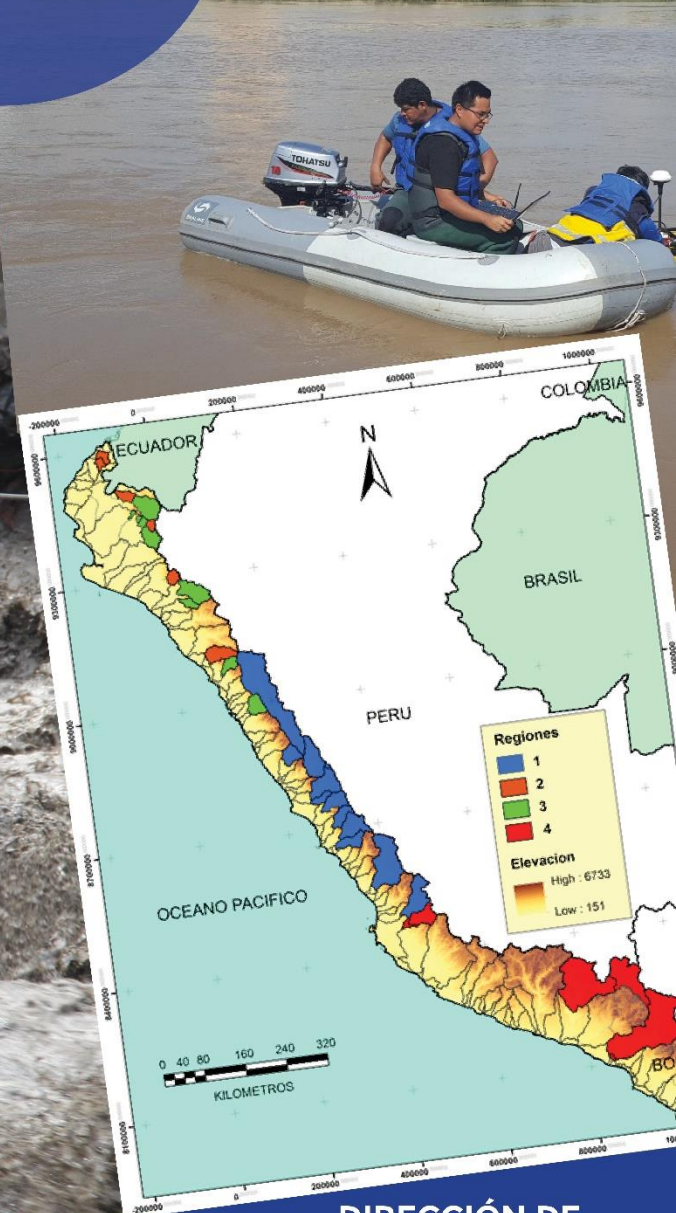




ESTIMACIÓN DE UMBRALES DE INUNDACIÓN EN LA REGIÓN HIDROGRÁFICA DEL PACÍFICO



ESTIMACIÓN DE UMBRALES DE INUNDACIÓN EN LA REGIÓN HIDROGRÁFICA DEL PACÍFICO



Dirección de
Hidrología



DIRECTORIO

Dr. Ken Takahashi Guevara
Presidente Ejecutivo del SENAMHI

Ing. Oscar Felipe Obando
Director de la Dirección de Hidrología del SENAMHI

Ph. D. Waldo Sven Lavado Casimiro
Subdirector de Estudios e Investigaciones Hidrológicas del SENAMHI

REVISOR DEL ESTUDIO

Ph. D. Waldo Sven Lavado Casimiro
Subdirector de Estudios e Investigaciones Hidrológicas del SENAMHI

RESPONSABLE DEL ESTUDIO

Christian Albert Montesinos Cáceres
Especialista en Hidrología del SENAMHI

Lima - Perú

Junio - 2018

Citar como:

Montesinos, C. Estimación de Umbrales de Inundación en la Región Hidrográfica del Pacífico. Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). Dirección de Hidrología (DHI). Junio, 2018.

Índice

Introducción	1
1. Objetivos	3
1.1. Objetivo General	3
1.2. Objetivos Específicos.....	3
2. Revisión Bibliográfica	4
2.1. Análisis de frecuencia.....	4
2.2. Selección del modelo probabilístico	4
2.3. Pruebas de bondad de ajuste.....	5
2.4. Cuantiles y período de retorno	5
2.5. Análisis regional de frecuencias	5
2.6. Análisis regional de frecuencia mediante el método de los L-Moments.....	6
2.6.1. Etapas del análisis regional de frecuencias basado en los L-Momentos	8
2.6.1.1. Análisis de datos.....	8
2.6.1.2. Cálculo de los L-Moments de las estaciones.....	8
2.6.1.3. Identificación de regiones homogéneas	12
2.6.1.4. Medida de la discordancia de una estación dentro de una región homogénea	12
2.6.1.5. Medida de heterogeneidad para aceptación de las regiones homogéneas propuestas (Tomado Núñez, <i>et al.</i> 2010)	14
2.6.1.6. Selección de distribuciones de probabilidad y cálculo de los cuantiles.....	16
3. Materiales y métodos	18
3.1. Descripción del área de estudio	18
3.2. Materiales	19
3.3. Metodología.....	20
3.3.1. Análisis exploratorio de datos.....	21
3.3.2. Supuestos del análisis regional de frecuencias	21
3.3.3. Filtrado de las observaciones empleando la medida de discordancia.....	21
3.3.4. Identificación de regiones homogéneas	22
3.3.5. Selección de una función de distribución de frecuencia para la región	22
3.3.6. Estimación de los cuantiles regionales de frecuencia y locales	22
4. Resultados y discusiones	23
4.1. Información de datos	23
4.1.1. Análisis exploratorio de datos (AED).....	24
4.1.2. Supuestos del análisis regional de frecuencias	24
4.1.2.1. Análisis de tendencia.....	25
4.1.2.2. Análisis de independencia serial	28
4.1.3. Formación de regiones homogéneas.....	29
4.1.4. Detección de datos usando medidas de discordancia.	32
4.1.5. Prueba de homogeneidad regional.....	33
4.1.6. Selección de la distribución regional de mejor ajuste y cálculo de los cuantiles.....	35
4.1.6.1. Medidas de bondad de ajuste.....	36

4.1.7.	Desarrollo de la fórmula regional de cuantiles para cuencas hidrográficas	38
4.1.8.	Desarrollo de la relación regional entre el caudal promedio anual máximo y el área de las cuencas.	39
4.1.8.1.	Desarrollo de formula regional de cuantiles para cuencas no monitoreadas.	41
4.1.9.	Determinación de umbrales de inundación.....	42
5.	Conclusiones.....	44
6.	Bibliografía	45

Lista de Figuras

Figura 01. Gráfico de comparación entre distribuciones comúnmente utilizadas mediante el método de los L-Moments.....	10
Figura 02. Diagrama representativo del indicador de discordancia basado en los L- Moments L-CV y LSK.	13
Figura 03. Ubicación de las cuencas hidrográficas seleccionadas y los sitios de medición de la corriente	18
Figura 04. Estaciones con 15 años como mínimo de longitud de registro.	23
Figura 05. Gráficos boxplots para valores de outliers	25
Figura 06. Gráficos de series de tiempo para las estaciones con tendencia significativa a un nivel de confianza 5%.....	27
Figura 07. Resumen de autocorrelaciones significativas detectadas para varios retardos a un nivel de 5% por el test de Ljun Box.....	28
Figura 08. Ratios de los L-momentos de las estaciones.....	31
Figura 09. A) Clasificación de estaciones hidrométricas mediante agrupación jerárquica y enlace de Ward, B) Clasificación de estaciones mediante K-means.....	32
Figura 10. Regiones conformadas para la Vertiente del Pacífico.....	35
Figura 11. Diagrama de L-momentos para las cuatro sub-regiones.....	37
Figura 12. Curva de crecimiento regional para las distribuciones con mejor ajuste.....	39
Figura 13. Umbrales de niveles de agua – Estación Santo Domingo.....	43

Lista de Tablas

Tabla 01: Valores críticos del indicador de discordancia D_i	14
Tabla 02: Valores críticos de heterogeneidad originales y actualizados.....	15
Tabla 03: Clasificación de regiones homogéneas y prueba de discordancia	33
Tabla 04: Resultados de la prueba de Homogeneidad, para las regiones definidas.....	34
Tabla 05: Estadístico ZDIST de varias distribuciones para cada región.....	37

Lista de Cuadros

Cuadro 1: Red de estaciones Hidrométricas	19
Cuadro 02: Resumen del test de Mann Kendall para el análisis de tendencias	25
Cuadro 03: Valores de los ratios de L-momentos de las estaciones	30
Cuadro 04: Parámetros y cuantiles regionales para las distribuciones con mejor ajuste por región....	38
Cuadro 05: Caudales promedio máximos anuales de las estaciones hidrométricas utilizadas	40
Cuadro 06: Cuantiles de varios periodos de retorno para diferentes áreas de drenaje de la región 01. ...	41
Cuadro 07: Valores de los niveles de alerta establecido por el SENAMHI.	42
Cuadro 08: Valores de periodos de retorno relacionados con los umbrales de inundación.	43

Introducción

Las inundaciones son un peligro natural importante en varios países y representan un riesgo recurrente, este peligro de inundaciones se transforma en muchas zonas en patente riesgo, como consecuencias del uso del territorio y de los asentamientos de población que la sociedad realiza en zonas potencialmente inundables. Los efectos destructivos que las inundaciones originan en el territorio hacen que, ante esa eventualidad, resulte necesario desarrollar herramientas que nos brinden información respecto a umbrales de inundación.

Se requieren estimaciones apropiadas de la magnitud y frecuencia de los eventos de flujo de caudal máximo para los proyectos de estructuras hidráulicas, así como para la gestión de recursos hídricos. Las estructuras hidráulicas como puentes, presas, alcantarillas, sistemas de recolección y drenaje, entre otros, requieren información sobre flujos de caudal de diseño, que están relacionados con eventos de flujo de caudal máximo asociados con períodos de retorno predefinidos (BESKOW et al., 2015; KUMAR et al., 2003). En la gestión de los recursos hídricos, la comprensión del comportamiento hidrológico de una cuenca dada en relación con los eventos de inundación ayuda a mitigar el impacto de los desastres naturales y el manejo del área inundada.

Los flujos de caudal de diseño sobreestimados pueden aumentar los costos de construcción; por otro lado, la subestimación de estos valores lleva a un pobre dimensionamiento de las estructuras hidráulicas. Idealmente, los flujos máximos se estiman a partir de series históricas originadas a partir del monitoreo hidrológico en la salida de interés. Sin embargo, nos encontramos con algunas limitantes; tal es el caso de la densidad insatisfactoria de las estaciones de medición que operan en nuestro país, lo que lleva al uso de métodos indirectos para estimar variables hidrológicas.

En este contexto, la regionalización hidrológica se puede aplicar en sitios medidos, con el fin de agregar información a las series existentes, así como transferirlas a lugares no monitoreados para satisfacer las demandas de datos. La aplicación del método regional de L-momento puede conducir a la selección de diferentes distribuciones de probabilidad para diversas regiones dentro de un área de estudio dada. Este trabajo aplica la metodología propuesta por Hosking y Wallis (1997) a los datos de inundación de la serie máxima anual de cuencas de la vertiente del Pacífico Peruano.

Los objetivos del estudio son identificar regiones homogéneas basadas en los L-momentos, distribuciones de frecuencia regionales adecuadas para las regiones utilizando diagramas de relación de L-momentos, estimar los cuantiles de inundación para diferentes periodos de retorno.

Diversas distribuciones de frecuencia, a saber, logística generalizada (GLOG), valor extremo general (GEV), normalización general (GNOR), Pearson tipo III (PE3), y Pareto generalizado (GPAR), son investigados.

Para la determinación de los cuantiles de inundación de varios periodos de retorno para las cuencas hidrográficas medidas en el área de estudio, es necesario desarrollar una fórmula de inundación regional utilizando distribuciones basadas en los L-Moments.

Además, para estimar las inundaciones de los períodos de retorno deseados para las cuencas hidrográficas no monitoreadas, será de interés desarrollar una fórmula de inundación regional combinando la fórmula de inundación regional con la relación regional entre la inundación pico media anual y el área de la cuenca hidrográfica; para finalmente en función de esta metodología determinar, los umbrales de inundación.

1. Objetivos

1.1. Objetivo General

- El objetivo principal de este estudio fue desarrollar una herramienta que determine los umbrales de inundación, basado en los L-momentos para generar funciones regionales destinada a estimar flujos máximos de corriente asociados con períodos de retorno de interés en cuencas de la Vertiente del Pacífico.

1.2. Objetivos Específicos

- Identificar las regiones hidrológicamente homogéneas en la Vertiente del Pacífico.
- Identificar una apropiada función de distribución que relacione la magnitud de las precipitaciones máximas con su frecuencia de ocurrencia en la Vertiente del Pacífico.
- Desarrollar una fórmula de inundación regional utilizando distribuciones de frecuencia basadas en los L-moments.

2. Revisión Bibliográfica

2.1. Análisis de frecuencia

El análisis de frecuencias es un procedimiento para estimar la frecuencia de ocurrencia o probabilidad de ocurrencia de eventos pasados o futuros. De este modo la representación gráfica de la probabilidad, con o sin suposiciones de distribuciones de probabilidad, es un método de análisis de frecuencias (Monsalve, 1995).

El objetivo del análisis de frecuencia de información hidrológica es relacionar la magnitud de los eventos con su frecuencia de ocurrencia, mediante el uso de distribuciones de probabilidad (Chow, 1964).

El análisis de frecuencias de datos hidrológicos requiere que los datos sean homogéneos e independientes. La restricción de homogeneidad asegura que todas las observaciones provengan de la misma población, por ejemplo, que la estación hidrométrica en un río no haya sido movida, que la hoya hidrográfica no se haya urbanizado, o que no se hayan colocado estructuras hidráulicas sobre la corriente principal o sus más importantes tributarios (Monsalve, 1995).

La consideración de eventos hidrológicos extremos como procesos aleatorios, implica que la variabilidad climática natural no afecta la ocurrencia de estos eventos. Un proceso estacionario significa que representa eventos independientes e idénticamente distribuidos por un modelo probabilístico que no varía a través del tiempo (Bradley & Zhao, 1997).

2.2. Selección del modelo probabilístico

Los criterios más importantes para la selección de un modelo de distribución de probabilidades en un análisis de frecuencia son que exista una teoría sólida que describa al fenómeno y que el modelo debe extraer la máxima información de los datos (Kite, 1977).

La capacidad de un modelo probabilístico para ajustarse a los datos hidrológicos depende de la flexibilidad y de la forma de la función de distribución de probabilidades. Mientras más parámetros formen un modelo, más versátil es su función de distribución de probabilidades y mejor se puede ajustar (Benjamín & Cornell, 1981).

No existe una base teórica sólida para justificar la elección de una función específica de distribución de probabilidades, tampoco existe un procedimiento teórico para determinar a un modelo probabilístico como el mejor en un análisis de frecuencia, comparando diferentes modelos probabilísticos. Es por esto que comúnmente se verifica la conveniencia o conformidad entre distintas distribuciones para seleccionar entre estas a una, basándose en el ajuste de datos disponibles, facilidad de cómputo o cálculo y la consistencia con diferentes tamaños de muestra (Zalina et al., 2002).

2.3. Pruebas de bondad de ajuste

El objetivo de una prueba de bondad de ajuste es analizar si los datos se desvían una cantidad estadísticamente significativa respecto del modelo de probabilístico (Benjamín & Cornell, 1981). Este procedimiento permite rechazar un modelo por no ser adecuado, pero no permite demostrar que un modelo probabilístico es el más idóneo (Varas, 1982). Es por esto que se recomienda utilizar pruebas de bondad de ajuste como una forma de analizar si una determinada distribución es aceptada o rechazada con un determinado nivel de significancia, pero no para seleccionar la distribución de mejor ajuste (Jara, 1986).

2.4. Cuantiles y período de retorno

Los cuantiles se definen como los valores de la variable asociados a una probabilidad de excedencia $P(X \geq x)$ o no excedencia $P(X \leq x)$, es decir, se asocian diferentes valores de la variable a diferentes probabilidades de que se exceda o no un determinado valor de la variable. (Stedinger et al., 1993). El periodo de retorno es una forma de expresar estas probabilidades. Corresponde al periodo para el cual, en promedio, la variable es igualada o excedida. Sin embargo, esto no implica que no puedan producirse dos eventos de una magnitud igual o superior en un mismo año (Varas & Bois, 1998).

Cuando la variable aleatoria representa un evento máximo anual, el periodo de retorno es el valor esperado del número de años que transcurren en promedio hasta que ocurra un evento de magnitud igual o superior. El término del período de retorno conduce a concluir que existe una cantidad determinada de años entre los eventos de tales magnitudes, cuando en efecto la probabilidad cualquier período sigue siendo la misma, independiente que ocurra tal evento en el año anterior o en un año reciente (Stedinger et al., 1993). Las expresiones para calcular los periodos de retorno son las siguientes:

$$T = \frac{1}{P} \quad (1)$$

En donde:

T: Periodo de retorno en años.

P: Probabilidad de excedencia.

2.5. Análisis regional de frecuencias

Para asociar los resultados obtenidos de un análisis de frecuencia de una estación o un grupo de estaciones a un área determinada, se debe aplicar el procedimiento conocido como análisis regional o regionalización. Este procedimiento consiste básicamente en definir regiones estadísticamente homogéneas en una zona determinada con el fin de asociar estaciones a estas regiones y poder estimar las frecuencias sobre la base de los datos hidrológicos. En cada región los resultados de los análisis de frecuencia de las estaciones pueden ser promediados para representar

la frecuencia característica de toda la región (Chow, 1964). Para determinar que un grupo de estaciones conforman una región, las estaciones deben cumplir el criterio de homogeneidad, es decir, los cuantiles estimados en todas las estaciones difieren solo por un factor de escala específico de cada zona (Bradley y Zhao, 1997).

Los análisis regionales de frecuencia son un gran desafío para la hidrología, ya que es poco común que se disponga de información suficiente y adecuada para estimar las frecuencias de eventos extremos en estaciones o lugares específicos y posteriormente asociarlas a regiones (Stedinger et al., 1993).

2.6. Análisis regional de frecuencia mediante el método de los L-Moments

Un análisis de frecuencia es una estimación de cuan frecuente un evento suele ocurrir. Usualmente la estimación de frecuencias de eventos extremos es de importancia, debido a que existen diferentes fuentes de incertidumbre de los procesos físicos que dan origen a los datos observados. Los métodos estadísticos reconocen la existencia de la incertidumbre y de sus efectos, con el fin de cuantificarlos (Hosking & Wallis, 1997).

Los procedimientos para realizar un análisis de frecuencia estadístico de un conjunto de datos son bien conocidos. Sin embargo suelen existir varias muestras de datos disponibles para un análisis. Estas observaciones pueden ser, por ejemplo, de orígenes meteorológicos o ambientales, de la misma variable y obtenidos en diferentes sitios de medición bajo diferentes condiciones. Si las distribuciones presentan similitud en las diferentes observaciones (datos), es posible obtener resultados más precisos utilizando todas las muestras en un análisis. En análisis ambientales este procedimiento es conocido como análisis regional de frecuencia, debido a que los datos utilizados corresponden a observaciones de la misma variable en una serie de lugares de medición dentro de una región o zona determinada.

Así, los principios y procedimientos del análisis regional de frecuencia son aplicables cuando existan muestras, de la misma variable, provenientes de diferentes lugares dentro de una misma región (Hosking & Wallis, 1997).

Suponiendo que las observaciones se realizan en intervalos regulares de tiempo en una región determinada, se denomina a Q como la magnitud del evento que ocurre en un momento dado en un lugar específico. Así se considera a Q como una variable aleatoria que teóricamente puede tomar cualquier valor entre cero y el infinito. El fundamento del análisis estadístico de frecuencia se basa en la distribución de probabilidad de los posibles valores de Q (Hosking & Wallis, 1997). Se denomina (x) la probabilidad de que el valor de Q no sea mayor que x , y se expresa como:

$$F(x) = P(Q \leq x) \quad (2)$$

(x) es la función de probabilidad de la distribución de frecuencia. La función inversa (F), es la función de densidad, asociada a los cuantiles de la distribución de frecuencia. Esta función inversa expresa la magnitud de un evento particular en términos de la

probabilidad de no excedencia F . El cuantil asociado al periodo de retorno T , denominado Q_T , es la magnitud de un evento tan extremo que tiene probabilidad $1/T$ de ser excedido por otro evento (Hosking & Wallis, 1997). Para eventos extremos de alta magnitud que se ubican en lo alto de la curva de la distribución de frecuencia, Q_T se expresa como:

$$Q_T = x \left(1 - \frac{1}{T} \right) \quad (3)$$

$$F(Q_T) = 1 - \frac{1}{T} \quad (4)$$

Para eventos extremos de baja magnitud que se ubican en la parte baja de la curva de la distribución de frecuencia, las expresiones de Q_T son las siguientes:

$$Q_T = x \left(\frac{1}{T} \right) \quad (5)$$

$$F(Q_T) = \frac{1}{T} \quad (6)$$

El objetivo del análisis de frecuencia es obtener estimaciones de cuantiles Q_T y probabilidades asociadas a diferentes periodos de retorno. Este periodo de retorno puede ser, por ejemplo, igual a la vida útil de una obra hidráulica. Generalmente el objetivo del análisis de frecuencia es estimar Q_T para un rango de periodos de retorno o para estimar la función de probabilidad. Si existen datos disponibles de un determinado sitio de monitoreo o estación de interés, los datos observados proveen una muestra de Q . En diversos análisis ambientales el tamaño de la muestra raramente es suficiente para estimar cuantiles de manera confiable. Generalmente, un cuantil asociado a un determinado periodo de retorno puede ser estimado de manera confiable utilizando un registro de datos de longitud n solo si $T \leq n$. Sin embargo, en diferentes aplicaciones de la ingeniería, basados en registros anuales, la condición antes mencionada no se satisface. Se han desarrollado diferentes metodologías para abordar este problema, utilizando fuentes de datos alternativos o adicionales. Una de estas metodologías es el análisis regional de frecuencia (Álvarez *et al.* 1999).

El análisis regional de frecuencias se caracteriza por aumentar la cantidad de datos para el análisis en una determinada región de estudio utilizando datos de diferentes estaciones, siendo los datos de estos sitios de medición similares en las distribuciones de frecuencia. Sin embargo, esto representa ciertos inconvenientes, ya que las distribuciones de frecuencia de los datos de diferentes estaciones no son exactamente iguales y además la magnitud de los eventos en las distintas estaciones puede no ser estadísticamente independiente. A pesar de estos inconvenientes, aplicando un análisis regional de frecuencia se pueden obtener estimaciones de cuantiles suficientemente precisas para diferentes aplicaciones (Hosking & Wallis, 1997). En este contexto existen diferentes investigaciones relacionadas con las propiedades de los algoritmos regionales de los momentos de probabilidad ponderados, que es un procedimiento de análisis regional de frecuencia basado en

el método estadístico denominado momentos de probabilidad ponderados (Greis & Wood, 1981). El método de los algoritmos regionales de momentos ponderados por probabilidad entrega resultados precisos (Cunnane, 1988). Los *LMoments* derivan de los momentos ponderados por probabilidad e incrementan la precisión de los cálculos (Hosking y Wallis, 1997).

2.6.1. Etapas del análisis regional de frecuencias basado en los L-Momentos

El procedimiento a seguir para la aplicación del análisis regional de frecuencias basado en los L-Moments contempla 4 etapas (Hosking & Wallis, 1997):

- a. Análisis de datos: Preparación de datos y construcción de una base de datos de la variable de interés.
- b. Cálculo de los L-Moments para cada una de las estaciones.
- c. Delimitación de regiones homogéneas.
- d. Selección de distribuciones de probabilidad y cálculo de los cuantiles.

A continuación se describen los pasos y procedimientos asociados al análisis regional de frecuencias mediante el método de los L-Moments.

2.6.1.1. Análisis de datos

Uno de los primeros pasos en el análisis regional de frecuencias basado en los L-Moments consiste en la preparación de datos mediante la revisión y análisis de estos. Los diferentes tipos de errores asociados a los datos deben ser identificados para posteriormente eliminar los datos erróneos o anómalos. En este contexto, se debe analizar la homogeneidad de las muestras, asumiendo que el fenómeno físico que se desea explicar mediante un modelo probabilístico es estacionario a través del tiempo y los datos no presentan correlación (Hosking & Wallis, 1997).

Por lo general existen datos disponibles de algunos o varios sitios o estaciones para realizar un análisis regional de frecuencias. En este contexto deberían realizarse diferentes verificaciones de datos. Una importante corresponde a la revisión de los datos de cada estación separadamente. Esto permite identificar valores atípicos que pueden ser explicados por errores al transcribir o transcribir los datos de una fuente a otra. Otra verificación es la revisión de la posible existencia de tendencias y cambios en los datos. Esto se puede analizar comparando los datos de diferentes estaciones mediante la aplicación de una medida de discordancia. Si una estación en particular presenta discordancia con respecto a un grupo de estaciones similares pueden existir errores en los datos (Wallis *et al.*, 1991).

2.6.1.2. Cálculo de los L-Moments de las estaciones.

Los L-moments constituyen un sistema alternativo a los momentos convencionales, para describir las formas de las funciones de distribución (Hosking, 1990). A continuación se muestra las estimaciones de los primeros

cuatro momentos de probabilidad ponderada (PWMS) propuesto por Hosking & Wallis, 1997:

$$\beta_0 = \frac{1}{N} \sum_{j=1}^N X(j) \quad (7)$$

$$\beta_1 = \frac{1}{N} \sum_{j=1}^{N-1} \left[\frac{N-j}{N(N-1)} \right] X(j) \quad (8)$$

$$\beta_2 = \frac{1}{N} \sum_{j=1}^{N-2} \left[\frac{(N-j)(N-j-1)}{N(N-1)(N-2)} \right] X(j) \quad (9)$$

$$\beta_3 = \frac{1}{N} \sum_{j=1}^{N-3} \left[\frac{(N-j)(N-j-1)(N-j-2)}{N(N-1)(N-2)(N-3)} \right] X(j) \quad (10)$$

Donde, $X(j)$ representa la serie máxima anual, con $X(1)$ siendo el valor más bajo y $X(N)$ el valor más alto, respectivamente.

La expresión general de los L-momentos desarrollados para los primeros cuatro términos son:

$$l_1 = \beta_0 \quad (11)$$

$$l_2 = 2\beta_1 - \beta_0 \quad (12)$$

$$l_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (13)$$

$$l_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (14)$$

Los L-Moments poseen diferentes ventajas sobre los momentos convencionales, ya que pueden caracterizar a una variedad más amplia de distribuciones y son más robustos frente a valores atípicos o anormales. Además, diferentes estudios que han comparado a los *L-Moments* con los momentos convencionales, han establecido que al utilizar los L-Moments se obtienen estimaciones más precisas, es decir, se obtienen estimaciones más cercanas a los valores observados (Hosking & Wallis, 1997).

a) Propiedades de los L-Moments

Los L-Moments de primer y segundo orden (l_1 y l_2), y las razones entre los L-Moments, son las expresiones que se utilizan para estimar los parámetros de las distribuciones de probabilidad asociadas a los L-Moments.

Las principales propiedades de estas expresiones según Hosking (1988), son las siguientes:

1) Existencia: Si el promedio de una distribución existe, entonces todos los L-Moments de esa distribución existen.

2) Unicidad: Si el promedio de una distribución existe, entonces los L-Moments caracterizan solamente a esa distribución. No existen dos distribuciones con los mismos valores de los L-Moments.

Sobre la base de las características y propiedades antes mencionadas de los L-Moments, estos se utilizan para ajustar muestras a distintos tipos de distribuciones. Para representar gráficamente las diferencias entre otras distribuciones ajustadas, comúnmente se utiliza un gráfico que incluye todas las distribuciones ajustadas, como lo muestra la Figura 01, donde se observan diferentes distribuciones en un gráfico de L-Asimetría (τ_3) versus L-Curtosis (τ_4).

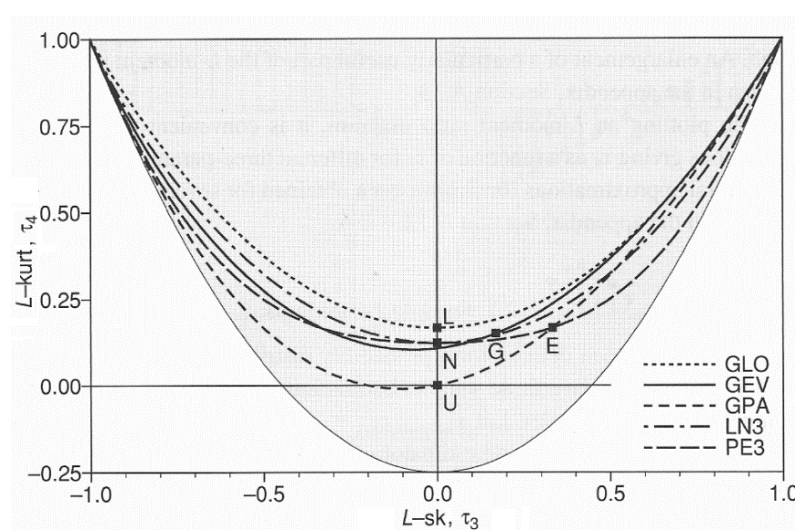


Figura 01. Gráfico de comparación entre distribuciones comúnmente utilizadas mediante el método de los L-Moments.

Fuente: Hosking y Wallis, 1997.

Figura basada en L-Asimetría y L-Curtosis. Las distribuciones representadas son: E: Exponencial, G: Gumbel, L: Logística, N: Normal, U: Uniforme, GPA: Generalizada de Pareto, GEV: Generalizada de valores extremos, GLO: Generalizada Logística, LN3: Log normal y PE3: Pearson Tipo III.

Como se muestra en la Figura 01, las distribuciones de dos parámetros son representadas mediante puntos. Estas distribuciones solo varían en sus parámetros de escala y de localización. Las distribuciones representadas mediante líneas punteadas corresponden a distribuciones de tres parámetros (localización, escala y forma). Las distribuciones que tienen más de un parámetro de forma abarcan comúnmente dos áreas dimensionales del gráfico (Hosking y Wallis, 1997).

b) L- Moments muestrales

Los L-Moments están definidos para distribuciones de probabilidad, pero en la práctica generalmente son estimados de una muestra de datos finita. El cálculo de los L-Moments se obtiene de una muestra de tamaño “n” ordenada en forma ascendente (Landwehr et al., 1979).

Un estimador de los momentos ponderados por probabilidad l_r es:

$$l_r = N^{-1} \left(\frac{N-1}{r} \right)^{-1} \sum_{j=r+1}^N \left(\frac{j-1}{r} \right) X_{j:n} \quad (14)$$

Esta expresión desarrollada para los primeros términos es:

$$l_1 = \left(\frac{N}{1} \right)^{-1} \sum_{i=1}^N X_{(i)} \quad (15)$$

$$l_2 = \frac{1}{2} \left(\frac{N}{2} \right)^{-1} \sum_{i=1}^N \left\{ \left(\frac{i-1}{1} \right) - \left(\frac{N-i}{1} \right) \right\} X_{(i)} \quad (16)$$

$$l_3 = \frac{1}{3} \left(\frac{N}{3} \right)^{-1} \sum_{i=1}^N \left\{ \left(\frac{i-1}{2} \right) - 2 \left(\frac{i-1}{1} \right) \left(\frac{N-i}{1} \right) + \left(\frac{N-i}{2} \right) \right\} X_{(i)} \quad (17)$$

$$l_4 = \frac{1}{4} \left(\frac{N}{4} \right)^{-1} \sum_{i=1}^N \left\{ \left(\frac{i-1}{3} \right) - 3 \left(\frac{i-1}{2} \right) \left(\frac{N-i}{1} \right) + 3 \left(\frac{i-1}{1} \right) \left(\frac{N-i}{2} \right) - \left(\frac{N-i}{3} \right) \right\} X_{(i)} \quad (18)$$

En donde:

l_r : PWD (Momentos ponderados por probabilidad) de orden r.

N: Número total de observaciones.

X_i : elemento i-esimo de la serie.

El primer L-Moments es el promedio o una medida de localización y puede tomar cualquier valor. El segundo L- Moments es una medida de dispersión de los datos respecto a la media y toma valores mayores o igual a cero. Dividiendo los L-Moments de mayor orden por la medida de dispersión, se obtiene las siguientes razones:

$$\tau_r = \frac{l_r}{l_2} \quad (19)$$

Estas razones son valores adimensionales, independientes de las unidades de medida de los datos. τ_3 corresponde a la medida de la asimetría y τ_4 corresponde a la medida de la curtosis, denominados respectivamente L-

Asimetría y L-Curtosis. El valor análogo del coeficiente de variación en los LMoments, denominado L-CV, está definido por la siguiente expresión:

- Coeficiente de variación L-CV:

$$\tau = \frac{l_2}{l_1} \quad (20)$$

- Coeficiente de Asimetría L-SK:

$$\tau_3 = \frac{l_3}{l_2} \quad (21)$$

- Coeficiente de Curtosis L-CK:

$$\tau_4 = \frac{l_4}{l_2} \quad (22)$$

Los L-Moments se pueden interpretar directamente como medidas de escala y forma de una distribución de probabilidad. Así los L-Moments pueden ser utilizados para estimar parámetros para ajustar un tipo de distribución a una muestra (Jaiswal et al, 2003).

2.6.1.3. 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).

2.6.1.4. Medida de la discordancia de una estación dentro de una región homogénea

Dado un conjunto de estaciones, un adecuado criterio en la formulación de regiones homogéneas consiste en que las estaciones que la integran no presenten propiedades que las hagan discordantes con el resto de las estaciones dentro de la región homogénea. Hosking & Wallis (1997) presentan una medida de discordancia, la cual permite evaluar el grado en que los L-Moments de una estación, se apartan significativamente del patrón medio de los L-Moments regionales. Para ello, los autores construyen una elipse concéntrica cuyas coordenadas corresponden a un par de puntos (por

ejemplo, sobre una gráfica L-CV vs L-SK) con valores correspondientes a los LCV y L-SK regionales. La elipse representa el eje mayor y menor, escogidos para dar el mejor ajuste a los datos, determinados por la matriz muestral de covarianza de los L-Moments *-ratios* de los sitios.

Formalmente, ese grado de alejamiento del centro de la elipse, se mide en términos de una prueba de discordancia D , la cual sirve para identificar estaciones cuyos valores muestrales de los L-Moments difieren significativamente de las otras estaciones de la supuesta región homogénea. Esta medida de discordancia puede ser utilizada en el proceso de control de calidad de los datos e indica los sitios dentro de la región donde los datos deben ser sometidos a un análisis de calidad más profundo.

Se considera que el vector de L-Moments (L-CV, L-SK, L-CK) de una estación determinada constituye un punto en un espacio tridimensional, por lo que un grupo de estaciones producirá una nube de puntos en este espacio. Cualquier punto que se ubique lejos del centro de gravedad del conjunto de éstos deberá ser considerado como discordante.

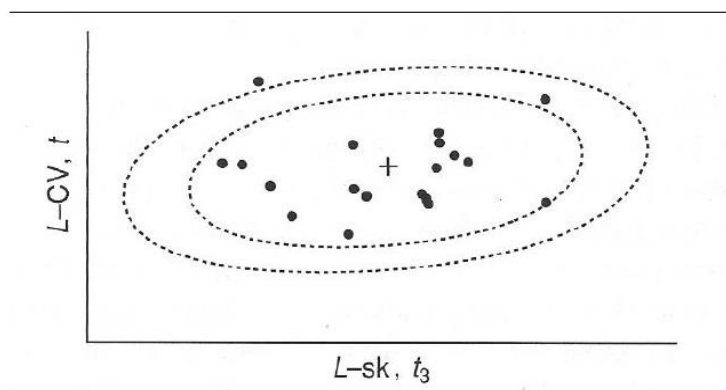


Figura 02. Diagrama representativo del indicador de discordancia basado en los L-Moments L-CV y LSK.

Fuente: Hosking y Wallis, 1997.

Para saber si una estación es discordante, es necesario conocer el número de estaciones dentro de la región homogénea que se está analizando. Para un determinado número de estaciones existe un valor crítico de D . Sobre ese valor, la estación es considerada discordante.

En la Figura 02 se observa un diagrama basado en L-CV y L-SK. El centro de la nube de puntos se identifica con el signo +, el cual está formado por las coordenadas de los promedio de L-CV y L-SK. Además se observan dos elipses concéntricas, las que ejemplifican las regiones críticas de discordancia dadas por la matriz de covarianza muestral de los ratios muestrales de los L-Moments de las estaciones. Los puntos (estaciones) discordantes son las que se ubican fuera de la elipse externa. El indicador de discordancia basado en los L-Moments, según lo planteado por Hosking & Wallis (1997), se expresa como:

$$D_i = \frac{1}{3} \cdot (u_i - U)^T \cdot A^{-1} \cdot (u_i - U) \quad (23)$$

Dónde:

$$u_i = [\tau^i \cdot \tau_3^i \cdot \tau_4^i]^T \quad (24)$$

$$U = N^{-1} \sum_{i=1}^N u_i \quad (25)$$

$$A = (N - 1)^{-1} \sum_{i=1}^N (u_i - U)(u_i - U)^T \quad (26)$$

N representa el número de estaciones que formen a un grupo en análisis. El vector transpuesto formado por τ , τ_3 y τ_4 del sitio i se expresa como u_i . Por otra parte, U es el promedio del grupo de estaciones y A es la matriz de covarianza de la muestra. La región crítica del indicador de discordancia D_i depende del número de estaciones que formen una región (Hosking y Wallis, 1997). Los valores de D_i sugeridos para considerar una estación discordante se muestran en la Tabla 01.

Tabla 01: Valores críticos del indicador de discordancia D_i .

Número de estaciones en la región	Valor crítico D_i
5	1,333
6	1,648
7	1,917
8	2,140
9	2,329
10	2,491
11	2,632
12	2,757
13	2,869
14	2,971
15	3,000

Fuente: Hosking y Wallis, 1997

2.6.1.5. Medida de heterogeneidad para aceptación de las regiones homogéneas propuestas (Tomado Núñez, *et al.* 2010)

El supuesto básico del análisis regional de frecuencias es que la distribución subyacente para un conjunto de sitios es la misma, dentro de una región homogénea. Por lo tanto, se requiere de una medida de este grado de homogeneidad, o de manera complementaria, en términos de la heterogeneidad que esta región presenta. Medidas de heterogeneidad como la estadística $H1$ han sido desarrolladas por Hosking & Wallis (1997) como un indicador del grado de heterogeneidad en los *LMoments-ratios* para un grupo de sitios/estaciones. La estadística $H1$ mide la variabilidad relativa del coeficiente de Variación (L-CV) observado a partir de la muestra, y es utilizada

para analizar la heterogeneidad en cada una de las regiones homogéneas propuestas (Wallis et al, 2007).

Existen dos medidas adicionales de heterogeneidad, denominadas $H2$ y $H3$. Sin embargo, no se recomienda su utilización debido a que se ha comprobado que presentan una notable falta de potencia estadística (Viglioni et al, 2007). El fundamento teórico y procedimiento de cálculo de la medida de heterogeneidad $H1$ está ampliamente descrita por Hosking y Wallis (1986).

El criterio de homogeneidad utilizado se basa en Wallis et al (2007), quienes sugieren los siguientes rangos, a diferencia de los rangos originales descritos por Hosking y Wallis (1997), presentados en la Tabla 02:

Tabla 02: Valores críticos de heterogeneidad originales y actualizados

Heterogeneidad	Hosking y Wallis (1997)	Wallis et al (2007)
Homogénea	$H < 1$	$H < 2$
Posiblemente heterogénea	$1 < H < 2$	$2 < H < 3$
Heterogénea	$H > 2$	$H > 3$

Fuente: Hosking & Wallis (1997) y Wallis et al (2007)

Esta determinación se fundamenta en que los límites propuestos originalmente por Hosking y Wallis (1997), sólo tenían en cuenta la variabilidad estadística. Sin embargo, las mediciones de precipitación habitualmente contienen otras fuentes de variación, tales como el movimiento de estaciones durante sus años de funcionamiento, cambio de operadores, datos faltantes surgidos de reportes inconsistentes, falta de atención para medir la precisión y condiciones específicas del sitio, debido a construcciones, crecimiento de árboles alrededor de la estación, todos factores que incrementan en algún grado la heterogeneidad natural de las regiones, por muy homogéneas que sean desde el punto de vista meramente estadístico.

La desviación estándar ponderada de los L-CV muestrales se define como:

$$V = \left\{ \frac{\sum_{i=1}^N (t^{(i)} - t^R)^2}{\sum_{i=1}^N N_i} \right\}^{1/2} \tag{27}$$

Las simulaciones basadas en la distribución kappa permiten obtener valores de V para cada región simulada. Además con las simulaciones se puede determinar el promedio y la desviación estándar de V (μ_v y σ_v). Según lo planteado por Hosking y Wallis (1997), el indicador de homogeneidad se expresa como:

$$H = \frac{V - \mu_v}{\sigma_v} \tag{28}$$

Se considera una región homogénea si $H < 2$, posiblemente heterogénea si $2 < H < 3$ y heterogénea si $H > 3$. Sin embargo estos valores no son estrictos (Wallis, 2007).

2.6.1.6. Selección de distribuciones de probabilidad y cálculo de los cuantiles

Para la selección de la mejor distribución de probabilidad hay diferentes métodos para evaluar la bondad de ajuste de una distribución asociada a los datos de una muestra, como por ejemplo, el test chi cuadrado y el test Kolmogorov-Smirnov. Estos se basan en los momentos ponderados por probabilidad y en los *L-Moments* y pueden ser adaptados para utilizarse en análisis de regiones específicas. Para el caso de los L- Moments existe un estadígrafo basado en la dispersión de τ_3 y τ_4 (Cong *et al.*, 1993), el cual se relaciona con diferentes distribuciones. Según lo planteado por Hosking & Wallis (1997), se recomienda utilizar un estadígrafo que relaciona directamente los promedios regionales de los *LMoments*, denominado ZDIST. Este se basa en los *L-Moments* (L-SK y L-CK) (Anli *et al.*, 2007). Mediante una prueba de bondad de ajuste basada en el ZDIST se comprueba objetivamente cuando un conjunto de distribuciones se ajusta a los datos aceptablemente (Álvarez, *et al.*, 1999). De esta forma se puede desestimar el uso de ciertas distribuciones que no se ajusten aceptablemente, pero este indicador por sí solo no tiene la capacidad para identificar a una distribución como la mejor (Hosking & Wallis, 1997).

La bondad de ajuste basada en el ZDIST determina si la medida de L-SK y L-CK simulados de una determinada región homogénea se ajustan adecuadamente a L-SK y L-CK de los datos observados. Es así como según lo planteado por Hosking & Wallis (1997), la expresión del estadígrafo Z es la siguiente:

$$Z^{DIST} = \frac{(\tau_4^{DIST} - \bar{t}_4 + B_4)}{\sigma_4} \quad (29)$$

$$\sigma_4 = \left[(N_{sim} - 1)^{-1} \left\{ \sum_{m=1}^{N_{sim}} (t_4^{(m)} - t_4^R)^2 - N_{sim} B_4^2 \right\} \right]^{\frac{1}{2}} \quad (30)$$

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} (t_4^{(m)} - t_4^R) \quad (31)$$

En donde:

σ_4 , Desviación estándar de L-curtosis regional

B_4 , L-Curtosis regional

τ_4^{DIST} , L-Curtosis de la distribución

\bar{t}_4 , L-Curtosis regional

Las expresiones relacionadas con las distribuciones se obtienen de acuerdo a la simulación para cada una de las cinco distribuciones que se van a utilizar. Según lo planteado por Hosking y Wallis (1997), se considera que el ajuste de una distribución es aceptable o adecuado si el valor del estadístico Z es cercano a cero. En este contexto, un valor razonable es el asociado al grado de significancia del 95%, es decir, $|Z_{DIST}| \leq 1,64$.

La estimación de los cuantiles se realiza una vez escogida una función de distribución, aplicando la prueba de bondad de mejor ajuste. Se procede a la estimación de los cuantiles asociados a un cierto período de retorno, para cada una de las estaciones.

3. Materiales y métodos

3.1. Descripción del área de estudio

La zona de estudio comprende 43 áreas de drenaje ubicadas dentro de la vertiente del Pacífico Peruano, que fueron delimitados a partir de la ubicación de estaciones hidrométricas; Los ríos generalmente fluyen de este a oeste desde los Andes hacia el Océano Pacífico y presentan pendientes desnudas y empinadas de 4 a 9% con cascadas intermedias de entre 500 y 16,000 km²; que provocan un aumento significativo, inundaciones y erosión durante grandes episodios de lluvia. Esta región presenta una amplia gama de tipos climáticos desde clima árido-tropical a tundra, lo que da como resultado gradientes inalterados y latitudinales en las condiciones climáticas.

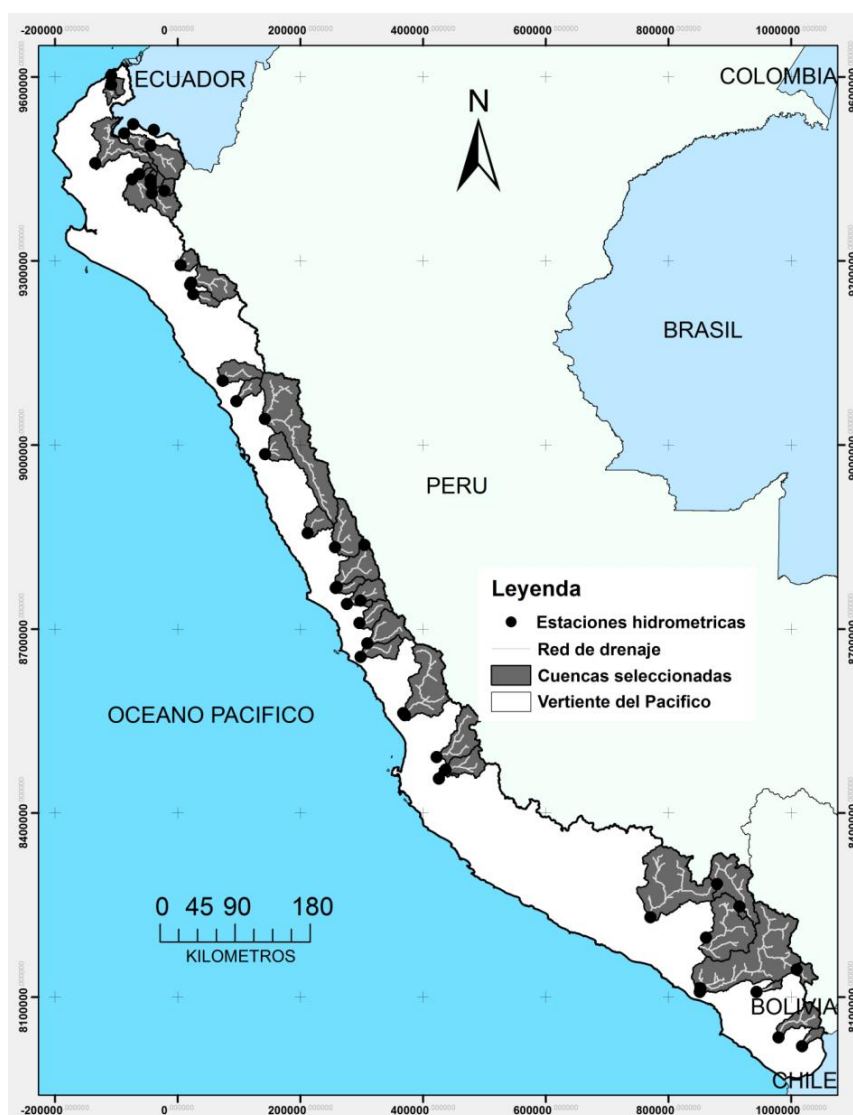


Figura 03. Ubicación de las cuencas hidrográficas seleccionadas y los sitios de medición de la corriente

Fuente: Elaboración propia

Mientras que cerca de la costa, las condiciones climáticas secas están fuertemente restringidas por las condiciones oceánicas que se caracterizan por una surgencia permanente al sur de $\sim 5^\circ$ S, las montañas experimentan más la influencia de variaciones estacionales en los patrones de circulación a gran escala (es decir, Zona de convergencia Intertropical, ITCZ) y el anticiclón Southern Pacific (SPA). Además, en escala de tiempo interanual, los eventos de lluvia anómala sobre el Pacífico Peruano están relacionados con la Oscilación del Sur El Niño (ENSO) (Rau et al., 2016).

3.2. Materiales

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 altitudinal en las áreas de drenaje elegidas, 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/>.

La delimitación de las cuencas seleccionadas en el estudio se realizó con la herramienta Arc Hydro Tools. Así mismo; se ha determinado los parámetros fisiográficos (Área de drenaje, longitud de la red de drenaje, pendiente del cauce y pendiente media de las cuencas) con el apoyo del programa ArcGis.

3.2.2 Información de datos hidrométricos observados

Se han identificado estaciones hidrométricas administradas por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) y la Autoridad Nacional del Agua (ANA), emplazadas a lo largo de la vertiente del Pacífico. En el Cuadro 01 se presenta la red de estaciones utilizadas en el presente estudio; las cuales se encuentran libres de supuestos, con información de caudales diarios superior a 15 años y sus características de ubicación.

Cuadro 01: Red de estaciones hidrométricas

N°	Estación	Id Station	Ubicación		Área (Km ²)
	Nombre		Latitud	Longitud	
1	Cahua	1	-10.55	-77.22	2974.84
2	Malvados	4	-10.34	-77.63	1372.25
3	Puente magdalena	111151	-11.34	-76.85	188.23
4	Puente tumbes	200201	-3.58	-80.47	581.28
5	El tigre	200202	-3.72	-80.47	423.96
6	Puente Sullana	200304	-4.87	-80.71	3377.99
7	Paraje Grande Quiroz	200311	-4.62	-79.9	2324.84
8	Los encuentros	200313	-4.43	-80.28	830.76
9	Puente Paltashaco	200403	-5.12	-79.9	142.31

10	Malacasi	200414	-5.32	-79.88	1430.03
11	Chillique	200415	-5.03	-80.07	158.91
12	San pedro	200416	-5.11	-80.17	2387.28
13	Teódulo Peña	200423	-5.18	-79.9	341.56
14	Barrios	200424	-5.28	-79.7	422.83
15	Puchaca	200801	-6.37	-79.47	739.75
16	Carhuaquero	200903	-6.66	-79.33	16.14
17	Racamuri	200907	-6.63	-79.32	2371.72
18	Batan	201001	-6.8	-79.28	633.68
19	Quirihuac	201401	-8.08	-78.87	1879.65
20	Huacapongo	201501	-8.38	-78.67	909.58
21	San jacinto	201901	-9.17	-78.25	1376.59
22	Laguna Surasaca	202601	-10.52	-76.78	97.33
23	Sayán	202608	-11.13	-77.2	2753.60
24	Casa blanca	202609	-11.15	-77.22	711.52
25	Santo Domingo	202701	-11.38	-77.05	1660.04
26	Puente Huarabi	202806	-11.67	-76.87	1272.46
27	Chacrasana	202908	-11.96	-76.75	2392.20
28	Manchay bajo	203005	-12.17	-76.85	1455.13
29	Socsi	203302	-12.99	-76.22	5790.85
30	Letrayoc	203602	-13.65	-75.72	3088.81
31	Huamani	203702	-13.83	-75.58	1827.90
32	La Achirana	203703	-13.97	-75.68	308.14
33	Puente Sibayo	204604	-15.49	-71.46	4494.64
34	Huatiapa	204617	-15.99	-72.47	8744.48
35	Sumbay	204702	-15.82	-71.12	171.57
36	Charcani	204710	-16.28	-71.62	5957.90
37	Ilabaya	204806	-17.06	-70.83	338.83
38	Chucarapi	204903	-17.08	-71.7	114.32
39	La tranca	205103	-17.73	-70.48	1920.83
40	Aguas calientes	205201	-17.85	-70.12	532.17
41	Condorcerro	211402	-8.63	-78.25	10403.98
42	Puente Santa Rosa	240108	-17.03	-71.69	12144.35
43	Pasto grande	1318612001	-16.72	-70.22	802.31

Fuente: Elaboración propia

3.3. Metodología

Una vez conformada las series de datos anuales de caudales máximos diarios para todas las estaciones, para cumplir con los objetivos se continua con la siguiente secuencia de trabajo, lo cual consiste en realizar el análisis exploratorio para detectar los datos atípicos, verificar los supuestos del análisis regional principalmente para determinar la estacionariedad, homogeneidad e independencia serial de los datos, realizar el análisis regional de frecuencias aplicando la técnica de L-moments, obtención de los cuantiles de inundación para diferentes periodos de retorno y finalmente determinar los umbrales de inundación en función del área de drenaje de las cuencas y la frecuencia (periodo de retorno).

3.3.1. Análisis exploratorio de datos

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

Cabe señalar que en las series anuales de caudales máximos diarios, 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 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. La homogeneidad de las series (libre de cambios en la media) se corrobora con el test no paramétrico de Pettit. Mientras la independencia serial y la aleatoriedad que se debe a procesos naturales se verifica con el test de Ljun Box. Las series que cumplan los tres supuestos pasan a la siguiente etapa, y los que no cumplen serán excluidos del ARF basado en L - moments.

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-moments, que consiste en identificar las estaciones que son groseramente discordantes con el grupo como un todo. Se determina la medida de discordancia D para cada estación lo cual mide la similitud entre las distribuciones de frecuencia de la estaciones, visualizada en términos de sus L-moments, detectando las estaciones que difieren significativamente del resto en función a sus L-CV, L-asimetría y L-curtosis. Los valores de D calculados serán comparados con un valor crítico que depende del número de estaciones que conforman la región. Para su cálculo se utiliza el paquete ImomRFA en el programa R.

La medida de discordancia 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) 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. (b) En caso se encuentre varias estaciones discordantes se continuará con la formación de regiones homogéneas, sin excluir las estaciones discordantes. Una vez conformada las regiones; 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-moments y el estadístico Z^{DIST} a un nivel de confianza del 90 por ciento.

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 - moments, implementada en el paquete ImomRFA.

Los cuantiles se definen como los montos de caudales 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.

4. Resultados y discusiones

4.1. Información de datos

Para el desarrollo de la primera etapa del análisis de datos de caudales diarios, se realizó una revisión de los mismos y la conformación de la base de datos. Además se hizo una comprobación de la confiabilidad de los datos, identificando los valores atípicos o anormales (outliers). Para la creación de la base de datos, uno de los criterios iniciales utilizados fue que las estaciones con registros menores a 15 años fueron descartadas.

Como punto de partida se identificó en total 150 estaciones hidrométricas administradas por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) y la Autoridad Nacional del agua (ANA), situados dentro de la vertiente del Pacífico; se recopiló información de caudales diarios para luego conformar las series de caudales máximos que serán útiles en el estudio de análisis regional de frecuencias de los caudales máximos.

En la Figura 04 se presenta las estaciones que cumplen con el criterio de 15 años como mínimo de longitud de registro de caudal por cada estación (66 estaciones). En relación a algunos autores que recomiendan sobre la base de longitud mínima de registro de variables hidrológicas, por ejemplo la IACWD (Interagency Advisory Comité on wáter Data), recomienda utilizar como mínimo 25 años de longitud de registros (IACWD, 1982). Sin embargo en este estudio la mayoría de estación no cumple con dicho criterio.

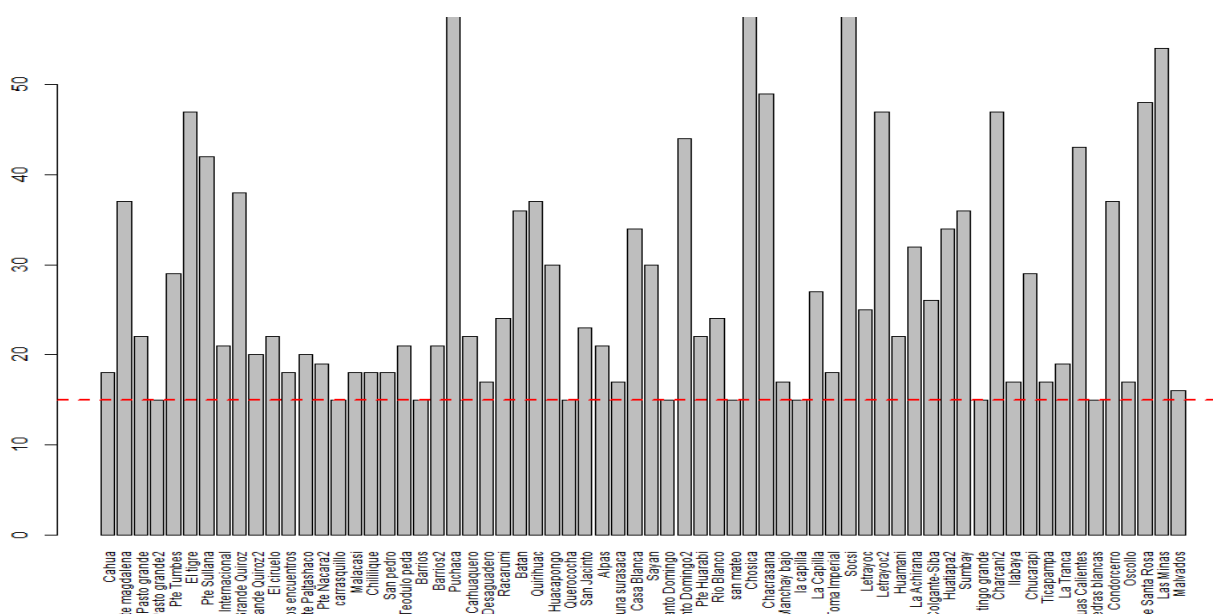


Figura 04. Estaciones con 15 años como mínimo de longitud de registro.

Fuente: SENAMHI.

4.1.1. Análisis exploratorio de datos (AED)

La mayoría de las técnicas de AED son de carácter gráfico con algunas técnicas cuantitativas. La razón de la fuerte dependencia de los gráficos es que, por su propia naturaleza, el papel principal de AED es una exploración de mente abierta, donde los gráficos permiten una mejor visión de los datos, obteniendo nuevos conocimientos y revelando sus secretos estructurales. En combinación con las capacidades de reconocimiento de patrones del analista, los gráficos constituyen una herramienta poderosa para llevar esto a cabo.

El Box Plot muestra los tres cuartiles, en una caja rectangular alineados horizontal o verticalmente. La caja, junto con los valores mínimos y máximos, que se muestran en los extremos a ambos lados de la caja de los puntos medios de sus extremidades, constituye la gráfica de caja y bigotes, como a veces se llama. Las señales numéricas se organizan así: mínimo y el máximo. Juntos constituyen un resumen de cinco números. Los valores mínimo y máximo podrán ser sustituidos por el 5 y 95.

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.

Los valores demasiado alejados del promedio fueron eliminados para evitar contaminar demasiado las estimaciones con datos evidentemente erróneos; Además se corroboró que aquellos valores atípicos que ocurrieron durante el evento El Niño fueran considerados en el estudio.

De la Figura 05, se puede catalogar que las estaciones: Puente Tumbes, el Tigre, Pardo de Zela, Puente Nácara, Sánchez Cerro, Puente Carretera, y Huatiapa; presentan datos atípicos por superar el rango intercuartil según los diagramas de boxplots. También se visualiza que las observaciones en todas las estaciones no tienen una distribución simétrica con respecto a la mediana.

4.1.2. Supuestos del análisis regional de frecuencias

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

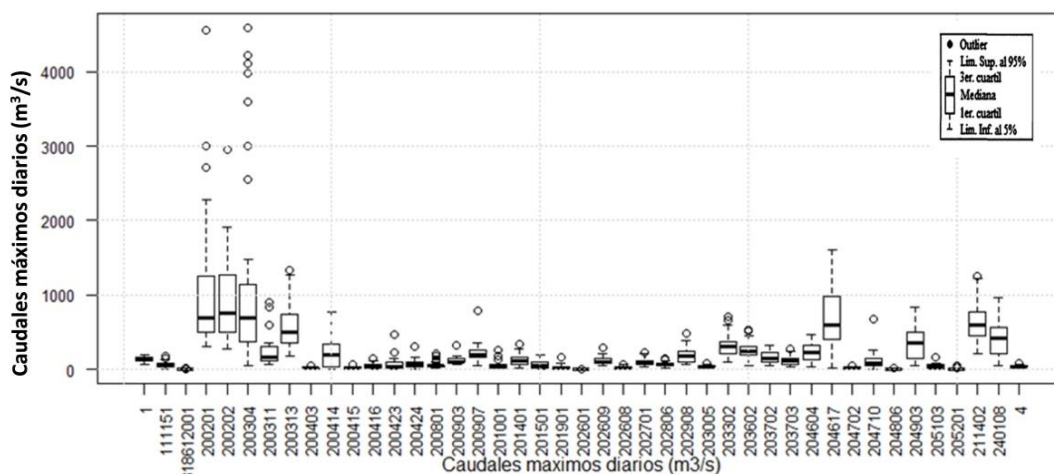


Figura 05. Gráficos boxplots para valores de outliers
Fuente: Elaboración propia

4.1.2.1. 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 preblanqueo (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.

El Cuadro 02 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: HO (Hipótesis Nula): No existe una tendencia en la serie
Ha (Hipótesis Alternativa): Hay una tendencia en la serie

Cuadro 02: Resumen del test de Mann Kendall para el análisis de tendencias

Estación	Tau Kendall	P-value	Pendiente Sen	Intercepto	Hipótesis Nula
Cahua	0.07	0.71	-0.63	188.80	Aceptado
Puente Magdalena	-0.21	0.08	-0.63	79.57	Aceptado
Pasto Grande	-0.20	0.22	-0.02	5.50	Aceptado
Puente Tumbes	0.17	0.21	14.96	-61.00	Aceptado
El Tigre	0.20	0.05	8.21	263.28	Aceptado
Puente Sullana	-0.18	0.10	-14.00	1246.35	Aceptado

Puente Internacional	0.17	0.31	2.72	84.38	Aceptado
Paraje Grande Quiroz	-0.18	0.29	-1.81	232.71	Aceptado
El Ciruelo	0.10	0.57	1.85	330.95	Aceptado
Los Encuentros	0.16	0.39	7.13	12.63	Aceptado
Puente Paltashaco	-0.09	0.62	-0.30	30.14	Aceptado
Puente Nácara2	-0.10	0.60	-8.48	769.18	Aceptado
Malacasi	-0.07	0.71	-4.14	450.42	Aceptado
Chillique	-0.21	0.27	-0.28	39.56	Aceptado
San Pedro	-0.07	0.71	0.15	23.53	Aceptado
Teódulo Peña	0.05	0.77	-0.45	66.00	Aceptado
Barrios2	-0.02	0.92	-0.35	91.43	Aceptado
Puchaca	0.03	0.77	0.07	43.04	Aceptado
Racarumi	0.16	0.29	3.25	-31.71	Aceptado
Quirihuac	-0.10	0.41	-0.30	129.07	Aceptado
Huacapongo	-0.15	0.25	-0.71	79.33	Aceptado
San Jacinto	0.27	0.08	1.38	-33.45	Aceptado
Alpas	-0.29	0.07	-6.38	535.94	Aceptado
Laguna Surasaca	-0.22	0.26	-0.03	4.19	Aceptado
Casa Blanca	-0.08	0.53	-0.50	117.24	Aceptado
Santo Domingo2	0.01	0.90	0.04	88.18	Aceptado
Puente Huarabi	0.21	0.19	1.00	61.68	Aceptado
Chacrasana	0.03	0.77	0.30	170.07	Aceptado
Manchay bajo	-0.37	0.05	-0.75	57.36	Aceptado
La Capilla	-0.26	0.07	-3.28	242.07	Aceptado
Toma Imperial	-0.19	0.30	-11.35	910.82	Aceptado
Socsi	-0.06	0.44	-0.63	320.03	Aceptado
Letrayoc	-0.12	0.44	-2.08	405.25	Aceptado
Letrayoc2	0.10	0.33	1.36	216.07	Aceptado
Huamani	0.00	1.00	-0.48	160.87	Aceptado
La Achirana	-0.19	0.13	-1.29	175.57	Aceptado
Puente Colgante-Sibayo	-0.13	0.39	-5.00	453.41	Aceptado
Huatiapa2	-0.14	0.27	-6.00	782.75	Aceptado
Sumbay	0.01	0.98	0.10	16.68	Aceptado
Ilabaya	0.18	0.34	-0.02	5.65	Aceptado
Chucarapi	0.01	0.98	3.45	204.90	Aceptado
La Tranca	0.18	0.32	1.86	-76.82	Aceptado
Aguas Calientes	-0.03	0.76	-0.02	5.27	Aceptado
Condorcerro	-0.07	0.57	0.12	584.16	Aceptado
Puente Santa Rosa	-0.13	0.21	-2.48	577.29	Aceptado
Malvados	0.33	0.09	0.90	-45.14	Aceptado
Carhuaquero	0.47	0.00	4.06	-45.84	Rechazado
Desaguadero	-0.67	0.00	-2.07	130.67	Rechazado
Batan	0.25	0.04	0.75	-0.15	Rechazado
Sayán	0.35	0.01	0.63	1.90	Rechazado
Río Blanco	-0.43	0.00	-0.47	41.86	Rechazado
Chosica	-0.21	0.01	-0.88	161.63	Rechazado
Charcani2	-0.21	0.04	-1.24	121.36	Rechazado
Ticapampa	0.68	0.00	0.27	-10.72	Rechazado
Oscollo	-0.47	0.01	-1.66	87.64	Rechazado
Las Minas	0.23	0.02	0.24	6.23	Rechazado

Fuente: Elaboración propia

De acuerdo al Cuadro 02, la hipótesis nula: no existe una tendencia en la serie fue rechazada en 10 estaciones, cabe indicar que la estacionariedad está influenciada por la cantidad de datos que presenta cada estación; por lo que existe la posibilidad que algunas series presenten tendencias por su limitada longitud de tiempo.

A continuación se presenta las estaciones que presentan tendencia.

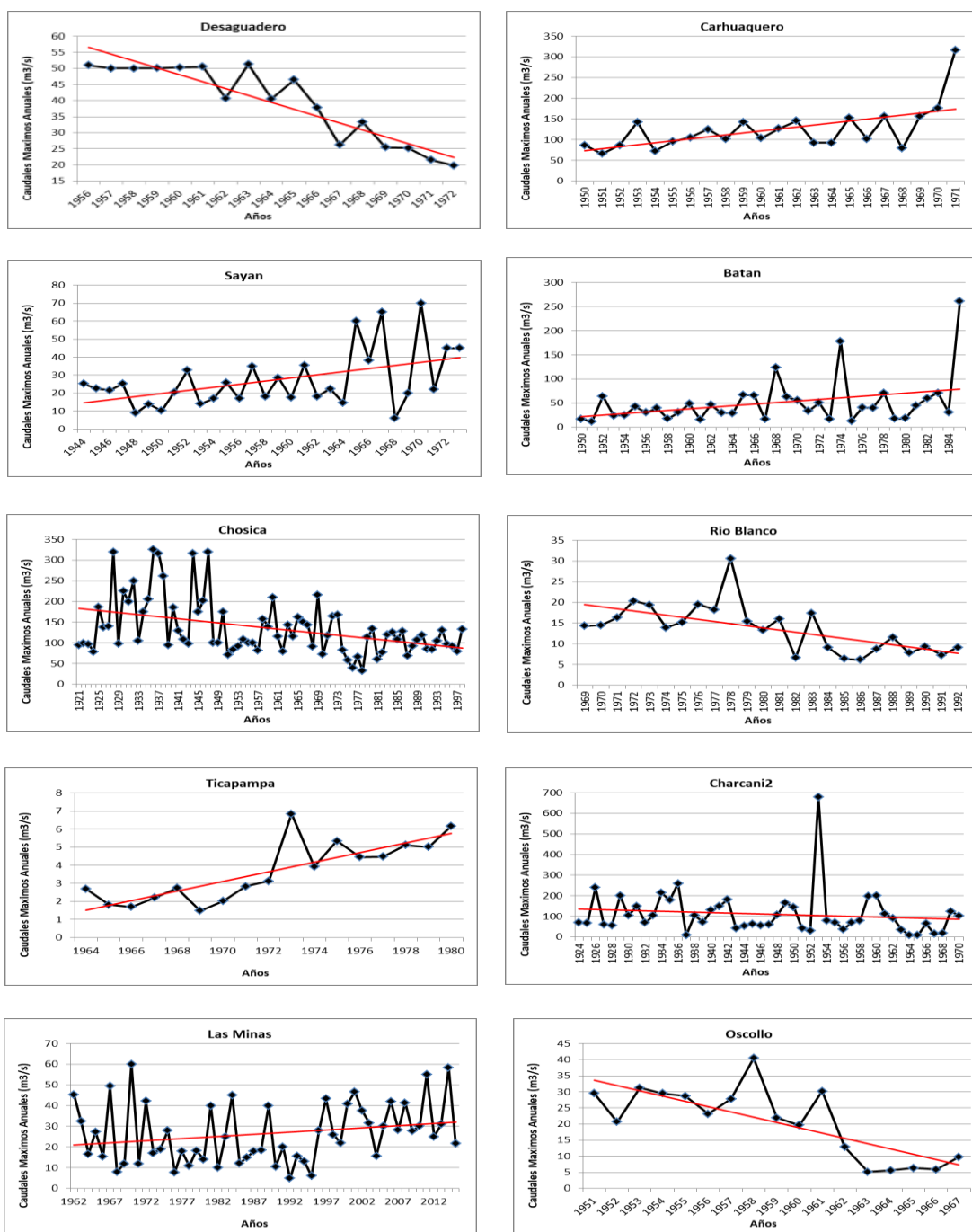


Figura 06. Gráficos de series de tiempo para las estaciones con tendencia significativa a un nivel de confianza 5%.

Fuente: Elaboración propia

4.1.2.2. 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 07, se aprecia el resumen de la prueba de Ljung y Box donde muestra las estaciones con y sin autocorrelación significativa.

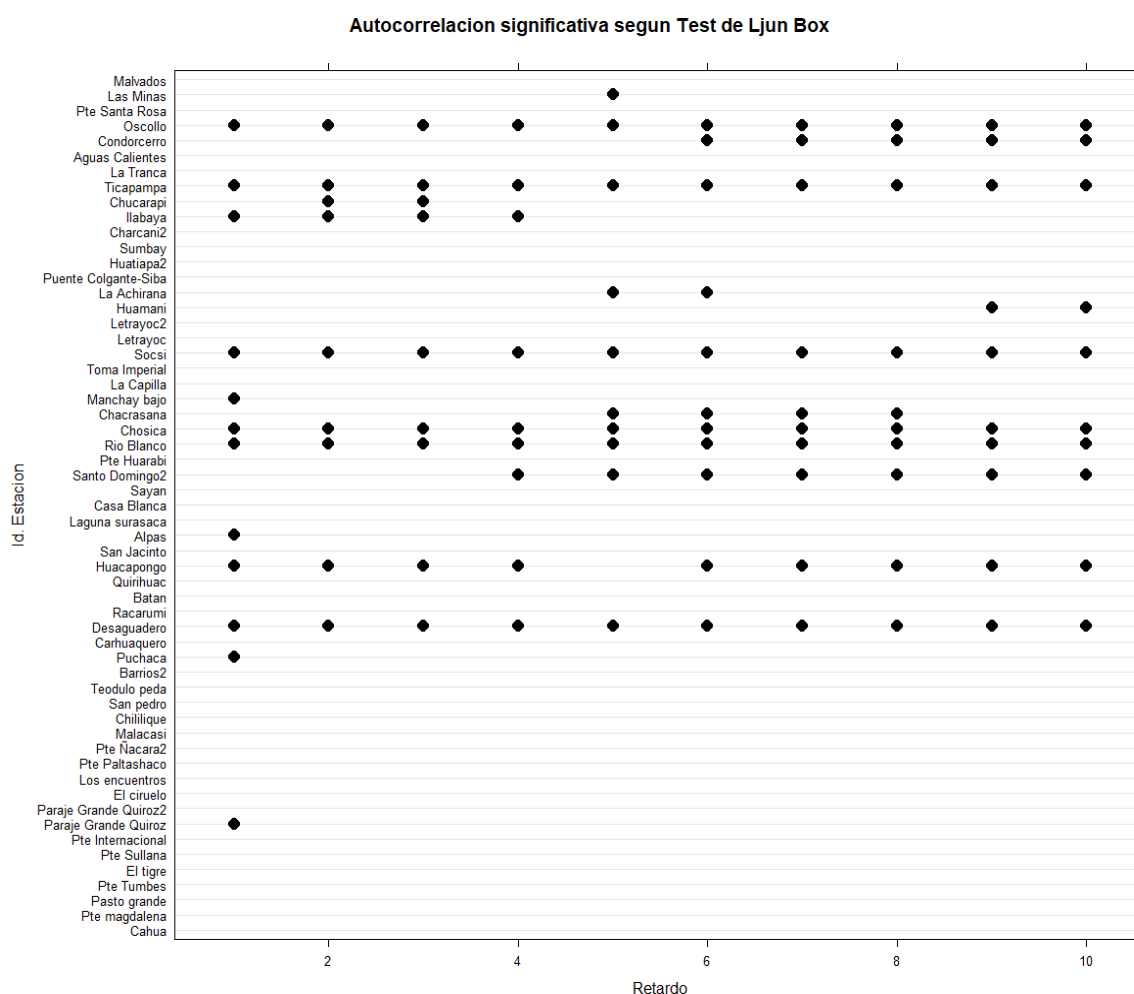


Figura 07. Resumen de autocorrelaciones significativas detectadas para varios retardos a un nivel de 5% por el test de Ljun Box.

Fuente: Elaboración propia

Según lo analizado, las siguientes estaciones presentan autocorrelación significativa: Desaguadero, Huacapongo, Santo Domingo, Río Blanco, Chosica, Chacrasana, Socsi, Ilabaya, Ticapampa, Condorcerro y Oscollo.

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

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 será 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.

De acuerdo la evaluación de los caudales máximos descrito anteriormente, se logró identificar 43 estaciones hidrométricas que cumplen con los criterios y análisis descritos anteriormente, los mismo que son detallados en el cuadro 03.

4.1.3. Formación de regiones homogéneas

Inicialmente en esta etapa del estudio se consideró las 43 estaciones como una sola región hidrológicamente homogénea. Para el filtrado de las estaciones discordantes, la medida de discordancia, D_i se ha calculado para cada estación (Cuadro 03). De acuerdo al número de estaciones que se presenta en el estudio, se estableció el valor crítico D_i es 3.

En la Figura 08 se aprecia que las estaciones de Teódulo Peña, Aguas calientes, Puente Sullana, Ilabaya, la tranca y Pasto Grande; parecen ser discordantes con el resto de las estaciones ya que sus valores se encuentran alejados del promedio regional (punto negro). Sin embargo ninguna de ellas excede el valor crítico D_i (Cuadro 03), por lo que serán examinadas en su respectivo grupo o región homogénea.

Cuadro 03: Valores de los ratios de L-momentos de las estaciones

Estación nombre	Id Station	Longitud n	Media l_1	L-CV t	L-Asimetría t_3	L-Curtosis t-4	Discordancia D
Cahua	1	18	134.7111	0.1502	0.0948	0.1264	1.1398
Malvados	4	16	40.3856	0.1793	0.3679	0.2335	1.9790
Puente magdalena	111151	37	63.6578	0.3583	0.3084	0.1941	0.0128
Puente tumbes	200201	29	1052.4166	0.4200	0.5178	0.3073	0.7991
El tigre	200202	47	910.9855	0.3099	0.2201	0.0878	0.5656
Puente Sullana	200304	42	1149.6426	0.5199	0.4572	0.2403	0.8337
Paraje Grande Quiroz	200311	38	236.8018	0.4157	0.5038	0.3024	0.6919
Los encuentros	200313	18	638.0600	0.3203	0.2941	0.0859	1.4416
Puente Paltashaco	200403	20	16.6350	0.4077	0.3296	0.1480	0.3859
Malacasi	200414	18	215.1556	0.4986	0.2405	0.0834	1.1710
Chillique	200415	18	23.8778	0.3282	0.2677	0.2938	1.1551
San pedro	200416	18	46.2611	0.5243	0.3481	0.0859	1.7338
Teódulo Peña	200423	21	74.0048	0.6016	0.5957	0.4236	2.9013
Barrios	200424	21	80.3714	0.4416	0.3153	0.2133	0.3583
Puchaca	200801	67	58.0772	0.3360	0.4472	0.2792	0.6152
Carhuaquero	200903	22	123.7395	0.2101	0.3164	0.2430	0.8121
Racamuri	200907	24	226.8342	0.2751	0.3495	0.3833	1.8401
Batan	201001	36	50.3744	0.4185	0.4506	0.3383	0.6358
Quirihuac	201401	37	120.6641	0.3328	0.1752	0.1217	0.2515
Huacapongo	201501	30	65.9050	0.4720	0.3036	0.0895	1.0555
San Jacinto	201901	23	33.4022	0.4612	0.4136	0.3079	0.7482
Laguna Surasaca	202601	17	2.7559	0.2629	0.4795	0.3514	1.4015
Sayán	202608	30	27.1963	0.3149	0.3010	0.2038	0.0571
Casa Blanca	202609	34	116.9403	0.2238	0.2768	0.1707	0.8045
Santo Domingo	202701	44	95.3648	0.2654	0.1603	0.1646	0.4269
Puente Huarabi	202806	22	73.2000	0.2237	0.1911	0.2917	1.9745
Chacrasana	202908	49	190.0412	0.2843	0.2553	0.1324	0.4452
Manchay bajo	203005	17	39.6788	0.2371	0.3646	0.1680	2.0534
Socsi	203302	75	315.7333	0.2373	0.1938	0.1390	0.4021
Letrayoc	203602	47	261.7823	0.2217	0.1405	0.1584	0.6032
Huamani	203702	22	160.7227	0.2906	0.1153	0.1065	0.5175
La Achirana	203703	32	118.5200	0.2901	0.1480	0.0864	0.3985
Puente Sibayo	204604	26	233.9850	0.3194	0.0129	0.0289	1.4632
Huatiapa	204617	34	648.4118	0.3055	0.1719	0.1350	0.2396
Sumbay	204702	36	21.4322	0.2603	0.0663	0.1357	1.2638
Charcani	204710	47	110.0634	0.4299	0.3474	0.2684	0.5191
Ilabaya	204806	17	6.6794	0.5607	0.4176	0.1944	1.2364
Chucarapi	204903	29	359.0028	0.3693	0.1016	0.0406	0.9109
La Tranca	205103	19	40.1021	0.5491	0.3309	0.1269	1.3643
Aguas Calientes	205201	43	7.5405	0.5140	0.5521	0.3414	1.1560
Condorcerro	211402	37	612.4846	0.2284	0.1382	0.1256	0.5026
Puente Santa Rosa	240108	48	401.8575	0.3294	0.0442	0.0688	1.3261
Pasto Grande	1318612001	22	4.5936	0.3338	0.6079	0.5070	2.8062

Fuente: Elaboración propia

Cabe señalar que se ha determinado la medida de heterogeneidad considerando la región como un todo, obteniéndose valores ($H_1=10.95$, $H_2=7.76$ y $H_3=3.94$) que declaran la región entera como heterogénea por lo que corresponde la necesidad de un análisis de conglomerados de regiones y una evaluación adicional basada en métricas como D_{ij} y H_i ($i=1, 2, 3, 4$).

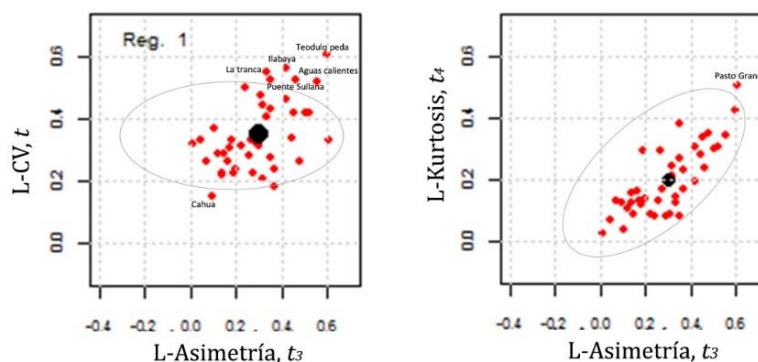


Figura 08. Ratios de los L-momentos de las estaciones.

Fuente: Elaboración propia

Las técnicas de agrupación se han utilizado ampliamente en hidrología, así como en una variedad de otras disciplinas (Theodoridis y Koutroubas 1999, Jain y Dubes 1988, Kaufman y Rousseeuw 1990, Everitt 1993). Para este estudio, como primera aproximación se utilizó el método de agrupamiento jerárquico de Ward y el método de particionamiento K-means para dividir los datos en regiones; en función de las primeras cinco estadísticas de L-momentos de los sitios individuales.

Además siguiendo el criterio de cuenca coherente y según lo establecido por J. R. Wallis et al (2007), se propuso la identificación de regiones utilizando predictores fisiográficos y meteorológicos. Por lo tanto, en el presente estudio se utilizó las siguientes características de captación: coordenadas geográficas de los centroides de las cuencas (\bar{X} y \bar{Y}), la elevación media, área de la cuenca y precipitación media anual de la cuenca de drenaje.

Posteriormente se utilizó el método de agrupación jerárquica de Ward y K-means para dividir el área de estudio en tres grupos. Para cada grupo, se realizó la prueba de discordancia (D_i) y se dibujaron los diagramas de los L-momentos. Donde, los grupos no mostraron homogeneidad suficientemente basada en pruebas de homogeneidad de los L-momentos, por lo que se formaron cuatro grupos y se repitieron las pruebas de homogeneidad (Figura 09).

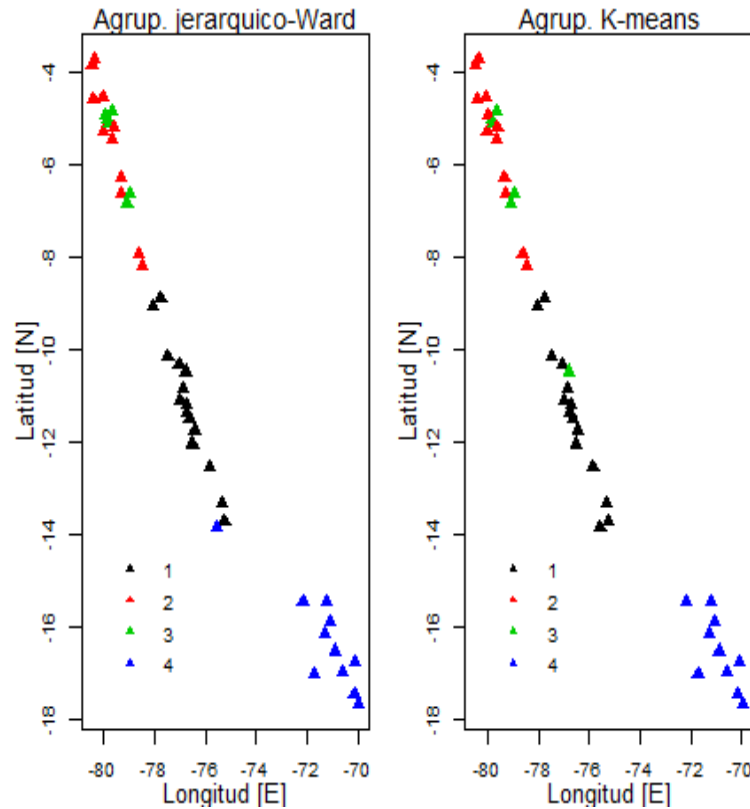


Figura 09. A) Clasificación de estaciones hidrométricas mediante agrupación jerárquica y enlace de Ward, B) Clasificación de estaciones mediante K-means.
Fuente: Elaboración propia

Además de los algoritmos de agrupamiento, algunas estaciones que no parecían pertenecer a un grupo se movieron manualmente a otro grupo. Las estadísticas de discordancia y heterogeneidad se calcularon para cada región candidata, y estos procedimientos se repitieron hasta que cada región contenía un grupo de estaciones que, para todos los L-moments proporcionaron resultados aceptables en las pruebas de homogeneidad.

4.1.4. Detección de datos usando medidas de discordancia.

Una medida de discordancia, D_i , que se define en términos de L-momentos se utiliza para detectar datos erróneos y verificar que los datos sean apropiados para realizar el análisis de frecuencia de inundación regional. Supongamos que $u_i = [t^{(1)}, t^{(3)}, t^{(4)}]^T$ es un vector que contiene los ratios de L - momentos (L - C_v , L - C_s y L - C_k) para el sitio i (Hosking y Wallis 1993). La medida de discordancia para el sitio i es:

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

$$\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i \quad (32)$$

$$S = \frac{1}{(N - 1)} \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \tag{33}$$

Donde N es el número de estaciones, u_i es el vector de L-moments, \bar{u} es el promedio regional, S es la matriz de covarianza de la muestra, n es el número total de sitios, y T denota la transposición de un vector o matriz. Si el $D_{\text{estadístico}}$ de un sitio excede 3 cuando el número de sitios en una región es mayor a 15, se considera que sus datos son discordantes con respecto al resto de los datos regionales. Los grandes valores de D_i indican sitios que son los más discordantes del grupo como un todo y son más dignos de investigación por la presencia de errores de datos (Hosking y Wallis 1993). En general, cualquier sitio con $D_i > 3$ puede considerarse como discordante.

Los resultados y análisis que se muestra en esta etapa son para las series de datos de caudales máximos que pasaron el análisis de supuestos (43 estaciones).

Tabla 03: Clasificación de regiones homogéneas y prueba de discordancia

N°	REGION 1		REGION 2		REGION 3		REGION 4	
	Estaciones	D	Estaciones	D	Estaciones	D	Estaciones	D
1	Cahua	1.591	Puente Tumbes	0.930	Malacasi	1.428	Huamani	0.374
2	Malvados	1.461	El Tigre	0.423	Huacapongo	0.627	La Achirana	1.186
3	Sayán	2.086	Los Encuentros	1.205	San Jacinto	1.342	Puente colgante	0.569
4	Casa Blanca	1.181	Barrios	1.213	Chillique	1.088	Sibayo	0.166
5	Santo Domingo	0.207	Puchaca	1.244	Paraje grande Quiroz	2.096	Huatiapa	1.627
6	Puente Huarabi	0.569	Quirihuac	0.983	Puente Paltashaco	1.269	Chucarapi	1.097
7	Chacrasana	2.112			Teóduo Peña	0.194	Puente Santa Rosa	0.998
8	Manchay bajo	0.853			Racarumi	0.788	Pasto Grande	1.980
9	Socsi	1.044			Batan	0.164		
10	Letrayoc	0.186						
11	Condorcerro	0.331						
12	Laguna Surasaca	0.374						

Fuente: Elaboración propia

4.1.5. Prueba de homogeneidad regional

Para generar regiones 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 ξ (localización), α (escala), k y h (forma); se obtiene ajustando a los ratios regionales medios de L-moments.

La distribución Kappa de cuatro parámetros tiene la forma: $f(x) = \alpha^{-1} [1 - k(x - \xi) / \alpha]^{1/k} [F(x)]^{1-h}$. Para el presente se ha utilizado 1000 simulaciones para obtener medidas de heterogeneidad más precisas.

Para verificar que las cuatro regiones propuestas sean homogéneas, se procedió a realizar la prueba del estadístico H_1 , teniendo en cuenta el criterio de clasificación de Wallis et al, que se observa en la Tabla 02.

Tabla 04: Resultados de la prueba de Homogeneidad, para las regiones definidas.

Medida de Heterogeneidad	Regiones			
	Región 01	Región 02	Región 03	Región 04
H_1	0.916	0.606	0.969	0.858
H_2	0.350	2.358	0.066	3.837
H_3	-0.412	1.879	-0.009	3.990

Fuente: Elaboración propia

El grado de homogeneidad de las sub-regiones propuestas se basó en la medida estadística H_j definida por Hosking y Wallis (1997). La estadística H_j mide la variabilidad relativa de L-Cv observado a partir de la muestra, y es utilizada para analizar la heterogeneidad en cada una de las 4 sub-regiones homogéneas propuestas. Según Hosking y Wallis (1997) H_j tiene mejor poder discriminatorio para detectar la homogeneidad de la región que H_2 y H_3 . Esto también fue comprobado por (Viglione, 2004), quien comparó la potencia asociada a estos 3 estadísticos mediante una simulación de Montecarlo.

La sub-región se declara heterogénea si el valor de H_j es suficientemente elevado. Hosking y Wallis (1997) sugiere considerar como “aceptablemente homogénea” si $H_j < 1$, “posiblemente heterogénea” si $1 < H_j < 2$, y “definitivamente heterogénea” si $H_j > 2$. Según Hosking y Wallis (1997),

En la Figura 10 se observa que las regiones 01 y 04 presenta una notable distribución espacial, lo cual hace posible tener un panorama visual de los límites espaciales de ambas regiones; caso contrario con lo ocurrido en las regiones 02 y 03 donde no se observa una homogenización clara en cuanto a la distribución espacial. Las medidas de discordancia para cada estación y las estadísticas de heterogeneidad para cada una de las cuatro regiones se resumen en las Tablas 03 y 04, respectivamente.

Es de importancia indicar, que las áreas de drenaje delimitados a través de las estaciones la tranca, Aguas calientes, Ilabaya, Sumbay, Carhuaquero, Puente Sullana y Puente Magdalena quedaron fuera del estudio, por desfavorecer los resultados de las pruebas de discordancia e heterogeneidad.

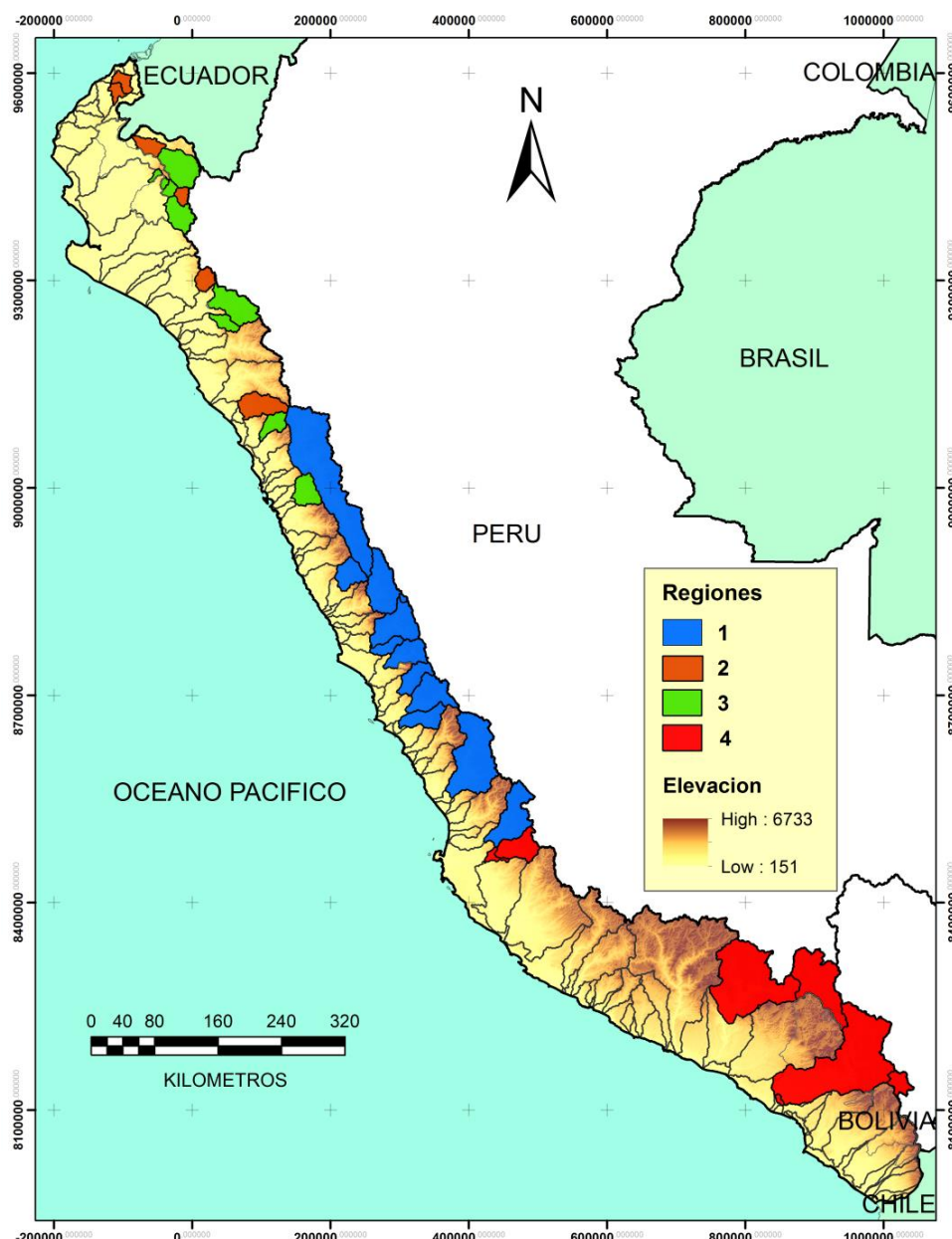


Figura 10. Regiones conformadas para la Vertiente del Pacífico.

Fuente: Elaboración propia

4.1.6. Selección de la distribución regional de mejor ajuste y cálculo de los cuantiles

Los diagramas de los L-moments se utilizaron como un método visual simple y rápido para identificar si un grupo de sitios satisface la homogeneidad en función sus valores $L-C_v$ y $L-C_s$, así como a las distribuciones candidatas de identidad para la región(s). Los valores en el sitio y los promedios regionales de L-Asimetría y L-Kurtosis se trazaron en un diagrama de L- moments ratios, junto con las curvas teóricas para las distribuciones candidatas como la logística generalizada (GLO), generalizada del

valor extremo (GEV), generalizado normal (GNO), generalizada de Pareto (GPS) y Pearson tipo III (PE3).

Si el punto de los promedios regionales se encuentra cerca de la curva correspondiente a una distribución dada, esta distribución puede ser una selección aceptable para la distribución padre en la región. Las distancias que separan los puntos de muestra de la curva para una distribución determinada se pueden tomar como una medida de bondad de ajuste. Además, si el diagrama de los L-moments indica que las estadísticas de un sitio en particular están lejos del resto del grupo, este sitio puede ser eliminado para re-análisis en un grupo diferente. Se debe tener en cuenta que este método gráfico de elegir la mejor distribución depende de la homogeneidad de los datos regionales (peel et al., 2001).

Los valores promedio regionales de L-Asimetría ($L-C_3$) y L-Kurtosis ($L-C_4$) para cada sitio se calcularon y trazaron en la Fig. 11 se incluye la relación entre la población $L-C_3$ y $L-C_4$ para un rango de distribuciones comúnmente aplicadas en el análisis de frecuencias de inundación.

Las relaciones teóricas se encuentran resumidas en Hosking y Wallis (1997). Como los L-momentos de las muestras son imparciales, los puntos de las muestras deben distribuirse por encima y por debajo de la línea teórica de una distribución adecuada.

4.1.6.1. Medidas de bondad de ajuste

En el último paso del análisis de regionalización, se evaluaron las medidas de bondad de ajuste para un conjunto de distribuciones candidatas para cada una de las cuatro sub-regiones: Región (1), homogénea (consta de 14 sitios); Región (2), homogénea (consta de 8 sitios); Región (3), homogénea (consta de 11 sitios); y finalmente la región (4), homogénea (consta de 7 sitios).

La identificación de las distribuciones de probabilidad depende de aquellas que presentan el mejor ajuste en los datos para diferentes estaciones en una región. Para dicho objetivo, se utilizó la medida o el estadístico Z^{DIST} en cada región, teniendo en cuenta que el valor absoluto de Z^{DIST} debe ser menor o igual a 1.64.

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-Moments 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 para identificar la distribución regional de mejor ajuste.

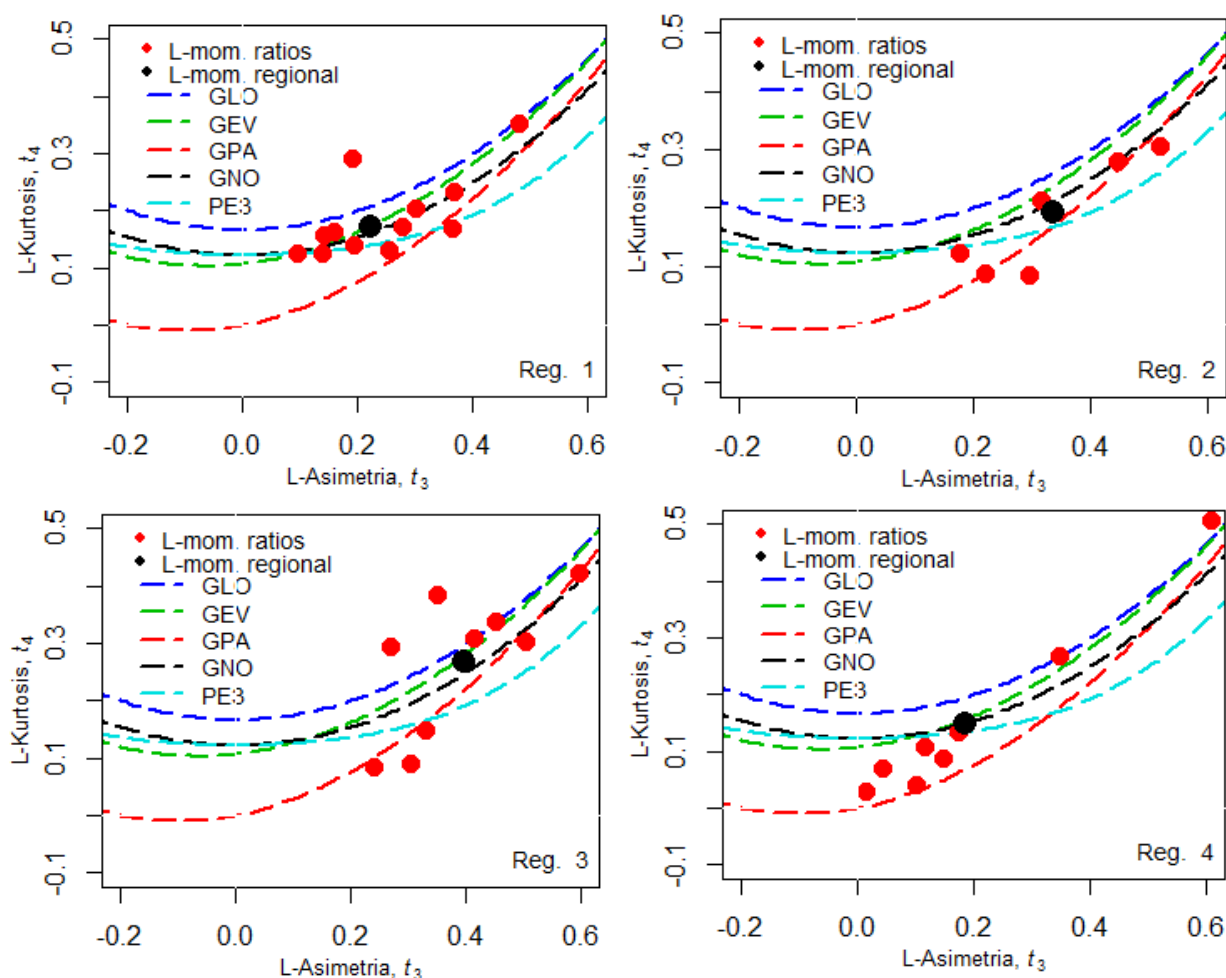


Figura 11. Diagrama de L-momentos para las cuatro sub-regiones

Fuente: Elaboración propia

Con base en el valor crítico de $Z_{\text{crítico}}$ DIST de bondad de ajuste estadístico, se seleccionaron la distribución GEV para la región 1; la distribución GNO para la región 2; la distribución GLO para la región 3 y finalmente la distribución GNO para la región 4 (tabla 06).

Tabla 05: Estadístico ZDIST de varias distribuciones para cada región.

Región	Z_{GLO}	Z_{GEV}	Z_{GNO}	Z_{PE3}	Z_{GPA}
1	1.486	-0.059	-0.579	-1.570	-3.772
2	1.738	1.086	0.365	-0.887	-0.845
3	0.061	-0.347	-1.105	-2.415	-1.757
4	1.818	0.252	-0.030	-0.673	-3.300

Fuente: Elaboración propia

4.1.7. Desarrollo de la fórmula regional de cuantiles para cuencas hidrográficas

Como muestra de desarrollo del presente ítem se tomara como referencia a la región 01 determinada en los pasos anteriores, la misma que presentó la distribución GEV como la distribución más robusta para dicha región; por lo tanto, las fórmulas regionales de cuantiles se desarrollara utilizando esta distribución.

La forma de la fórmula regional para una distribución GEV se expresa como:

$$\frac{Q_T}{\bar{Q}} = \frac{\xi + \alpha \left[1 - \left\{ -\ln \left(1 - \frac{1}{T} \right) \right\}^k \right]}{k} \tag{34}$$

Aquí, Q_T es la estimación de los cuantiles del periodo de retorno de T-años, \bar{Q} es el caudal máximo medio anual de la cuenca, ξ , α y k son los parámetros de la distribución. Los valores de los parámetros regionales de la distribución GEV para la región 01 son: $\xi= 0.7860$, $\alpha=0.3225$ y $k=-0.0804$ (cuadro 04). Sustituyendo valores para estos parámetros regionales en la ecuación 35, la fórmula del cuantil regional para la estimación de cuantiles de varios periodos de retorno para las cuencas hidrográficas medidas de la región 01 se expresa de la siguiente manera:

$$Q_T = \left[-3.225 + 4.011 \left\{ -\ln \left(1 - \frac{1}{T} \right) \right\}^{-0.0804} \right] \times \bar{Q} \tag{35}$$

La fórmula anterior se puede usar para estimar los cuantiles de los periodos de retorno deseados para las cuencas hidrográficas de la región 01. Alternativamente, los cuantiles de varios períodos de retorno también pueden calcularse multiplicando el caudal máximo anual medio de la cuenca (\bar{Q}) por el valor correspondiente de los factores de crecimiento (QT/\bar{Q}), que se desarrollan utilizando la ecuación 34, y se muestra en el Cuadro 04 de manera similar para todas la regiones conformadas, las curvas de crecimiento regional también se desarrollaron para cada una de las distribuciones elegidas como se muestra en la Figura 12.

Cuadro 04: Parámetros y cuantiles regionales para las distribuciones con mejor ajuste por región.

Distrib	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												
GEV	0.7860	0.3225	-0.0804	0.906	1.300	1.581	1.868	2.264	2.581	2.915	3.386	3.765
Región 2												
GNO	0.7976	0.5027	-0.7083	0.797	1.376	1.847	2.363	3.128	3.775	4.488	5.539	6.423
Región 3												
GLO	0.7403	0.3261	-0.3974	0.740	1.343	1.885	2.564	3.774	5.017	6.647	9.615	12.695
Región 4												
GNO	0.8874	0.5671	-0.3826	0.887	1.450	1.825	2.186	2.657	3.015	3.376	3.864	4.240

Fuente: Elaboración propia

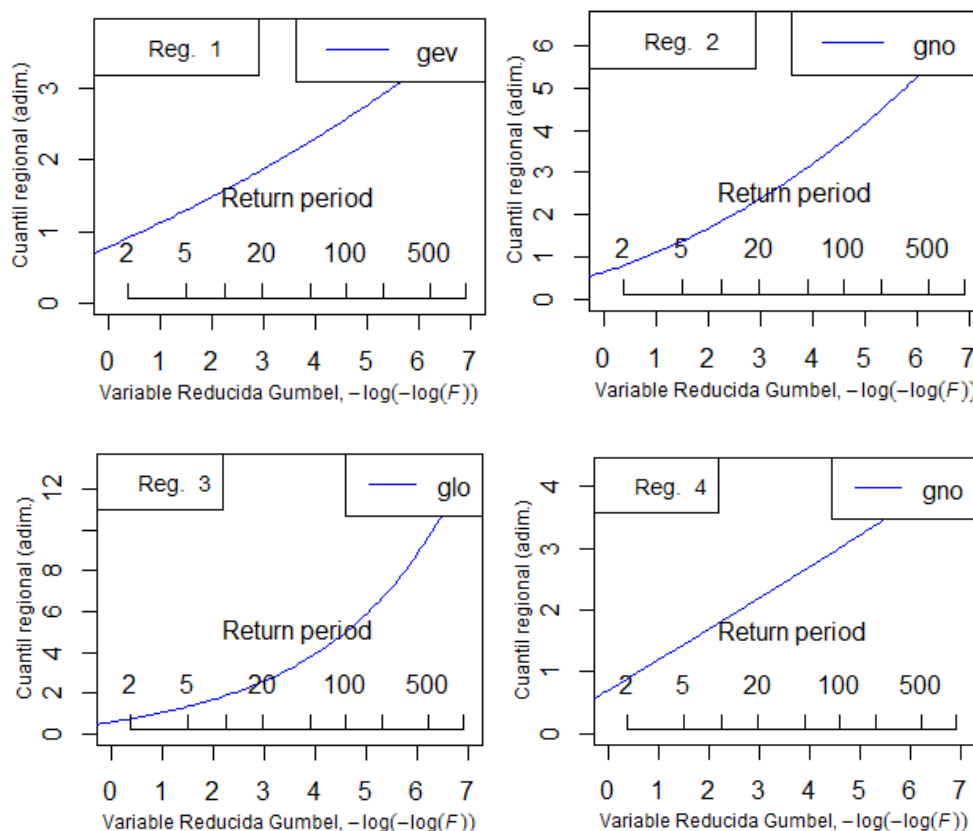


Figura 12. Curva de crecimiento regional para las distribuciones con mejor ajuste.
Fuente: Elaboración propia

4.1.8. Desarrollo de la relación regional entre el caudal promedio anual máximo y el área de las cuencas.

Para la estimación de la inundación del periodo de retorno de T-años en un lugar, se requiere la estimación del caudal máximo medio anual. Para cuencas hidrográficas no evaluadas en un lugar, no se puede calcular una media en ausencia de los datos de flujo observados. En tal situación, se necesita una relación entre los caudales máxima medios anuales de las cuencas hidrográficas medidas en la región y sus características fisiográficas y climáticas pertinentes para la estimación de la inundación pico media anual. Se desarrolló la siguiente relación regional en términos de área de cuencas hidrográficas para la estimación de inundaciones máximas anuales medias para las cuencas hidrográficas no monitoreadas:

$$\bar{Q} = 1.8627(A)^{0.4979} \tag{36}$$

Aquí, A es el área de la cuenca en (km²) y \bar{Q} es el caudal máximo anual medio (m³/s). Esta relación se desarrolla en base al análisis de regresión, utilizando un enfoque de mínimos cuadrados. Para esta relación, el coeficiente de correlación (0.90).

Cuadro 05: Caudales promedio máximos anuales de las estaciones hidrométricas utilizadas

ID estaciones	Nombre estaciones	Caudales máximos promedio anual	Área de la cuenca (km ²)	Tamaño de la muestra (Años)
Región 01				
1	Cahua	134.71	2974.84	18
4	Malvados	40.38	1372.25	16
202601	Laguna Surasaca	2.75	97.33	17
202608	Sayán	27.19	2753.60	30
202609	Casa Blanca	116.94	71.52	34
202701	Santo Domingo	95.36	1660.04	44
202806	Puente Huarabi	73.20	1272.46	22
202908	Chacrasana	190.04	2392.20	49
203005	Manchay bajo	39.67	1455.13	17
203302	Socsi	315.73	5790.85	55
203602	Letrayoc	261.78	3088.81	47
211402	Condorcerro	612.48	10403.98	37
Región 02				
200201	Puente Tumbes	1052.41655	581.28	29
200202	El Tigre	910.985532	423.96	47
200313	Los Encuentros	638.06	830.76	18
200424	Barrios	80.3714286	422.83	21
200801	Puchaca	58.0771642	739.75	57
201401	Quirihuac	120.664054	1879.65	37
Región 03				
200414	Malacasi	215.155556	1430.03	18
201501	Huacapongo	65.905	909.58	30
201901	San Jacinto	33.402173	1376.6	23
200415	Chillique	23.8777778	158.91	18
200311	Paraje Grande Quiroz	236.801842	2324.84	38
200403	Puente Paltashaco	16.635	142.31	20
200423	Teódulo Peña	74.0047619	341.56	21
200907	Racarumi	226.834167	2371.71	24
201001	Batan	50.374444	633.68	36
Región 04				
203702	Huamani	160.722727	1827.9	22
203703	La Achirana	118.52	308.14	32
204604	Puente colgante Sibayo	233.985	4494.64	26
204617	Huatiapa	648.411765	8744.48	34
204710	Charcani	110.063404	5957.9	47
204903	Chucarapi	359.002759	114.32	29
240108	Puente Santa Rosa	401.8575	12144.35	48
1318612001	Pasto grande	4.59363636	802.31	22

4.1.8.1. Desarrollo de fórmula regional de cuantiles para cuencas no monitoreadas.

Para el desarrollo de la fórmula regional de inundaciones para la estimación de cuantiles de varios periodos de retorno para cuencas hidrográficas no evaluadas, la fórmula de inundación regional dada en la ecuación 35 se combinó con la relación regional entre caudal máximo medio anual y el área de la cuenca, dada en la ecuación 36. La siguiente fórmula regional fue desarrollada.

$$Q_T = \left[-6.007 + 7.471 \left\{ -\ln \left(1 - \frac{1}{T} \right) \right\}^{-0.0804} \right] A^{0.4979} \quad (37)$$

Donde Q_T es la estimación de los cuantiles (m^3/s) para el periodo de retorno de T-años, y A es el área de la cuenca (km^2). La fórmula de cuantiles regional anterior, su forma tabular (Cuadro 06) puede usarse para estimar las inundaciones de los periodos de retorno deseados para las cuencas no monitoreadas de la región 01.

Cuadro 06: Cuantiles de varios periodos de retorno para diferentes áreas de drenaje de la región 01.

Área de cuencas (km^2)	Periodos de retorno (Años)								
	2	5	10	25	50	100	200	500	1000
	Inundaciones de varios periodos de retorno (m^3/s)								
10	5.31	7.62	9.27	11.50	13.27	15.13	17.09	19.84	22.07
20	7.50	10.76	13.09	16.24	18.74	21.36	24.13	28.02	31.16
30	9.18	13.17	16.02	19.88	22.93	26.14	29.53	34.29	38.13
40	10.59	15.20	18.49	22.94	26.47	30.17	34.07	39.57	44.00
50	11.83	16.98	20.66	25.63	29.58	33.72	38.08	44.22	49.17
60	12.96	18.60	22.62	28.07	32.39	36.92	41.70	48.42	53.85
70	13.99	20.08	24.43	30.31	34.97	39.86	45.02	52.29	58.14
80	14.95	21.46	26.11	32.39	37.37	42.61	48.12	55.88	62.14
90	15.86	22.76	27.68	34.35	39.63	45.18	51.02	59.26	65.89
100	16.71	23.98	29.17	36.20	41.76	47.61	53.77	62.45	69.44
150	20.45	29.35	35.70	44.29	51.11	58.26	65.80	76.42	84.97
200	23.60	33.87	41.20	51.12	58.98	67.23	75.94	88.18	98.06
250	26.37	37.85	46.04	57.12	65.91	75.14	84.86	98.55	109.58
300	28.88	41.44	50.41	62.55	72.17	82.28	92.92	107.91	120.00
350	31.18	44.75	54.44	67.54	77.93	88.84	100.34	116.52	129.57
400	33.33	47.83	58.18	72.18	83.29	94.95	107.23	124.53	138.48
450	35.34	50.71	61.69	76.54	88.32	100.68	113.71	132.05	146.84
500	37.24	53.45	65.01	80.67	93.07	106.10	119.83	139.16	154.75
600	40.78	58.52	71.19	88.33	101.92	116.19	131.22	152.39	169.45
700	44.04	63.19	76.87	95.38	110.05	125.45	141.69	164.54	182.97

800	47.06	67.54	82.16	101.94	117.61	134.08	151.43	175.86	195.55
900	49.90	71.62	87.12	108.09	124.72	142.18	160.58	186.48	207.36
1000	52.59	75.47	91.81	113.91	131.43	149.83	169.22	196.52	218.53
1200	57.59	82.65	100.53	124.74	143.92	164.07	185.30	215.20	239.30
1400	62.18	89.24	108.56	134.69	155.41	177.16	200.09	232.36	258.39
1600	66.46	95.37	116.02	143.95	166.09	189.34	213.84	248.34	276.15
1800	70.47	101.13	123.02	152.64	176.12	200.78	226.76	263.34	292.83
2000	74.27	106.58	129.65	160.86	185.61	211.59	238.97	277.52	308.60

Fuente: Elaboración propia

4.1.9. Determinación de umbrales de inundación.

Como caso práctico de desarrollo de los umbrales de inundación fue necesario tomar como referencia los valores de niveles de alerta establecido por la Sub Dirección de Predicción Hidrológica - SPH del Servicio Nacional de Meteorología e Hidrología del Perú - SENAMHI, las mismas que se encuentran clasificadas en tres niveles de alerta y dicha información está disponible para las estaciones monitoreadas por el SENAMHI.

En el Cuadro 07 se detalla valores de diferentes niveles de alerta, para cinco estaciones que se encuentran ubicadas dentro de la región homogénea 01 del presente estudio.

Cuadro 07: Valores de los niveles de alerta establecido por el SENAMHI.

Estaciones	Umbrales		
	Nivel alerta rojo	Nivel alerta naranja	Nivel alerta amarillo
Santo domingo	140	115	65
Cahua	170	120	70
Socsi	250	180	130
Letrayoc	310	146	110
Condorcerro	620	480	370

Nivel alerta rojo	Umbral correspondiente al nivel de desborde del río
Nivel alerta naranja	Valor de la crecida máxima ordinaria
Nivel alerta amarillo	Valor máximo de la serie normal de caudal promedio diario

Fuente: SENAMHI, DHI-SPH.

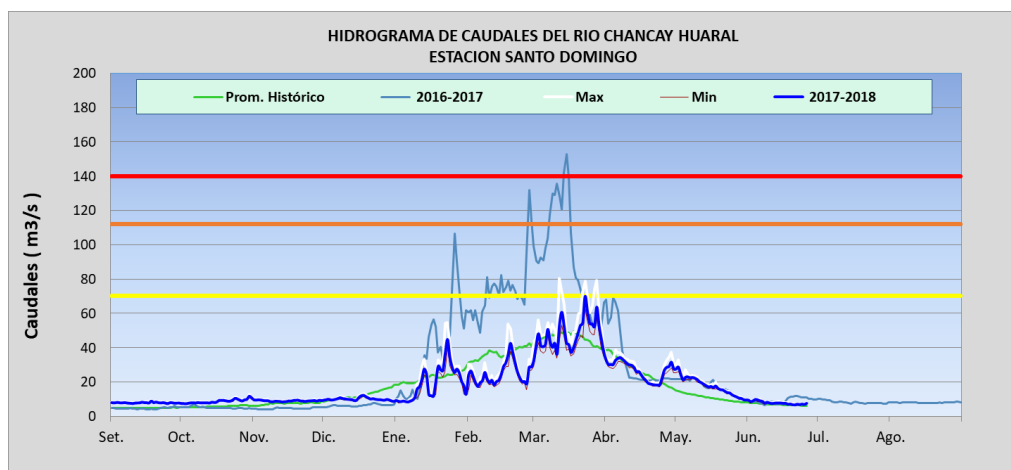


Figura 13. Umbrales de niveles de agua – Estación Santo Domingo.
Fuente: SENAMHI, DHI-SPH.

Este procedimiento se basa en la estimación de los cuantiles de inundación (magnitud) para diferentes periodos de retorno (frecuencia), que se encuentran manifiestos en la ecuación N° 37 (formula regional para cuencas no monitoreadas). Para determinar los umbrales de inundación fue necesario conocer las características fisiográficas de las cuencas; tal es el caso del área de drenaje, que se utilizó como variable de entrada a la ecuación tabular regional. Es así que en función del área de la cuenca y el periodo de retorno se dispuso de los cuantiles de inundación (cuadro XX).

Una vez establecidos los valores de los cuantiles de inundación se realizó la comparación con los valores de los umbrales definidos por el SENAMHI; es decir que se buscó la similitud de ambos para finalmente encontrar el valor óptimo del periodo de retorno.

Cuadro 08: Valores de periodos de retorno relacionados con los umbrales de inundación.

Estación	Área (km ²)	Periodos de retorno		
Santo domingo	1660	20	9	2
Cahua	2975	13	4	1.5
Socsi	5791	17	5	2
Letrayoc	3089	250	7	3
Condorcerro	10404	450	100	25

Fuente: Elaboración propia

En ese sentido, podríamos indicar que los valores de periodo de retorno encontrados para las estaciones Santo Domingo, Cahua, Socsi, Letrayoc y Condorcerro se encuentran claramente relacionados a los umbrales de inundación. Este procedimiento deberá ser aplicado para las demás regiones homogéneas determinadas en el presente estudio.

5. Conclusiones

Sobre la base de este estudio, se extraen las siguientes conclusiones

- En el presente documento se aplicó la metodología de los L-Moments para regionalizar eventos máximos. Para este análisis se utilizaron las estaciones con un mínimo de 15 años de registro, por lo tanto, se seleccionaron 43 estaciones con registros de datos anuales de caudales máximos.
- Los resultados de la presente investigación pueden ser utilizados con un alto grado de confiabilidad en estudios de prefactibilidad y factibilidad y con cierta reserva en la etapa de diseño.
- El método de L-moments a través de la medida de heterogeneidad, permite definir objetivamente que la vertiente del Pacífico conforma cuatro regiones hidrológicamente homogéneas.
- El diagrama convencional de ratios de L-moments y el test de bondad de ajuste denominado Z estadístico permiten identificar que existen distribuciones de probabilidad de mejor ajuste a los eventos de caudales máximos regionales, siendo GEV la mejor distribución para la región (1); la distribución GNO para la región (2); También se encuentra la distribución GLO para la región (3); y para la región (4) el mejor ajuste fue GNO.
- Para la estimación de cuantiles de inundación de los períodos de retorno deseados para cuencas hidrográficas no monitoreadas del área de estudio, se recomienda utilizar la fórmula de inundación regional desarrollada en el presente documento en su forma tabular.
- Se determinó los umbrales de inundación en función del periodo de retorno y el área de drenaje de las cuencas estudiadas.

6. Bibliografía

- Álvarez, M., Puertas, J., & Soto, B. (1999).** Análisis regional de las precipitaciones máximas en Galicia mediante el método del índice de avenida. Ingeniería del Agua.
- Benjamín, J., Cornell, C. (1981).** Probabilidad y estadística en ingeniería civil. Traducido por Sin, A. y Prieto, V.
- Bradley, A., Zhao, H. (1997).** Climatic variations in extreme precipitation in the Midwest (Submitted to Journal of Hydrologic Engineering). University of Iowa.
- Chow, V. (1964).** Statistical and probability analysis of hydrologic data; part I: frequency analysis. In: Handbook of applied hydrology; a compendium of water resources technology.
- Cunnane, C. (1988).** Methods and merits of regional flood frequency analysis. Journal of Hydrology.
- Greis, N., Wood, E. (1981).** Regional flood frequency estimation and network design. Water Resources research.
- Hosking, J. R.M., Wallis, J.R. (1986).** The value of historical data in flood frequency analysis. Water Resources Research.
- Hosking, J. R.M., Wallis, J.R. (1988).** The effect of intersite dependence on regional flood frequency analysis. Water Resources Research.
- Hosking, J.R.M. (1990).** L-Moments: Analysis and estimation of distribution using linear combinations of order statistics. Journal of the Royal Statistics Society.
- Hosking, J. R.M. & Wallis, J.R. (1993).** Some statistics useful in regional frequency analysis, Water Resource.
- Hosking, J. R.M. & Wallis, J. (1997).** Regional frequency analysis. Cambridge University Press.
- Jara, J. (1986).** Análisis de frecuencias bayesiano de series de datos hidrológicos. Memoria Ingeniero Civil. Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas.
- Kite, G. (1977).** Frequency and risk analyses in hydrology. Fort Collins, Colorado. Water Resources Publications.
- Monsalve G (2005).** Hidrología en la ingeniería. Bogotá, D.C.

Stedinger, J., Vogel, R. & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. Handbook of hydrology.

Varas, E. (1982). Fundamentos de probabilidad y estadística para hidrología. Universidad Nacional Agraria La Molina. Facultad de Ingeniería Agrícola. Departamento de Recursos de Agua y Tierra. Lima.

Wallis, J., M. Schaefer, B. Barker. & G. Taylor. (2007): Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington States. Hydrol. Earth Syst. Sci.

Zalina, M., Desa, M., Nguyen, V. & Kassim, A. (2002). Selecting a probability distribution for extreme rainfall series in Malaysia. Water Science and Technology.