realizaciones de las variables aleatorias $Z^\alpha(s_i)$ y $Z^\beta(s_i)$ en los puntos particulares s_i dentro de un campo de S . La hipótesis intrínseca (Chauvet 1999) asume que, para una variable aleatoria $Z(s_i)$: (i) el valor esperado de $Z(s_i)$ no depende de la posición s_i , y (ii) la varianza de $[Z(s_i) - Z(s_i + h)]$ no depende de la posición s_i en S para cualquier vector de separación h .

Entonces la función semivariograma da una medida de la correlación espacial de una variable aleatoria o las variables, como una función de la distancia de separación. Muestra de semivariogramas y variogramas cruzados fueron estimados por la función:

$$\gamma_{\alpha\beta}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [Z^\alpha(s_i + h) - Z^\alpha(s_i)][Z^\beta(s_i + h) - Z^\beta(s_i)]$$

Donde $\gamma_{\alpha\beta}$ es la semivarianza de Z^α (cuando $\alpha = \beta$) o semivarianza cruzada de Z_α y Z_β (cuando $\alpha \neq \beta$) en la distancia de separación h ; y $n(h)$ es el número de pares de puntos en un intervalo de distancia $(h + \Delta h)$.

El variograma cruzado debe ser parte de un modelo lineal de co-regionalización (LMC). Un LMC requiere que cada estructura en el variograma cruzado, se incluye en el modelo de continuidad espacial para las variables primarias y secundarias. Un modelo autorizado matemático puede ser entonces ajustado al variograma experimental y los coeficientes de este modelo se pueden utilizar para Kriging.

b. Validación modelo

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. 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 METODOLOGÍA

3.1 Descripción del área de estudio

La zona de estudio corresponde al ámbito geográfico de las cuencas de los ríos Rímac, Chillón, Lurín y parte Alta del Mantaro (Figura 1).

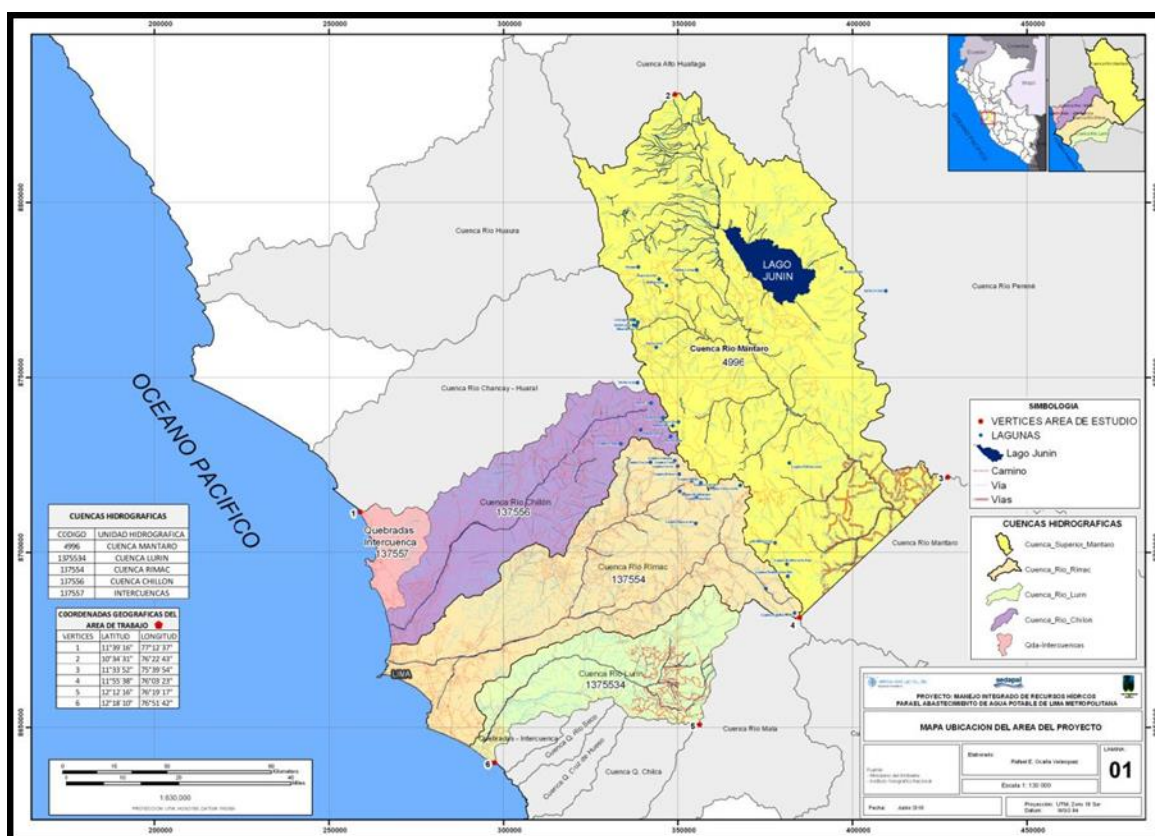


Figura 1: Mapa de ubicación de la zona de estudios (cuencas del río Chillón, Rímac, Lurín y parte alta de Mantaro)

Lagos et. al., (2009) menciona que en los Andes Peruanos, Existe una gran variabilidad estacional e interanual de las precipitaciones en toda la región andina. el periodo de lluvias coincide con la estación de primavera y el verano del hemisferio sur. Se inicia en setiembre y finaliza en abril. El inicio y la intensidad de las lluvias varía de un año a otro y en algunos años la variabilidad interanual puede estar asociada con eventos extremos, tales como lluvias excesivas o déficit de lluvias, ocasionando desastres.

Existen tres subregiones en la región andina con características climáticas ligeramente distintas, las subregiones norte, centro y sur. El comportamiento de las lluvias varían de subregión en subregión, la máxima precipitación del ciclo anual ocurre en enero en la subregión sur, en febrero en la subregión centro y en marzo en la subregión norte.

3.2 Materiales

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

3.2.1 Información básica de sistema de información geográfica

Para identificar el área de influencia se ha recopilado inicialmente la información de sistemas hidrográficos de la cuenca del Perú 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

Se han identificado en total 84 estaciones administradas por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), emplazadas en las cuencas de los ríos Chillón, Rímac, Lurín y Alto Mantaro. En el **Cuadro 4** se presenta la red de estaciones, sus características de ubicación y elevación, así como la cuenca en la que se ubican.

Para las 84 estaciones identificadas dentro y circundantes a la zona de estudio, se han recopilado los registros de precipitación diaria para luego conformar las series de precipitación máxima diaria que serán útiles en el estudio de análisis regional de frecuencias de las precipitaciones máximas. La distribución espacial de las estaciones y el modelo digital de elevación empleada se muestra en la Figura 2.

Cuadro 4: Red de estaciones Meteorológicas

N°	Id_Estación	Nombre	Cuenca	Longitud (°)	Latitud (°)	Altitud (msnm)
1	140500	Aeropuerto internacional	Rimac	-77.11667	-12	13
2	151204	Arahuay	Chillon	-76.7	-11.61667	2800
3	151205	Canchacalla	Rimac	-76.53111	-11.84472	2554
4	151207	Huancata	Mala	-76.21667	-12.21667	2700
5	151208	Gorgor	Pativilca	-77.03333	-10.56667	3070
6	151209	Chosica	Rimac	-76.73333	-11.93333	850
7	151210	Rio blanco	Rimac	-76.25889	-11.73444	3550
8	151211	Ambar	Supé	-77.28333	-10.75	2100
9	151212	San pedro de pilas	Omas	-76.21667	-12.45	2600
10	151213	Sheque	Rimac	-76.49861	-11.66639	3214
11	151214	Langa	Lurin	-76.4	-12.1	2860
12	155107	Pampa libre	Huaura	-76.96667	-10.86667	1800
13	155111	Tingo	Rimac	-76.48333	-11.61667	4200
14	155112	Pariacancha	Chillon	-76.5	-11.38333	3800
15	155113	Mina colqui	Rimac	-76.48333	-11.58333	4600
16	155115	Carhuacayan	Mantaro	-76.28333	-11.2	4150
17	155116	Alpamarca	Mantaro	-76.45	-11.21667	4637
18	155117	La quisha	Rimac	-76.48333	-11.61667	4200
19	155121	Yantac	Mantaro	-76.4	-11.33333	4600
20	155122	Autisha	Rimac	-76.60639	-11.735	2171
21	155200	Paccho	Huaura	-76.93333	-10.95	3250
22	155201	Andajes	Huaura	-76.9	-10.78333	3950
23	155202	Santa cruz	Huaura	-76.63333	-11.2	3700
24	155203	Carac	Chancay-lima	-76.78333	-11.18333	2600
25	155204	Pachangara	Huaura	-76.81667	-10.78333	3600
26	155205	Pallac	Chancay-huaral	-76.8	-11.35	2333
27	155206	Laguna surasaca	Huaura	-76.78333	-10.51667	4400
28	155207	Pachamachay	Huaura	-76.83333	-11.05	4200
29	155209	Huamantanga	Chillon	-76.75	-11.5	3392
30	155212	Parquin	Huaura	-76.71667	-10.96667	3590
31	155213	Santa eulalia	Rimac	-76.66667	-11.91778	982
32	155214	Pirca	Chancay-huaral	-76.65	-11.23333	3255
33	155217	Lachaqui	Chillon	-76.61667	-11.55	3668
34	155218	Huaros	Chillon	-76.56667	-11.4	3585
35	155219	Tupe	Huaura	-76.65	-11	4450
36	155223	Carampoma	Rimac	-76.51528	-11.655	3489
37	155224	Santiago de tuna	Lurin	-76.51667	-11.98333	2921
38	155225	San jose de parac	Rimac	-76.25806	-11.80028	3866
39	155228	San pedro de chuclu	Mantaro	-75.5	-11.75	3380
40	155235	Yauli	Mantaro	-76.08333	-11.66667	4141
41	155291	Laguna cochaquillo	Huaura	-76.66667	-10.78333	4400
42	155446	Casapalca	Rimac	-76.23333	-11.64778	4214
43	155450	Yauricocha	Cañete	-75.7225	-12.31639	4675
44	155514	Milloc	Rimac	-76.35	-11.57111	4398
45	156100	Antioquia	Lurin	-76.5	-12.08333	1839
46	156102	San lazaro de escomarca	Lurin	-76.35	-12.18333	3600
47	156103	Huañec	Mala	-76.13333	-12.28333	3205
48	156104	Ayaviri	Mala	-76.13333	-12.38333	3228
49	156106	Tanta	Cañete	-76.01667	-12.11667	4323
50	156109	Carania	Cañete	-75.87194	-12.34417	3875

51	156110	Huangascar	Cañete	-75.83361	-12.89833	2533
52	156111	Vilca	Cañete	-75.82611	-12.11444	3864
53	475	Yanahuanca	Alto huallaga	-76.51389	-10.49056	3473
54	501	Alcantarilla	Huaura	-77.55	-11.05	120
55	502	Surasaca	Rimac	-76.78333	-10.51667	4400
56	503	Jauja	Mantaro	-75.47444	-11.78361	3322
57	531	Isla don martin	Huaura	-77.66667	-11.01667	8
58	532	Camay	Huaura	-77.64889	-10.91278	65
59	534	Lomas de lachay	Interc. Del pacifico	-77.36667	-11.36667	300
60	535	Andahuasi	Huaura	-77.23333	-11.13333	470
61	536	Santa rosa	Huaura	-77.38333	-11.21667	485
62	539	Huayan	Chancay-huaral	-77.11667	-11.45	350
63	540	Cajatambo	Pativilca	-76.98333	-10.46667	3350
64	541	Oyon	Huaura	-76.76667	-10.66667	3641
65	542	Picoy	Huaura	-76.71667	-10.88333	2900
66	543	Ñaña	Rimac	-76.83861	-11.98833	566
67	546	Donoso	Chancay-huaral	-77.23333	-11.46667	180
68	547	Canta	Chillon	-76.62583	-11.47111	2832
69	548	Matucana	Rimac	-76.37778	-11.83889	2479
70	549	Marcapomacocha	Mantaro	-76.325	-11.40444	4479
71	550	Milpo	Mantaro	-76.21667	-10.6	4100
72	554	Tarma	Perene	-75.69139	-11.39667	3000
73	555	Huasahuasi	Perene	-75.63194	-11.26944	2750
74	569	Atacocha	Huallaga	-76.21667	-10.56667	4100
75	593	Cerro de pasco	Mantaro	-76.26417	-10.69333	4260
76	604	La oroya	Mantaro	-75.9575	-11.57167	3860
77	616	Cañete	Cañete	-76.33028	-13.07472	158
78	617	Modelo campo de marte	Rimac	-77.03333	-12.08333	110
79	618	Huarangal	Chillon	-77.1	-11.78333	410
80	629	Oroya mayupampa	Mantaro	-75.9	-11.51667	3750
81	631	Calango (la capilla ii)	Mala	-76.49306	-12.52167	442
82	633	Huaro-chiri	Mala	-76.23333	-12.13333	3154
83	636	Yauyos	Cañete	-75.90833	-12.49167	2327
84	638	Pacaran	Cañete	-76.06667	-12.83333	700

Cabe señalar que en los posteriores cálculos en este estudio la identificación de las estaciones se realiza en función al Id. De Estación.

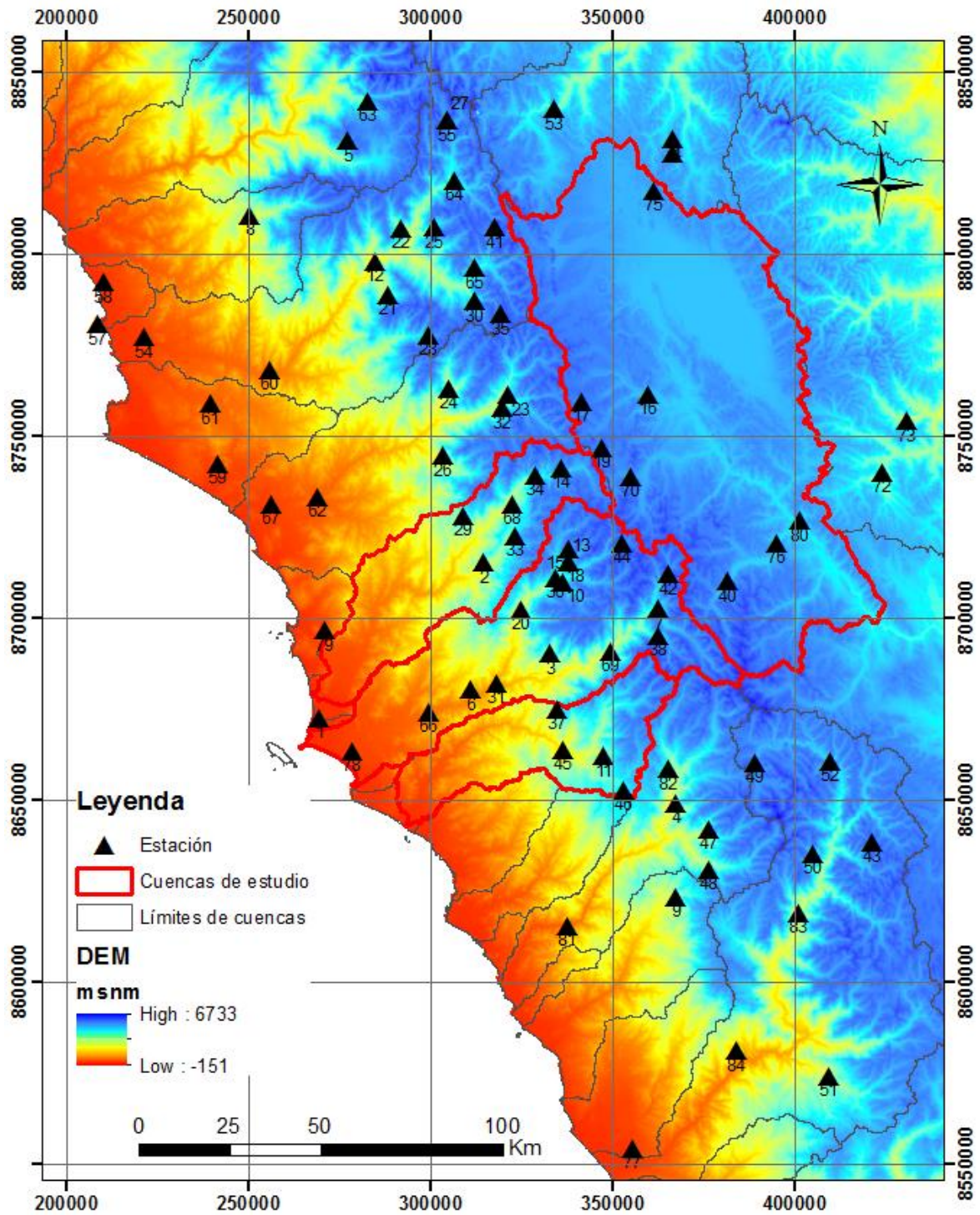


Figura 2: Ubicación de las estaciones meteorológicas.

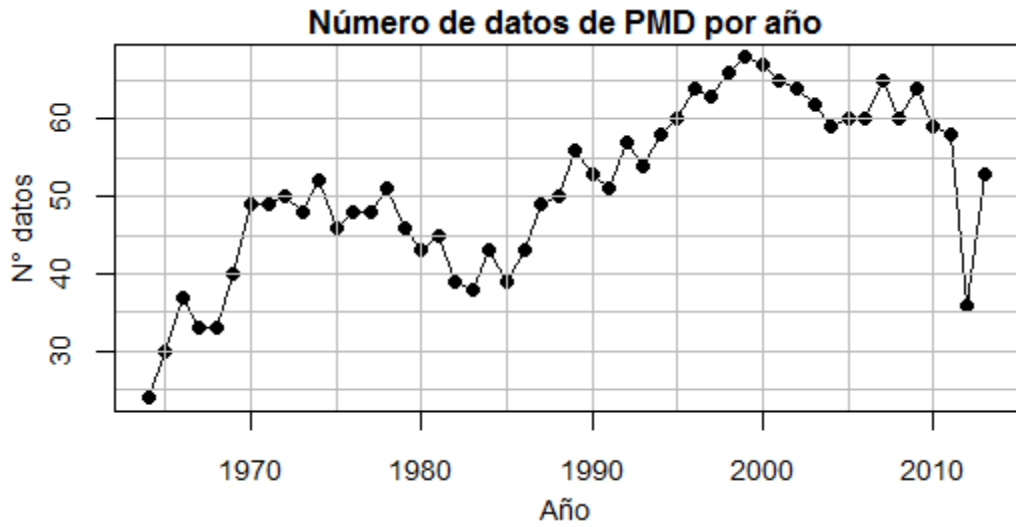


Figura 3: Estaciones con registro de precipitación máxima diaria en el tiempo.

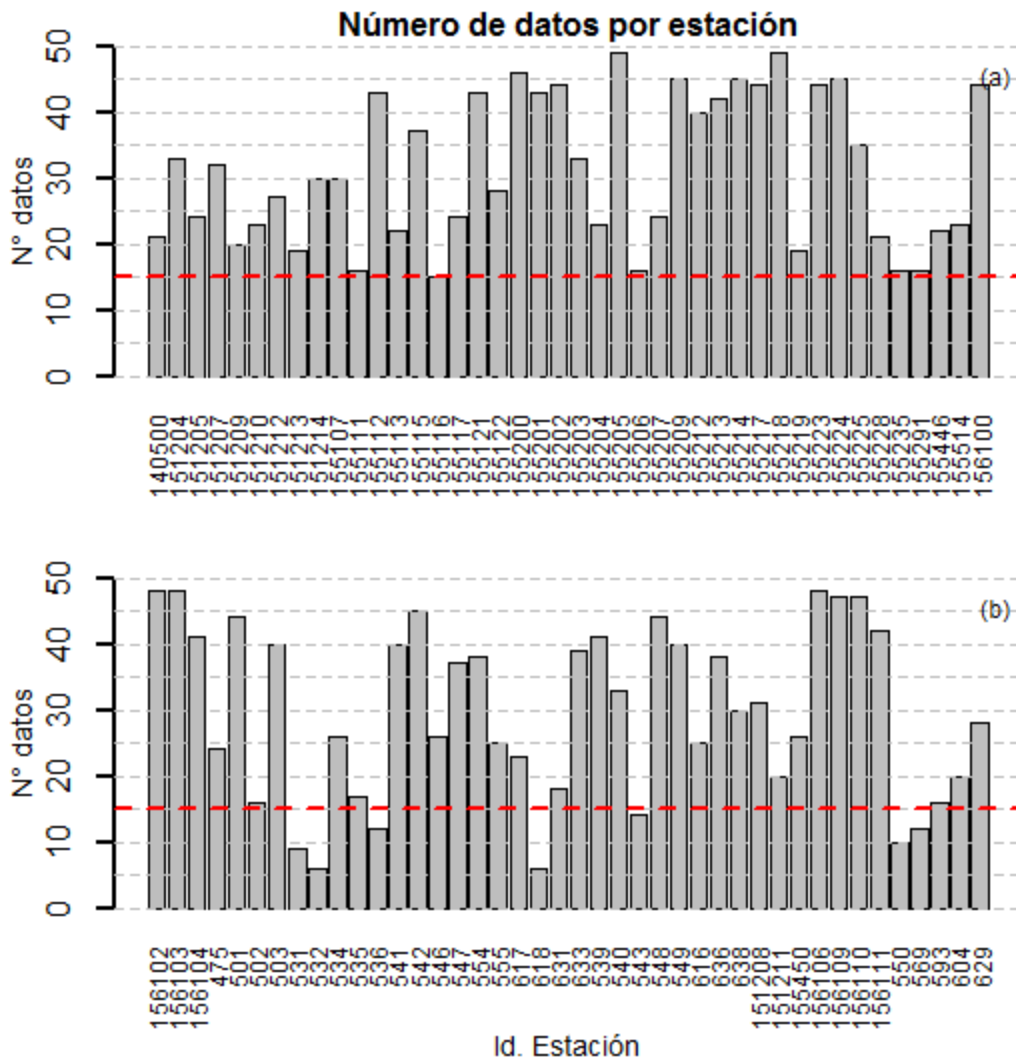


Figura 4: Cantidad de registro de precipitación máxima diaria por estación.

En la Figura 3 se describe la cantidad de estaciones que presentan registro de precipitación máxima diaria en el tiempo. Este gráfico muestra la cantidad de información disponible en el tiempo donde a partir de los décadas de 1990 se incrementa la cantidad de información. También se observa que no existe más de 2 años donde se disponga de información en todas las estaciones.

En la figura se presenta la cantidad de información disponible por cada estación donde se puede apreciar la longitud de registro variable por cada estación, por lo que 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). Sin embargo en este estudio por tratarse de un análisis regional de frecuencias de precipitaciones máximas se considera como mínimo 15 años; la cual deberán cumplir las estaciones después del análisis de datos atípicos.

3.2.3 Programas

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

- **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:

*zyp*¹, El paquete *zyp* contiene una implementación eficiente del método de la pendiente de Sen (Sen, 1968), además de la implementación de análisis de tendencias en los datos climáticos de Mann Kendall modificado Yue-Pilon de (Yue, 2002).

*FitAR*², Se ha empleado para probar el test de Ljun Box en el análisis de autocorrelaciones de series de tiempo.

*lmomRFA*³, funciones para el análisis regional de frecuencia utilizando los métodos de Hosking y Wallis (1997).

*lmom*⁴, 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.

*Gstat*⁵, 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 que se muestra en la Figura 5. 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, 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.

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

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

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

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

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

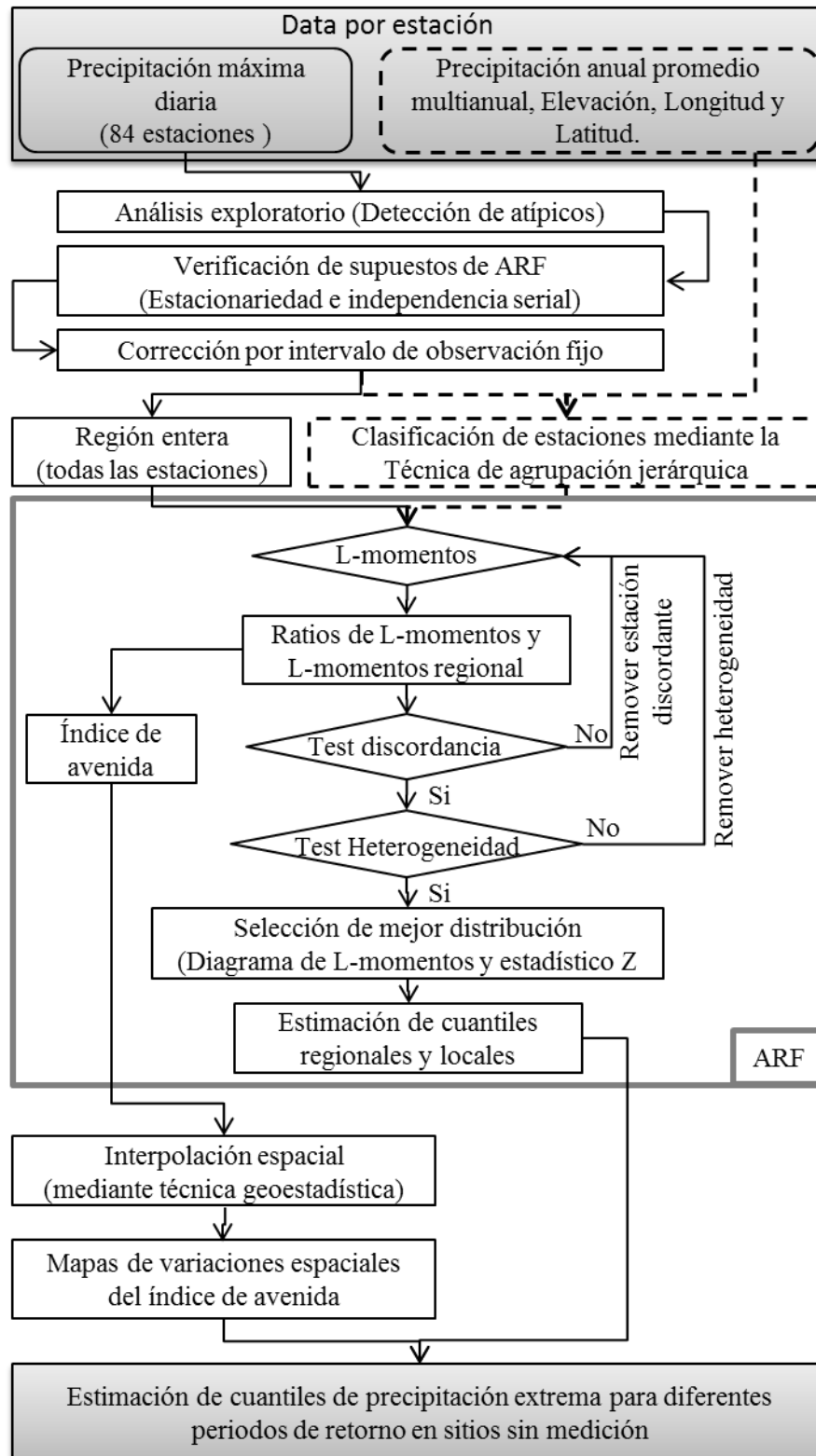


Figura 5: Diagrama esquemático de la metodología.

3.3.1 Análisis exploratorio de datos

Esta etapa consiste en la detección de datos atípicos (outliers) de manera cualitativa, mediante los gráficos de Box plots. Los posibles outliers que serán detectados con los gráficos de Box plots, se contrasta con el comportamiento de los eventos de las estaciones contiguas; de aquí se pone en juicio los datos atípicos detectados para ser excluidos o incluidos en la siguiente etapa de análisis.

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, 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. 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 dos 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_{OIFF}), debido que la medida y reporte de las observaciones analizadas son realizados en intervalos de tiempo fijo. El factor de corrección C_{OIFF} 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 $C_{OIFF} = 1.13$; 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 y no se cumpla la condición de homogeneidad como una región única 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.

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 ImomRFA 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 ImomRFA.

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 ImomRFA 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 1000 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 Mapeo de índice de avenida

Se obtendrán mapas de regionalización espacial del índice de avenida empleando la técnica de interpolación Co-Kriging. La performance de la técnica se verifica con una validación cruzada y medidas de precisión.

El paquete gstat se ha empleado para el mapeo, tanto como para el modelado de variogramas, predicción y validación cruzada. Cabe señalar que todos los paquetes mencionados para su uso fueron codificados en el programa R.

IV. RESULTADOS Y DISCUSIÓN

4.1 Análisis exploratorio de datos

En esta etapa se ha realizado los gráficos de Box plots, para detectar las estaciones con registros dudosos y los datos atípicos (outliers) de manera cualitativa.

4.1.1 Diagramas de Boxplots

En la Figura 6, 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 dos estaciones (Id: 155219 y 155117) que presentan registros dudosos por lo que fueron excluidas en este estudio.

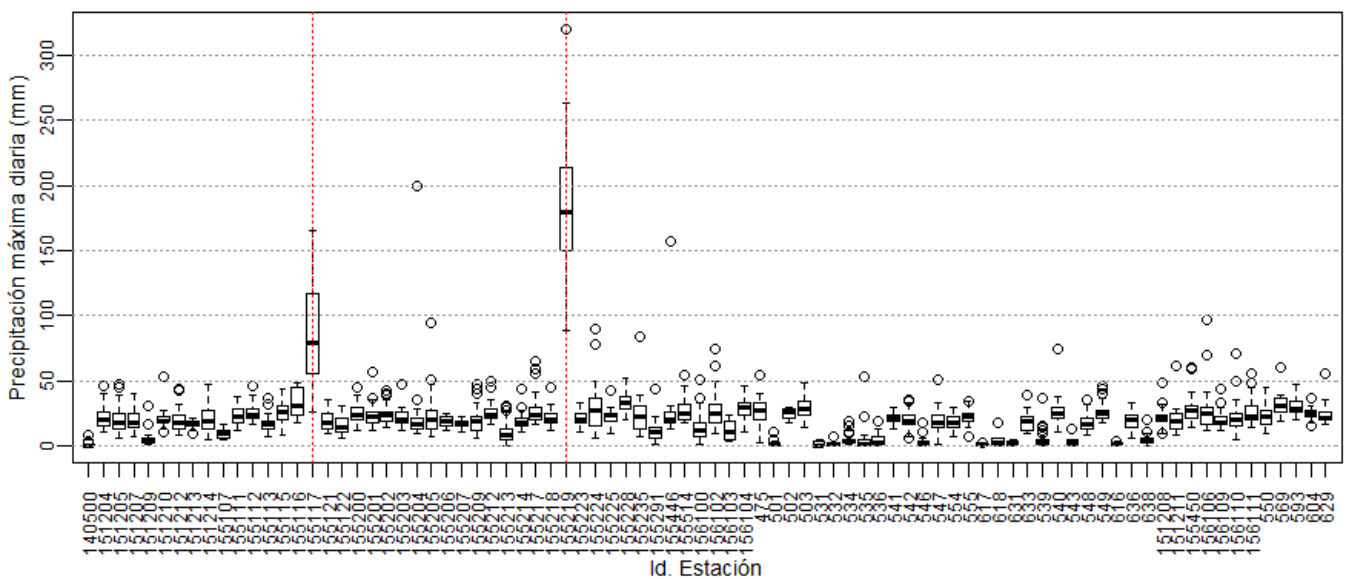


Figura 6: Gráficos box plots para detección de outliers.

Excluidas las estaciones se realiza el análisis de datos atípicos como se muestra en la siguiente figura, donde se ha detectado en total 20 datos atípicos las mismas fueron excluidas por superar evidentemente el rango intercuartil según los diagramas de box plots.

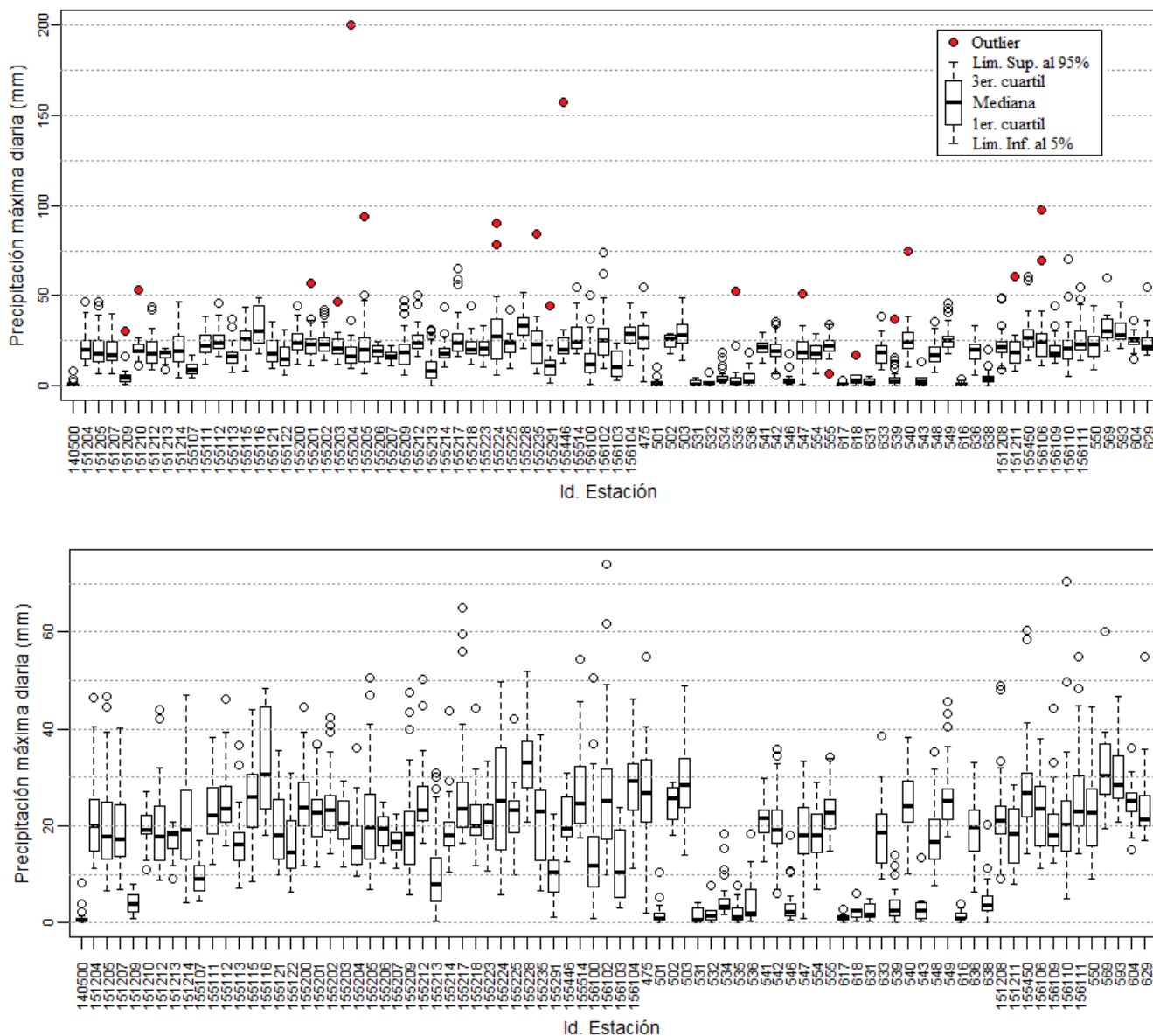


Figura 7: Superior: Gráficos box plots con datos atípicos, Inferior: Gráficos box plots de series libre de atípicos.

Las series de PMD libres de datos atípicos son filtradas de tal manera que cumplan con un mínimo de 15 años de registro.

A continuación se muestra los gráficos de las estaciones que fueron seleccionadas para el análisis de supuestos, en total se tiene 71 estaciones con registros mayores que 15 años de registro.

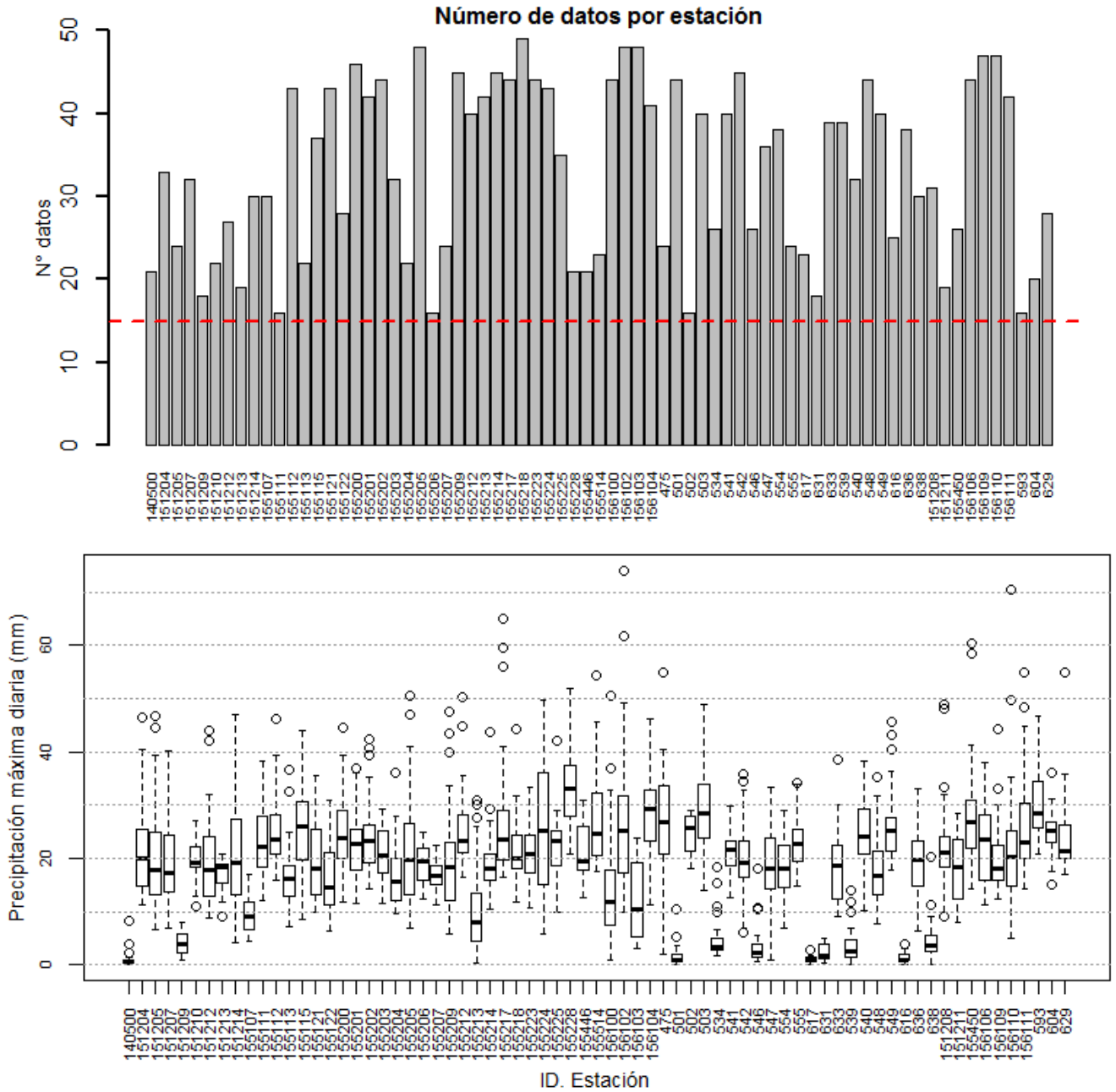


Figura 8: Superior: Bar plots de estaciones con longitud de registro mayor a 15 años. Inferior: Gráficos box plots de series libre de atípicos con longitud de registro mayor a 15 años.

4.2 Supuestos del análisis regional de frecuencias

Se analiza principalmente dos supuestos del análisis regional de frecuencias: estacionariedad (libre de tendencias) e independencia serial de los datos.

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 5 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

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

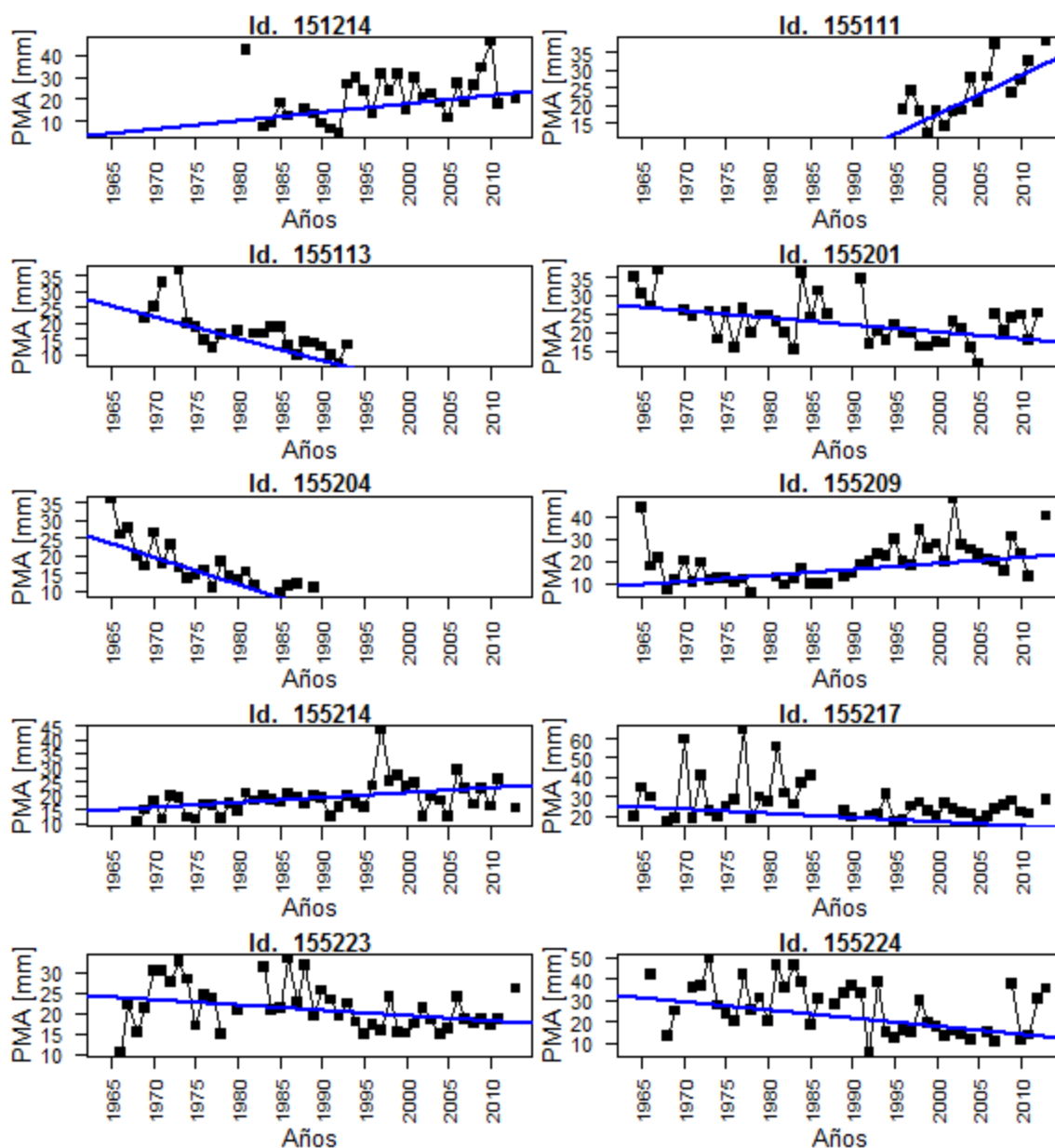
Cuadro 5: Resumen del test de Mann Kendall para el análisis de tendencias

Estación	Tau Kendall	p-value	Pendiente Sen	Intercepto	Hipótesis Nula
140500	-0.276	0.166	-0.044	0.604	Aceptado
151204	0.054	0.683	-0.063	19.952	Aceptado
151205	0.238	0.139	0.534	-0.147	Aceptado
151207	-0.010	0.955	0.112	12.735	Aceptado
151209	0.121	0.584	0.072	1.250	Aceptado
151210	-0.029	0.889	0.004	18.250	Aceptado
151212	0.073	0.624	0.189	15.033	Aceptado
151213	0.150	0.444	0.038	16.640	Aceptado
151214	0.292	0.038	0.385	4.147	Rechazado
155107	0.133	0.362	-0.031	9.396	Aceptado
155111	0.564	0.009	1.109	-23.813	Rechazado
155112	-0.200	0.067	-0.169	28.089	Aceptado
155113	-0.582	0.001	-0.684	26.500	Rechazado
155115	0.049	0.708	0.026	24.079	Aceptado
155121	-0.193	0.078	-0.096	20.420	Aceptado
155122	0.247	0.088	0.297	4.935	Aceptado
155200	0.189	0.075	0.094	21.759	Aceptado
155201	-0.264	0.022	-0.192	27.327	Rechazado
155202	-0.004	0.981	-0.024	21.455	Aceptado

155203	-0.169	0.213	-0.063	22.797	Aceptado
155204	-0.637	0.000	-0.766	25.150	Rechazado
155205	-0.065	0.538	0.026	19.629	Aceptado
155206	0.029	0.921	0.170	10.586	Aceptado
155207	0.286	0.075	0.132	11.110	Aceptado
155209	0.429	0.000	0.277	8.923	Rechazado
155212	-0.022	0.865	-0.038	23.850	Aceptado
155213	-0.183	0.114	-0.254	13.088	Aceptado
155214	0.300	0.005	0.171	14.657	Rechazado
155217	-0.263	0.019	-0.217	25.577	Rechazado
155218	-0.108	0.287	-0.094	22.403	Aceptado
155223	-0.315	0.004	-0.125	24.089	Rechazado
155224	-0.384	0.001	-0.373	31.779	Rechazado
155225	0.177	0.158	0.056	21.825	Aceptado
155228	-0.088	0.650	-0.329	39.500	Aceptado
155446	-0.298	0.080	-0.269	31.521	Aceptado
155514	0.132	0.484	-0.127	28.460	Aceptado
156100	-0.285	0.008	-0.268	16.572	Rechazado
156102	-0.057	0.591	-0.088	26.694	Aceptado
156103	0.557	0.000	0.282	3.992	Rechazado
156104	-0.038	0.754	0.060	30.056	Aceptado
475	0.392	0.021	0.158	11.900	Rechazado
501	-0.118	0.281	-0.027	1.188	Aceptado
502	-0.154	0.502	-0.146	27.310	Aceptado
503	0.069	0.570	0.039	26.965	Aceptado
534	0.116	0.496	0.033	2.901	Aceptado
541	0.009	0.953	0.001	21.441	Aceptado
542	-0.181	0.090	-0.099	21.714	Aceptado
546	-0.283	0.056	-0.162	6.942	Aceptado
547	0.076	0.568	0.054	16.317	Aceptado
554	0.326	0.008	0.037	15.092	Rechazado
555	0.064	0.726	0.126	17.832	Aceptado
617	-0.067	0.695	-0.005	1.000	Aceptado
631	0.256	0.246	0.032	1.496	Aceptado
633	0.481	0.000	0.290	12.204	Rechazado
539	-0.021	0.887	-0.044	2.987	Aceptado
540	-0.040	0.797	0.031	24.828	Aceptado
548	0.024	0.841	0.012	15.336	Aceptado
549	-0.156	0.191	-0.099	26.787	Aceptado
616	0.357	0.036	0.062	-0.163	Rechazado
636	-0.270	0.031	-0.108	21.517	Rechazado
638	0.056	0.735	-0.045	3.387	Aceptado
151208	-0.088	0.532	0.047	22.791	Aceptado
151211	-0.128	0.583	-0.098	19.656	Aceptado
155450	-0.109	0.472	-0.150	35.333	Aceptado
156106	0.217	0.053	0.101	19.784	Aceptado
156109	0.169	0.108	-0.053	18.680	Aceptado
156110	-0.044	0.678	-0.135	21.114	Aceptado
156111	-0.092	0.438	0.040	22.761	Aceptado
593	-0.086	0.692	-0.095	36.555	Aceptado
604	-0.029	0.889	0.077	22.075	Aceptado
629	0.336	0.027	0.402	19.166	Rechazado

De acuerdo al Cuadro 5, la hipótesis nula: No existe una tendencia en la serie fue rechazada en 18 estaciones, que representan el 25% de las series analizadas. Cabe señalar que la estacionariedad está influenciada por la cantidad de datos por lo que presenta cada estación, siendo en la mayoría de las estaciones menores a 30 años por lo que existe la posibilidad que algunas series presenten tendencia por su limitada longitud de registro.

A continuación se muestra las estaciones que presentan tendencia.



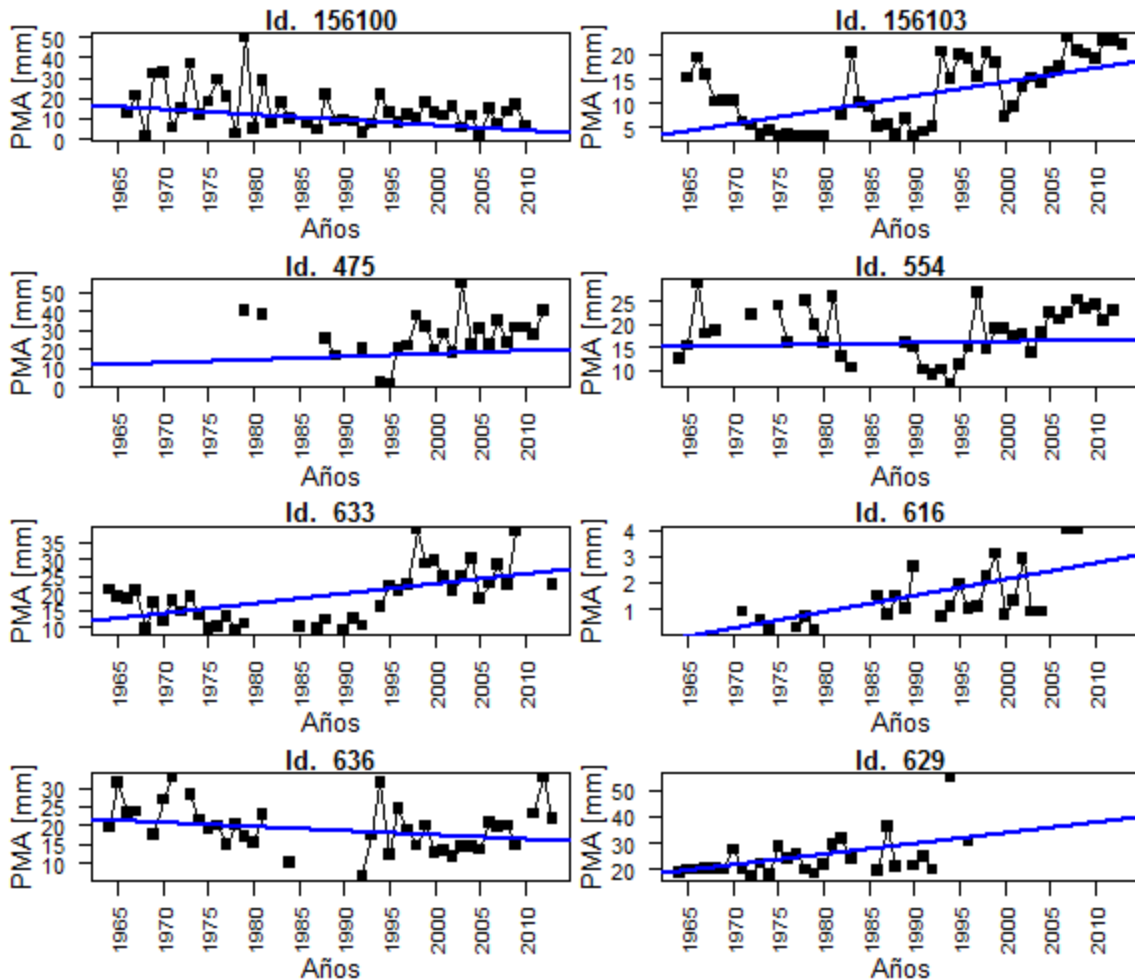


Figura 9: Gráficos de series de tiempo para las estaciones con tendencia significativa a un nivel de confianza de 5%.

4.2.2 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 para todas las estaciones.

En la Figura 10, se aprecia el resumen de la prueba de Ljun y Box donde muestra las estaciones con y sin autocorrelación significativa y en la Figura 11 se muestra los correlogramas para las estaciones en las cuales se detectó la autocorrelación significativa.

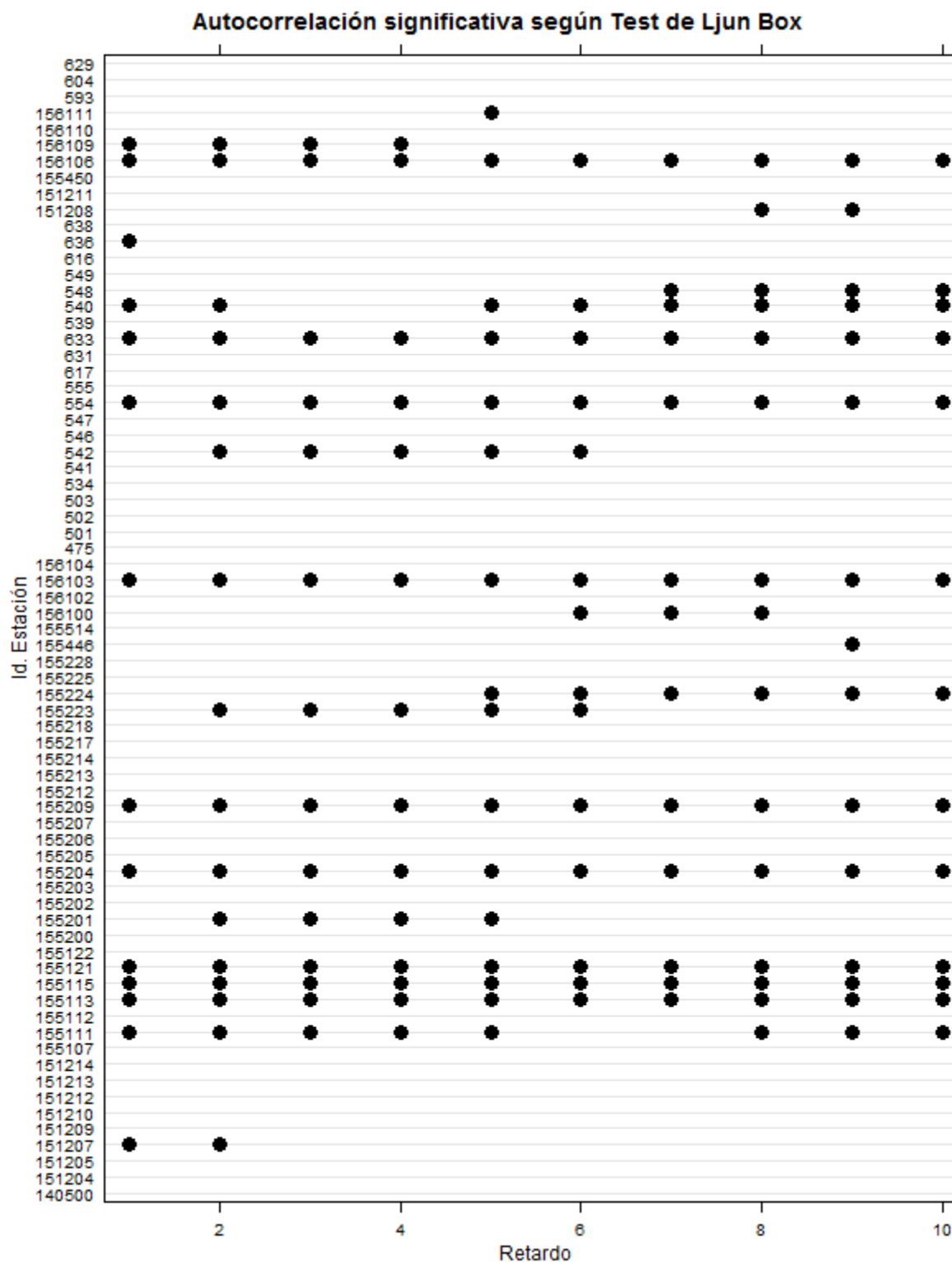


Figura 10: Resumen de autocorrelaciones significativas detectadas para varios retardos a un nivel de 5% por el test de Ljun Box.

Las estaciones que presentan autocorrelaciones significativas que se indican según su Id, son: 156109, 155223, 542, 155224, 155111, 540, 155113, 155115, 155121, 155204, 155209, 156103, 554, 633 y 156106. En total 15 que representa el 21 % de las estaciones analizadas.

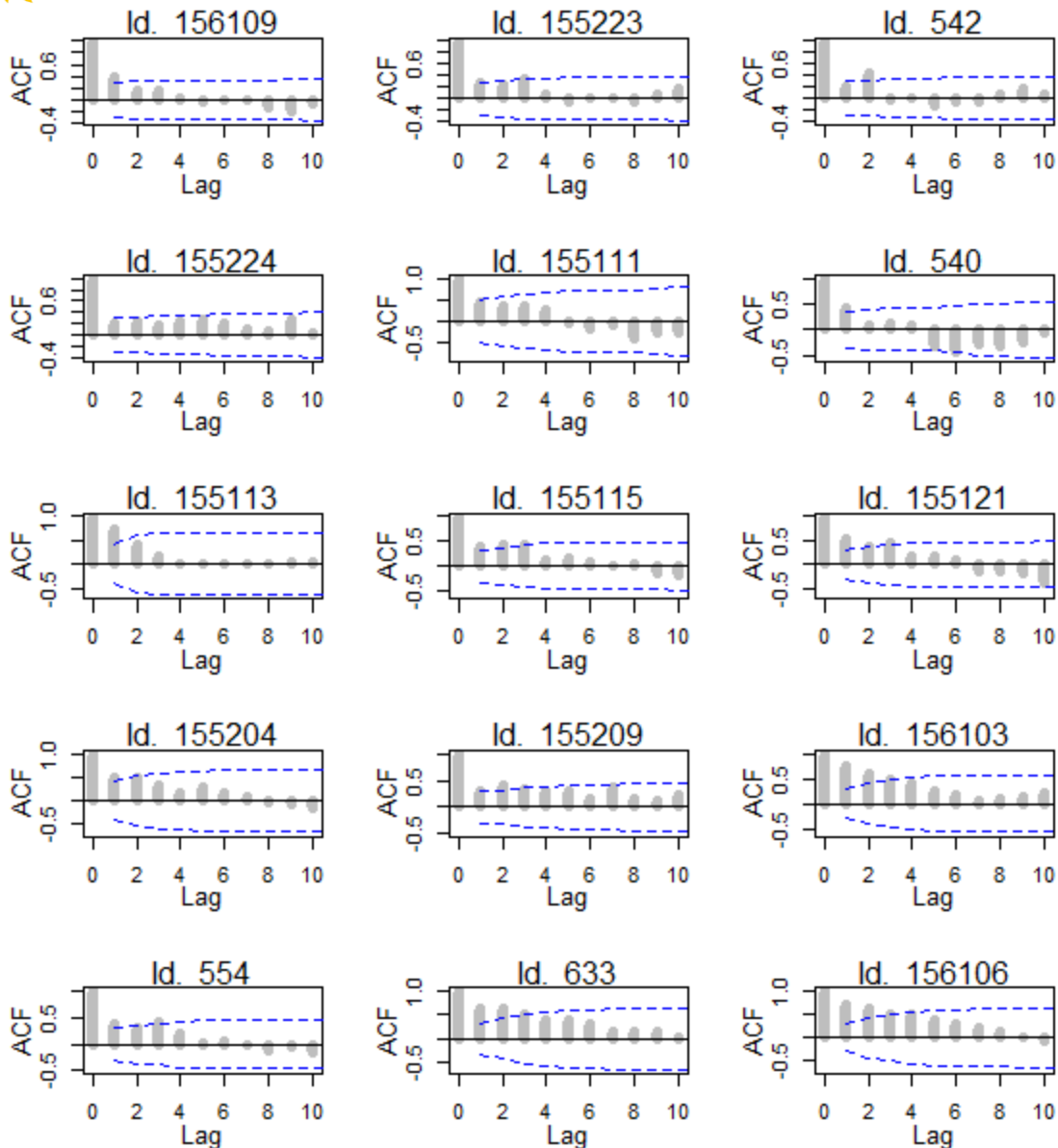


Figura 11: Autocorrelaciones detectadas por el test de Ljun Box.

Cabe señalar que las estaciones que presentan autocorrelación significativa con varios retrasos, no presentan independencia serial; siendo la ocurrencia de los eventos extremos posiblemente gobernada por procesos físicos no aleatorios en estas quince estaciones.

El enunciado para el análisis de frecuencias es que se cumpla los supuestos, de aquí, las estaciones que no cumplen fueron excluidas para el análisis de frecuencias. El siguiente cuadro muestra las estaciones aptas para el análisis regional de frecuencias.

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.

Cuadro 6: Lista de estaciones aptas para el ARF.

N°	Id Station	Nombre	Cuenca	Longitud [E°]	Latitud [N°]	Altitud [msnm]	PPMA [mm]
1	501	ALCANTARILLA	HUAURA	-77.55	-11.05	120	5.68
2	502	SURASACA	RIMAC	-76.78	-10.52	4400	886.72
3	503	JAUJA	MANTARO	-75.47	-11.78	3322	677.64
4	534	LOMAS DE LACHAY	INTERC. DEL PACIFICO	-77.37	-11.37	300	143.96
5	539	HUAYAN	CHANCAY-HUARAL	-77.12	-11.45	350	14.27
6	541	OYON	HUAURA	-76.77	-10.67	3641	353.63
7	546	DONOSO	CHANCAY-HUARAL	-77.23	-11.47	180	11.75
8	547	CANTA	CHILLON	-76.63	-11.47	2832	375.17
9	548	MATUCANA	RIMAC	-76.38	-11.84	2479	306.22
10	549	MARCAPOMACOCHA	MANTARO	-76.33	-11.40	4479	957.61
11	555	HUASAHUASI	PERENE	-75.63	-11.27	2750	414.39
12	593	CERRO DE PASCO	MANTARO	-76.26	-10.69	4260	487.20
13	604	LA OROYA	MANTARO	-75.96	-11.57	3860	628.17
14	617	MODELO CAMPO DE MARTE	RIMAC	-77.03	-12.08	110	6.08
15	631	CALANGO (LA CAPILLA II)	MALA	-76.49	-12.52	442	4.38
16	638	PACARAN	CAÑETE	-76.07	-12.83	700	17.91
17	140500	AEROPUERTO INTERNACIONAL	RIMAC	-77.12	-12.00	13	11.32
18	151204	ARAHUAY	CHILLON	-76.70	-11.62	2800	336.05
19	151205	CANCHACALLA	RIMAC	-76.53	-11.84	2554	309.24
20	151207	HUANCATA	MALA	-76.22	-12.22	2700	456.57
21	151208	GORGOR	PATIVILCA	-77.03	-10.57	3070	231.21
22	151209	CHOSICA	RIMAC	-76.73	-11.93	850	39.21
23	151210	RIO BLANCO	RIMAC	-76.26	-11.73	3550	532.24
24	151211	AMBAR	SUPE	-77.28	-10.75	2100	277.27
25	151212	SAN PEDRO DE PILAS	OMAS	-76.22	-12.45	2600	237.20
26	151213	SHEQUE	RIMAC	-76.50	-11.67	3214	504.93
27	155107	PAMPA LIBRE	HUAURA	-76.97	-10.87	1800	261.92
28	155112	PARIACANCHA	CHILLON	-76.50	-11.38	3800	698.19
29	155122	AUTISHA	RIMAC	-76.61	-11.74	2171	231.63
30	155200	PACCHO	HUAURA	-76.93	-10.95	3250	277.44
31	155202	SANTA CRUZ	HUAURA	-76.63	-11.20	3700	577.59
32	155203	CARAC	CHANCAY-LIMA	-76.78	-11.18	2600	377.34
33	155205	PALLAC	CHANCAY-HUARAL	-76.80	-11.35	2333	300.59
34	155206	LAGUNA SURASACA	HUAURA	-76.78	-10.52	4400	849.38
35	155207	PACHAMACHAY	HUAURA	-76.83	-11.05	4200	376.59
36	155212	PARQUIN	HUAURA	-76.72	-10.97	3590	692.10
37	155213	SANTA EULALIA	RIMAC	-76.67	-11.92	982	39.54
38	155218	HUARIOS	CHILLON	-76.57	-11.40	3585	536.98
39	155225	SAN JOSE DE PARAC	RIMAC	-76.26	-11.80	3866	615.69
40	155228	SAN PEDRO DE CHUCLU	MANTARO	-75.50	-11.75	3380	699.60
41	155446	CASAPALCA	RIMAC	-76.23	-11.65	4214	680.93
42	155450	YURICOCHA	CAÑETE	-75.72	-12.32	4675	991.62
43	155514	MILLOC	RIMAC	-76.35	-11.57	4398	857.64
44	156102	SAN LAZARO DE ESCOMARCA	LURIN	-76.35	-12.18	3600	536.53
45	156104	AYAVIRI	MALA	-76.13	-12.38	3228	548.45
46	156110	HUANGASCAR	CAÑETE	-75.83	-12.90	2533	262.66
47	156111	VILCA	CAÑETE	-75.83	-12.11	3864	419.56

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 (47 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 7 y Figura 12.

Cuadro 7: Valores de los ratios de L – momentos de las estaciones y medida de discordancia considerando región entera.

Estación N°	Estación nombre	Longitud n	Media l_1	L-CV t	L-Asimetría t_3	L-Curtosis t_4	Discordancia D*	D
1	501	44	1.70	0.493	0.452	0.304	1.701	2.384
2	502	16	27.71	0.091	-0.220	-0.054	2.026	2.719
3	503	40	33.46	0.141	0.155	0.158	0.303	0.275
4	534	26	5.65	0.383	0.538	0.349	1.510	2.066
5	539	39	3.85	0.455	0.345	0.229	1.364	1.833
6	541	40	24.05	0.112	0.042	0.136	0.314	0.375
7	546	26	4.12	0.489	0.491	0.338	1.742	2.493
8	547	36	21.92	0.205	0.071	0.133	0.136	0.220
9	548	44	20.21	0.202	0.151	0.142	0.036	0.030
10	549	40	29.88	0.138	0.245	0.183	0.840	0.866
11	555	24	26.39	0.133	0.101	0.147	0.225	0.219
12	593	16	34.04	0.129	0.235	0.169	0.898	0.942
13	604	20	28.06	0.108	-0.028	0.270	2.080	3.302
14	617	23	1.17	0.324	0.120	0.179	0.817	1.164
15	631	18	2.64	0.359	0.142	-0.049	3.562	----
16	638	30	5.24	0.389	0.370	0.325	0.973	1.487
17	140500	21	1.31	0.615	0.679	0.511	4.826	----
18	151204	33	24.16	0.216	0.225	0.133	0.268	0.341
19	151205	24	23.65	0.285	0.280	0.166	0.236	0.343
20	151207	32	21.52	0.220	0.171	0.140	0.048	0.056
21	151208	31	25.80	0.204	0.234	0.298	0.799	1.089
22	151209	18	4.74	0.307	0.040	0.019	1.843	2.300
23	151210	22	22.14	0.119	-0.027	0.199	0.986	1.582
24	151211	19	20.14	0.213	0.023	-0.025	1.473	1.980
25	151212	27	22.62	0.255	0.241	0.133	0.270	0.384
26	151213	19	19.24	0.101	-0.336	0.189	5.624	----
27	155107	30	10.93	0.197	0.113	0.096	0.177	0.230
28	155112	43	28.48	0.145	0.233	0.162	0.710	0.750
29	155122	28	18.61	0.237	0.145	0.078	0.405	0.561
30	155200	46	28.25	0.156	0.096	0.155	0.122	0.138
31	155202	44	27.25	0.156	0.212	0.159	0.458	0.460
32	155203	32	23.89	0.138	-0.013	0.036	0.565	0.750
33	155205	48	23.96	0.260	0.199	0.134	0.151	0.216
34	155206	16	21.58	0.118	-0.052	0.028	0.677	0.904
35	155207	24	18.95	0.092	0.067	0.170	0.553	0.630
36	155212	40	28.67	0.146	0.305	0.226	1.259	1.331

37	155213	42	11.56	0.424	0.283	0.165	1.311	1.700
38	155218	49	24.18	0.159	0.177	0.193	0.277	0.258
39	155225	35	25.43	0.144	-0.005	0.203	0.845	1.393
40	155228	21	37.86	0.143	0.082	0.123	0.166	0.174
41	155446	21	23.88	0.149	0.125	0.066	0.557	0.721
42	155450	26	32.60	0.211	0.251	0.261	0.436	0.529
43	155514	23	31.48	0.187	0.308	0.154	0.976	1.202
44	156102	48	29.54	0.252	0.226	0.205	0.045	0.075
45	156104	41	31.37	0.150	-0.099	0.173	1.446	2.355
46	156110	47	25.09	0.245	0.241	0.246	0.205	0.298
47	156111	42	29.59	0.187	0.298	0.175	0.760	0.875

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 las estaciones discordantes.

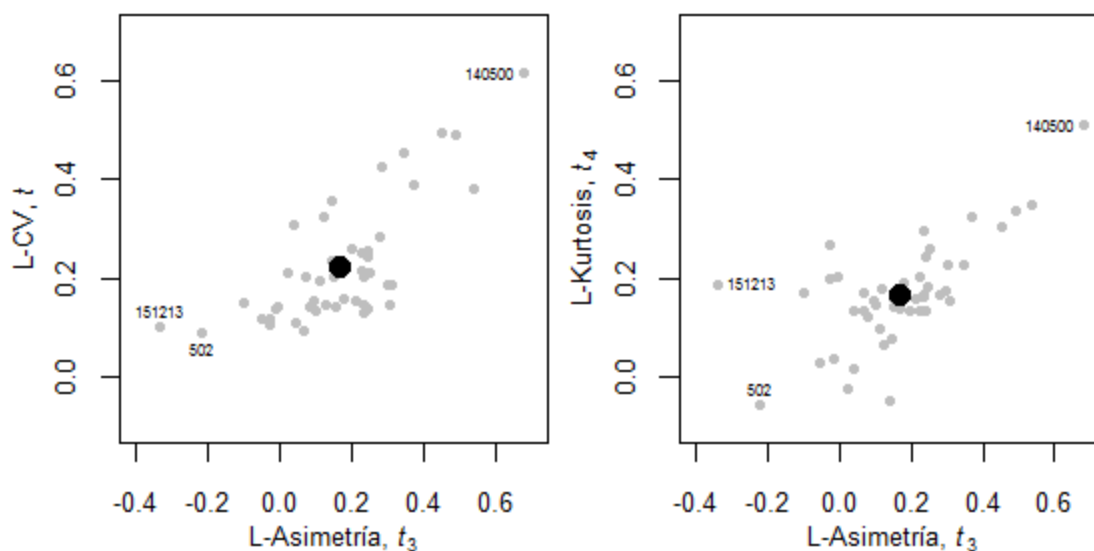


Figura 12: Ratios de los L – momentos de las estaciones.

Hosking y Wallis (1997) sugirieron que una estación es considerada como discordante si su valor D_i excede el valor crítico (D_i es 3 para más de 15 estaciones analizadas en el grupo). De acuerdo a la metodología propuesta primero se considera el análisis considerando todo el conjunto de las 47 estaciones como una sola región. Para el filtrado de la estación discordante, la medida de discordancia D^* se ha calculado para cada estación. Donde se aprecia que las estaciones de 151213, 502 y 140500 (Figura 12); evidentemente son discordantes con el resto de las estaciones ya que sus valores se encuentran alejados del promedio regional (punto en negrita). Estas discordancias se deben por presentar diferentes patrones de la L-CV, L-asimetría y L-kurtosis con respecto al grupo por lo que se excluyen definitivamente estas estaciones y son determinadas por segunda vez la medida de discordancia D .

Aparentemente la estación 604 (Oroya) en el segundo análisis resulta ser discordante, sin embargo esta será examinada en su respectivo grupo o región homogénea.

Cabe señalar que se ha determinado la medida de heterogeneidad considerando la región como un todo, obteniéndose ($H_1=21.22$, $H_2=7.00$ y $H_3=2.61$) valores que declaran la región entera como heterogénea por lo que se procede a la formación de regiones homogéneas.

4.4 Heterogeneidad regional y la formación de regiones homogéneas

Para la formación de 2 o más regiones homogéneas se han utilizado la técnica de agrupamiento jerárquico y el método de particionamiento K-means como una primera aproximación para clasificar las estaciones inicialmente en 6 grupos utilizando como variables predictores la longitud, latitud, elevación, L-CV, L-Asimetría y L-Curtosis de cada estación. Cabe señalar que la primera aproximación resultó ser mejor mediante la técnica de clasificación no jerárquica con el enlace de Ward; a partir del cual se obtuvo las regiones definitivas (Figura 13).

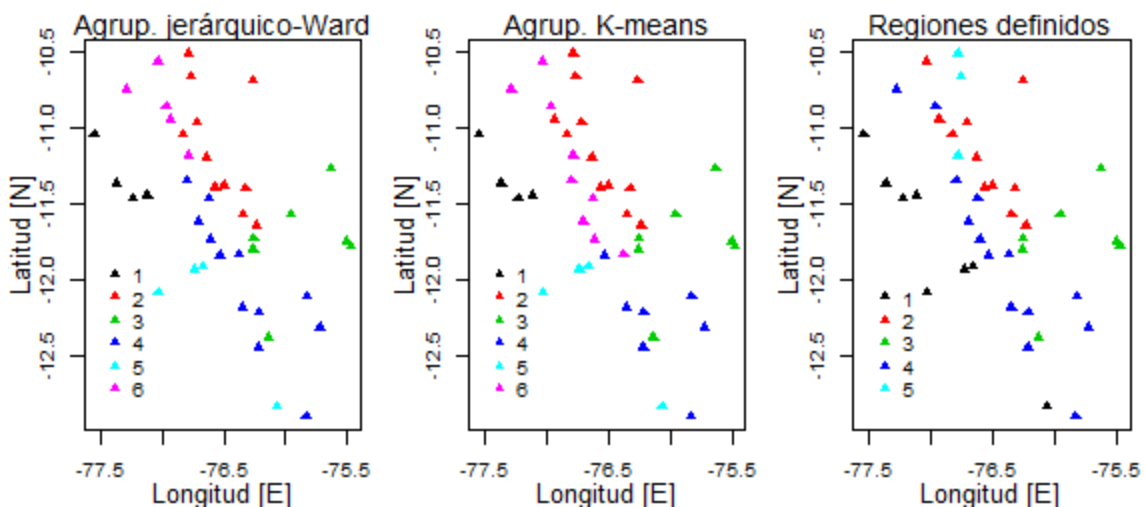


Figura 13: Izquierda: Clasificación de estaciones mediante agrupación jerárquica y enlace de Ward. Centro: Clasificación de estaciones mediante K-means. Derecha: Regiones definitivas.

Para cada una de las regiones definidas se ha determinado la medida de discordancia D y heterogeneidad H de Hosking y Wallis para validar la condición de homogeneidad de las cinco regiones definidas.

Cuadro 8: Medida de discordancia para las estaciones de cada región

Nº	Región 1		Región 2		Región 3		Región 4		Región 5	
	Id. Estación	D	Id. Estación	D	Id. Estación	D	Id. Estación	D	Id. Estación	D
1	501	0.60	549	0.33	503	1.14	547	1.51	541	1.00
2	534	2.14	593	0.49	555	0.49	548	0.35	155203	1.00
3	539	0.33	151208	2.54	604	1.46	151204	0.49	155206	1.00
4	546	0.54	155112	0.17	151210	0.86	151205	1.63		
5	617	1.45	155200	1.00	155225	0.93	151207	0.04		
6	638	0.84	155202	0.07	155228	0.69	151211	2.12		
7	151209	1.61	155207	2.12	156104	1.42	151212	0.65		
8	155213	0.49	155212	1.06			155107	0.54		
9			155218	0.18			155122	0.41		
10			155446	1.71			155205	0.51		
11			155514	1.33			155450	1.32		
12							156102	0.59		
13							156110	1.04		
14							156111	2.82		
	Dcrít.= 2.14		Dcrít.= 2.63		Dcrít.= 1.92		Dcrít.= 2.97		Dcrít.= 1.33	

La medida de discordancia en todas las regiones resultaron ser menores que el D crítico por lo que se procede a determinar la medida de heterogeneidad.

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 10): x (localización), a (escala), k y h (forma); se obtienen ajustando a los ratios regionales medios de L-momentos (Cuadro 9).

La distribución Kappa de cuatro parámetros definidos en la ecuación (35) tiene la forma: $f(x) = a^{-1} \xi [1 - k(x-x)/a]^{k-1} \xi F(x) \xi^{1-h}$. Para el presente estudio se ha realizado 1000 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 9.

Cuadro 9: Estadísticas de medias ponderadas de L-momentos regionales

Región	L ^R	L-momentos regional		
		T ^R	T ₃ ^R	T ₄ ^R
1	1	0.4218	0.3491	0.2473
2	1	0.1525	0.2036	0.1800
3	1	0.1368	0.0245	0.1795
4	1	0.2278	0.1942	0.1528
5	1	0.1222	0.0048	0.0801

Cuadro 10: Parámetros de la distribución Kappa para la región

Región	xi	Kappa parámetros		
		alpha	k	h
1	0.5931	0.4438	-0.2641	-0.0206
2	0.8938	0.1844	-0.1057	-0.2388
3	0.9945	0.1367	-0.0245	-1.0000
4	0.7858	0.3368	-0.0105	0.0986
5	0.8844	0.2739	0.4118	0.2624

Hecha la simulación de Monte Carlo para cada región se obtiene: v_1 que es la desviación estándar ponderada observada de los valores t (L-CV), m_{v_1} y s_{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 , m_{v_2} y s_{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 , m_{v_3} y s_{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 11).

Cuadro 11: Medidas de heterogeneidad

Región	Resultados de la simulación de montecarlo									Heterogeneidad		
	V_1	V_2	V_3	σ_{V_1}	σ_{V_2}	σ_{V_3}	S_{V_1}	σ_{V_2}	σ_{V_3}	H_1	H_2	H_3
1	0.0606	0.1271	0.1417	0.0552	0.1025	0.1258	0.0168	0.0274	0.0346	0.32	0.90	0.46
2	0.0243	0.0638	0.0744	0.0208	0.0757	0.0993	0.0050	0.0169	0.0191	0.68	-0.71	-1.30
3	0.0128	0.0828	0.0919	0.0199	0.0792	0.1012	0.0061	0.0227	0.0229	-1.17	0.16	-0.41
4	0.0272	0.0659	0.0796	0.0268	0.0715	0.0907	0.0054	0.0137	0.0158	0.06	-0.40	-0.70
5	0.0118	0.0362	0.0617	0.0110	0.0509	0.0670	0.0056	0.0242	0.0251	0.15	-0.61	-0.21

Del Cuadro 11, 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 las regiones definidas se consideran hidrológicamente homogéneas respecto a la frecuencia del régimen de las precipitaciones máximas.

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 14) 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 12), 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 14 para cada región. El punto para la región 1 se encuentra cerca de la curva de distribución GEV. Mientras de la región 2 se encuentra más próximo a la curva teórica de GEV y GLO, región 3 (GLO), región 4 (GEV y GNO) y región 5 (GEV). Estas aproximaciones cualitativas son verificadas mediante una bondad de ajuste denominado Z-estadístico.

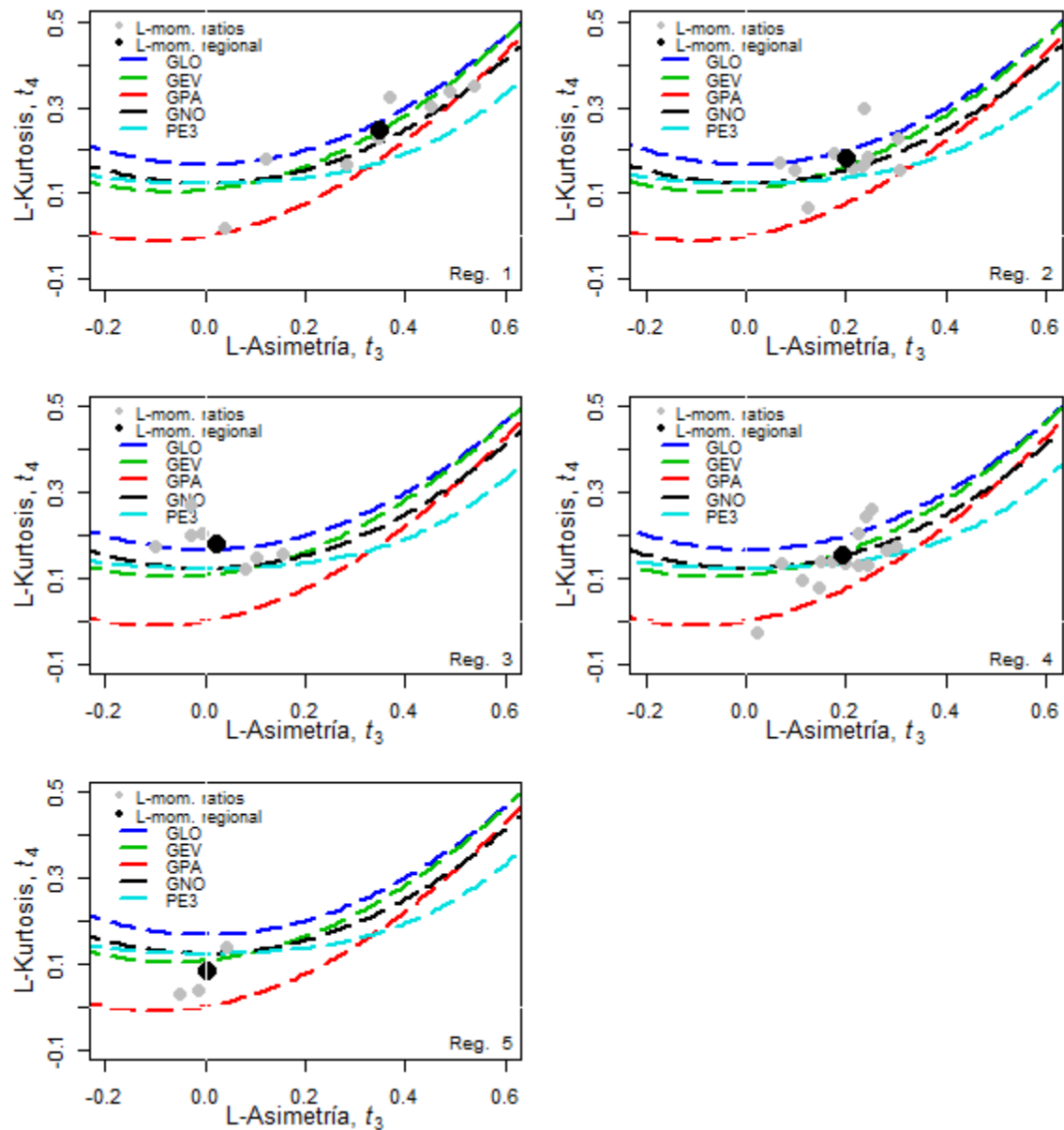


Figura 14: Diagrama de L-momentos ratio con el L-momentos ratio regional para las distribuciones candidatas.

Cuadro 12: Estadístico Z^{DIST} de varias distribuciones para cada región

Región	Z^{GLO}	Z^{GEV^*}	Z^{GNO^*}	Z^{PE3}	Z^{GPA}
1	0.06	-0.50	-1.19	-2.40	-2.22
2	0.71	-0.84	-1.23	-2.03	-4.46
3	-0.95	-2.99	-2.53	-2.54	-6.77
4	2.34	0.32	-0.11	-1.03	-4.31
5	2.75	0.95	1.40	1.40	-2.33

* 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 12), lo cual muestra que las distribuciones de mejor ajuste para cada región son las mismas que se encuentran con la letra en negrita en el Cuadro 12, y arreglando el valor de Z^{DIST} en orden ascendente para seleccionar la distribución, que da un mejor ajuste para cada región se tiene: región 1 (GLO), región 2 (GLO), región 3 (GLO), región 4 (GNO), y región 5 (GEV).

4.6 Estimación de los cuantiles de frecuencia

4.6.1 Estimación de la curva de crecimiento regional

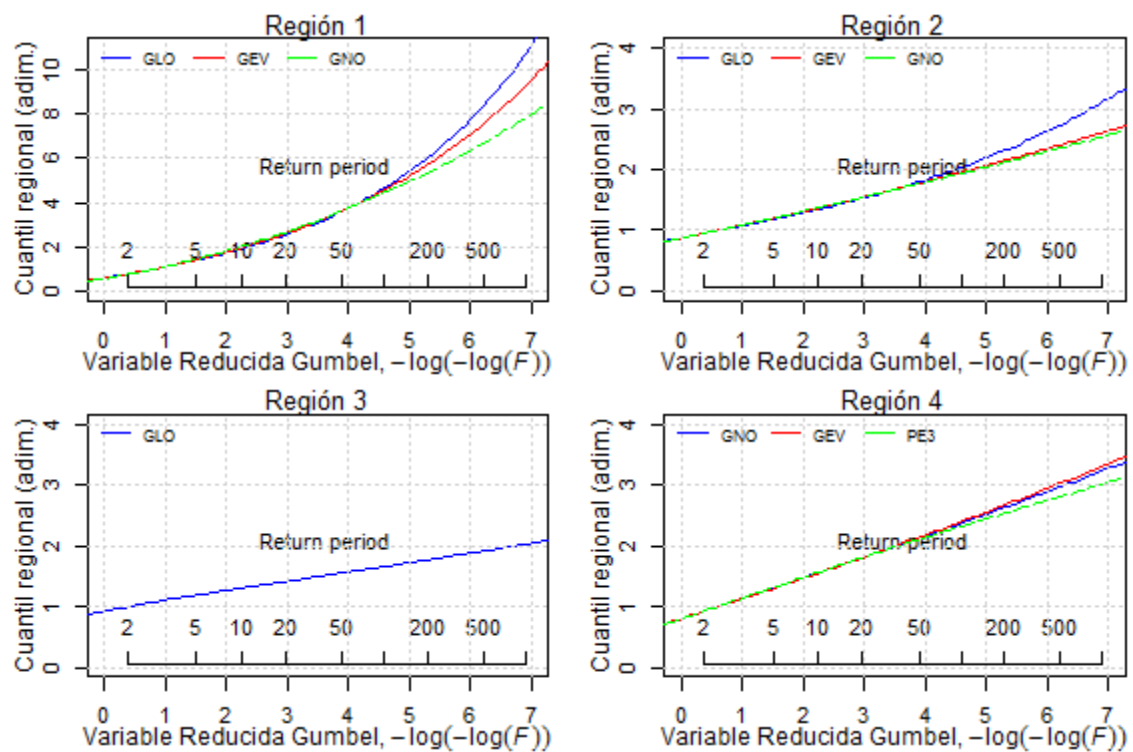
La estimación de los cuantiles regionales $\hat{q}(F)$, para varias probabilidades de no excedencia, así como los parámetros de las distribuciones candidatas se presentan en el Cuadro 13. Los parámetros de las distribuciones se obtiene ajustando sus L-momentos a los L-momentos promedios regionales de la muestra de datos de todas las estaciones que forman la región.

Las curvas de crecimiento regional para cada región se presentan en la Figura 15. Estos pueden ser interpretados como, por ejemplo $\hat{q}_{GEV}(0.99)$ de la región 5, la cantidad de lluvia que se producirá una vez cada 100 años y es 1.4836 veces más grande que el promedio para todos los sitios en la región 5. Las curvas de crecimiento regional candidatas para cada región son aproximadamente igual para periodo de retorno de hasta 50 años. Luego se muestran diferencias que se incrementan en grandes periodos de retorno.

Cuadro 13: Parámetros y cuantiles regionales para las distribuciones candidatas por región.

Distribución	Parámetros			Cuantiles regionales para probabilidades de no excedencia F								
	xi (x)	alpha (α)	k	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
Región 1												
GLO	0.7720	0.3422	-0.3491	0.9609	1.1972	1.3489	1.4911	1.6704	1.8012	1.9287	2.0926	2.2133
GEV	0.5878	0.4476	-0.2610	0.7600	1.4096	1.9585	2.5962	3.6212	4.5703	5.7046	7.5537	9.2764
GNO	0.7480	0.5961	-0.7361	0.7480	1.4428	2.0182	2.6560	3.6104	4.4264	5.3312	6.6754	7.8136
Región 2												
GLO	0.9500	0.1423	-0.2036	0.9500	1.1778	1.3442	1.5238	1.7946	2.0322	2.3043	2.7269	3.1028
GEV	0.8680	0.2092	-0.0518	0.9454	1.1943	1.3673	1.5397	1.7725	1.9546	2.1426	2.4013	2.6051
GNO	0.9448	0.2510	-0.4209	0.9448	1.1982	1.3711	1.5401	1.7638	1.9359	2.1116	2.3509	2.5379
Región 3												
GLO	0.9945	0.1367	-0.0245	0.9945	1.1873	1.3031	1.4119	1.5527	1.6594	1.7672	1.9119	2.0233
Región 4												
GNO	0.9212	0.3775	-0.4010	0.9212	1.2991	1.5536	1.8004	2.1249	2.3727	2.6244	2.9653	3.2303
GEV	0.8048	0.3170	-0.0374	0.9218	1.2940	1.5492	1.8008	2.1368	2.3963	2.6618	3.0227	3.3038
PE3	1.0000	0.4214	1.1756	0.9193	1.3101	1.5650	1.8033	2.1026	2.3208	2.5340	2.8097	3.0146
Región 5												
GEV	0.9232	0.2148	0.2751	0.9980	1.1871	1.2835	1.3590	1.4370	1.4836	1.5219	1.5625	1.5871
GNO	0.9989	0.2166	-0.0099	0.9989	1.1819	1.2782	1.3581	1.4482	1.5086	1.5639	1.6312	1.6785
PE3	1.0000	0.2166	0.0297	0.9989	1.1820	1.2782	1.3581	1.4482	1.5086	1.5639	1.6312	1.6785

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



Continúa...

Continuación

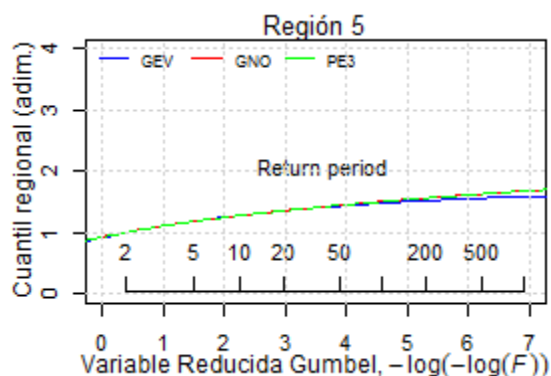


Figura 15: Curva de crecimiento regional para las distribuciones candidatas

Las frecuencias de las precipitaciones máximas en cada región, presentan para su modelización distribuciones candidatas por ello en este estudio para seleccionar la distribución que permita determinar las estimaciones más robustas de los cuantiles regionales, se realiza la simulación de Monte Carlo y partir de ella algunas medidas de precisión para su 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 de cada región, se han llevado a cabo simulaciones sobre la base de las distribuciones candidatas, 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. En este estudio es insignificante la correlación entre los pares de estaciones que pertenecen a una región, por lo que para la simulación de acuerdo con el Cuadro 3, no se considera la dependencia entre estaciones.

En la región 1 para las distribuciones candidatas GLO, 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, 8 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 0.3073 a 0.4933 para la posible heterogeneidad, estos valores son el mínimo y máximo respectivamente de L-CV del grupo.
- L-Cs = 0.3491 es lo mismo en cada estación simulado. Este valor es la media ponderada regional de L-asimetría (Cuadro 9).

La región se ha simulado $M = 1000$ 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 la medida de precisión de la raíz de error cuadrático medio (RECM) relativo ($R^R(F)$). Cabe señalar que para las otras regiones se sigue el mismo procedimiento de simulación realizada para la región 1.

Las medidas de precisión se muestran en el Cuadro 14. Los resultados de la simulación para la región 1, muestran que en general para periodos de retorno de 2 – 20 años el $R^R(F)$ es más bajo para la distribución GLO, sin embargo para periodos de retorno mayores a 50 años la distribución GNO presenta los menores errores relativos. En la región 2 se presenta resultados similares que al de la región 1. La región 3 solo tiene una distribución candidata. La región 4 y región 5 presenta resultados similares, donde para periodos de retorno menores a 50 años la distribución GEV es mejor, mientras para mayores de 50 años PE3 es más robusta. En consecuencia, se resume en el Cuadro 15 las distribuciones para cada región y para cada periodo de retorno.

Cuadro 14: Resultados de la simulación para la curva de crecimiento regional

Distri- bución	F = TR =	0.5 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
Región 1										
GLO	$R^R(F)$	0.0722	0.0473	0.0725	0.0970	0.1280	0.1514	0.1753	0.2078	0.2331
	LEI	0.9247	0.9460	0.8977	0.8483	0.7908	0.7486	0.7077	0.6557	0.6168
	LES	1.1384	1.0881	1.1246	1.1571	1.1961	1.2239	1.2521	1.2889	1.3173
GEV	$R^R(F)$	0.0686	0.0490	0.0744	0.0975	0.1270	0.1499	0.1738	0.2074	0.2341
	LEI	0.9196	0.9405	0.8962	0.8510	0.7966	0.7557	0.7155	0.6631	0.6250
	LES	1.1310	1.0909	1.1282	1.1614	1.2027	1.2327	1.2673	1.3077	1.3463
GNO	$R^R(F)$	0.0695	0.0479	0.0768	0.0999	0.1250	0.1417	0.1571	0.1760	0.1896
	LEI	0.9119	0.9357	0.8895	0.8501	0.8070	0.7795	0.7552	0.7241	0.7039
	LES	1.1311	1.0847	1.1303	1.1682	1.2090	1.2376	1.2612	1.2919	1.3138
Región 2										
GLO	$R^R(F)$	0.0161	0.0360	0.0614	0.0838	0.1110	0.1304	0.1491	0.1734	0.1915
	LEI	0.9781	0.9487	0.9118	0.8802	0.8418	0.8149	0.7885	0.7551	0.7301
	LES	1.0292	1.0601	1.1030	1.1426	1.1914	1.2282	1.2622	1.3089	1.3429
GEV	$R^R(F)$	0.0160	0.0386	0.0641	0.0850	0.1085	0.1244	0.1394	0.1583	0.1723
	LEI	0.9769	0.9442	0.9090	0.8809	0.8484	0.8267	0.8067	0.7798	0.7605
	LES	1.0282	1.0640	1.1081	1.1461	1.1896	1.2205	1.2492	1.2858	1.3122

GNO	R ^R (F)	0.0163	0.0392	0.0647	0.0851	0.1073	0.1218	0.1348	0.1503	0.1611
	LEI	0.9763	0.9431	0.9079	0.8806	0.8509	0.8327	0.8165	0.7969	0.7834
	LES	1.0286	1.0648	1.1090	1.1465	1.1884	1.2163	1.2425	1.2728	1.2941
Región 3										
GLO	R ^R (F)	0.0087	0.0190	0.0279	0.0363	0.0476	0.0565	0.0659	0.0792	0.0900
	LEI	0.9854	0.9704	0.9570	0.9444	0.9279	0.9155	0.9030	0.8861	0.8734
	LES	1.0140	1.0313	1.0468	1.0625	1.0830	1.1005	1.1180	1.1421	1.1631
Región 4										
GEV	R ^R (F)	0.0164	0.0314	0.0491	0.0633	0.0796	0.0911	0.1023	0.1172	0.1287
	LEI	0.9762	0.9545	0.9282	0.9062	0.8799	0.8611	0.8432	0.8198	0.8030
	LES	1.0293	1.0523	1.0817	1.1057	1.1345	1.1537	1.1745	1.2013	1.2213
GNO	R ^R (F)	0.0168	0.0316	0.0495	0.0635	0.0784	0.0882	0.0970	0.1077	0.1153
	LEI	0.9753	0.9538	0.9271	0.9058	0.8827	0.8682	0.8540	0.8377	0.8261
	LES	1.0298	1.0522	1.0822	1.1062	1.1330	1.1503	1.1664	1.1864	1.2003
PE3	R ^R (F)	0.0169	0.0323	0.0503	0.0634	0.0764	0.0841	0.0906	0.0979	0.1026
	LEI	0.9744	0.9523	0.9259	0.9068	0.8875	0.8763	0.8673	0.8565	0.8494
	LES	1.0296	1.0530	1.0838	1.1066	1.1301	1.1442	1.1566	1.1699	1.1789
Región 5										
GEV	R ^R (F)	0.0101	0.0193	0.0270	0.0347	0.0450	0.0527	0.0604	0.0702	0.0773
	LEI	0.9832	0.9686	0.9574	0.9466	0.9330	0.9242	0.9152	0.9058	0.8985
	LES	1.0163	1.0313	1.0455	1.0604	1.0799	1.0939	1.1114	1.1309	1.1444
GNO	R ^R (F)	0.0105	0.0190	0.0273	0.0352	0.0452	0.0526	0.0598	0.0692	0.0761
	LEI	0.9827	0.9691	0.9572	0.9461	0.9320	0.9226	0.9132	0.9028	0.8949
	LES	1.0170	1.0313	1.0467	1.0618	1.0812	1.0946	1.1077	1.1267	1.1407
PE3	R ^R (F)	0.0105	0.0190	0.0273	0.0352	0.0451	0.0522	0.0591	0.0679	0.0742
	LEI	0.9826	0.9694	0.9574	0.9461	0.9310	0.9213	0.9121	0.9006	0.8912
	LES	1.0171	1.0316	1.0471	1.0618	1.0801	1.0929	1.1050	1.1208	1.1330

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

R^R (F)=RECM relativo, LEI=límite de error inferior, LES=límite de error superior

Cuadro 15: Resumen de las distribuciones más robustas para cada región

Región	Distribución más robusta									
	TR =	2	5	10	20	50	100	200	500	1000
1	GLO					GNO				
2	GLO					GNO				
3	GLO									
4	GEV					PE3				
5	GEV						PE3			

En este trabajo se definió la distribución más robusta para cada región a aquello que resulto ser más robusto para grandes periodos de retorno ya que las diferencias para bajos periodos de retorno (< 50 años) según la medida de precisión es insignificante, además en la práctica se requiere conocer cuantiles de diseño para periodos de retorno mayores a la longitud de registro.

Las ecuaciones de las distribuciones se encuentran en el primer anexo adjunto a este estudio.

4.6.3 Cuantiles de frecuencia regional y local

Seleccionada las distribuciones de mejor ajuste según el Z- estadístico y determinada sus parámetros (Cuadro 16), 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 16: Resultado de cuantiles para la curva de crecimiento regional de cada región.

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
Región 1: GNO									
Cuantil	0.7480	1.4428	2.0182	2.6560	3.6104	4.4264	5.3312	6.6754	7.8136
RECM	0.0520	0.0691	0.1550	0.2653	0.4514	0.6271	0.8373	1.1749	1.4811
LEI	0.6613	1.3301	1.7856	2.2736	2.9863	3.5767	4.2272	5.1670	5.9472
LES	0.8203	1.5420	2.2691	3.1244	4.4739	5.6783	7.0588	9.2192	11.1010
Región 2: GNO									
Cuantil	0.9448	1.1982	1.3711	1.5401	1.7638	1.9359	2.1116	2.3509	2.5379
RECM	0.0154	0.0469	0.0887	0.1310	0.1893	0.2358	0.2846	0.3534	0.4087
LEI	0.9185	1.1253	1.2363	1.3433	1.4842	1.5916	1.6996	1.8470	1.9611
LES	0.9677	1.2705	1.5101	1.7489	2.0728	2.3247	2.5864	2.9501	3.2395
Región 3: GLO									
Cuantil	0.9945	1.1873	1.3031	1.4119	1.5527	1.6594	1.7672	1.9119	2.0233
RECM	0.0087	0.0226	0.0363	0.0513	0.0739	0.0938	0.1165	0.1515	0.1821
LEI	0.9808	1.1512	1.2449	1.3288	1.4338	1.5078	1.5806	1.6741	1.7396
LES	1.0092	1.2234	1.3617	1.4950	1.6734	1.8126	1.9570	2.1576	2.3167
Región 4: PE3									
Cuantil	0.9193	1.3101	1.5650	1.8033	2.1026	2.3208	2.5340	2.8097	3.0146
RECM	0.0156	0.0424	0.0787	0.1143	0.1606	0.1952	0.2297	0.2750	0.3092
LEI	0.8929	1.2441	1.4440	1.6296	1.8606	2.0283	2.1909	2.4016	2.5572
LES	0.9435	1.3757	1.6902	1.9885	2.3691	2.6484	2.9217	3.2805	3.5489
Región 5: PE3									
Cuantil	0.9989	1.1820	1.2782	1.3581	1.4482	1.5086	1.5639	1.6312	1.6785
RECM	0.0105	0.0225	0.0349	0.0478	0.0653	0.0788	0.0924	0.1107	0.1246
LEI	0.9822	1.1458	1.2208	1.2791	1.3408	1.3803	1.4153	1.4553	1.4815
LES	1.0166	1.2192	1.3351	1.4355	1.5555	1.6374	1.7146	1.8112	1.8834

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

El Cuadro 16 y la Figura 16, 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:

$$\hat{Q}_i(F) = I_1^{(i)} \hat{q}(F).$$

Donde los valores adimensionales $\hat{q}(F)$ estimados para la curva de crecimiento regional (Cuadro 16), se multiplica por el índice de avenida $l_1^{(i)}$, considerado como la media de las observaciones de cada estación, la misma se muestra en el Cuadro 17.

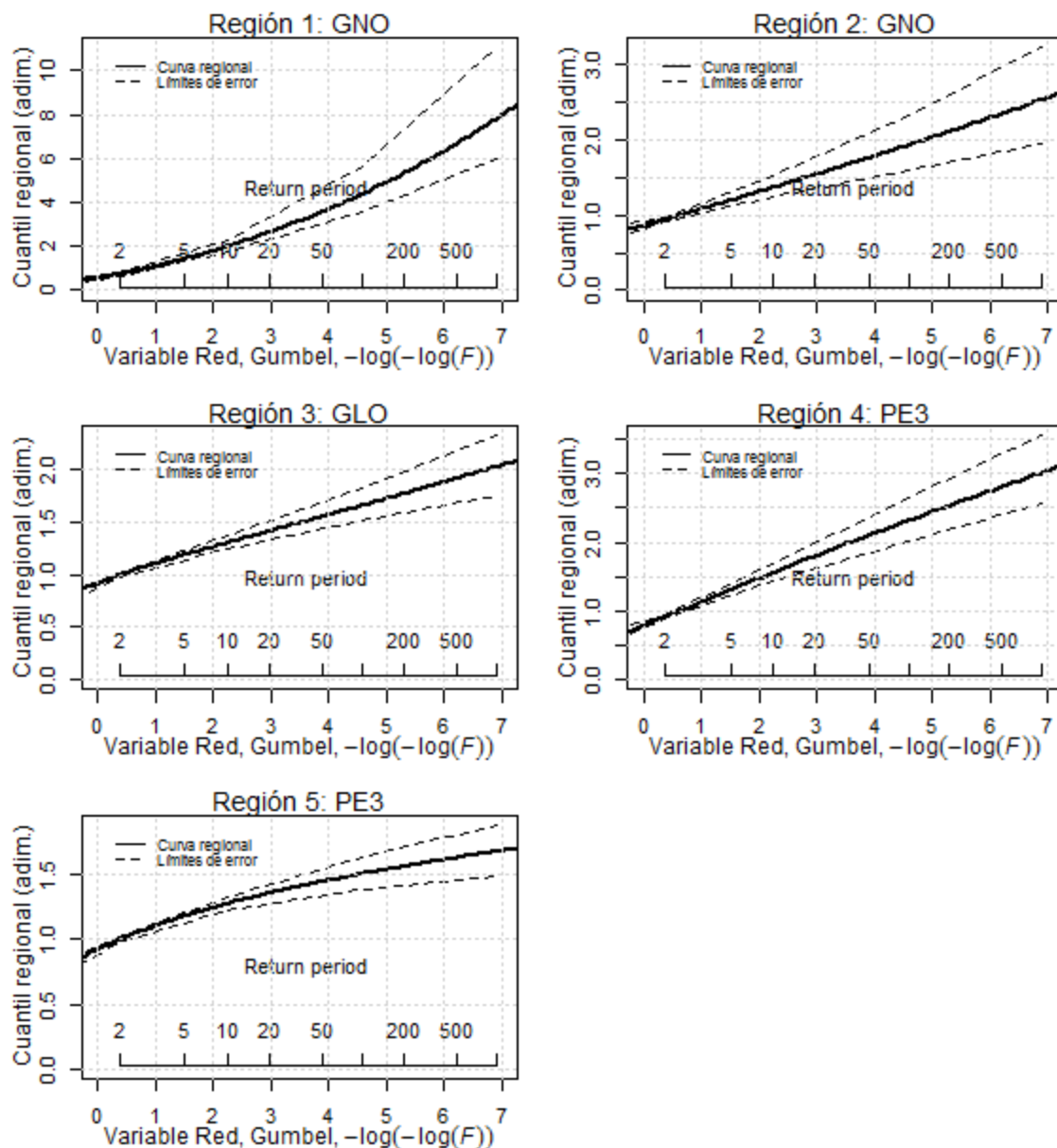


Figura 16: Representación gráfica de la curva de crecimiento regional

Cuadro 17: Índice de avenida de las estaciones de la región

N°	Región 1		Región 2		Región 3		Región 4		Región 5	
	Id. Estación	I_1	Id. Estación	I_1	Id. Estación	I_1	Id. Estación	I_1	Id. Estación	I_1
1	501	1.70	549	29.88	503	33.46	547	21.92	541	24.05
2	534	5.65	593	34.04	555	26.39	548	20.21	155203	23.89
3	539	3.85	151208	25.80	604	28.06	151204	24.16	155206	21.58
4	546	4.12	155112	28.48	151210	22.14	151205	23.65		
5	617	1.17	155200	28.25	155225	25.43	151207	21.52		
6	638	5.24	155202	27.25	155228	37.86	151211	20.14		
7	151209	4.74	155207	18.95	156104	31.37	151212	22.62		
8	155213	11.56	155212	28.67			155107	10.93		
9			155218	24.18			155122	18.61		
10			155446	23.88			155205	23.96		
11			155514	31.48			155450	32.60		
12							156102	29.54		
13							156110	25.09		
14							156111	29.59		

I_1 : índice de avenida

Los resultados de los cuantiles estimados a partir del ARF se presentan en las figuras adjuntas en el Anexo 8.2.

4.8 Mapeo del índice de avenida

Para estimar cuantiles de precipitación máxima de 24 horas de duración asociado a diferentes periodos de retorno, esencialmente en sitios, 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 estudio se ha regionalizado espacialmente utilizando una técnica de mapeo geoestadístico multivariado la misma que se denomina *Co-Kriging* (*Co-Ko*), la elección de esta técnica es porque existe buena relación entre el índice de avenida y la elevación y consecuentemente las estimaciones serán más precisas que las técnicas univariadas como la técnica de interpolación de ponderado de distancia inversa (PDI) y el Kriging ordinario (KO). La validación cruzada fue realizada para determinar la eficiencia de la técnica de mapeo.

4.8.1 Mapeo de índice de avenida por Co - Kriging (Co-Ko)

Para el mapeo por Co-Ko, como en todo análisis geoestadístico, es necesario realizar un análisis exploratorio del comportamiento espacial del índice de avenida.

a. Gráficos de histogramas

En la Figura 17, se muestran los histogramas de frecuencias del índice de avenida de las estaciones que se encuentran en las cuencas del río Rímac, Chillón, Lurín, parte alta de Mantaro y cuencas vecinas. Esta figura claramente nos muestra que son más frecuentes las estaciones que presentan un índice de avenida entre 20 y 30 mm.

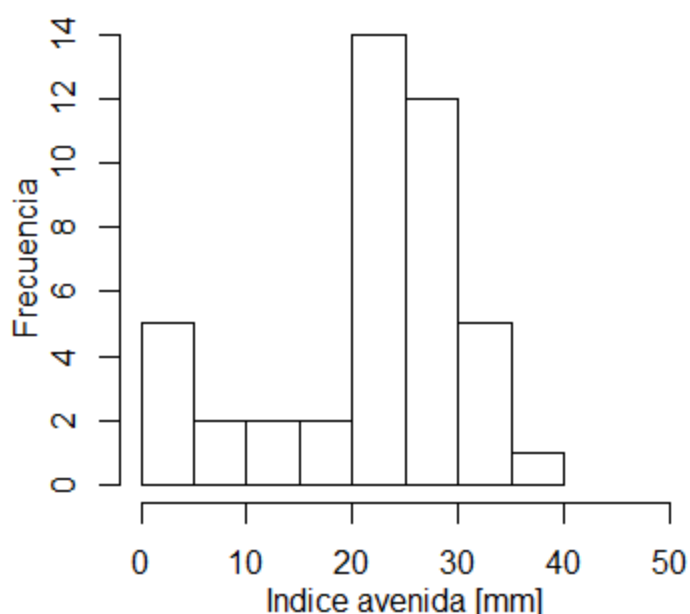


Figura 17: Histograma de frecuencias del índice de avenida.

b. Gráficos de nube de semivariancia

Los gráficos de la nube de semivariancia son importantes en el análisis exploratorio de datos espaciales, porque da conocer la variancia que existe para cada par de puntos (estación) en función a la distancia de separación de las mismas. De esta manera se observa la dependencia espacial para justificar el uso de la técnica geoestadística de la familia Kriging. De la Figura 18 se observa que existe dependencia espacial del índice de avenida, donde a distancias separación menores entre las estaciones se tiene mayor dependencia por presentar menores semivariancias y esta dependencia va disminuyendo a grandes distancias.

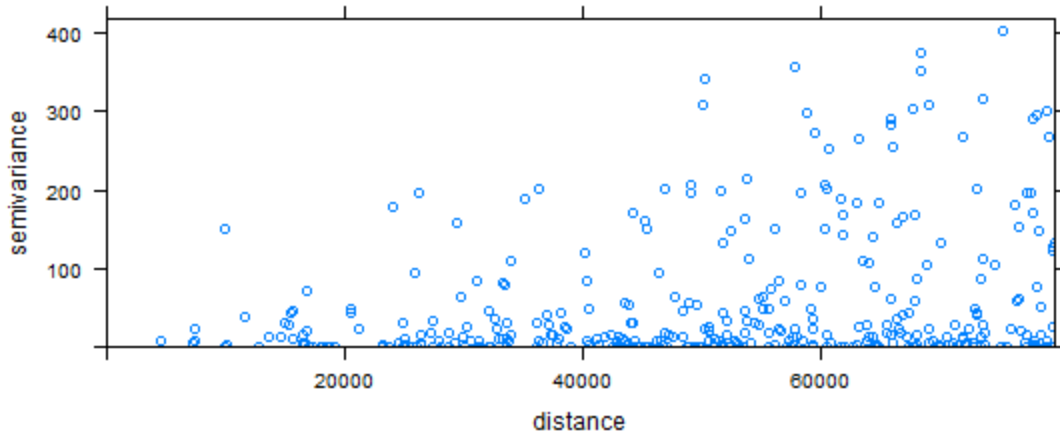


Figura 18: Gráfica de nube de semivariancia de índice de avenida.

c. Relación de índice de avenida versus elevación

En la Figura 19 se observa la relación de índice de avenida de las estaciones estudiadas versus su elevación, se tiene buena relación con coeficientes de determinación de 0.77 la cual nos indica que la variabilidad del índice de avenida es explicada casi en un 80% por la elevación. Es entonces verificada la existencia de la relación entre estas dos variables por lo que para hacer uso de la técnica multivariada Ko-Co se utiliza como covariable la elevación.

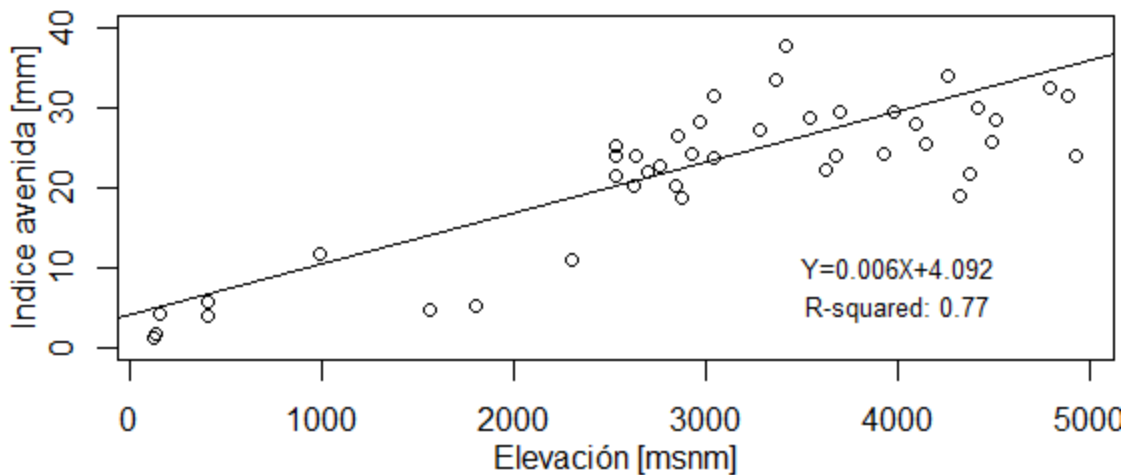


Figura 19: Gráfica de relación de índice de avenida versus elevación.

d. Análisis estructural

En esta etapa y par este estudio se realiza únicamente el análisis estructural considerando que el proceso físico de la variación espacial del índice de avenida es isotrópico, es decir

que los variogramas que caracterizan la dependencia espacial es igual en todas las direcciones a la cual se denomina variogramas omnidireccionales.

El análisis estructural para el Co-Ko consiste en ajustar un modelo teórico al semivariograma experimental de la variable principal (índice de avenida), covariable (elevación) y al semivariograma del producto de ambas variables denominado semivariograma cruzado como se observa por ejemplo en la Figura 20. Cabe señalar que los parámetros del modelo teórico en el Co-Ko son estimadas por el método de modelo lineal de co-regionalización (MLC).

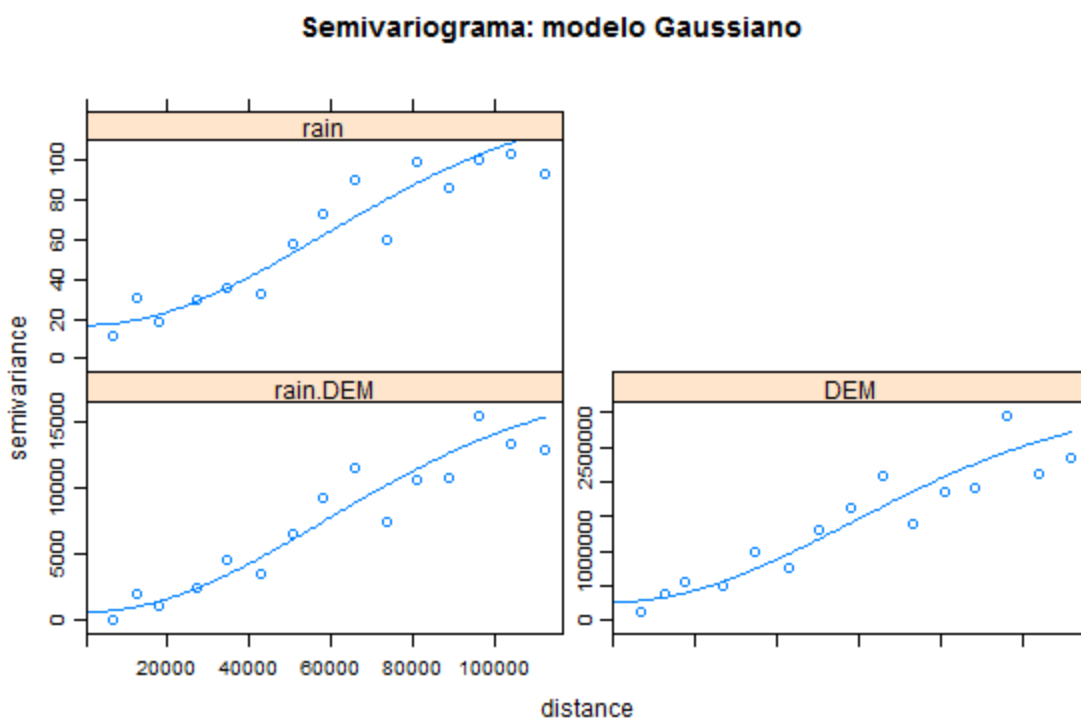


Figura 20: Gráfica de análisis estructural mediante semivariogramas: Ajuste de semivariograma cruzado por el modelo teórico Gaussiano.

Cuadro 18: Parámetros del semivariograma empírico del modelo Gaussiano

N°	Model	Nugget	Sill	Range	SSErr	RSQR	RMSD
	Modelo	Pepita	Meseta	Rango	SSErr	R ²	RECM
Semivariograma cruzado de índice avenida – elevación							
1	Gau	649.53	17391.11	82249.62	8.52E-01	0.69	5.25
Semivariograma de índice avenida							
2	Gau	16.91	115.64	82249.62	2.64E-05	0.69	5.25
Semivariograma de elevación							
3	Gau	268880.25	2903910.91	82249.62	1.55E+04	0.69	5.25

En el cuadro anterior se muestra respectivamente los parámetros del mejor semivariograma cruzado (índice de avenida–Elevación), semivariograma de índice de avenida y semivariograma de Elevación. El parámetro pepita es la semivariancia no explicada por el modelo teórico, la meseta es la máxima semivariancia al alcance del rango. Mientras el parámetro rango ha sido estimada en aproximadamente 82 km, esto quiere decir que en la interpolación para conocer el valor de una celda intervendrán las estaciones que se encuentran dentro del rango para su estimación. Sin embargo en este estudio se ha limitado la cantidad máxima del número de estaciones que intervienen en la interpolación igual a 5 estaciones.

SS_{Err} (Sum of Squares error) es la bondad de ajuste del modelo teórico a los respectivos semivariogramas experimentales, determinados mediante el modelo lineal de co-regionalización (MLC). RECM y R^2 son las medidas de precisión resultado de la validación cruzada.

e. Validación cruzada

La eficiencia de la técnica de interpolación se evalúa mediante una validación cruzada determinando para ello algunas medidas de precisión como la raíz del error cuadrático medio (RECM=5.25 mm) y el coeficiente de determinación ($R^2=0.69$).

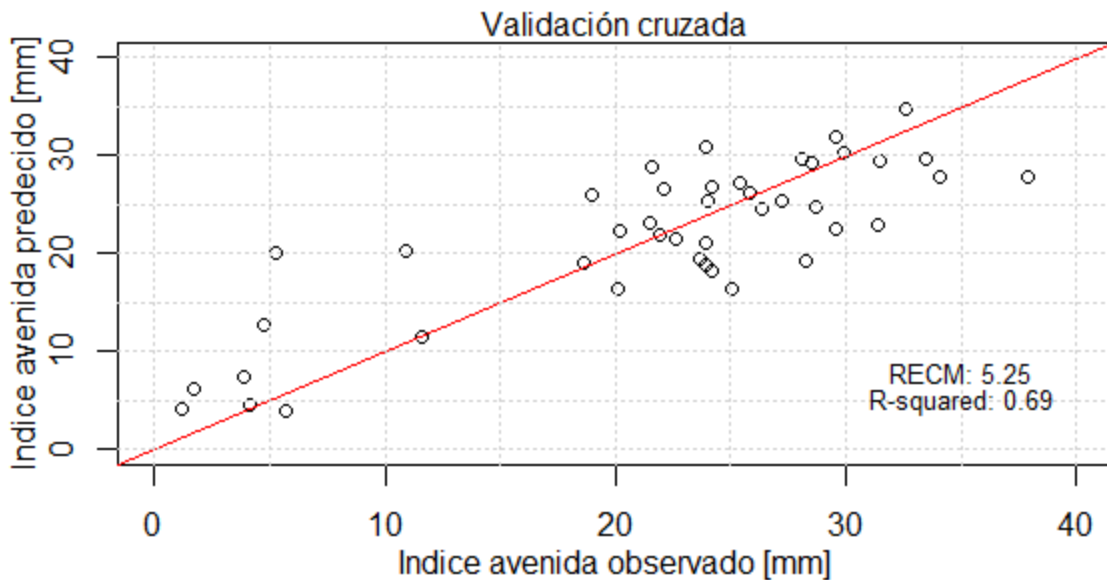


Figura 21: Gráfico de cuantiles de índice de avenida observada y estimada en mm

e. Mapas de precipitación

En este ítem se presentan los mapas de índice de avenida válidos para su estimación de la misma en lugares no medidos. Así mismo se presenta el mapa de incertidumbre para conocer los sitios con menores errores de estimación para una estimación confiable. Cabe señalar que la interpolación se realizó en celdas o grillado con resolución de 1000 x 1000 m.

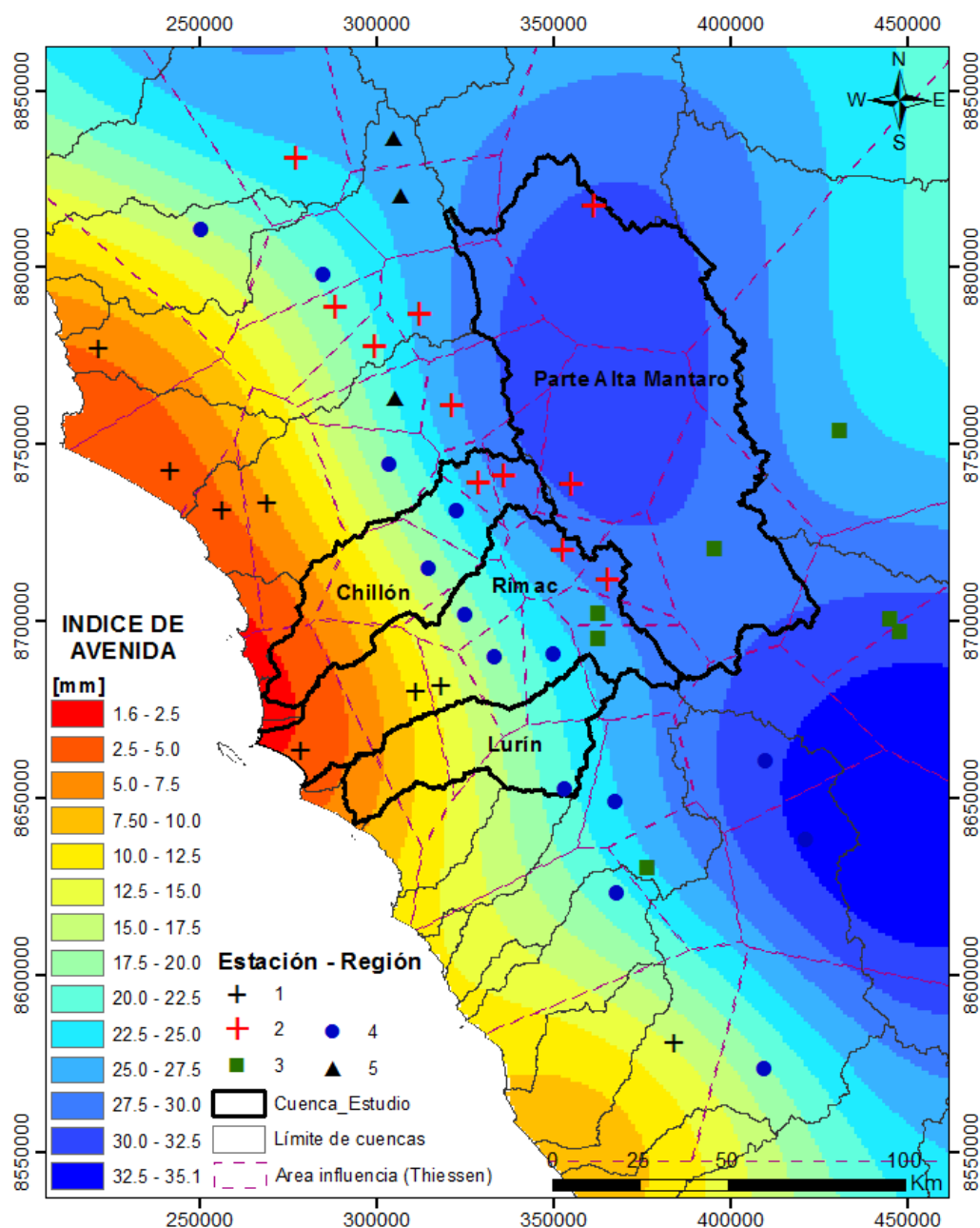


Figura 22: Mapa de índice de avenida.

En la Figura 22 se observa claramente que el índice es mayor en la parte alta de las cuencas, alcanzando valores de 35 mm y ésta disminuye en la dirección de E a O hasta

alcanzar valores de 1.6 mm en la costa peruana. Así mismo se observa el área de influencia a través de polígonos de Thiessen para cada estación con el cual se ha hará uso correcto de las regiones homogéneas.

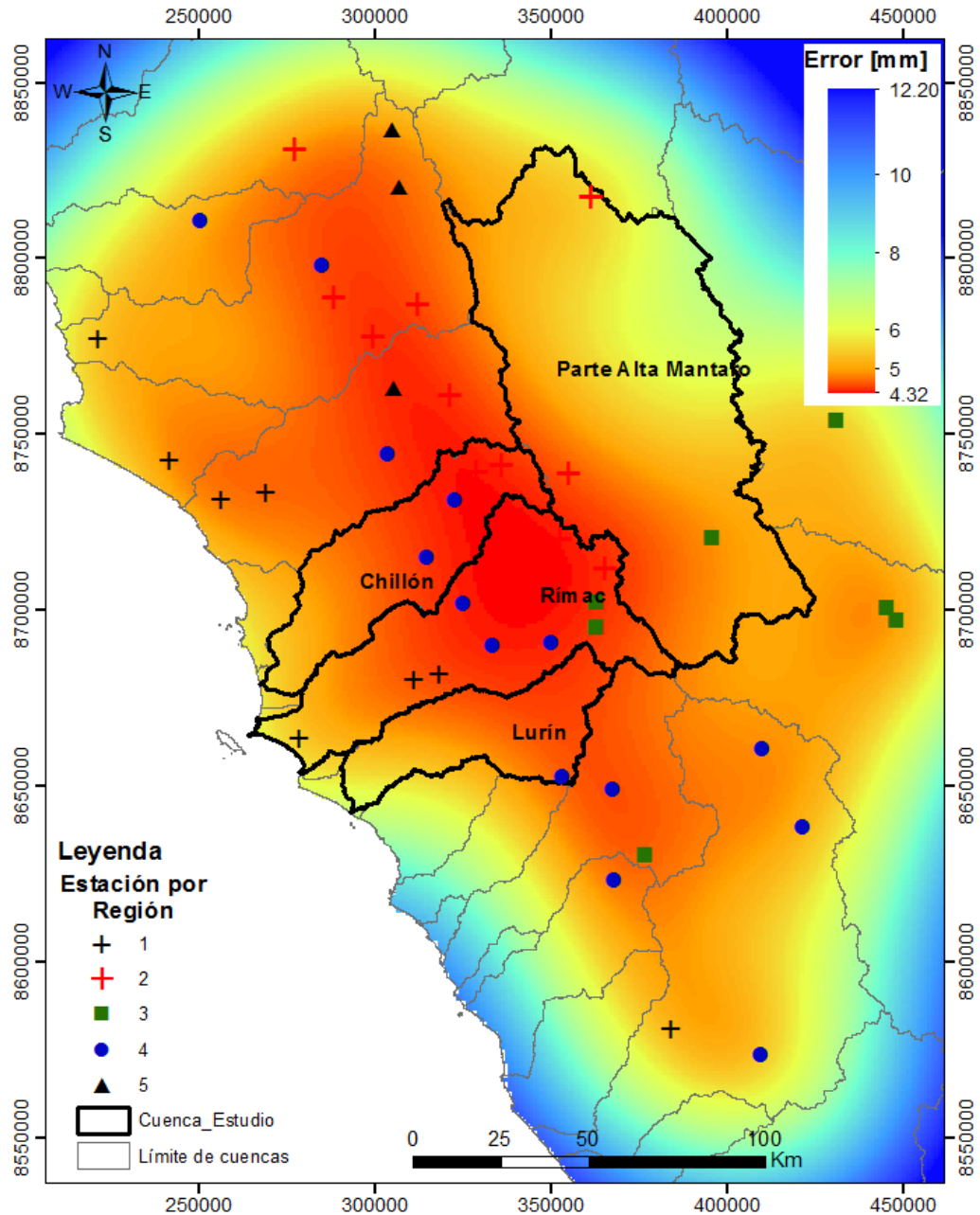


Figura 23: Mapa de incertidumbre del índice de avenida.

Este mapa muestra la incertidumbre de la técnica de interpolación, se aprecia menores incertidumbres en sitios próximos a las estaciones de medición, por lo que se recomienda a los usuarios finales que el uso confiable del mapa de índice de avenida es para los sitios con incertidumbres menores a 6 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 las cuencas de los ríos Chillón, Rímac, Lurín, parte Alta de Mantaro y Cuencas vecinas, 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 en la región de estudio se identifica 5 regiones hidrológicamente homogéneas.
- Para cada región homogénea se ha determinado la función de distribución de probabilidad más robusta para estimar los cuantiles de diseño en grandes periodos de retorno (>50 años), siendo para la región 1 y 2 la distribución normal generalizado (GNO), región 3 (Logístico generalizado - GLO), región 4 y 5 (Pearson tipo 3).
- 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 técnica de interpolación geoestadística Co-Kriging. Válido para estimar el valor del índice de avenida en sitios sin medición. Así mismo se ha generado el mapa de incertidumbre de la técnica de interpolación del índice de avenida para identificar los sitios con menor incertidumbre para su uso correcto y confiable de este estudio por los usuarios finales.

VI. RECOMENDACIONES

Se dan algunas recomendaciones y sugerencias para el uso correcto del estudio.

- Se recomienda que para estimar el cuantil de diseño en sitios con o sin medición primero se debe identificar la región homogénea en la que se ubica el punto de interés utilizando para ello el mapa de índice de avenida (Figura 22) donde se presenta el área de influencia de cada estación y región homogénea. Estimar el cuantil regional y el índice de avenida como se muestra en la Figura 24, luego haciendo uso de la ecuación: $\hat{Q}_i(F) = l_1^{(i)} \hat{q}(F)$. se estima el cuantil de diseño en el sitio de interés.

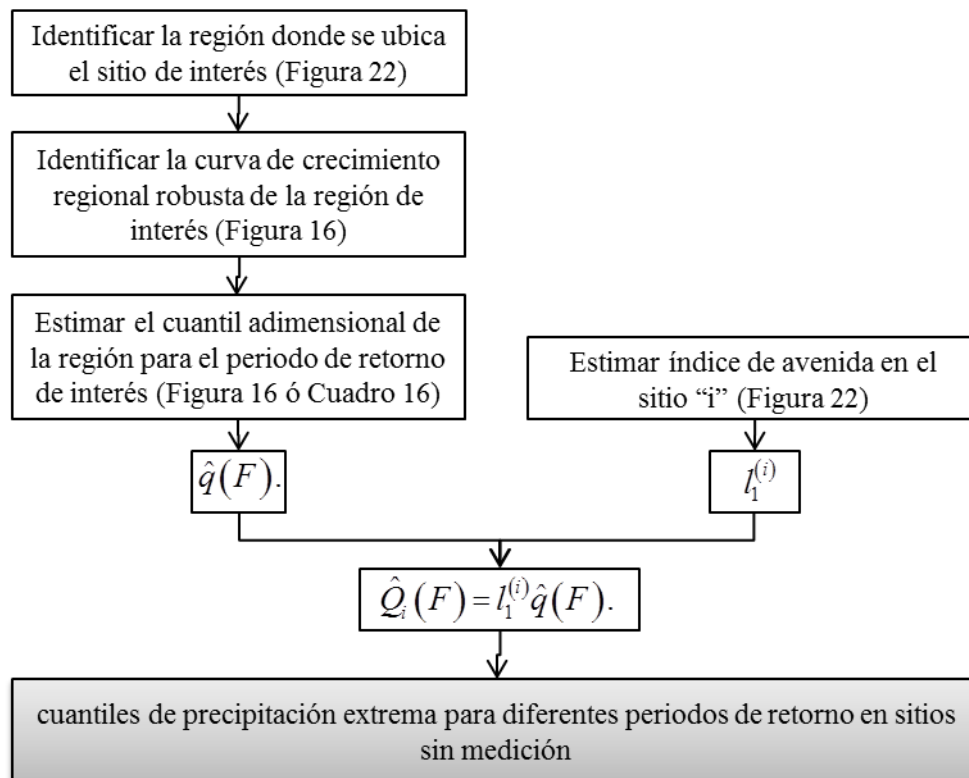


Figura 24: Diagrama de uso práctico del estudio.

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.
- 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.
- 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.
- Feki, H. and Slimani, M., 2006. Analyse structurale de la pluviométrie en Tunisie. In: *WATMED 3, Troisième conférence internationale sur les ressources en eau dans le bassin méditerranéen*, Liban.
- 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.
- Lagos, P., Silva, Y. y Nickl, E. 2009 El niño y la precipitación en los andes del Perú. http://www.met.igp.gob.pe/publicaciones/2000_2007/Lagosetal.pdf
- 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.
- Mann, H. B., 1945. Nonparametric tests against trend. Econometrica 13(1): 245-259.
- Matheron, G., 1970. La théorie des variables régionalisées et ses applications. Fontainebleau: Ecole Nationale Supérieure des Mines de Paris, Cahiers du centre de morphologie mathématique, fascicule 5.
- 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.
- Nalder, I. A., and Wein, R. W., 1998. "Spatial Interpolation of Climatic Normals: Test of a New Method in the Canadian Boreal Forest," Agriculture and Forest Meteorology, Vol. 92, No. 4, pp. 211-225.
- Pebesma, E. J., 2004. Multivariable geostatistics in S: the gstat package. Computers y Geosciences 30(7), 683-691

- 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.
- 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.
- Stein, A., Van Eijnsbergen, A. C., and Barendregt, L. G., 1991. “Cokriging Nonstationary Data,” *Mathematical Geology,* Vol. 23, No. 5, pp. 703-719. doi:10.1007/BF02082532
- 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.
- 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.
- Paciorek, C.J., 2003. “Nonstationary Gaussian Processes for Regression and Spatial Modelling.” Ph.D. Dissertation, Carnegie Mellon University, Pittsburgh.
- 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.
- 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.
- 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
- Xiong, Y., Chen, W., Apley, D. and Ding, X., 2007. “A Nonstationary Covariance-Based Kriging Method for Meta-modeling in Engineering Design,” *International Journal for Numerical Methods in Engineering*, Vol. 71, No. 6, pp. 733-756. doi:10.1002/nme.1969

VIII ANEXO

8.1 Distribuciones teóricas y sus relaciones

Cuadro 19: 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 $\alpha \neq 0$. 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 20: 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)	$x(F) = \varepsilon + \alpha \cdot \frac{1 - [1 - F(x)]^k}{k}$	$\lambda_1 = \xi + \frac{\alpha}{1+k}$ $\lambda_2 = \frac{\alpha}{(1+k) \cdot (2+k)}$ $\tau_3 = \frac{1-k}{3+k}, \tau_4 = \frac{(1-k) \cdot (2-k)}{(3+k) \cdot (4+k)}$	$k = \frac{(1-3\tau_3)}{1+\tau_3}$ $\alpha = (1+k)(2+k)\lambda_2$ $\xi = \lambda_1 - (2+k)\lambda_2$
Valor extremo Generalizado (GEV)	$x(F) = \varepsilon + \alpha \cdot \frac{1 - [-\ln F(x)]^k}{k}$	$\lambda_1 = \xi + \alpha \cdot \frac{1 - \Gamma(1+k)}{k}$ $\lambda_2 = \alpha \cdot \frac{(1-2^{-k}) \cdot \Gamma(1+k)}{k}$ $\tau_3 = \frac{2 \cdot (1-3^{-k})}{1-2^{-k}} - 3$ $\tau_4 = \frac{1-6 \cdot 2^{-k} + 10 \cdot 3^{-k} - 5 \cdot 4^{-k}}{1-2^{-k}}$	$c = \frac{2}{3+\tau_3} - \frac{\ln 2}{\ln 3}$ $k = 7.8590c + 2.9554c^2$ $\alpha = \frac{\lambda_2 k}{(1-2^{-k}) \Gamma(1+k)}$ $\xi = \lambda_1 - \frac{\alpha [1 - \Gamma(1+k)]}{k}$
Logístico Generalizado (GLO)	$x(F) = \varepsilon + \alpha \cdot \frac{1 - \left[\frac{1-F(x)}{F(x)} \right]^k}{k}$	$\lambda_1 = \xi + \alpha \cdot \frac{1 - \Gamma(1+k) \cdot \Gamma(1-k)}{k}$ $\lambda_2 = \alpha \cdot \Gamma(1+k) \cdot \Gamma(1-k)$ $\tau_3 = -k$ $\tau_4 = \frac{1+5k^2}{6}$	$k = -\tau_3$ $\alpha = \frac{\lambda_2 \sin k\pi}{k\pi}$ $\xi = \lambda_1 - \alpha \left(\frac{1}{k} - \frac{\pi}{\sin k\pi} \right)$

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

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

Cuadro 21: 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 22: 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

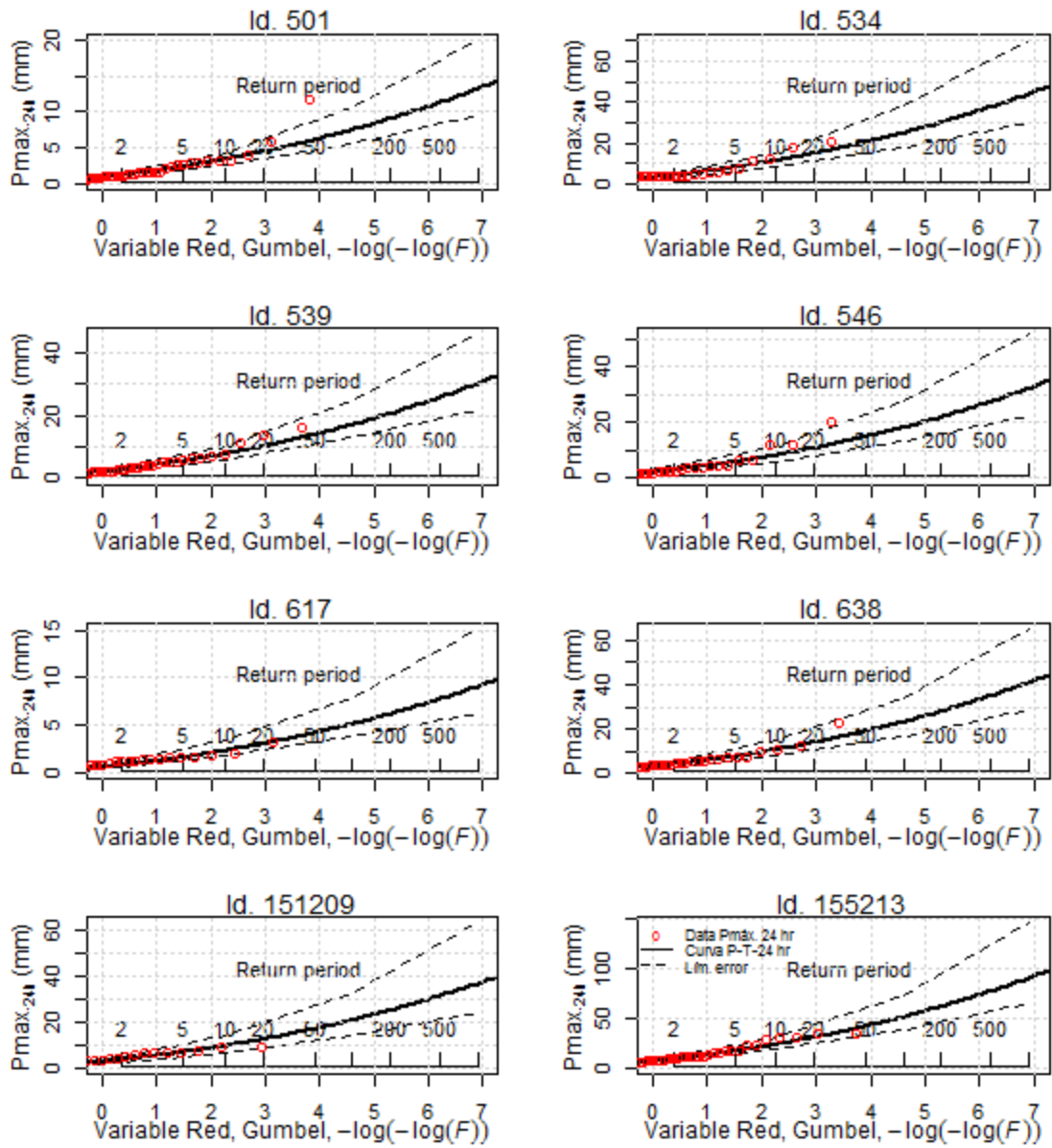


Figura 25: Cuantiles locales para las estaciones de la región 1.

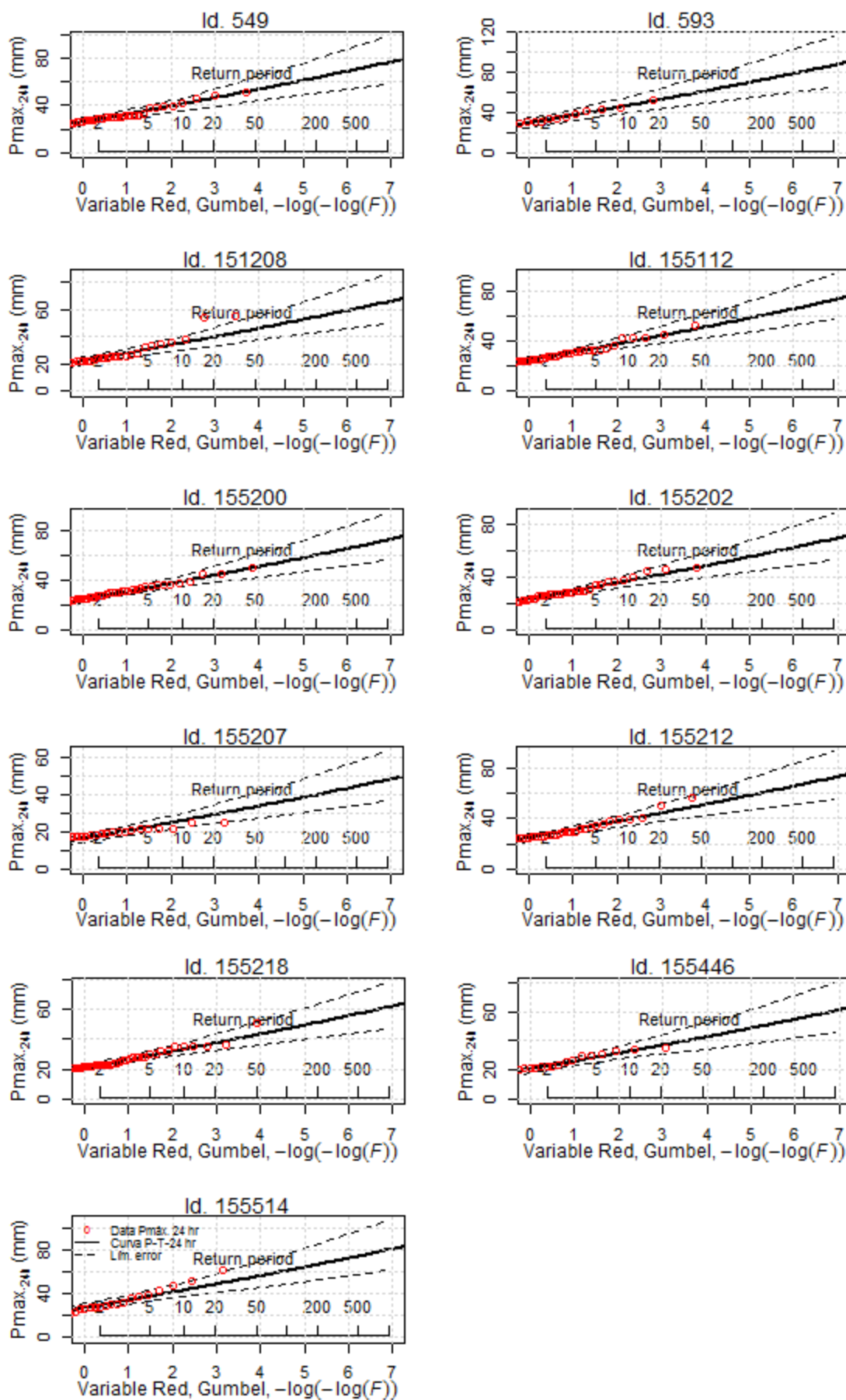


Figura 26: Cuantiles locales para las estaciones de la región 2.

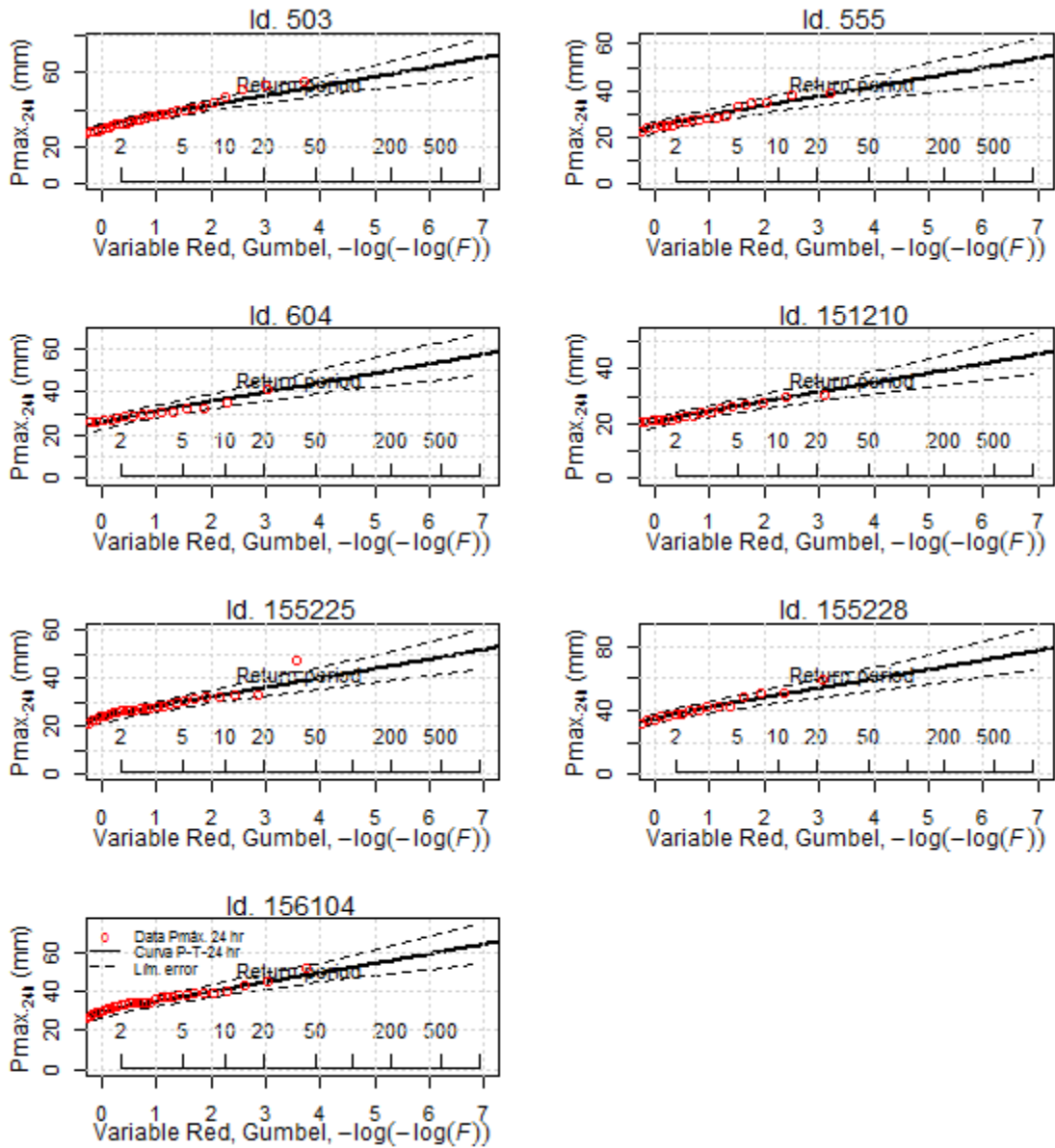
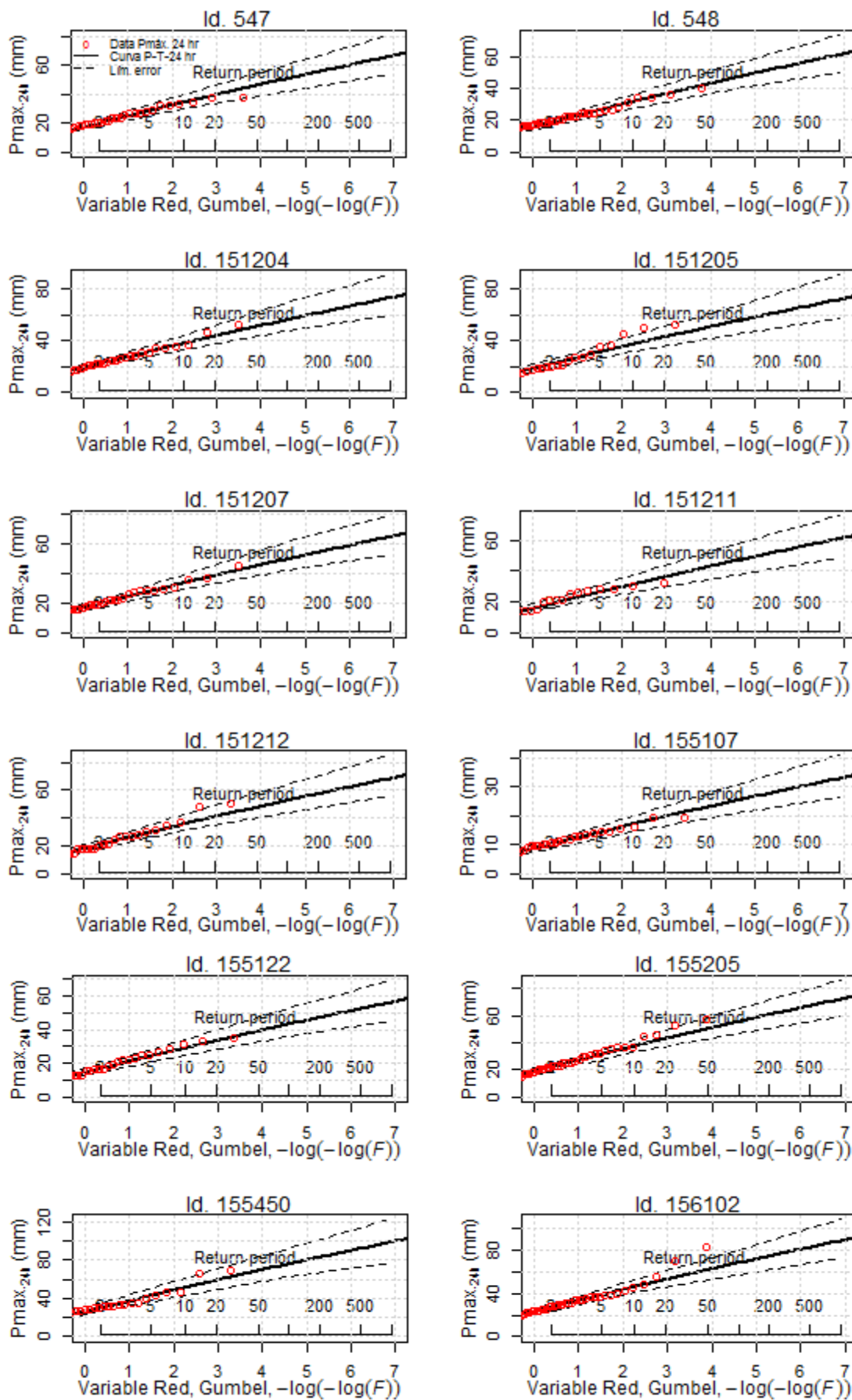


Figura 27: Cuantiles locales para las estaciones de la región 3.



Continúa...

Continuación

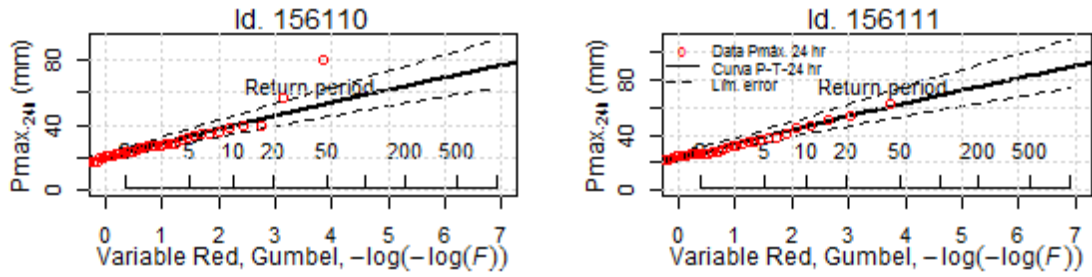


Figura 28: Cuantiles locales para las estaciones de la región 4.

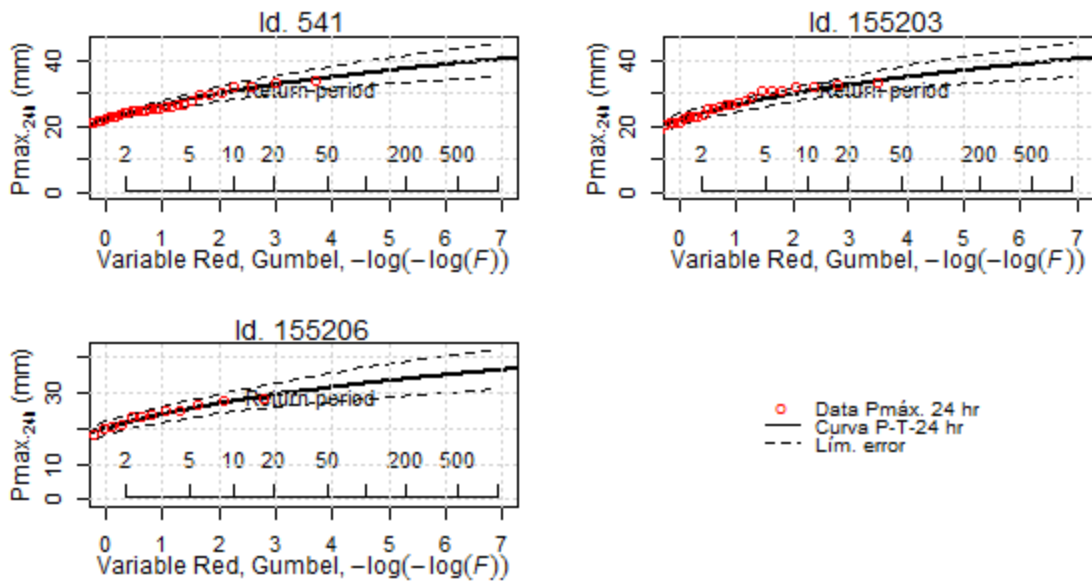


Figura 29: Cuantiles locales para las estaciones de la región 5.