

UNIVERSIDAD NACIONAL DEL ALTIPLANO
ESCUELA DE POSGRADO
DOCTORADO EN CIENCIA, TECNOLOGÍA Y MEDIO
AMBIENTE



TESIS

ANÁLISIS REGIONAL DE FRECUENCIA PARA LA ESTIMACIÓN DE
CAUDALES MÁXIMOS EN LA CUENCA AMAZÓNICA DEL PERÚ

PRESENTADA POR:

EFRAIN LUJANO LAURA

PARA OPTAR EL GRADO ACADÉMICO DE:

DOCTORIS SCIENTIAE EN CIENCIA, TECNOLOGÍA Y MEDIO
AMBIENTE

PUNO, PERÚ

2019

UNIVERSIDAD NACIONAL DEL ALTIPLANO

ESCUELA DE POSGRADO



DOCTORADO EN CIENCIA, TECNOLOGÍA Y MEDIO
AMBIENTE

TESIS

ANÁLISIS REGIONAL DE FRECUENCIA PARA LA ESTIMACIÓN DE
CAUDALES MÁXIMOS EN LA CUENCA AMAZÓNICA DEL PERÚ

PRESENTADA POR:

EFRAIN LUJANO LAURA

PARA OPTAR EL GRADO ACADÉMICO DE:

DOCTORIS SCIENTIAE EN CIENCIA, TECNOLOGÍA Y MEDIO AMBIENTE

APROBADA POR EL SIGUIENTE JURADO:

PRESIDENTE



.....
Dr. EDUARDO FLORES CONDORI

PRIMER MIEMBRO



.....
Dr. CIRO MARINO TRAVERSO ARGUEDAS

SEGUNDO MIEMBRO

.....
Dr. ANGEL CANALES GUTIÉRREZ

ASESOR DE TESIS



.....
Dr. GERMAN BELLIZARIO QUISPE

Puno, 03 de octubre de 2019

ÁREA: Ciencia Tecnología y Medio Ambiente.

TEMA: Estimación de caudales en el Perú.

LÍNEA: Evaluación Tecnológica y del Medio Ambiente

DEDICATORIA

Uno de mis grandes anhelos hecho realidad. Lo dedico con mucho cariño, amor y respeto:

- A mi padre, don LUIS RAFAEL y a mi adorada madre doña DOMITILA, por su aliento, paciencia y consejos que hicieron posible el logro de mi noble profesión.
- A mis hermanos: APOLINARIO y RENE con mucho cariño y afecto, por su apoyo y contribución en mi formación profesional.
- A SILVIA y JEANIE MARIET, por brindarme su apoyo incondicionalmente y sobre todo al cariño mostrado en los momentos más críticos.
- A Dios por darme la vida y la oportunidad de adquirir conocimiento y sabiduría

AGRADECIMIENTOS

- A la Escuela de Posgrado de la Universidad Nacional del Altiplano, Programa de Doctorado en Ciencia, Tecnología y Medio Ambiente.
- Al Servicio Nacional de Meteorología e Hidrología y a SO HYBAM por facilitar la información hidrológica.
- Al Dr. Eduardo Flores Condori, al Dr. Ángel Canales Gutiérrez y al Dr. Ciro Marino Traverso Arguedas miembros del jurado, por sus críticas constructivas que permitieron mejorar el contenido de la investigación.
- Mi sincera gratitud por su acertada dirección a mi asesor Dr. Waldo Sven Lavado Casimiro y al Dr. Germán Belizario Quispe, quienes motivaron y dedicaron su tiempo para la realización del trabajo de investigación.
- A todas las personas, amigos y familiares que, de una u otra manera, directa e indirectamente, motivaron y contribuyeron en la ejecución y culminación de la tesis.

ÍNDICE GENERAL

	Pág.
DEDICATORIA	i
AGRADECIMIENTOS	ii
ÍNDICE GENERAL	iii
ÍNDICE DE TABLAS	vi
ÍNDICE DE FIGURAS	vii
ACRÓNIMOS	viii
RESUMEN	ix
ABSTRACT.....	x
INTRODUCCIÓN	1

CAPÍTULO I

REVISIÓN DE LITERATURA

1.1 Marco teórico	3
1.1.1 Periodo de retorno	3
1.1.2 Análisis de frecuencia.....	3
1.1.3 Análisis regional de frecuencia	4
1.1.3.1 Momentos-L.....	5
1.1.3.2 Filtrado de datos.....	7
1.1.3.3 Identificación de regiones homogéneas	8
1.1.3.4 Elección de una distribución de frecuencia	12
1.1.3.5 Estimación de cuantiles regionales	12
1.1.4 Regionalización del índice de avenida	13
1.2 Antecedentes	15

CAPÍTULO II

PLANTEAMIENTO DEL PROBLEMA

2.1 Identificación del problema	17
2.2 Enunciado del problema	18
2.3 Justificación	18
2.4 Objetivos.....	19
2.4.1 Objetivo general	19
2.4.2 Objetivos específicos.....	19

CAPÍTULO III**MATERIALES Y MÉTODOS**

3.1	Área de estudio	20
3.2	Materiales.....	21
3.2.1	Información cartográfica	21
3.2.2	Datos hidrométricos.....	21
3.2.3	Datos climáticos	22
3.2.4	Programas informáticos.....	22
3.3	Métodos	23
3.3.1	Regionalización de la curva de crecimiento para la estimación de cuantiles de caudales máximos instantáneos en la cuenca amazónica del Perú	23
3.3.1.1	Estimación de parámetros morfoclimáticos.....	23
3.3.1.2	Estimación de caudales máximos instantáneos	23
3.3.1.3	Filtrado de datos.....	24
3.3.1.4	Identificación de regiones homogéneas.....	24
3.3.1.5	Selección de la distribución de frecuencia.....	25
3.3.1.6	Estimación de cuantiles	26
3.3.2	Regionalización del índice de avenida a partir de parámetros morfoclimáticos en la cuenca amazónica del Perú.....	26
3.3.2.1	Elección de mejores parámetros	26
3.3.2.2	Elección de mejores regresiones.....	26

CAPÍTULO IV**RESULTADOS Y DISCUSIÓN**

4.1	Regionalización de la curva de crecimiento para la estimación de cuantiles de caudales máximos instantáneos en la cuenca amazónica del Perú	29
4.1.1	Estimación de parámetros morfoclimáticos	29
4.1.2	Estimación de caudales máximos instantáneos	31
4.1.3	Filtrado de datos	32
4.1.4	Identificación de regiones homogéneas.....	35
4.1.5	Selección de la distribución regional de frecuencia	39
4.1.5.1	Medidas de bondad de ajuste	40
4.1.5.2	Estimación de cuantiles	41

4.2	Regionalización del índice de avenida a partir de parámetros morfoclimáticos en la cuenca amazónica del Perú	43
4.2.1	Elección de mejores parámetros	43
4.2.2	Elección de mejores regresiones	43
4.2.2.1	Multicolinealidad	45
4.2.2.2	Homocedasticidad y normalidad de residuos	45
	CONCLUSIONES	48
	RECOMENDACIONES	49
	BIBLIOGRAFÍA	50

ÍNDICE DE TABLAS

	Pág.
1. Distribuciones de probabilidad utilizados en hidrología	4
2. Valores críticos para la estadística de discordancia D_i	8
3. Momentos-L para distribuciones estadísticas	14
4. Características generales de las estaciones hidrométricas	22
5. Parámetros morfoclimáticos y características hidrológicas	31
6. Funciones para estimar caudales máximos instantáneos	32
7. Medida de discordancia regional de las estaciones hidrométricas.....	34
8. Prueba de heterogeneidad regional del grupo de estaciones hidrométricas.....	36
9. Mejores regresiones con parámetros morfoclimáticos	37
10. Estadísticos Z^{DIST} para las distribuciones de frecuencia.....	40
11. Criterios de selección de modelos de distribución de frecuencia	41
12. Prueba de bondad de ajuste y parámetros de la función de distribución GEV	41
13. Mejores regresiones entre Q y sus transformadas con parámetros morfométricos ..	44
14. Coeficientes de regresión del índice de avenida.....	45
15. Factor de Inflación de la Varianza (VIF) de las variables independientes	45

ÍNDICE DE FIGURAS

	Pág.
1. Diagrama de relación de momentos-L.....	5
2. Definición de la medida de discordancia	8
3. Dendograma de aglomeración de elementos	9
4. Optimización de grupos con un algoritmo de reasignación de elementos.....	10
5. Ubicación del área de estudio y distribución espacial de estaciones hidrométricas ...	21
6. Parámetros morfoclimáticos de las cuencas	30
7. Hidrograma de caudal máximo instantáneo para las estaciones hidrométricas	33
8. Dispersión de las relaciones de momentos-L de las muestras	34
9. Estaciones separadas en el espacio CA-L – CV-L	35
10. Regiones homogéneas identificadas	38
11. Trazado de funciones de distribución en el plano 'CA-L – CK-L'	39
12. Función de distribución GEV y estaciones región 1.....	41
13. a) Curva de crecimiento regional y b) regiones homogéneas	42
14. Coeficiente de correlación de Pearson.....	43
15. Trazo del diagnóstico del modelo índice de avenida	46

ACRÓNIMOS

AFL	:	Análisis de frecuencia local
ARF	:	Análisis regional de frecuencia
AD	:	Anderson Darling
ADC	:	Criterio de Anderson-Darling
AIC	:	Criterio de información de Akaike
AICc	:	Criterio de Información de Akaike corregido
BIC	:	Criterio de Información Bayesiano
D	:	Medida de discordancia
GEV	:	Valor extremo generalizado
GLO	:	Logística generalizada
GPA	:	Pareto generalizada
H	:	Medida de heterogeneidad
L-CA	:	Coefficiente de Asimetría-L
L-CK	:	Coefficiente de Curtosis-L
L-CV	:	Coefficiente de Variacion-L
GNO	:	Normal generalizada
PEII	:	Pearson tipo III
T	:	Periodo de retorno
VIF	:	Factor de Inflación de la varianza
Z^{DIST}	:	Estadístico Z

RESUMEN

La estimación del caudal de diseño en cuencas no instrumentadas para la planeación hidráulica, es un problema habitual en hidrología. La poca densidad de estaciones hidrométricas y la reducida longitud de series hidrológicas hacen que los profesionales utilicen metodologías no apropiadas para generar información en sitios sin registro. El análisis regional de frecuencia justifica su aplicación para estimar variables ambientales en sitios con o sin registro, reduciendo de esta manera los efectos que podrían generarse en el planeamiento hidráulico. El objetivo de la investigación fue realizar el análisis regional de frecuencia para la estimación de caudales máximos en la cuenca amazónica del Perú. El conjunto de datos utilizados, fueron mediciones de caudales máximos diarios recolectados de 10 estaciones hidrométricas, las mismas que se encuentran dentro del área de estudio. La metodología consistió en la regionalización de la curva de crecimiento para la estimación de cuantiles de caudales máximos instantáneos, así como la regionalización del índice de avenida a partir de parámetros morfoclimáticos. Los resultados, evidencian la identificación de tres regiones homogéneas determinadas mediante un proceso mixto, entre el algoritmo de Ward, una técnica para la reasignación de los elementos y la medida de heterogeneidad. Dos regiones (región 2 y 3) fueron excluidas del análisis por estar formado con una reducida cantidad de estaciones hidrométricas. Así entonces, la función de distribución valor extremo generalizado demostró adecuarse mejor a los datos de los sitios de la región 1, determinando de esta manera la curva de crecimiento regional. Por otro lado, el índice de avenida demostró estar en función del área de la cuenca, pues fue la variable independiente más significativa para la transferencia de información hidrológica. Se concluye que la curva de crecimiento regional junto con el índice de avenida puede ser utilizada de manera confiable para estimar los caudales máximos instantáneos para diferentes periodos de retorno en la cuenca amazónica del Perú. Estos hallazgos pueden brindar apoyo técnico a las instituciones públicas y privadas que están involucrados en la gestión ambiental, la gestión del riesgo de desastres, el control de inundaciones y el planeamiento hidráulico.

Palabras clave: Caudales máximos, cuenca amazónica, índice de avenida, momentos-L y regionalización hidrológica.

ABSTRACT

The design flow estimate in non-instrumented basins for hydraulic planning is a common problem in hydrology. The low density of hydrometric stations and the reduced length of hydrological series causes professionals to use inappropriate methodologies to generate information at sites without registration. The regional frequency analysis justifies its application to estimate environmental variables at sites with or without registration, thus reducing the effects that could be generated in hydraulic planning. The objective of the research was to carry out the regional frequency analysis for the estimation of maximum flows in the Amazon basin of Peru. The set of data used, were measurements of daily maximum flows collected from 10 hydrometric stations, the same that are within the study area. The methodology consisted of the regionalization of the growth curve for the estimation of quantiles of instantaneous maximum flows, as well as the regionalization of the avenue index from morphoclimatic parameters. The results show the identification of three homogeneous regions determined by a mixed process, between Ward algorithm, a technique for the reallocation of the elements and the measure of heterogeneity. Two regions (region 2 and 3) were excluded from the analysis because they were formed with a small number of hydrometric stations. Thus, the generalized extreme value distribution function proved to be better suited to the data of the sites in region 1, thus determining the regional growth curve. On the other hand, the avenue index proved to be a function of the area of the basin, as it was the most significant independent variable for the transfer of hydrological information. It is concluded that the regional growth curve together with the avenue index can be used reliably to estimate the maximum instantaneous flows for different return periods in the Amazon basin of Peru. These findings can provide technical support to public and private institutions that are involved in environmental management, disasters risk management, flood control and hydraulic planning.

Keywords: Maximum flows, amazon basin, index flood method, L-moments and hydrological regionalization.

INTRODUCCIÓN

La investigación intitulada “Análisis regional de frecuencia para la estimación de caudales máximos en la cuenca amazónica del Perú”, fue realizado con la finalidad de implementar modelos regionales que permitan estimar caudales máximos de diseño en cuencas sin información hidrológica. En la última década la cuenca amazónica, ha sido afectado por eventos extremos de inundaciones en 2009, 2012 y 2014 (Espinoza *et al*, 2012; Marengo y Espinoza, 2015), ocasionando cuantiosas pérdidas económicas. En el análisis de inundaciones los valores extremos de interés se ubican en la cola derecha de la distribución y estarían vinculados al impacto del cambio climático (Mujumdar y Nagesh Kumar, 2012). Probablemente a lo largo del siglo XXI los episodios de precipitaciones extremas que tienen un impacto directo en los caudales, sean más frecuentes e intensos en muchas regiones del mundo, bajo el futuro escenario del cambio climático (IPCC, 2014). En ese entender, los modelos regionales que planteamos, ayudarán a las entidades públicas y privadas, que están relacionados con el recurso hídrico, a formular planes para la gestión del riesgo de desastres, zonificación de llanuras de inundación y el diseño de estructuras hidráulicas.

Las estaciones hidrológicas, proporcionan mediciones de caudales en un espacio geográfico puntual, con los cuales es posible realizar predicciones de caudales máximos para periodos de retorno mediante el análisis de frecuencia local (AFL) (OMM, 2011) para el dimensionamiento de estructuras hidráulicas. El AFL de datos hidrológicos extremos, permite realizar predicciones confiables asociados a altos periodos de retorno (Campos, 2016), los cuales son necesarios para el diseño de una obra de ingeniería hidráulica. La literatura científica ha demostrado que los procedimientos para el análisis de frecuencia de una muestra de datos o AFL, están bien establecidos (Viglione, 2007; Hosking y Wallis 1997).

Sin embargo, la información que se requiere, no siempre está disponible en el espacio y tiempo por lo que se debe utilizar otras metodologías adecuadas. Algunas técnicas de interpolación espacial son utilizados para estimar variable hidrometeorológicas en sitios sin información por ejemplo, la precipitación, la temperatura, la evapotranspiración, etc., sin embargo los caudales máximos son acumulativos, no siendo posible utilizar técnicas de interpolación (Viglione, 2007). Entonces el análisis regional de frecuencia (ARF) es una metodología adecuada cuando nos encontramos en esta situación, pues su propósito

es utilizar registros de varios sitios para caracterizar las variables hidrológicas en sitios no instrumentados, permitiendo de esta manera obtener las variables estudiadas mediante datos no supervisadas como los parámetros morfométricos de las cuencas, con inferencias más precisas, cuando las muestras de datos se encuentren dentro de una misma región homogénea (Hosking y Wallis, 1997). La ventaja del ARF es que las distribuciones con 3 o incluso más parámetros se pueden estimar de manera más predecible de lo que sería posible con el AFL (Viglione, 2007).

Por tal motivo, el objetivo del estudio fue realizar el análisis regional de frecuencia para la estimación de caudales máximos en la cuenca amazónica del Perú y para ello se ha desarrollado básicamente una revisión de literatura, bajo los fundamentos del ARF, que han permitido dar a conocer la existencia de bases científicas relacionados al tema (capítulo I), caracterizando la problemática que habitualmente se asocia a la zona de estudio a través de una descripción detallada, preguntas de investigación, justificación, objetivos e hipótesis (capítulo II). En base a ello se propuso el procedimiento metodológico, desarrollado por objetivos específicos (capítulo III), presentando los resultados y discusiones en el capítulo IV, que básicamente ha consistido en explicar, comentar, comparar y discutir, corroborando la investigación con autores citados en la revisión de literatura.

CAPÍTULO I

REVISIÓN DE LITERATURA

1.1 Marco teórico

1.1.1 Periodo de retorno

El periodo de retorno T se define como el intervalo promedio de tiempo, en el que un evento de magnitud x sea igualado o excedido por lo menos una vez en ese intervalo de tiempo en años (OMM, 2011; Escalante y Reyes, 2002; Stedinger *et al.*,1992). Así, si un evento igual o mayor a x , ocurre una vez en T años, su probabilidad de ocurrencia P , es igual a $1/T$, en T casos, es decir:

$$P(X \geq x) = \frac{1}{T}$$

donde $P(X \geq x)$, es la probabilidad de ocurrencia de un evento mayor o igual a x y T es el periodo de retorno.

1.1.2 Análisis de frecuencia

Es la estimación de la frecuencia con la que ocurre un evento determinado a lo largo del tiempo (Viglione, 2007). Para que los resultados del análisis de frecuencia sean teóricamente válidos, las series de datos deberán satisfacer determinados criterios estadísticos, como los de aleatoriedad, independencia, homogeneidad y estacionalidad (OMM, 2011). La guía de prácticas hidrológicas OMM (1994), indica que las distribuciones de probabilidades de mayor uso en hidrología, son las presentadas en la Tabla 1 y la elección de una distribución, depende del método de estimación de parámetros. Tradicionalmente, el método de momentos ordinarios ha

sido muy usado en hidrología a pesar de que se ha reconocido que es ineficiente en comparación con el método de máxima verosimilitud (MV).

Hosking y Wallis (1997) ha encontrado que el método de momentos ponderados por probabilidad (MPP) que introdujo Greenwood *et al.* (1979) en muchos casos es conveniente aplicarlo, pues tiene propiedades estadísticas comparables con el método de MV en cuanto se refiere al tamaño de las muestras, sin embargo, un método más robusto, basado en las estadísticas de momentos-L constituye una mejora considerable con respecto a los métodos de momentos convencionales, MV y MPP.

Tabla 1

Distribuciones de probabilidad utilizados en hidrología

Nombre	Acrónimo	Referencia
Normal	(N)	(Gauss)
Log-normal	(LN)	(Hazen, 1914)
Pearson de tipo 3	(P3)	(Foster, 1924)
Valores extremos tipo 1	(EV1)	(Gumbel, 1941)
Valores extremos tipo 1	(EV2)	(Gumbel, 1941)
Valores extremos tipo 1	(EV3)	(Jenkinson, 1969)
Distribución gamma de tres parámetros		(Kritsky, Menkel, 1946)
Gamma	(G)	(Moran, 1957)
Log-Pearson tipo 3	(LP3)	(U.S. Water Resources Council, 1967, 1981)
Valores extremos generalizados	(GEV)	(Jenkinson, 1955)
Weibull		(Wu et Goodbrige, 1976)
Wakeby	(WAK)	(Houghton, 1978)
Boughton		(Boughton, 1980)
Valores extremos con dos componentes	(TCEV)	(Rossi, y col., 1984)
Logística logarítmica	(LLG)	(Ahmad, y col., 1988)
Logística generalizada	(GLG)	(Ahmad, 1988)

Fuente: OMM (1994)

1.1.3 Análisis regional de frecuencia

Una gran herramienta y método más exacto para el ARF es el método de momentos-L, pues involucra el filtrado de datos, la identificación de regiones homogénea, la selección de la distribución de frecuencia y la estimación de cuantiles (Hosking y Wallis, 1997). Además, abarca gran cantidad de muestras de datos de la misma variable en diferentes puntos de medición y reducen las incertidumbres en la estimación de cuantiles de eventos extremos (Viglione, 2007).

Para el ARF basado en momentos-L Fowler y Kilsby (2003) justifican la elección entre cuatro distribuciones, generalizada de valores extremos (GEV), Pearson III (PE3), logística generalizada (GLO) y normal generalizada (GNO). El coeficiente de asimetría-L y el coeficiente de curtosis-L son las cantidades útiles para resumir las distribuciones de probabilidad. Una distribución de dos parámetros con una ubicación y una escala se representan como un solo punto en el diagrama (L-Logística, N-Normal, U-Uniforme, G-Gumbel y E-Exponencial), mientras que una distribución de tres parámetros con ubicación, escala y forma como líneas continuas y discontinuas correspondientes a diferentes valores del parámetro de forma. Las distribuciones con más de un parámetro cubren áreas bidimensionales (Figura 1 y Tabla 3) (Hosking y Wallis, 1997).

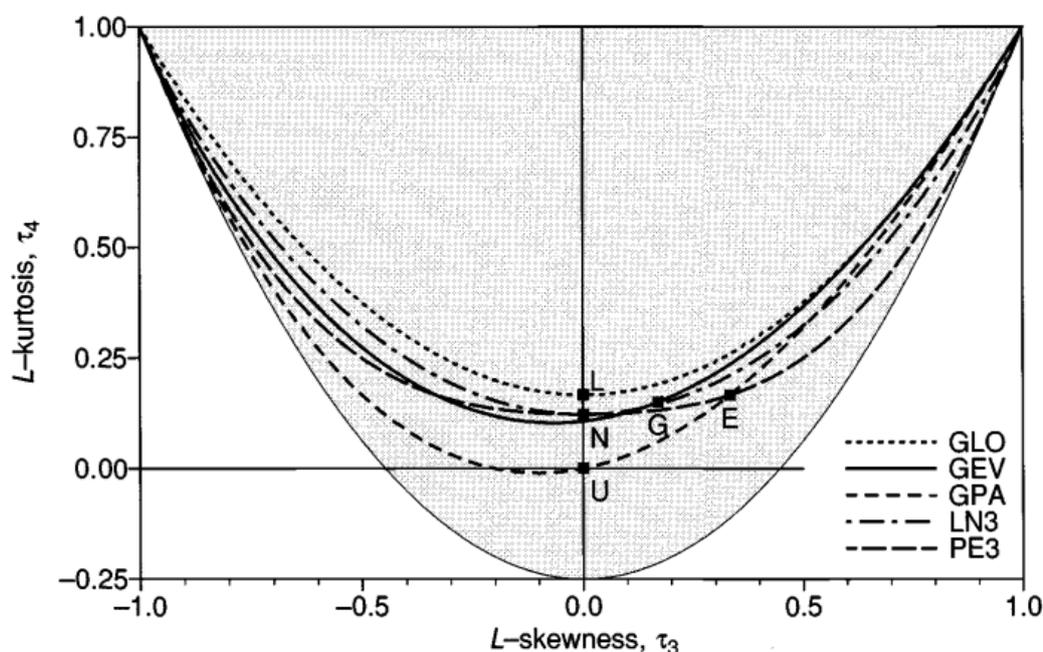


Figura 1. Diagrama de relación de momentos-L

Fuente: Hosking y Wallis (1997)

1.1.3.1 Momentos-L

Constituyen un sistema alternativo al método tradicional de momentos convencionales (Hosking, 1990) y surgen de combinaciones lineales de los momentos ponderados por probabilidad (MPP) introducidos por Greenwood *et al.* (1979). Hosking y Wallis (1997) la estimación se basa en una muestra de tamaño n , organizada en orden ascendente $x_{1:n} \leq x_{2:n} \leq$

$\dots \leq x_{n:n}$. Es conveniente comenzar con un estimador de MPP β_r . Un estimador imparcial de β_r es

$$\beta_r = n^{-1} \binom{n-1}{r} \sum_{j=r+1}^n \binom{j-1}{r} x_{j:n}$$

Alternativamente se puede escribir como

$$\beta_0 = n^{-1} \sum_{j=1}^n x_{j:n}$$

$$\beta_1 = n^{-1} \sum_{j=2}^n \frac{(j-1)}{(n-1)} x_{j:n}$$

$$\beta_2 = n^{-1} \sum_{j=3}^n \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{j:n}$$

$$\beta_3 = n^{-1} \sum_{j=4}^n \frac{(j-1)(j-2)(j-3)}{(n-1)(n-2)(n-3)} x_{j:n}$$

Análogamente los momentos-L de la muestra de están definidos por

$$l_1 = \beta_0$$

$$l_2 = 2\beta_1 - \beta_0$$

$$l_3 = 6\beta_2 - 6\beta_1 + \beta_0$$

$$l_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$$

donde l_1 es la ubicación-L o media de la distribución, l_2 es la escala-L. El coeficiente de variación-L (L-CV) se puede definir como:

$$t = \frac{l_2}{l_1}$$

Mientras que los coeficientes de asimetría-L (L-CA) y curtosis-L (L-CK) serían:

$$t_3 = \frac{l_3}{l_2} \quad y \quad t_4 = \frac{l_4}{l_2}$$

1.1.3.2 Filtrado de datos

Lo importante para cualquier análisis estadístico, es esencial examinar los datos (Viglione, 2007), a pesar de que los datos pueden ser confiables, es importante verificar que los datos fueron grabados o transcritos correctamente, y así también si los dispositivos de medición se han cambiado a una ubicación diferente con el tiempo, lo que significa que el ARF no será una base válida para estimar la distribución de probabilidad de mediciones futuras en el sitio. Para Hosking y Wallis (1997) en el contexto del ARF utilizando momentos-L, los valores incorrectos, valores atípicos, tendencias y cambios en la media de una muestra pueden reflejarse en los momentos-L de la muestra, por lo tanto, la medida de discordancia (D_i) se utiliza para reconocer sitios discordantes en una región homogénea.

a) Medida de discordancia

Según Hosking y Wallis (1997), si consideramos la dispersión de las estadísticas L-CV y L-CA de un grupo de estaciones, producirán una nube de puntos y el centro de la nube de puntos (+) es el promedio del grupo. Por lo tanto, se declara al sitio discordante a cualquier punto que esté lejos del centro de la nube (Figura 2). Se sugiere que un sitio sea considerado discordante si su D_i excede la medida crítica de discordancia D_c (Tabla 2). La medida de discordancia (D_i) para la estación i se define como:

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

donde $\bar{u} = N^{-1} \sum_{i=1}^N u_i$ es el promedio regional no ponderado de vectores u_i , $A = \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T$ es la matriz de sumas de cuadrados y productos cruzados, N es el número de estaciones en la región de estudio, $u_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}]^T$, es un vector de las relaciones de momentos-L para el sitio i -ésimo, $t^{(i)}$, $t_3^{(i)}$ y $t_4^{(i)}$ son el L-CV, L-CA y L-CK para la estación i respectivamente.

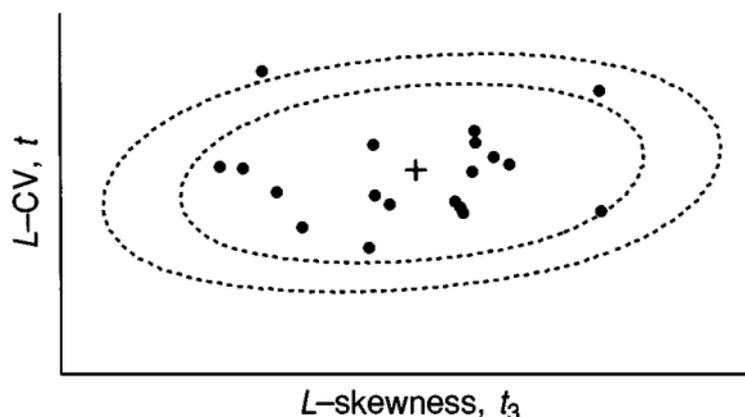


Figura 2. Definición de la medida de discordancia

Fuente: Hosking y Wallis (1997)

Tabla 2

Valores críticos para la estadística de discordancia D_i

Número de sitios	D_c	Número de sitios	D_c
5	1.333	10	2.491
6	1.648	11	2.632
7	1.917	12	2.757
8	2.140	13	2.869
9	2.329	14	2.971
		≥ 15	3.000

Fuente: Hosking y Wallis (1997)

1.1.3.3 Identificación de regiones homogéneas

La fase más compleja del ARF es la identificación de regiones homogéneas para estimar las curvas de crecimiento regional (Hosking y Wallis, 1997; Viglione, 2007). En esta etapa se puede definir la zona homogénea basado en las características geográficas (clima, vegetación, geología, geomorfología, etc.) o mediante pruebas estadísticas (Campos, 2006). En el pasado, las regiones de sitios contiguos se han elegido geográficamente siguiendo los límites administrativos, el término "región" sugiere un conjunto de sitios cercanos entre sí, pero la proximidad geográfica no es necesariamente un signo de similitud entre las distribuciones de frecuencia (Viglione, 2007). Algunos criterios que se utilizan para definir de manera más confiable regiones homogéneas se basan en el método de los trazos multidimensionales, método del coeficiente de variación (Escalante y Reyes, 2002), método de conglomerados y componentes principales

(Gottschalk, 1985), el enfoque de la región de influencia (ROI) (Burn, 1990) y el método de conglomerados de Ward (Rao y Srinivas, 2006).

a) Análisis de conglomerado

Es un método estándar de análisis estadístico multivariado para dividir regiones homogéneas (Hosking y Wallis, 1997), los grupos se forman a partir de características de sitio y no de estadísticas del sitio (Viglione, 2007). El aspecto más delicado, es la elección de las características del sitio (variables de clasificación), para la formación de regiones homogéneas. El análisis de conglomerados mide la similitud a través de la distancia euclidiana en el espacio de la variable de clasificación (Viglione, 2007), expresada como

$$d_{ij} = \sqrt{\frac{1}{p} \sum_{h=1}^p (x_{hi} - x_{hj})^2}$$

donde p es el número de variables de clasificación y x_{hi} es el valor de la variable h -ésima de la entidad i -ésima, estandarizada de modo que la muestra de los elementos de las variables tengan media 0 y varianza 1.

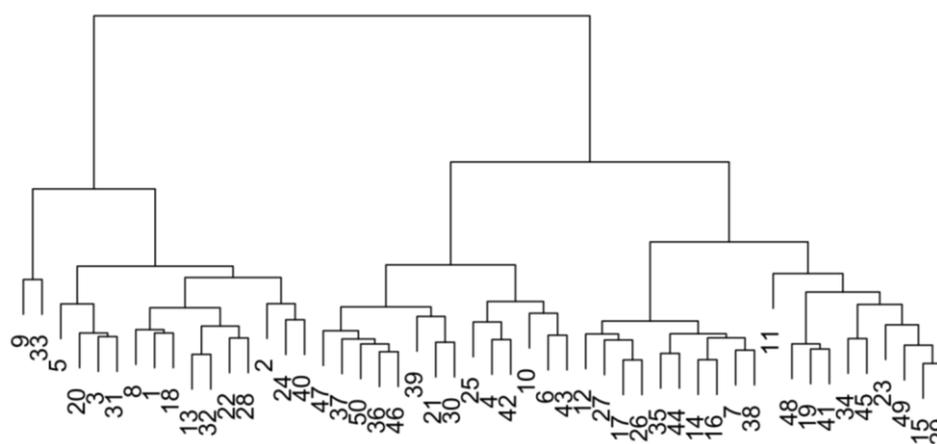


Figura 3. Dendrograma de aglomeración de elementos

Fuente: Viglione (2007)

El algoritmo de Ward (1963) es de tipo aglomerativo, es decir, inicia desde una situación en la que cada individuo constituye un grupo (Figura 3). Como todos los algoritmos jerárquicos, incluso Ward tiene la desventaja de no

admitir reasignaciones de elementos entre grupos, por lo que la configuración final no es necesariamente la óptima. Para superar este inconveniente, podemos usar un método mixto, que después de la subdivisión con el algoritmo de Ward, se debe utilizar una técnica para la reasignación de los elementos (Viglione, 2007) definida por

$$W = \sum_{i=1}^k \left(\sum_{j=1}^{n_i} \delta_{ij}^2 \right)$$

donde δ_{ij} es la distancia euclidiana entre el elemento j-ésimo del grupo i-ésimo y el centro de gravedad del grupo i-ésimo, n_i es el número de elementos del mismo y k es una subdivisión inicial obtenida con el algoritmo de Ward.

Sucesivamente se evalúa si el desplazamiento de un elemento entre dos grupos produce una disminución de W , en cuyo caso se aplica la corrección, hasta que alcanzamos una configuración en la que no hay modificaciones capaces de reducir aún más la dispersión de los grupos y estén más cerca (Figura 4) (Viglione, 2007). Así pues, el procedimiento se debe detener evaluando el número de grupos y para ello se utiliza la medida de heterogeneidad de Hosking y Wallis.

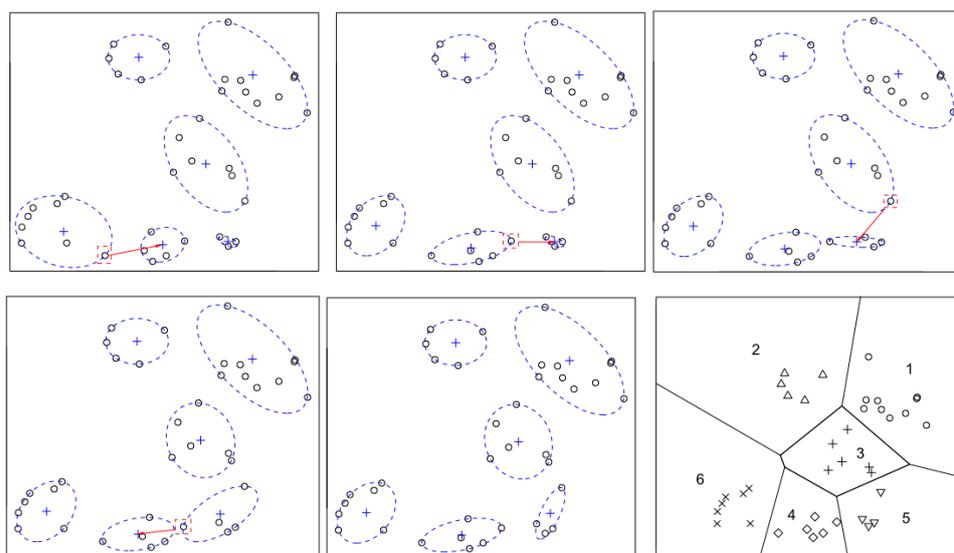


Figura 4. Optimización de grupos con un algoritmo de reasignación de elementos

Fuente: Viglione (2007)

b) Prueba de heterogeneidad regional

La homogeneidad de cada región se evalúa utilizando medidas de heterogeneidad H1, H2 y H3, cada una está basada en una medida diferente de dispersión entre sitios de las relaciones de momentos-L (L-CV, L-CA y L-CK). Hosking y Wallis (1993) observaron que H2 y H3 carecían del poder de discriminar entre regiones homogéneas y heterogéneas y que H1 basado en L-CV tenía un poder discriminante mucho mejor. Por lo tanto, H1 se recomienda como un indicador principal de heterogeneidad. En la práctica H1 es probablemente suficiente (Hosking y Wallis, 1997) y se define como:

$$H = \frac{(V - \mu_V)}{\sigma_V}$$

donde μ_V y σ_V son la media y desviación estándar de V , derivada de una gran cantidad de valores simulados (N_{sim}) de la región en estudio. La desviación estándar ponderada V , se calcula como:

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

donde N es la cantidad de sitios de una región homogénea, n_i el tamaño de muestra para la estación i , $t^{(i)}$, $t_3^{(i)}$ y $t_4^{(i)}$ denota la relación de momentos-L de la muestra, $t^{(R)}$, $t_3^{(R)}$ y $t_4^{(R)}$ son expresados como el promedio regional de L-CV, L-CA y L-CK, ponderada proporcionalmente a la longitud de registro de los sitios.

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

Se ajusta a una distribución kappa como su distribución de frecuencia las relaciones regionales promedio de momento-L 1, $t^{(R)}$, $t_3^{(R)}$ y $t_4^{(R)}$. Se simula una gran cantidad ($N_{sim} = 500$) de realizaciones de una región con N sitios. Se declara una región “aceptablemente homogénea” si $H < 1$, “posiblemente heterogénea” si $1 \leq H < 2$ y “definitivamente heterogénea” si $H \geq 2$ (Hosking y Wallis, 1997), o alternativamente “aceptablemente homogénea” si $H < 2$,

“posiblemente heterogénea” si $2 \leq H < 3$ y “definitivamente heterogénea” si $H \geq 3$ (Wallis *et al.*, 2007).

1.1.3.4 Elección de una distribución de frecuencia

a) Medida de bondad de ajuste

La distribución de frecuencia regional se elige en base a la prueba de bondad de ajuste, Z^{DIST} (Hosking y Wallis, 1997). Para cada distribución candidata Z^{DIST} se define,

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

donde τ_4^{DIST} es el coeficiente curtosis-L de la distribución ajustada, $DIST$ se refiere a GLO, GEV, GPA, GNO y PE3, la desviación estándar de t_4^R se calcula con,

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

y el sesgo de t_4^R queda definido por,

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

N_{sim} es el conjunto de datos regionales simulados, usando una distribución kappa (Tabla 3). Declare que el ajuste es adecuado si Z^{DIST} está suficientemente cerca de cero, siendo un criterio razonable $|Z^{DIST}| \leq 1.64$. Este criterio corresponde a la aceptación de la distribución hipotética a un nivel de confianza del 90 %.

1.1.3.5 Estimación de cuantiles regionales

El enfoque de índice de avenida (Darlymple, 1960) se usa para estimar los cuantiles de inundación correspondientes a períodos de retorno de T años.

$$Q_i(T) = \bar{Q}_i q(T)$$

donde $Q_i(T)$ es una estimación cuantil, \bar{Q} es el índice de avenida y $q(T)$ es la curva de crecimiento regional, que es una función cuantil escalada adimensional común a cada sitio en una región homogénea, $i = 1, 2, \dots, N$ denota los sitios y N el número de sitios. La media muestral de la serie de caudal máximo anual (l_1) es usada como índice de avenida (Hosking y Wallis, 1997; OMM, 2011; Viglione, 2007).

1.1.4 Regionalización del índice de avenida

El índice de avenida es el valor de la crecida correspondiente al periodo de retorno de 2 o 2.33 años o a la media de la muestra (OMM, 2011). La suposición clave de un procedimiento de índice de avenida es que los N sitios que forman una región homogénea tienen una distribución de frecuencia idéntica, denominada curva de crecimiento regional, aparte de un factor de escala específico de sitio, el índice de avenida (Hosking y Wallis, 1997). Para obtener valores en cuencas no instrumentadas, el uso de ecuaciones de regresión entre las características de la cuenca y el índice de avenida son usados (Viglione, 2007; Smith *et al.*, 2015; Viglione *et al.*, 2007)

Para una mayor eficiencia de modelos, se deben utilizar 04 transformaciones diferentes de la variable dependiente (\bar{Q}) para evitar la heterocedasticidad y la no normalidad de los residuos de las regresiones (Viglione, 2007; Viglione *et al.*, 2007). Si los residuos de los modelos de regresión muestran heterocedasticidad, el modelo pierde la eficiencia y confiabilidad.

$$\bar{Q} = \beta_0 + \beta_1 M_1 + \beta_2 M_2 + \dots + \beta_{p-1} M_{p-1} + \varepsilon$$

$$\sqrt{\bar{Q}} = \beta_0 + \beta_1 M_1 + \beta_2 M_2 + \dots + \beta_{p-1} M_{p-1} + \varepsilon$$

$$\sqrt[3]{\bar{Q}} = \beta_0 + \beta_1 M_1 + \beta_2 M_2 + \dots + \beta_{p-1} M_{p-1} + \varepsilon$$

$$\ln(\bar{Q}) = \beta_0 + \beta_1 M_1 + \beta_2 M_2 + \dots + \beta_{p-1} M_{p-1} + \varepsilon$$

donde \bar{Q} es el índice de avenida, M_i son los parámetros morfoclimáticos y β_i son los coeficientes de la regresión y ε es un término de error.

Tabla 3
Momentos-L para distribuciones estadísticas

Distribución	$f(x)$	$F(x)$	$x(F)$	Momentos-L
GNO	$\frac{e^{ky-y^2/2}}{\alpha\sqrt{2\pi}}$	$\Phi(y)$	No tiene forma analítica	$\lambda_1 = \xi + \alpha(1 - e^{k^2/2})/k, \lambda_2 = \frac{\alpha e^{k^2/2}}{k} [1 - 2\Phi(-k/\sqrt{2})],$ $\tau_3 \approx -k \frac{A_0 + A_1k^2 + A_2k^4 + A_3k^6}{1 + B_1k^2 + B_2k^4 + B_3k^6},$ $\tau_4 \approx \tau_3^0 - k^2 \frac{C_0 + C_1k^2 + C_2k^4 + C_3k^6}{1 + D_1k^2 + D_2k^4 + D_3k^6}$
GLO	$\frac{\alpha^{-1}e^{-(1-k)y}}{(1+e^{-y})^2}$	$\frac{1}{(1+e^{-y})}$	$\left\{ \begin{array}{l} \xi + \alpha \left[1 - \left\{ \frac{(1-F)/F^k}{k}, k \neq 0 \right\} \right. \\ \left. \xi - \alpha \log \left\{ \frac{(1-F)/F}{k}, k = 0 \right\} \right. \end{array} \right\}$	$\lambda_1 = \xi + \alpha(1/k - \pi/\sin k\pi), \lambda_2 = \alpha k\pi / \sin k\pi$ $\tau_3 = -k, \tau_4 = (1+5k^2)/6$
GEV	$\alpha^{-1}e^{-(1-k)y}e^{-y}$	e^{-y}	$\left\{ \begin{array}{l} \xi + \alpha \left[1 - (-\log F)^k \right] / k, k \neq 0 \\ \xi - \alpha \log(-\log F), k = 0 \end{array} \right\}$	$\lambda_1 = \xi + \alpha \left\{ 1 - \Gamma(1+k) \right\} / k, \lambda_2 = \alpha(1-2^{-k})\Gamma(1+k)/k$ $\tau_3 = 2(1-3^{-k})/(1-2^{-k}) - 3,$ $\tau_4 = \left\{ 5(1-4^{-k}) - 10(1-3^{-k}) + 6(1-2^{-k}) \right\} / (1-2^{-k})$
Kappa	$\alpha^{-1} [1 - k(x-\xi)/\alpha]^{1/k-1} [F(x)]^{1-h} [1 - h 1 - k(x-\xi)/\alpha ^{1/k}]^{1/h}$	$G\left(\alpha, \frac{x-\xi}{\beta}\right) / \Gamma(\alpha)$	$\xi + \frac{\alpha}{k} \left[1 - \left(\frac{1-F^h}{h} \right)^k \right]$	$\lambda_1 = \xi + \alpha(1-g_1)/k, \lambda_2 = \alpha(g_1 - g_2)/k,$ $\tau_3 = (-g_1 + 3g_2 - 2g_3)/(g_1 - g_2),$ $\tau_4 = (-g_1 + 6g_2 - 10g_3 + 5g_4)/(g_1 - g_2)$
PE3	$\frac{(x-\xi)^{\alpha-1} e^{-(\alpha-\xi)/\beta}}{\beta^\alpha \Gamma(\alpha)}$	$G\left(\alpha, \frac{x-\xi}{\beta}\right) / \Gamma(\alpha)$	No tiene forma analítica	$\lambda_1 = \xi + \alpha\beta, \lambda_2 = \pi^{-1/2} \beta \Gamma(\alpha + \frac{1}{2}) / \Gamma(\alpha),$ $\tau_3 = 6I_{1/3}(\alpha, 2\alpha) - 3$ $\tau_4 \approx \frac{C_0 + C_1\alpha^{-1} + C_2\alpha^{-2} + C_3\alpha^{-3}}{1 + D_1\alpha^{-1} + D_2\alpha^{-2}}$
GPA	$\alpha^{-1}e^{-(1-k)y}$	$1 - e^{-y}$	$\left\{ \begin{array}{l} \xi + \alpha \left[1 - (1-F)^k \right] / k, k \neq 0 \\ \left(\xi - \alpha \log(1-F) \right), k = 0 \end{array} \right\}$	$\lambda_1 = \xi + \alpha/(1+k), \lambda_2 = \alpha/\{(1+k)(2+k)\},$ $\tau_3 = (1-k)/(3+k), \tau_4 = (1-k)(2-k)/(3+k)(4+k)$

$f(x)$: función de densidad de probabilidad, $F(x)$ función de distribución acumulada, $x(F)$ función cuantil, parámetros: ξ (localización), α (escala), κ (forma), γ (forma), Φ es la función de distribución acumulada de una distribución normal, $A_n, B_n, C_n, D_n, n = 0, 1, \dots$ coeficientes de aproximación y Γ la función gamma. Fuente: (Hosking y Wallis, 1997).

1.2 Antecedentes

El ARF incluye el uso de momentos-L (Hosking y Wallis, 1997) junto con el método del índice de avenida (Dalrymple, 1960), el cual es una alternativa útil, y toma en consideración las deficiencias de los métodos de análisis local, aumentando la longitud de los registros disponibles mediante la transferencia de información de las diferentes estaciones que componen una supuesta región homogénea (Rodríguez y Marreno de León, 2011).

La metodología de análisis regional se ha aplicado con éxito en el modelado de inundaciones en una serie de estudios de caso de Estados Unidos (Vogel *et al.*, 1993), Australia (Vogel *et al.*, 1993; Zaman *et al.*, 2012), Nueva Zelanda (Madsen *et al.*, 1997), el Sur de África (Kachroo *et al.*, 2000), Turquía (Saf, 2009), Pakistán (Khan *et al.*, 2017), entre otros. Sin embargo Cunnane (1988) y GREHYS (1996) presentan comparaciones detalladas de las diversas metodologías de estimación regional para el análisis de frecuencias.

Un ARF exhaustivo, requiere de la identificación de regiones homogéneas y para ello Gottschalk (1985) aplicó el análisis de conglomerados y componentes principales e indica que el análisis de conglomerados es un método adecuado para esta etapa. El método de conglomerado de Ward ha sido estudiado ampliamente para la clasificación de diferentes datos climáticos e hidrológicos (Domroes *et al.*, 1998; Jackson y Weinand, 1995; Nathan y McMahon, 1990; Ramachandra Rao y Srinivas, 2006). Sin embargo, este procedimiento puede no ser apropiado, pues algunas características tienen mayor influencia en la forma de la distribución y se les debe dar mayor peso en la agrupación, es difícil elegir los pesos apropiados, pero, este no es un problema, ya que la validez de las regiones finales se prueba mediante la medida de heterogeneidad (Hosking y Wallis, 1997).

La utilización de análisis de conglomerados con las variables hidrológicas son las observaciones de grupos o variables en grupos sobre la base de la gran similitud de características hidrológicas, como las propiedades geográficas, físicas, estadística o estocástica (Hassan y Ping, 2012). Hosking y Wallis (1997) recomiendan que los procedimientos para la identificación de regiones homogéneas sean basados en estadísticas de sitio, los autores proponen una prueba estadística, materializado por la prueba de heterogeneidad y construido con base de los cocientes de momentos-L muestrales. Viglione (2007). Es preferible utilizar características del sitio en lugar de las

estadísticas del sitio, pues de este último los resultados del ARF no se pueden usar para sitios sin datos simplemente porque no se pueden estimar las estadísticas del sitio. Rodríguez y Marreno de León (2011) indican que al identificar regiones homogéneas existe la posibilidad de realizar estudios de frecuencia más certeros.

GREHYS (1996) en su trabajo de investigación utiliza la técnica de momentos-L en todas las etapas de análisis regional, incluyendo la delineación de región homogénea, identificación y prueba de distribución de frecuencias regionales y la estimación de cuantiles. La experiencia también muestra que, en comparación con momentos convencionales, momentos-L están menos sujetos a sesgo en la estimación (Hosking y Wallis, 1997; Vogel y Fennessey, 1993).

En su investigación Peel *et al.* (2001) demostraron que el uso de métodos gráficos con diagramas de relación momentos-L en el proceso de selección de distribución de datos regionales puede ser engañosa. Según Naghettini y Pinto (2007) la selección de distribución de datos homogéneos es mejor basada en el promedio de la muestra y no en una línea de mejor ajuste a través de los puntos de datos. Los resultados destacan la importancia de utilizar las pruebas de heterogeneidad junto con diagramas de relación momentos-L (Viglione, 2007; Hosking y Wallis, 1997).

Parida y Moalafhi (2008) encontraron en los análisis, que todo Botswana se comportó como una región homogénea y los datos siguieron una función de distribución valor extremo generalizado (GEV), lo que sugiere que los parámetros GEV regionales podrían ser transferidas a los sitios específicos con confianza. La distribución GEV es usada ampliamente en el ARF (Cunnane, 1989; Parida y Moalafhi, 2008).

Las aplicaciones de ARF en Perú, también fueron estudiadas con resultados satisfactorios en la estimación de eventos extremos como precipitaciones máximas y sequías (Acuña *et al.*, 2015; Fernández y Lavado, 2016; Lujano y Obando, 2015) y regionalización de caudales (Acuña *et al.*, 2011).

CAPÍTULO II

PLANTEAMIENTO DEL PROBLEMA

2.1 Identificación del problema

Un problema común en hidrología, es la estimación de caudales máximos para el diseño de estructuras hidráulicas en cuencas sin información hidrológica. Es cierto, que los eventos extremos de caudales máximos, son fenómenos recurrentes que ocasionan inundaciones, desbordes de ríos, destrucción de obras hidráulicas, destrucción de vías, erosión de suelos agrícolas, aniego de cultivos y aniego de viviendas. Estos eventos se han presentado durante mucho tiempo, provocando daños económicos-sociales y pérdida de vidas humanas. Probablemente a lo largo del siglo XXI los episodios de precipitaciones extremas que tienen un impacto directo en los caudales, sean más frecuentes e intensos en muchas regiones del mundo, bajo el futuro escenario del cambio climático (IPCC, 2014).

En la cuenca amazónica de Perú, la estimación de caudales máximos para diferentes periodos de retorno, se realiza normalmente con el análisis de frecuencia local, cuyo procedimiento de cálculo según Viglione (2007) y Hosking y Wallis (1997) están bien establecidos. Sin embargo, el uso tradicional del análisis de frecuencia local para estimar la avenida de diseño, no se justifica cuando se tiene una cantidad reducida de datos hidrológicos y la inexistencia de información en sitios no instrumentados. Entonces, el ARF es una alternativa de solución, cuando en el área de interés nos encontramos frente a una gran cantidad de estaciones hidrológicas, una diferente longitud de registro histórico y la falta de información en cuencas no instrumentadas. Asimismo, incluye información histórica con registros cortos y el uso de funciones de distribución de tres parámetros que son más flexibles que las funciones de distribución de dos parámetros.

El Servicio Nacional de Meteorología e Hidrología del Perú, cuenta con información hidrológica en puntos estratégicos dentro de la cuenca amazónica del Perú, el cual fue un inicio clave para el ARF.

2.2 Enunciado del problema

2.2.1 Pregunta general

¿Permite el análisis regional de frecuencia, estimar los caudales máximos en la cuenca amazónica del Perú?

2.2.2 Preguntas específicas

- ¿Es posible estimar cuantiles de caudales máximos instantáneos a partir de la curva de crecimiento regional en la cuenca amazónica del Perú?
- ¿Existe relación entre el índice de avenida y los parámetros morfoclimáticos en la cuenca amazónica del Perú?

2.3 Justificación

El diseño y funcionamiento de estructuras hidráulicas depende del conocimiento de las entradas del caudal máximo instantáneo para un periodo de retorno determinado. Sin embargo, estos valores son escasos y deben de predecirse a partir de una función de distribución teórica que mejor representa a la muestra de datos de caudales máximos instantáneos en un sitio de interés. El análisis de frecuencia se basa en obtener esa estimación para un tiempo de retorno, pero el caudal estimado no siempre es confiable si el tiempo de retorno es mayor que la cantidad de datos que se tiene en una estación. Pero el problema se agrava más cuando se requiere proyectar una obra de ingeniería hidráulica y no existe información disponible para conocer la frecuencia de valores extremos y los profesionales en el área, extrapolan la información con metodologías no adecuadas y ello, genera la subestimación del caudal de diseño que conlleva básicamente al subdimensionamiento de la obra de ingeniería hidráulica con un posible riesgo de falla, mientras que por el contrario se genera una sobreestimación del caudal de diseño que conlleva al sobredimensionamiento de la estructura hidráulica que por supuesto incrementan los costos de construcción.

En tal razón, existe la necesidad e importancia de realizar el ARF y contribuir con modelos regionales operativos para la sociedad, que, de una manera breve, rápida, fiable y a un costo económico, nos permita solucionar la estimación de caudales máximos de diseño en cuencas con o sin información hidrológica, dentro de la cuenca amazónica de Perú, cuyos resultados servirán como insumo para la gestión ambiental, la gestión del riesgo de desastres, el control de inundaciones, el planeamiento hidráulico y el diseño de estructuras hidráulicas.

2.4 Objetivos

2.4.1 Objetivo general

Realizar el análisis regional de frecuencia para la estimación de caudales máximos en la cuenca amazónica del Perú

2.4.2 Objetivos específicos

- Regionalizar la curva de crecimiento para la estimación de cuantiles de caudales máximos instantáneos en la cuenca amazónica del Perú.
- Regionalizar el índice de avenida a partir de parámetros morfoclimáticos en la cuenca amazónica del Perú.

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1 Área de estudio

El área de estudio, forma parte de la región hidrográfica del Amazonas (RHA) – Perú y Ecuador. Se caracteriza como un sistema de cuenca exorreica, con una superficie aproximada de 877,478.6 km², bordeando el límite oeste con la región hidrográfica del pacífico (RHP), por el norte con Colombia, mientras que por el este limita con Brasil. La región hidrográfica del Titicaca (RHT) y parte de la RHP son sus límites hidrográficos por el sur (Figura 5). El régimen hidrológico en la zona sur de la cuenca se generaliza en verano austral mientras que en la parte norte en otoño. Ello se debe a que la inclinación de la tierra favorece a la intensificación solar, formando sistemas nubosos que originan precipitaciones debido a la interacción de la Zona de Convergencia Intertropical (ZCIT), la Temperatura Superficial del Mar (TSM), el Monzón Sudamericano y la Alta de Bolivia.

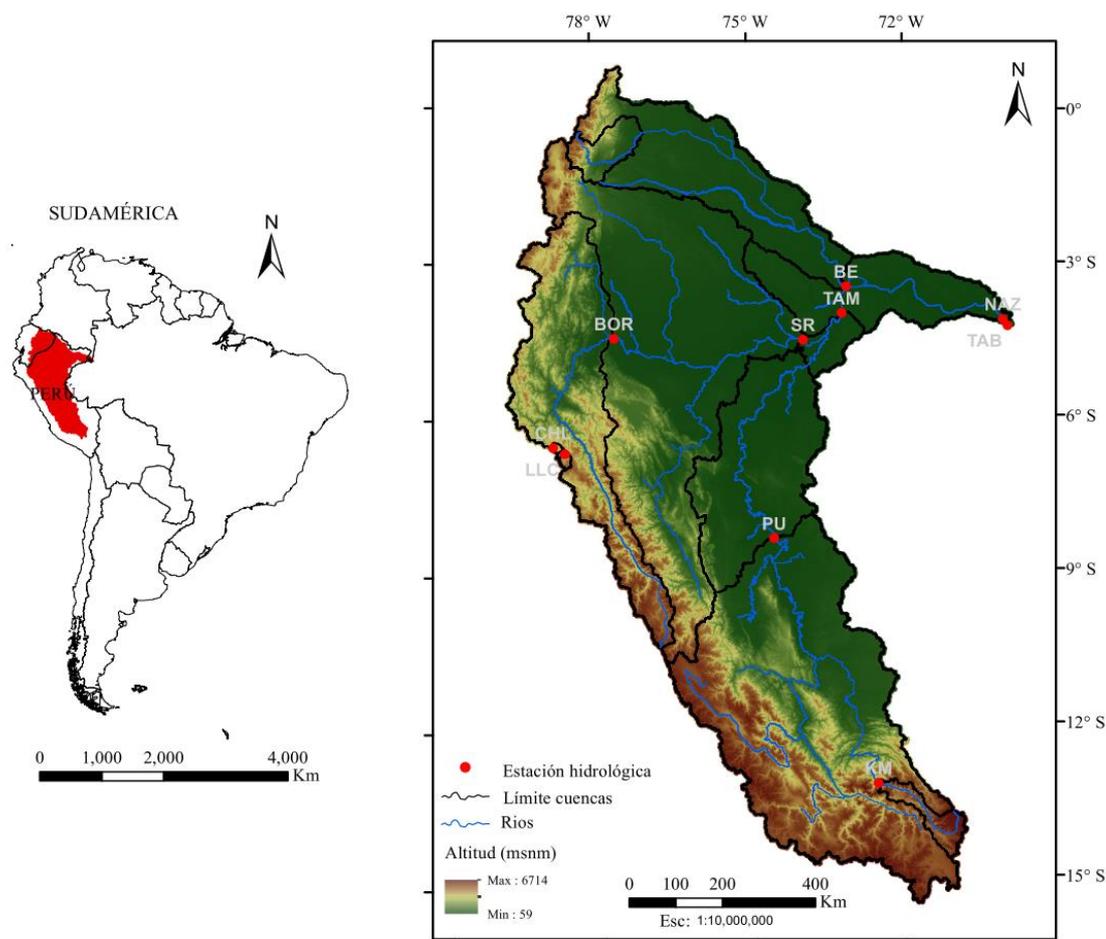


Figura 5. Ubicación del área de estudio y distribución espacial de estaciones hidrométricas

3.2 Materiales

3.2.1 Información cartográfica

Para caracterizar los parámetros morfométricos de las cuencas, fue necesario un modelo digital de elevación (DEM) de la Misión Topográfica del Radar Shuttle de la NASA (SRTM), obtenido de la plataforma google earth engine (GEE) (<https://earthengine.google.com/>), ID de la imagen CGIAR/SRTM90_V4 (Jarvis *et al.*, 2008), con una resolución espacial de ~90 m.

3.2.2 Datos hidrométricos

Se recopilaron datos de caudales máximos promedios diarios de 10 estaciones ubicadas dentro de la cuenca amazónica de Perú, de las cuales 06 estaciones hidrológicas (Tamishiyacu, Km 105, Chotano Lajas, Llaucano Corellama,

Bellavista, Pucallpa) pertenecen al Servicio Nacional de Meteorología e Hidrología del Perú, mientras que 04 estaciones (Tabatinga, Nazareth, Borja, Francisco de Orellana y San Regis) corresponden al Servicio de Observación SO HYBAM (anteriormente Observatorio de Investigación Ambiental) "Control geodinámico, hidrológico y biogeoquímico de la erosión / alteración y transporte de materiales en las cuencas del Amazonas, el Orinoco y el Congo", que se encuentra operativo desde 2003, respondiendo a una invitación del Ministerio de Educación Superior de Francia y la investigación, cuyo objetivo es proporcionar a la comunidad investigadora los datos científicos de alta calidad necesarios para comprender y modelar el comportamiento de los sistemas y su dinámica a largo plazo (<http://www.ore-hybam.org/index.php/eng>) (Tabla 4).

Tabla 4

Características generales de las estaciones hidrométricas

Cód.	Río	Estación	Lat. [°]	Long. [°]	Periodo
1	Amazonas	Tamishiyacu (TAM)	-4.000	-73.160	1985-2010
2	Ucayali	Km 105 (KM)	-13.183	-72.534	1958-2012
3	Chotano	Chotano Lajas (CHL)	-6.560	-78.741	1979-2008
4	Llaucano	Llaucano Corellama (LLC)	-6.687	-78.518	1980-2011
5	Napo	Bellavista (BE)	-3.480	-73.080	1990-2009
6	Ucayali	Pucallpa (PU)	-8.378	-74.533	1988-2009
7	Amazonas	Tabatinga (TAB)	-4.250	-69.933	1983-2017
8	Amazonas	Nazareth (NAZ)	-4.121	-70.036	1990-2004
9	Marañon	Borja (BO)	-4.470	-77.548	1987-2016
10	Marañon	San Regis (SR)	-4.516	-73.908	1999-2014

3.2.3 Datos climáticos

Se utilizaron datos climáticos grillados de precipitación PISCOp v2.1 beta (climatología 1981-2010), disponible en (<ftp://ftp.senamhi.gob.pe/>). PISCOp es un conjunto de datos de precipitación histórica que tiene información disponible para el área de estudio desde 1981 a la actualidad, con una resolución temporal diaria y una resolución espacial de 0.1° (Aybar *et al.*, 2017).

3.2.4 Programas informáticos

Utilizamos el programa RStudio, el cual es un entorno de desarrollo integrado (IDE) para R. RStudio está disponible para su descarga en código abierto en el enlace (<https://www.rstudio.com/>). Aunque no es posible su instalación directa, se debe primero descargar e instalar R. R (R Core Team, 2019) es un entorno de software

libre para computación estadística y gráficos, disponible en (<https://www.r-project.org/>). Los paquetes adicionales para el desarrollo del ARF son: Lmom (Hosking, 2019), LmomRFA (Hosking, 2019) y nsRFA (Viglione, 2018).

Para delimitar cuencas y estimar sus parámetros morfométricos fue necesario utilizar ArcGIS para escritorio (componente ArcMap) y la herramienta Arc Hydro. Arc Hydro fue desarrollado por el Centro de Investigaciones en recursos de agua de la Universidad de Texas, Austin. También se ha usado la herramienta hidrología que se encuentra dentro de análisis espacial de ArcMap.

3.3 Métodos

3.3.1 Regionalización de la curva de crecimiento para la estimación de cuantiles de caudales máximos instantáneos en la cuenca amazónica del Perú

3.3.1.1 Estimación de parámetros morfoclimáticos

Los parámetros morfoclimáticos de las cuencas (variables independientes), fueron estimados utilizando herramientas SIG, a partir del sitio de estaciones hidrológicas. Los parámetros morfométricos considerados fueron: el área de la cuenca (A) en km², altitud media de la cuenca (Emc) en m.s.n.m., pendiente media de la cuenca (Pmc) en %, longitud del río principal (Lrp) en km, pendiente del río principal (Pmrp) en %, elevación media del río principal (Emrp) en msnm., precipitación media de la cuenca (Pmc) en mm, radio de circularidad (Rc) e índice de compacidad (K) como valores adimensionales.

3.3.1.2 Estimación de caudales máximos instantáneos

Para la estimación de caudales máximos instantáneos, nos basamos en la relación propuesta por Fuller (1914). Para ello fueron necesarios los caudales máximos promedios diarios de las estaciones hidrométricas y el área de la cuenca. Fuller presentó la relación entre el caudal máximo promedio diario (Qm) y el máximo instantáneo (Qp), el cual responde a la función ajustada:

$$Q_p = Q_m \left(1 + \frac{2.66}{A^{0.3}}\right)$$

donde Q_p es el caudal máximo instantáneo (m^3/s), Q_m el caudal máximo promedio diario (m^3/s), y A el área de la cuenca (Km^2).

3.3.1.3 Filtrado de datos

En esta etapa analizamos primero todo el conjunto de datos para verificar la existencia de valores incorrectos, valores atípicos, tendencias y cambios en la media de una muestra utilizando la medida de discordancia (D_i). Un sitio es discordante si $D_i > D_c$.

3.3.1.4 Identificación de regiones homogéneas

a) Elección de la prueba de heterogeneidad regional

Se utilizó la dispersión de puntos entre L-CA y L-CV, si los valores del L-CA son inferiores a 0.23 los puntos de dispersión se ubican en el espacio de Hosking y Wallis (HW) y es la parte donde la prueba de heterogeneidad de H es más eficaz, en el caso de que la dispersión de puntos L-CA sean mayores a 0.23, es preferible utilizar la prueba de bootstrap de AD (Viglione, 2007).

b) Identificación de las variables de clasificación

La determinación de las variables de clasificación para la formación de las regiones se realizó identificando el mejor modelo de regresión lineal que vincule la matriz de las distancias de las curvas de crecimiento $\Delta_{q(F)}$ con las matrices de las distancias de los parámetros morfoclimáticos.

$$\Delta_{q(F)} = \beta_0 + \beta_1 \Delta_{x_1} + \beta_2 \Delta_{x_2} + \dots + \beta_{p-1} \Delta_{x_{p-1}} + \varepsilon$$

donde $\Delta_{q(F)}$ es la matriz de las distancias entre las curvas de crecimiento de la muestra, Δ_{x_1} son las matrices de distancias entre las variables candidatas para ser considerado como clasificación para la formación de regiones homogéneas, β_0 son los coeficientes de regresión. El mejor modelo

regresivo entre las matrices de distancia fue elegido mediante el coeficiente de determinación ajustado (R_{adj}^2).

También se evaluó el supuesto de multicolinealidad que indica la no existencia de una relación lineal entre los regresores (ausencia de multicolinealidad). La multicolinealidad fue evaluada mediante el factor de inflación de la varianza (VIF).

c) Algoritmo de Ward

La metodología para este propósito de división de regiones homogéneas se realizó con el algoritmo de jerárquico de Ward (1963), utilizando una técnica de reasignación de los elementos, que consiste en el refinamiento grupal con un método de minimización de la dispersión dentro de los conglomerados.

d) Prueba de heterogeneidad

Entonces, el procedimiento se detuvo evaluando la homogeneidad de cada región formada y para ello se tomó en consideración el umbral de $H1 \geq 2$, considerado como región heterogénea.

3.3.1.5 Selección de la distribución de frecuencia

Se evaluaron cinco funciones de distribución probabilística de tres parámetros GLO, GEV, GPA, GNO y PE3 y la estimación de parámetros se realizó mediante el método de momentos-L, método más eficiente cuando las muestras son de longitud pequeña o moderada.

Prueba de bondad de ajuste

Para la selección del modelo probabilístico, se recurrieron a las pruebas de bondad de ajuste Z^{DIST} , el criterio de información de Akaike (AIC), criterio de información de Akaike corregido (AICc), criterio de información Bayesiano (BIC) y el criterio de Anderson-Darling (ADC). Valores de $|Z^{DIST}| \leq 1.64$, y valores más bajos de AIC, AICc, BIC y ADC son recomendables para identificar la distribución principal, mientras que,

valores cercanos a cero para AD indican mejor adaptación de la distribución a la muestra.

3.3.1.6 Estimación de cuantiles

Se utilizó el índice de avenida (Darlymple, 1960), considerado como la media muestral de la serie de caudal máximo anual (l_1) (Hosking y Wallis, 1997; OMM, 2011; Viglione, 2007).

$$Q_i(T) = \bar{Q}_i q(T)$$

donde $Q_i(T)$ es una estimación cuantil, \bar{Q} es el índice de avenida y $q(T)$ es la curva de crecimiento regional, que es una función cuantil escalada adimensional común a cada sitio en una región homogénea, $i = 1, 2, \dots, N$ denota los sitios y N el número de sitios.

3.3.2 Regionalización del índice de avenida a partir de parámetros morfoclimáticos en la cuenca amazónica del Perú

3.3.2.1 Elección de mejores parámetros

Se utilizó una gráfica de matriz de correlación entre los parámetros morfoclimáticos y el índice de avenida. La significación estadística del coeficiente de correlación, fue evaluada mediante la prueba de t'Student con un nivel de confianza del 95% y un nivel de significancia del 5%.

3.3.2.2 Elección de mejores regresiones

Se realizó construyendo modelos de regresión lineal simple y múltiple que relaciona el índice de avenida (variable dependiente) con las características morfométricos (variables independientes). Para una mayor eficiencia de modelos, se utilizaron 04 transformaciones diferentes de la variable dependiente (\bar{Q}) para evitar la heterocedasticidad y la no normalidad de los residuos de las regresiones (Viglione, 2007; Viglione *et al.*, 2007). Si los residuos de los modelos de regresión muestran heterocedasticidad, el modelo pierde la eficiencia y confiabilidad.

Para la estimación de los coeficientes de regresión, se utilizó la técnica de mínimos cuadrados y la significancia de los mismo fueron evaluados

mediante la prueba t de Student al 5%, las variables explicativas que no eran significativas, fueron excluidos de los modelos. Los resultados de cada regresión se evaluaron a través del coeficiente de determinación ajustado R_{aj}^2 , definido por:

$$R_{aj}^2 = 1 - \frac{(n-1) \sum_{i=1}^n (\bar{Q}_i - \hat{Q}_{m,i})^2}{(n-p) \sum_{i=1}^n (\bar{Q}_i - \bar{Q}_m)^2}$$

donde n es el número de estaciones consideradas, p el número de coeficientes estimados, \bar{Q} el caudal máximo promedio multianual medido, \hat{Q}_m el caudal máximo promedio multianual estimado en el sitio i-ésimo, \bar{Q}_m el promedio del caudal máximo promedio multianual medido para todas las estaciones consideradas. Para las demás ecuaciones, se sustituirán las diferentes transformaciones de la variable dependiente.

El coeficiente de determinación ajustado, fue necesario para elegir el mejor modelo entre los que pertenece a una clase determinada, pero no se puede utilizar para comparar modelos de naturaleza diferente (Veza *et al.*, 2010). Para ello se llevará a cabo un método de validación cruzada, calculando la RMSE (Root Mean Square Error) en los residuos $\hat{Q}'_{m,i} - \bar{Q}_{m,i}$, donde $\hat{Q}'_{m,i}$ es el valor estimado de la variable dependiente i-ésima obtenida utilizando un modelo estimado con todas las observaciones excepto la i-ésima. Además, la validación cruzada es una emulación completa del caso de la aplicación del modelo en sitios no sometidos a prueba. El $RMSE_{CV}$ está definido por:

$$RMSE_{CV} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Q}'_{m,i} - \bar{Q}_{m,i})^2}$$

La combinación de modelos, deben satisfacer los supuesto generales como la ausencia de multicolinealidad entre los coeficientes del modelo de regresión, la homocedasticidad (variación del residuo es constante) y la normalidad de residuos (los residuales se distribuyen normal) (Veza *et al.*, 2010; Veza *et al.*, 2009).

a) Multicolinealidad

Uno de los supuestos, se basa en que la relación entre la variable dependiente y los regresores es lineal o aproximadamente lineal y que no existe una relación lineal entre los regresores (ausencia de multicolinealidad). La multicolinealidad fue evaluada mediante el Factor de Inflación de Variación (VIF):

$$VIF = (1 - R_j^2)^{-1}$$

VIF mayores a 5 o 10, es una indicación que los coeficientes de regresión asociados están mal estimados e implican problemas serios de multicolinealidad (Montgomery *et al.*, 2012; Viglione *et al.*, 2007).

b) Homocedasticidad y normalidad de los residuos

Para detectar la heterocedasticidad, se realizó una gráfica de los residuos contra los valores ajustados y además se efectuó la prueba de homocedasticidad de Harrison-McCabe (1979). Se rechaza la hipótesis nula si la estadística es demasiado pequeña ($p < 0.05$). Si se confirma heterocedasticidad, se puede utilizar el método de mínimos cuadrados ponderados (WLS) para estimar los parámetros y si además del problema de heterocedasticidad, las oscilaciones de los pares (x,y) se correlacionan entre sí, se debe utilizar es el de los mínimos cuadrados generalizados (GLS) (Viglione, 2007).

Para detectar la no normalidad de los residuos (H_0 : los residuales se distribuyen normal), se utilizó la prueba de Anderson-Darling (Laio, 2004).

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1 Regionalización de la curva de crecimiento para la estimación de cuantiles de caudales máximos instantáneos en la cuenca amazónica del Perú

4.1.1 Estimación de parámetros morfoclimáticos

Consideramos 09 características morfoclimáticas de las cuencas hidrográficas, que incluyen el área de la cuenca (Figura 6a), la altitud media de la cuenca (Figura 6b), la pendiente media de la cuenca (Figura 6c), longitud del río principal (Figura 6d), altitud media del río principal (Figura 6e), pendiente media del río principal (Figura 6f), precipitación total anual (Figura 6f), radio de circularidad e índice de compacidad. Los resultados cuantitativos de los parámetros morfoclimáticos son presentados en la Tabla 5. Según Tucci (2002), las características más utilizadas en la regionalización son: el área de drenaje, la longitud del río principal, la pendiente media del río principal, la densidad de drenaje y el desnivel de la cuenca, mientras que Viglione (2007) utiliza características morfométricas de las cuencas, entre ellas el área, la elevación media, pendiente media, longitud del río principal, pendiente del río principal, área por encima de 2000 msnm, orientación de la cuenca, centro de gravedad, radio de circularidad y características climáticas como el índice de Thornthwaite y el índice de Budyko.

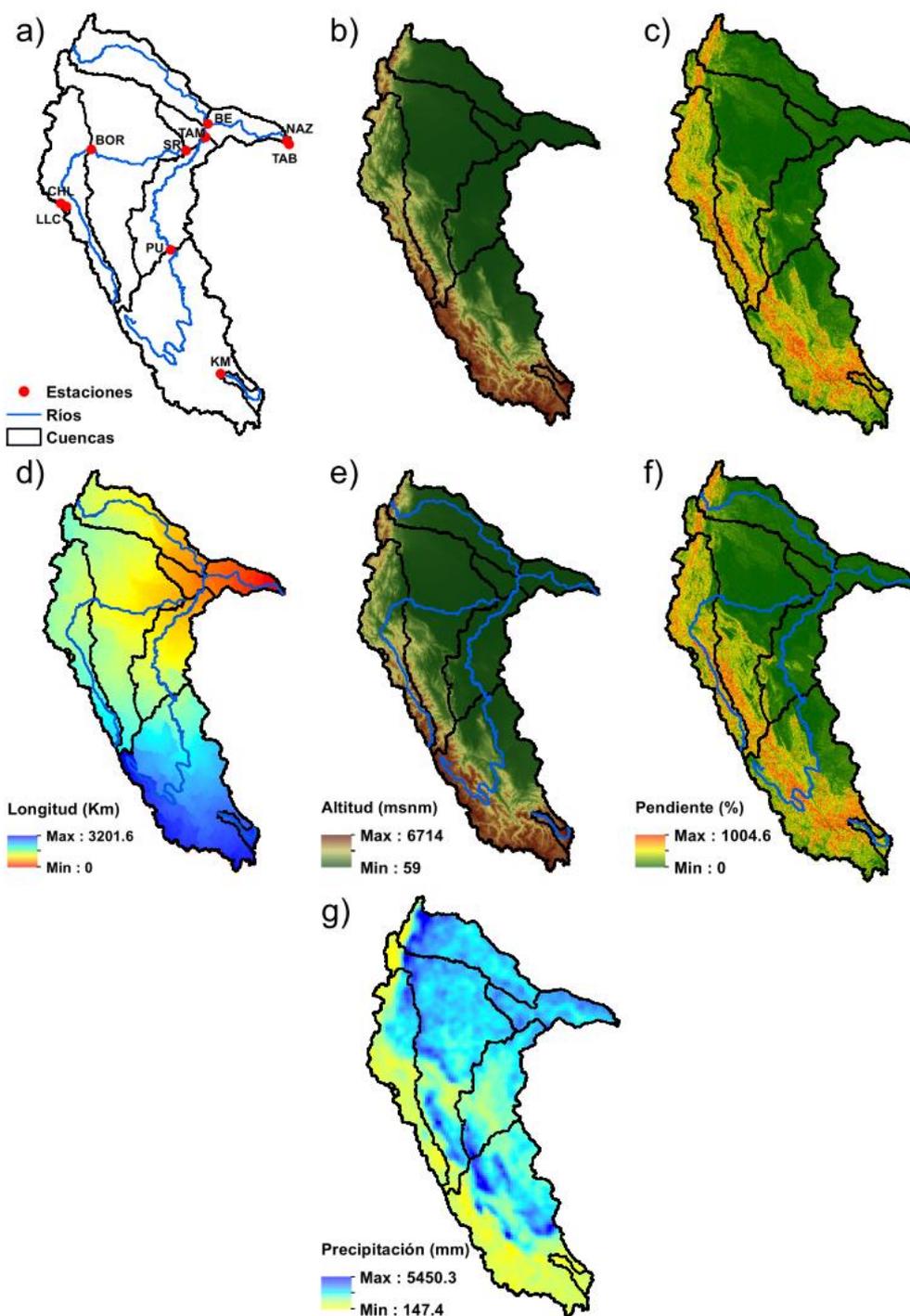


Figura 6. Parámetros morfoclimáticos de las cuencas

Tabla 5

Parámetros morfoclimáticos y características hidrológicas

Est.	\bar{Q} [m ³ /s]	Pma [mm]	A [km ²]	Emc [msnm]	Pmc [%]	Lrp [km]	Pmp [%]	Emrp [msnm]	Rc	K
TAM	48594.2	2048.2	719917.8	1489.9	20.4	2667.9	4.2	867.9	0.10	3.10
KM	640.8	804.2	9613.3	4193.8	32.3	325.8	10.9	3463.4	0.11	2.98
CHL	74.4	982.2	360.4	2871.3	18.9	30.6	8.8	2704.9	0.35	1.70
LLC	117.8	879.7	608.7	3477.2	23.3	42.3	11.2	3140.0	0.34	1.71
BE	11820.2	3041.1	99779.4	552.9	8.9	1187.7	3.4	176.1	0.11	3.01
PU	20164.4	1732.3	260890.0	2306.6	26.1	1636.3	5.7	1352.8	0.13	2.77
TAB	55092.9	2241.1	877478.6	1293.6	17.9	3195.5	3.8	738.1	0.09	3.31
NAZ	54278.8	2241.7	877066.5	1294.2	17.9	3176.9	3.8	741.7	0.09	3.28
BO	13178.4	1529.4	114529.8	2147.0	33.9	995.6	7.2	1497.9	0.08	3.44
SR	29042.3	2227.8	356882.9	1250.8	20.3	1572.1	5.4	985.1	0.12	2.83

4.1.2 Estimación de caudales máximos instantáneos

Consideramos estaciones hidrométricas que tienen un registro histórico ≥ 15 años. Según la OMM (2011) para definir una distribución de frecuencia hidrológica estable que permita estimar predicciones probabilísticas en un sitio, es necesario que el tamaño de la muestra sea suficientemente grande. Sin embargo, Hosking y Wallis (1997) indica que en el análisis regional para estimar las predicciones hidrológicas y para definir una distribución estable el tamaño de muestra debe ser ≥ 15 años, además, mencionan que utilizando el método de estimación de parámetros (momentos-L) pueden producir resultados muy fiables con tamaños de muestra pequeñas e incluso con valores atípicos.

La fórmula de Fuller fue utilizado para estimar los caudales instantáneos (Q_p) de 10 estaciones hidrométricas, las mismas que se encuentran dentro de la cuenca amazónica de Perú. Es posible deducir, que con un aumento en el tamaño de la cuenca el coeficiente para estimar Q_p disminuye, mientras que en pequeñas cuencas el coeficiente se incrementa (Tabla 6). Para todas las inundaciones Q_p debe ser mayor que Q_m , esto se debe que en las grandes cuencas, la tasa de escorrentía es alta durante al menos 24 horas porque la tormenta que la genera es de duración considerable, mientras que en pequeñas cuencas una tormenta puede causar inundaciones en pocas horas resultando un Q_p grande y Q_m moderado (Fuller, 1914).

Tabla 6

Funciones para estimar caudales máximos instantáneos

Cód.	Estación	Función
1	TAM	$Q_p = 1.047Q_m$
2	KM	$Q_p = 1.170Q_m$
3	CHL	$Q_p = 1.455Q_m$
4	LLC	$Q_p = 1.389Q_m$
5	BE	$Q_p = 1.084Q_m$
6	PU	$Q_p = 1.063Q_m$
7	TAB	$Q_p = 1.044Q_m$
8	NAZ	$Q_p = 1.044Q_m$
9	BO	$Q_p = 1.081Q_m$
10	SR	$Q_p = 1.057Q_m$

4.1.3 Filtrado de datos

La serie temporal de Q_p (Figura 7) obtenidos a partir de las funciones (Tabla 6), las muestras de las estaciones TAM y KM (Figura 7a y 7b) presentan fluctuaciones estables en el tiempo, mientras que, valores atípicos altos pueden diferenciarse en los datos de la estación KM (Figura 7b), CHL (Figura 7c) y BE (Figura 7e). Hosking y Wallis (1997) indican que estas fluctuaciones atípicas podrían haber afectado a un sitio, pero no a sus vecinos. También es posible apreciar valores faltantes en el registro de las estaciones KM (Figura 7b), LLC (Figura 7d), BO (Figura 7i) y supuestas tendencias en la serie de datos de las estaciones PU (Figura 7f) y SR (Figura 7j). Aparentemente cuando los registros son cortos, la variabilidad climática puede dar lugar fácilmente a una tendencia y puede desaparecer cuando se haya recopilado tanta información como sea posible (Kundzewicz y Robson, 2004). Para que los resultados del análisis de frecuencia en hidrología sean teóricamente válidos, cada muestra de datos debe satisfacer ciertos supuestos básicos, como el de aleatoriedad, independencia, homogeneidad y estacionalidad (OMM, 2011). Según Hosking y Wallis (1997) los valores de datos incorrectos, las anomalías, las tendencias y los cambios en la media de una muestra se reflejan en los momentos-L de la muestra, entonces una medida de la discordancia entre las relaciones de momento L de un sitio y las relaciones promedio de momento L de un grupo de sitios similares se aplica a los datos para identificar sitios discordantes, es decir, aquellos sitios que son extremadamente incompatibles con el grupo.

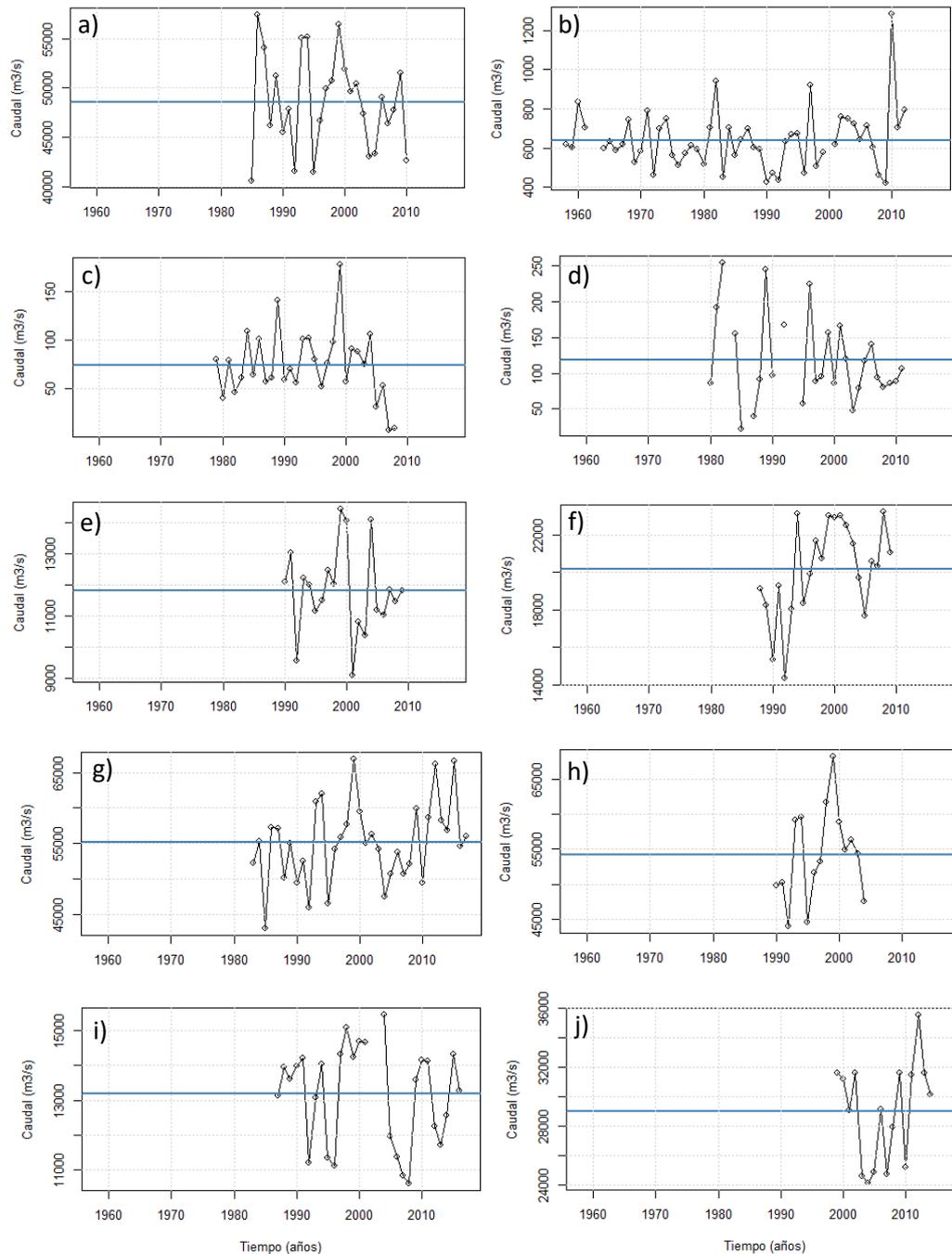


Figura 7. Hidrograma de caudal máximo instantáneo para las estaciones hidrométricas

a) Medida de discordancia regional

Una evaluación para todo el grupo de estaciones (TAM, KM, CHL, LLC, BE, PU, TAB, NAZ, BO y SR), se ha realizado mediante el diagrama de dispersión entre asimetría-L versus L-CV (Figura 8a) y asimetría-L versus curstosis-L (Figura 8b), aunque los sitios KM, CHL y LLC parecen estar alejados del grupo por tener valores altos de L-CV, las estadísticas de la medida de discordancia para cada sitio

(Di), indica que los valores son inferiores a la medida de discordancia crítica Dc. El valor de discordancia crítica (Dc) para un grupo de 10 sitios es 2.49, por lo tanto, deducimos que el grupo de estaciones en términos de los momentos-L, no son extremadamente discordantes, los valores de Di de cada sitio son inferiores a 2.49 (Tabla 7). Hosking y Wallis (1997) indican que una estación es discordante si la Di excede la Dc del grupo de estaciones, además en el contexto del ARF utilizando momentos-L, encontraron que al comparar las relaciones de los momentos-L de las muestras de diferentes sitios, los valores de datos incorrectos, valores atípicos, tendencias y cambios en la media, pueden reflejarse en los momentos-L de los datos de los sitios.

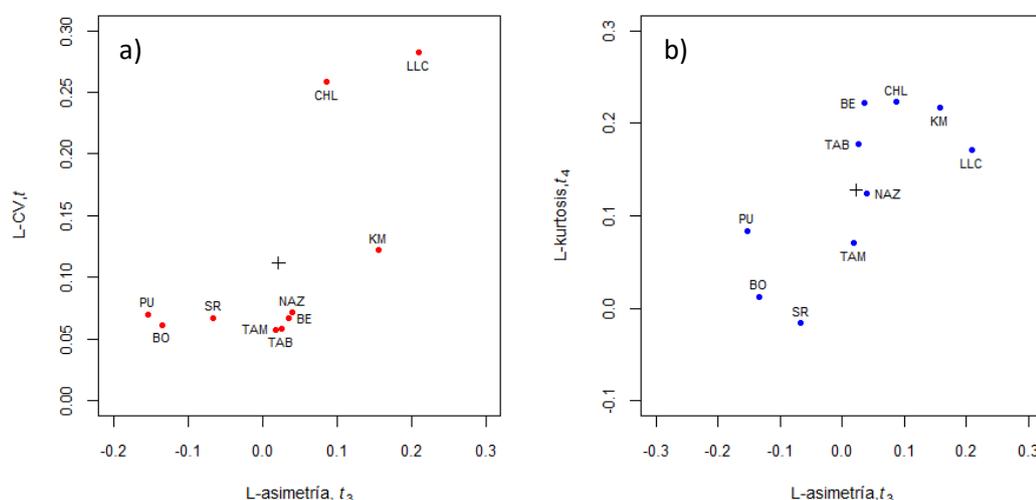


Figura 8. Dispersión de las relaciones de momentos-L de las muestras

Tabla 7

Medida de discordancia regional de las estaciones hidrométricas

Cód.	Estación	Di	Dc
1	Tamishiyacu	0.58	
2	Km 105	0.82	
3	Chotano	1.70	
4	Llaucano	1.86	
5	Bellavista	0.98	2.76
6	Pucallpa	1.32	
7	Tabatinga	0.50	
8	Nazareth	0.25	
9	Borja	0.84	
10	San Regis	1.15	

4.1.4 Identificación de regiones homogéneas

a) Elección de la prueba de homogeneidad regional

La dispersión de puntos entre L-CA y L-CV muestra las regiones separadas entre Hosking y Wallis (HW) y Anderson Darling (AD). La dispersión de puntos de las estaciones TAM, KM, CHL, LLC, BE, PU, TAB, NAZ, BO y SR, se encuentra dentro del espacio de HW, por lo tanto, utilizamos la prueba de heterogeneidad de HW para definir regiones hidrológicas homogéneas (Figura 9). Si los valores del L-CA son inferiores a 0.23 los puntos de dispersión se ubican en el espacio de HW y es la parte donde la prueba de heterogeneidad de HW es más eficaz, en el caso de que la dispersión de puntos L-CA sean mayores a 0.23, es preferible utilizar la prueba de bootstrap de AD (Viglione, 2007).

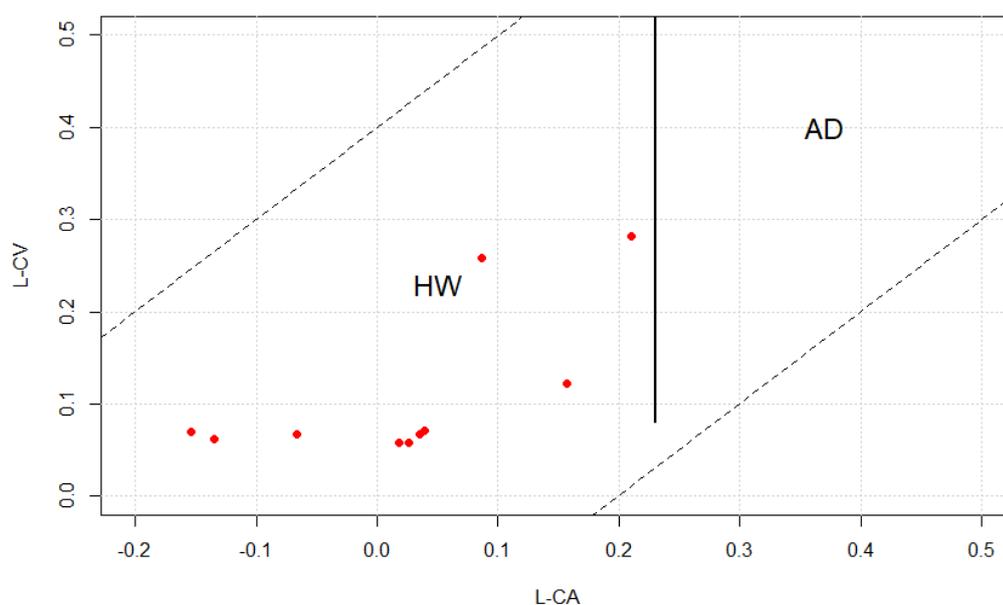


Figura 9. Estaciones separadas en el espacio L-CA – LCV

b) Medida de heterogeneidad regional

Verificando que no existen estaciones discordantes en el grupo de estaciones, evaluamos la homogeneidad, mediante la medida de heterogeneidad de HW, obteniendo un valor de $H1 = 15.62$ para el grupo de 10 estaciones hidrométricas (Tabla 8). Según Hosking y Wallis (1997) se declara una región “aceptablemente homogénea” si $H < 1$, “posiblemente heterogénea” si $1 \leq H < 2$ y “definitivamente heterogénea” si $H \geq 2$, mientras que Wallis *et al.* (2007) establecen que una región

es “aceptablemente homogénea” si $HI < 2$, “posiblemente heterogénea” si $2 \leq HI < 3$, y “definitivamente heterogénea” si $HI \geq 3$. Bajo estas premisas el conjunto de datos de 10 estaciones hidrométricas no pertenece a una supuesta región homogénea, por lo que se procedió a utilizar el algoritmo de Ward para separar estaciones con un comportamiento similar. El algoritmo de conglomerado de Ward ha sido estudiado ampliamente para la clasificación de diferentes datos climáticos e hidrológicos (Domroes *et al.*, 1998; Jackson y Weinand, 1995; Nathan y McMahon, 1990; Ramachandra Rao y Srinivas, 2006). Viglione (2007) indica que se debe de seleccionar variables de clasificación para poder introducirlos en el algoritmo de Ward.

Tabla 8

Prueba de heterogeneidad regional del grupo de estaciones hidrométricas

Cód.	Estación	H1
1	Tamishiyacu	
3	Km 105	
4	Chotano	
5	Llaucano	
6	Bellavista	15.62
7	Pucallpa	
8	Tabatinga	
9	Nazareth	
10	Borja	
12	San Regis	

c) Selección de variables de clasificación

Se encontró que las mejores regresiones múltiples entre las matrices de distancia de las características morfoclimáticas de las cuencas y las curvas de crecimiento regional fueron la elevación media del río principal ($Emrp$) y el radio de circularidad (Rc) con un coeficiente de determinación $R_{aj}^2 = 71.4\%$. Los tres primeros modelos (Tabla 9) aunque presentan mayor R_{aj}^2 tienen problemas de multicolinealidad al ser evaluado con el Factor de inflación de la varianza ($VIF > 5$). Entonces el modelo cuatro sería el más representativo, pues el $VIF < 5$ ($Emrp=1.198$ y $Rc=1.198$). Valores del VIF mayores a 5 o 10, indican que los coeficientes de regresión asociados están mal estimados e implican problemas serios de multicolinealidad (Montgomery *et al.*, 2012; Viglione *et al.*, 2007).

Tabla 9

Mejores regresiones con parámetros morfoclimáticos

Modelo	R_{aj}^2 [%]
Pma + Emrp + Rc	73.8
Emc + Emrp + Rc	73.0
A + Lrp + Rc	72.7
Emrp + Rc	71.4
Emrp + K	70.2
Pmrp + Rc	69.9
Rc	68.5
K	65.6
Emrp	24.3

R_{aj}^2 : Coeficiente de determinación ajustado.

d) Algoritmo de Ward

La división de las cuencas en 03 regiones homogéneas, fueron formados mediante el algoritmo de Ward (Figura 10). El algoritmo de Ward es considerado un método adecuado para una determinación preliminar de regiones homogéneas (Domroes *et al.*, 1998; Hosking y Wallis, 1997; Jackson y Weinand, 1995; Gottschalk, 1985). Las variables utilizadas fueron los parámetros morfométricos (Emrp y Rc). Hosking y Wallis (1997) indican que los datos a usar para la formación de regiones homogéneas, son las cantidades calculadas a partir de las mediciones de caudales denominados estadísticas en el sitio y otros descriptores del sitio como las características del sitio tales como la ubicación geográfica del sitio, la altitud, la precipitación media anual y la estación del año en la que se producen los eventos máximos, pero también pueden usarse características físicas, pedológicas, morfométricas y climática (Viglione, 2007).

e) Medida de heterogeneidad

Se conformaron 03 regiones homogéneas con las siguientes estadísticas: Región 1 ($H1 = -1.27$), Región 2 ($H1 = -0.09$) y R3 ($H1 = -0.64$). Si $H1 > 1$ Hosking y Wallis (1997) sugiere que se debe considerar una mayor subdivisión de la región, ya que podría mejorar la precisión de las estimaciones de los cuantiles. Por ser un criterio netamente estadístico Wallis *et al.* (2007) y Schaefer *et al.* (2006) consideran que una región sea considerada homogénea cuando $H1 < 2$. Por su parte Viglione (2007) también sugiere aceptar valores de $H1$ inferiores a 2 como homogénea para evitar formar grupos demasiado pequeños.

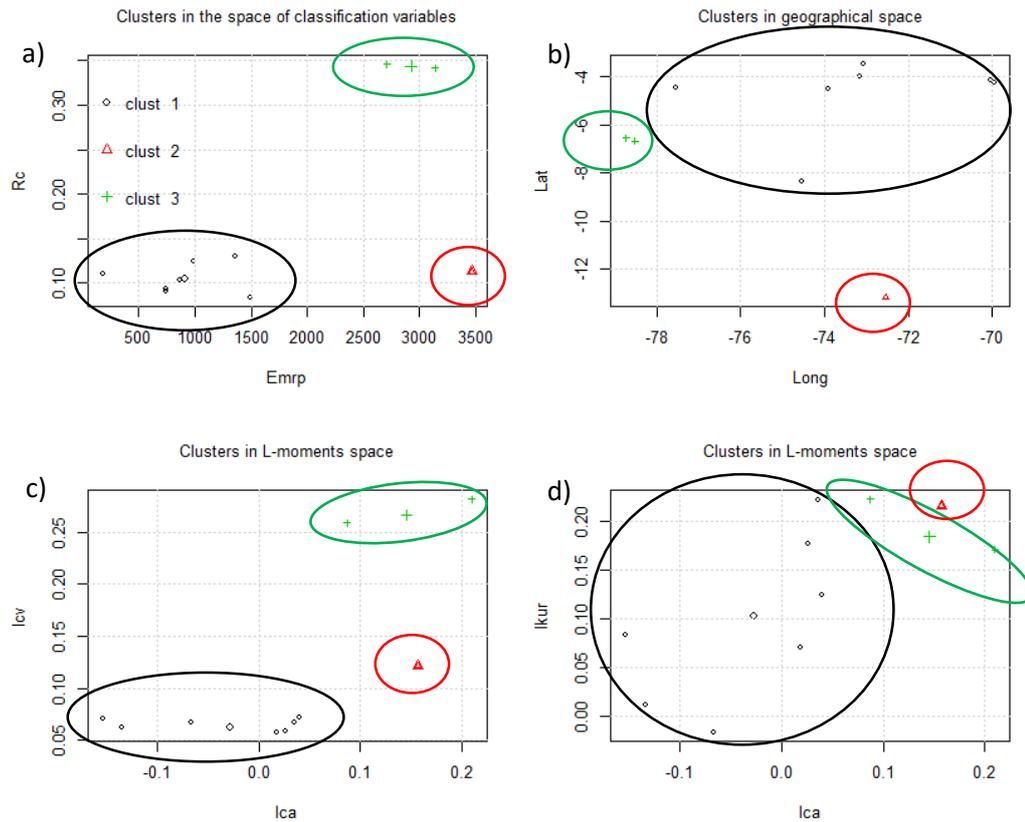


Figura 10. Regiones homogéneas identificadas

Las regiones homogéneas identificadas se encuentran marcadamente separadas entre sí (Figura 10), así tenemos presentado en el espacio de clasificación de parámetros morfométricos de las cuencas Emrp y Rc (Figura 10a), en la representación del espacio geográfico longitud y latitud (Figura 10b), en el espacio de las relaciones de momentos L-CA y L-CV (Figura 10c) y L-CA y L-CK (Figura 10d). La región 1 está conformada por las estaciones TAM, BE, PU, TAB, NAZ, BO y SR, mientras que la región 2 por la estación KM y la región 3 por las estaciones CHL y LLC. No hay un número correcto de grupos, pero se debe buscar un equilibrio entre el uso de regiones que son demasiado pequeñas o demasiado grandes. Las regiones que contienen pocos sitios lograrán poca mejora en la precisión de las estimaciones de cuantiles sobre el análisis in situ, mientras que, a medida que aumenta el tamaño de la región, la precisión es mayor, pero se obtiene poca ganancia en la precisión de la estimación de cuantiles al usar más de aproximadamente 20 sitios en una región (Hosking y Wallis, 1997; Viglione, 2007).

4.1.5 Selección de la distribución regional de frecuencia

Las distribuciones trazadas en la Figura 11, logística generalizada (GLO), valor extremo generalizado (GEV), Pareto generalizado (GPA), lognormal (LN3), Pearson tipo III (PE3) son distribuciones de tres parámetros más utilizadas en el ARF, asimismo se aprecian las distribuciones de dos parámetros exponencial (E), Gumbel (G), logística (L), normal (N) y uniforme (U). Según Hosking y Wallis (1997) las distribuciones de dos parámetros pueden causar sesgo en la cola de los cuantiles estimados si la forma de la distribución de frecuencia verdadera no está bien aproximada por la distribución ajustada. La mejor distribución de ajuste es uno que da estimaciones robustas para la curva de crecimiento regional, así como para los cuantiles en cada sitio.

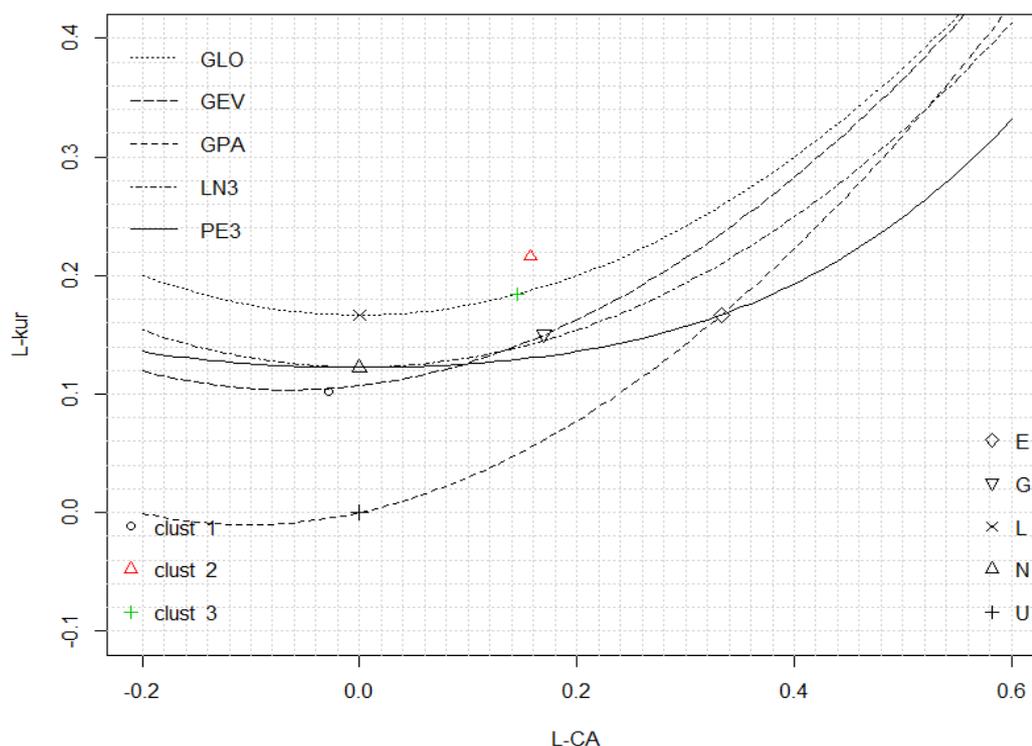


Figura 11. Trazado de funciones de distribución en el plano 'L-CA – L-CK'

Cualitativamente las distribuciones de frecuencia que mejor se ajustan al promedio regional de la relación de momentos-L (clus1, clus2 y clus3), deducimos que para la región 1 (clus 1) el promedio (punto) se encuentra próximo a la distribución GEV y poco distante a PE3 y LN3, mientras que la región 2 (clus 2) y región 3 (clus 3) la distribución GLO es la más representativa (Figura 11). Peel *et al.* (2001) indica que la selección de la distribución de frecuencia de grupos homogéneos, es mejor

basada en el promedio de la muestra y no en una línea de mejor ajuste a través de los puntos de datos.

4.1.5.1 Medidas de bondad de ajuste

Siguiendo las recomendaciones de Hosking y Wallis (1997) y Viglione (2007) en el que los grupos que contienen pocos sitios, lograrán poca mejora en la precisión de las estimaciones, por lo tanto, en los siguientes análisis solo será considerado la región 1, conformado por las estaciones TAM, BE, PU, TAB, NAZ, BO y SR

Para la región 1 la función de distribución que mejor se ajusta a los datos, es la distribución GEV por tener los menores valores de AIC, AICc, BIC, ADC y Z^{DIST} , seguido de la función de distribución Pearson tipo 3 (P3) (Tabla 10 y 11). Hosking y Wallis (1997) seleccionaron apropiadas distribuciones regionales de frecuencias utilizando la prueba de bondad de ajuste Z estadístico ($|Z^{DIST}| \leq 1.64$), lo que significa, que la verdadera distribución de la región debe aceptarse aproximadamente al 90% del tiempo, mientras que Viglione (2007) ha basado la selección de distribuciones con los valores más bajos de los criterios de AIC, AICc, BIC y ADC.

Al aplicar la prueba de bondad de idoneidad de AD, obtuvimos un valor de $A^2=0.358$ por lo tanto este valor es casi cercano a cero y concluimos que la distribución GEV de 3 parámetros ofrece un ajuste adecuado a los datos, asimismo el p-valor ($p(A^2) = 0.619$) es mayor que el nivel de significancia de 5%, por lo tanto, los datos siguen una distribución GEV (Tabla 12). La distribución GEV es usada ampliamente para el ARF (Cunnane, 1989; Parida y Moalafhi, 2008).

Tabla 10

Estadísticos Z^{DIST} para las distribuciones de frecuencia

Región	Z^{GLO}	Z^{GEV}	Z^{GNO}	Z^{PE3}	Z^{GPA}
1	2.65	0.25	0.98	0.96	-3.92

Tabla 11

Criterios de selección de modelos de distribución de frecuencia

Región	Modelo	AIC	AICc	BIC	ADC
1	GEV	-248.8	-284.6	-275.2	0.153
	P3	-283.3	-283.2	-273.8	0.176
	LN	-280.2	-280.1	-273.8	0.629

Los parámetros de la distribución GEV (Tabla 12), se estimaron mediante el uso del método de momentos-L (Hosking y Wallis, 1997). En la Figura 12 se aprecia la función de distribución GEV y las muestras individuales de las estaciones TAM (1), BE (5), PU (6), TAB (7), NAZ (8), BO (9), SR (10) y la línea continua la distribución GEV.

Tabla 12

Prueba de bondad de ajuste y parámetros de la función de distribución GEV

Región	Distribución	AD A ²	Valor p p(A ²)	Parámetros		
				ϵ	α	k
1	GEV	0.358	0.619	0.965	0.111	0.343

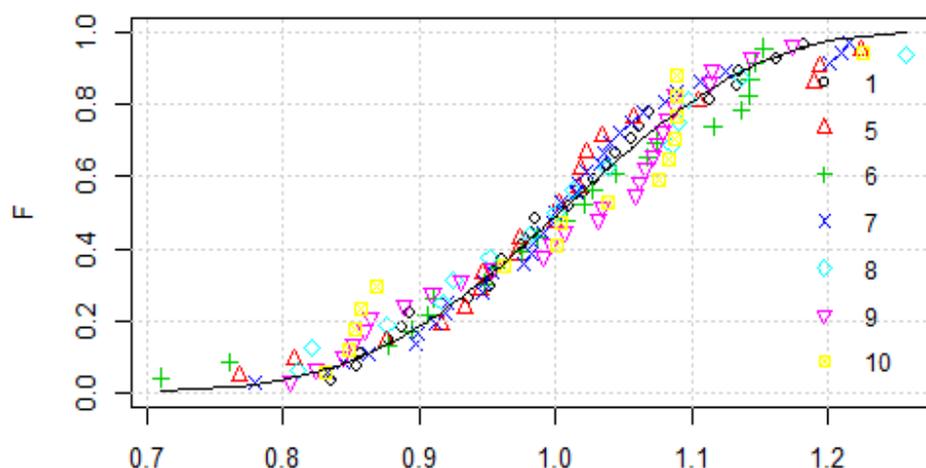


Figura 12. Función de distribución GEV y estaciones región 1

4.1.5.2 Estimación de cuantiles

La curva de crecimiento regional y límites de error para diferentes periodos de retorno, se ha elaborado mediante la función de distribución GEV (Figura 13a). Asimismo, las ecuaciones regionales pueden ser utilizado en cuencas sin información hidrológica dentro de la región 1 (Figura 13b) para estimar

el caudal máximo instantáneo para periodos de retorno 50 y 100 años, con RMSE 0.019 y 0.024 respectivamente. Entonces los parámetros GEV regionales pueden ser transferidos a los sitios específicos con confianza (Parida y Mohalafhi, 2008).

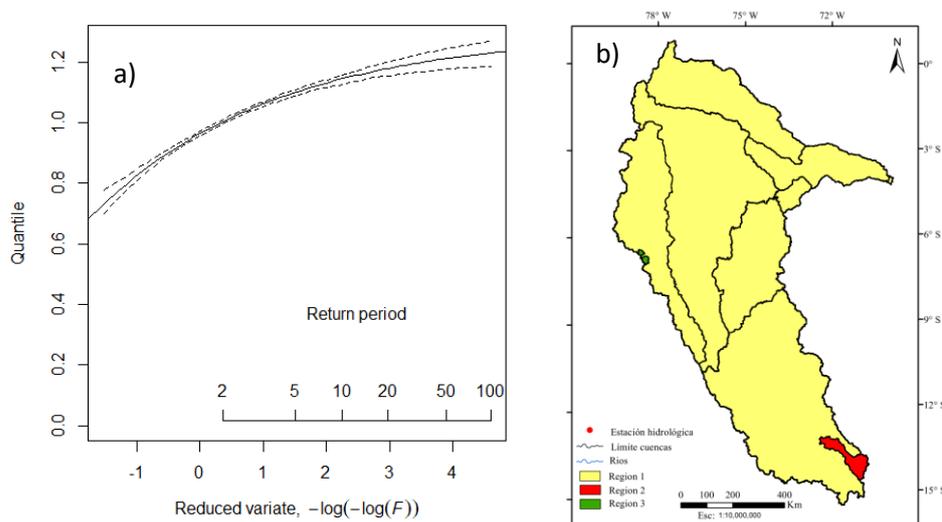


Figura 13. a) Curva de crecimiento regional y b) regiones homogéneas

Las ecuaciones regionales para estimar los cuantiles para periodos de retorno de 50 y 100 años son

$$Q(T_{50años}) = 1.21\bar{Q}$$

$$Q(T_{100años}) = 1.23\bar{Q}$$

y para otros periodos de retorno reemplazando los parámetros de localización, escala y forma ($\varepsilon = 0.965, \alpha = 0.111$ y $k = 0.343$) es

$$Q(T) = \bar{Q} \left[0.965 + \frac{0.111}{0.343} \left\{ 1 - \left(-\ln \left(1 - \frac{1}{T} \right) \right)^{0.343} \right\} \right]$$

Para los cuantiles adimensionales para diferentes periodos de retorno se tiene

$$\frac{Q(T)}{\bar{Q}} = q(T) = 0.965 + \frac{0.111}{0.343} \left\{ 1 - \left(-\ln \left(1 - \frac{1}{T} \right) \right)^{0.343} \right\}$$

4.2 Regionalización del índice de avenida a partir de parámetros morfoclimáticos en la cuenca amazónica del Perú

4.2.1 Elección de mejores parámetros

Una visualización gráfica de la matriz de correlación, con un nivel de confianza del 95% y nivel de significancia del 5%, nos indica que las variables Emc, Pmc, Pmr, Emrp, Rc y Cc tienen una correlación insignificante con el \bar{Q} . Las mejores correlaciones entre \bar{Q} son el A, Lrp. La correlación es una forma de verificar la existencia del grado de relación entre las variables dependientes e independientes (Figura 14).

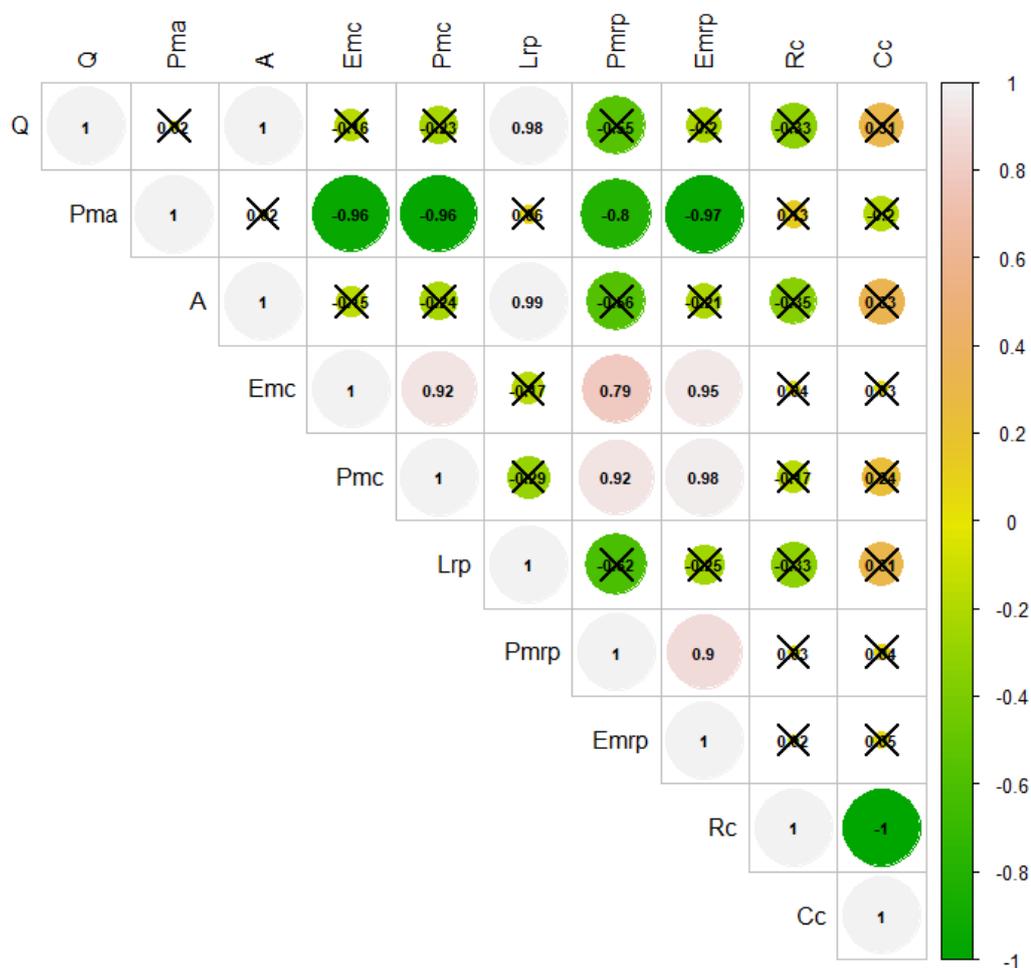


Figura 14. Coeficiente de correlación de Pearson

4.2.2 Elección de mejores regresiones

La variable dependiente \bar{Q} se ha analizado entre cuatro posibilidades, valores de caudal sin transformar \bar{Q} , valores de caudal transformado a logaritmo ($\log \bar{Q}$),

valores de caudal transformado a raíz cuadrada($\sqrt{\bar{Q}}$) y caudal transformado a raíz cubica ($\sqrt[3]{\bar{Q}}$). Los mejores resultados evaluados con R_{aj}^2 , RMSE y RMSEjk mostraron los caudales sin transformación (Tabla 13). Los modelos de regresion multiple son los más utilizados para estimar el índice de avenida en sitios sin datos medidos (Viglione, 2007), este enfoque vincula el índice de avenida con las características de la cuenca.

Tabla 13

Mejores regresiones entre \bar{Q} y sus transformadas con parámetros morfométricos

Dependiente	Independiente	R_{aj}^2 [%]	RMSE [m^3/s]	RMSEcv [m^3/s]
\bar{Q}	A + Lrp	99.7	788.2	1335.2
	A	99.3	1323.5	1737.9
	Lrp	95.6	3418.1	4409.3
$\log(\bar{Q})$	A + Lrp	95.1	2838.4	4991.6
	A	94.2	4059.9	5735.9
	Lrp	88.6	5161.2	7043.6
$\sqrt{\bar{Q}}$	A + Lrp	98.4	1681.9	2997.1
	A	97.7	2472.0	3340.1
	Lrp	92.9	3983.1	5213.1
$\sqrt[3]{\bar{Q}}$	A + Lrp	97.6	2035.9	3627.2
	A	96.7	2942.6	4020.6
	Lrp	91.7	4300.4	5683.6

R_{aj}^2 : Coeficiente de determinación ajustado, RMSE: Error medio cuadrático, RMSEcv: Error medio cuadrático de la validación cruzada

Elegimos al modelo ($\bar{Q}: f(A + Lrp)$) como el adecuado pues presenta un RMSE 788.2 m^3/s y RMSEcv = 1335.2 m^3/s , pero si evaluamos los coeficientes de regresión sobre la base de t Student podemos excluir aquellos modelos en que la variable explicativa no es significativa al 5%, así concluimos que la variable predictor Lrp (del modelo $\bar{Q}: f(A + Lrp)$), no es significativa pues el p-valor > 0.05, probablemente su adición en los modelos para estimar el índice de avenida no sea significativa. No en tanto para el modelo $\bar{Q}: f(A)$ la variable predictor A es significativa, pues el p-valor < 0.05, entonces los cambios en el valor de A están relacionados con cambios en el \bar{Q} (Tabla 14). Viglione (2007) obtiene mejores relaciones entre la altitud media de la cuenca, el centro de gravedad de la cuenca y el índice de Budiko con los caudales transformado en logaritmo. No en tanto Tucci (2002) encuentra una mejor relación entre el área de la cuenca y los caudales sin transformar.

Tabla 14

Coefficientes de regresión del índice de avenida

Modelo	Coefficientes	Std. Error	Valor t	Pr (> t)	
$\bar{Q} : f(A, Lrp)$	Intercepto	14620.00	2916.0	5.015	0.007 **
	A	0.081	0.01	8.440	0.001 **
	Lrp	-9.59	3.56	-2.698	0.054.
$\bar{Q} : f(A)$	Intercepto	6988.00	1057.0	6.613	0.001**
	A	0.055	0.002	29.908	7.84e-07***
$\bar{Q} : f(Lrp)$	Intercepto	-8429.82	3957.55	-2.13	0.0864.
	Lrp	20.18	1.77	11.39	9.11e-05***

Nivel de significancia: 0 '***', 0.001 '**', 0.01 '*', 0.05 '.', 0.1 ' ', 1

4.2.2.1 Multicolinealidad

En modelos de regresión con más de una variable predictora fue necesario realizar un estudio de residuos y también descartar multicolinealidad entre variables. Como se puede apreciar en la matriz de correlación grafica (Figura 14), existe correlación muy alta entre las variables independientes (A y Lrp). Según Viglione (2007) la presencia de alta correlación entre variable independientes (valores que están cerca de ± 1) deben considerarse problemáticos porque es un índice de multicolinealidad. En correlaciones altas las variables predictoras aportan información redundante, por tanto, realizamos un estudio de multicolinealidad y para ello utilizamos el Factor de inflación de la varianza (VIF). Los parámetros morfométricos A y Lrp tienen valores VIF > 5 por tanto, hay problema de multicolinealidad (Tabla 15). Valores del VIF mayores a 5 o 10, indican que los coeficientes de regresión asociados están mal estimados e implican problemas serios de multicolinealidad (Montgomery *et al.*, 2012; Viglione *et al.*, 2007).

Tabla 15

Factor de inflación de la varianza (VIF) de las variables independientes

Modelo	Variable independiente	VIF
$\bar{Q} : f(A, Lrp)$	A	60.7
	Lrp	60.7

4.2.2.2 Homocedasticidad y normalidad de residuos

Mediante la prueba de Harrison-McCabe se ha obtenido un p-value = 0.494 lo que indica que la varianza de los residuos es constante (Figura 15a).

Viglione (2007) indica que la presencia de patrones particulares en la disposición de los puntos puede ser un índice de heterocedasticidad (diversidad en la varianza). También verificamos el supuesto de normalidad de los residuos graficando en un papel de probabilidad normal (Figura 15b) y para hacer inferencias, utilizamos la prueba de normalidad de AD, el cual da un p-valor = 0.881 pues indica que los residuos se distribuyen normalmente. El resultado de los valores ajustados y observados y el resultado de los caudales predichos con los observados que incluye el intervalo de pronóstico del 95% son presentados en la Figura 15c y 15d respectivamente.

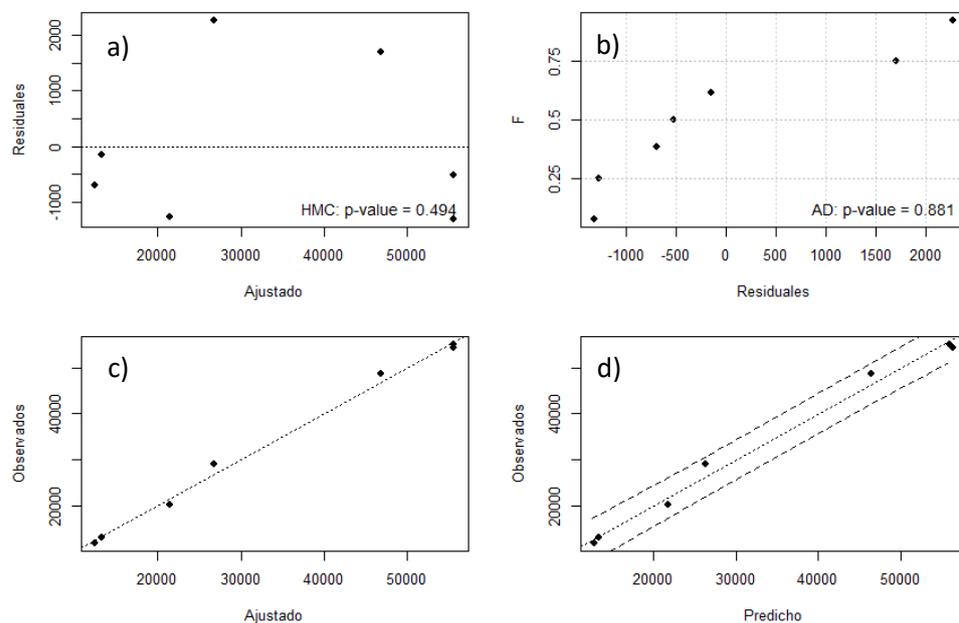


Figura 15. Trazo del diagnóstico del modelo índice de avenida

Entonces, de los modelos analizados consideramos que el modelo en función del área de la cuenca, el parámetro A (área) en todos los modelos muestra el p-valor más bajo (Tabla 14) y sería la mejor variable explicativa para estimar el \bar{Q} en cuencas sin medición. Además, según el R_{aj}^2 puede explicar el 99.3 % de la variabilidad observada en el \bar{Q} , con un $RMSE = 1323.5 \text{ m}^3/\text{s}$ y $RMSE_{jk} = 1737.9 \text{ m}^3/\text{s}$. Naghettini y Pinto (2007) indica que el \bar{Q} es un método para adimensionalizar cualquier dato obtenido en puntos distintos de una región homogénea, su valor considerado es el \bar{Q} (OMM, 2011; Viglione, 2007; Hosking y Wallis, 1997). El modelo índice de avenida, queda definido por

$$\bar{Q} = 6988 + 0.055(A)$$

La ecuación es válida para cuencas con áreas comprendidas entre 99,779.4
 $\text{Km}^2 \leq A (\text{Km}^2) \leq 877,478.6 \text{ Km}^2$.

CONCLUSIONES

- Tres regiones homogéneas fueron definidas mediante un proceso mixto entre el algoritmo de Ward, una técnica de reasignación de los elementos y la medida de heterogeneidad, de las cuales dos de ellas (región 2 y 3) fueron excluidos por estar conformado con una reducida cantidad de estaciones hidrométricas. La función GEV demostró adecuarse mejor a la muestra de datos de la región 1, pues presentó valores más bajos al ser evaluados con la prueba bondad de ajuste Z^{DIST} y los criterios de información de Akaike (AIC), criterio de información de Akaike corregido (AICc), criterio de información Bayesiano (BIC) y el criterio de Anderson-Darling (ADC). En tal razón, la curva de crecimiento regional determinada mediante la función GEV permite de manera confiable estimar los caudales máximos instantáneos para diferentes periodos de retorno en la región 1, junto con el índice de avenida.
- El caudal máximo promedio anual \bar{Q} , fue considerado como el valor índice de avenida para cada estación hidrométrica. El área de la cuenca demostró ser la variable independiente más significativa para la transferencia de información hidrológica a cuencas sin información dentro de la cuenca amazónica del Perú – región 1, con un coeficiente de determinación ajustado ($R_{aj}^2=99.3\%$), raíz del error medio cuadrático (RMSE = 1323.169 m³/s) y raíz del error medio cuadrático de la validación cruzada (RMSEcv=1737.864 m³/s).

RECOMENDACIONES

- Se recomienda que la ecuación obtenida para estimar el índice de avenida es válida para cuencas con áreas comprendidas entre $99,779.4 \text{ Km}^2 \leq A (\text{Km}^2) \leq 877,478.6 \text{ Km}^2$, dentro de la cuenca amazónica del Perú.
- Realizar una evaluación de las técnicas de análisis de frecuencia regional basado en el método del índice de avenida y métodos de regresión y verificar la incertidumbre de modelos.
- Evaluar el comportamiento de caudales de avenidas en cuencas húmedas y secas, para relacionar el caudal máximo y los parámetros morfoclimáticos de las cuencas en Perú.

BIBLIOGRAFÍA

- Acuña, J., Felipe, O. G., y Fernández, C. (2015). Análisis regional de frecuencia de precipitación anual para la determinación de mapas de sequías en las cuencas Chillón, Rímac, Lurín y Alto Mantaro. *Revista Peruana Geo-Atmosférica*, 4, 104–115.
- Acuña, J. (2010). *Análisis Regional de Caudales Medios Mensuales para Diversos Niveles de Persistencia de los Ríos Peruanos Pertenecientes en la Vertiente del Pacífico* (Tesis de maestría). Universidad Nacional Agraria La Molina, Lima, Perú.
- Aybar, C., Lavado-Casimiro, W., Huerta, A., Fernández, C., Vega, F., Sabino, E., y Felipe-Obando, O. (2017). Uso del Producto Grillado “PISCO” de precipitación en Estudios, Investigaciones y Sistemas Operacionales de Monitoreo y Pronóstico Hidrometeorológico. Nota Técnica 001 SENAMHI-DHI-2017, Lima, Perú.
- Burn D. (1990). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26, 2257–2265.
- Campos, D. F. (2006). *Análisis probabilístico univariado de datos hidrológicos*. México: Asociación Mexicana de Hidráulica.
- Campos, D. F. (2016). Ajuste de las distribuciones GVE, LOG y PAG con momentos L de orden mayor. *Ingeniería, Investigación y Tecnología*, 17(1), 131–142. <https://doi.org/10.1016/j.riit.2016.01.012>
- Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, 100(1–3), 269–290. [https://doi.org/10.1016/0022-1694\(88\)90188-6](https://doi.org/10.1016/0022-1694(88)90188-6)
- Dalrymple, T. (1960). Flood-frequency analyses, Manual of Hydrology: Part 3. *Water*

Supply Paper.

- Domroes, M., Kaviani, M., y Schaefer, D. (1998). An Analysis of Regional and Intra-annual Precipitation Variability over Iran using Multivariate Statistical Methods. *Theoretical and Applied Climatology*, 61(3), 151–159. <https://doi.org/10.1007/s007040050060>
- Escalante, C. A., y Reyes, L. (2002). *Técnicas Estadísticas en Hidrología*. México, UNAM, Facultad de ingeniería.
- Espinoza, J.C., Ronchail, J. Guyot, J. L., Junquas, C., Drapeau, G., Martinez, J.M., Santini, W., Vauchel, P., Lavado, W., Ordoñez, J., y Espinoza, R. (2012). From drought to flooding: understanding the abrupt 2010–11 hydrological annual cycle in the Amazonas River and tributaries. *Environmental Research Letters*, 7.
- Fernández, C. A., y Lavado, W. S. (2016). Regional maximum rainfall analysis using L-moments at the Titicaca Lake drainage, Peru. *Theoretical and Applied Climatology*, 1–13. <https://doi.org/10.1007/s00704-016-1845-3>
- Fowler, H.J., y Kilsby, C.G. (2003). A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *International Journal of Climatology*, 23, 1313–1334. doi:<https://doi.org/10.1002/joc.943>
- Fuller, W. E. (1914). Flood Flows. *Transactions of the American Society of Civil Engineers*, 77, 564– 617.
- Gottschalk, L. (1985). Hydrological regionalization of Sweden. *Hydrological Sciences Journal*, 30(1), 65–83. <https://doi.org/10.1080/02626668509490972>
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., y Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5), 1049–1054.
- GREHYS (1996). Inter-comparison of regional flood frequency procedures for Canadian rivers. *Journal of Hydrology*, 186(1–4), 85–103. [https://doi.org/10.1016/S0022-1694\(96\)03043-0](https://doi.org/10.1016/S0022-1694(96)03043-0)
- Hassan, B. G. H., y Ping, F. (2012). Regional Rainfall Frequency Analysis for the Luanhe Basin – by Using L-moments and Cluster Techniques. *APCBEE Procedia*, 1, 126–

135. <https://doi.org/10.1016/j.apcbee.2012.03.021>

- Hosking, J. (1990) L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society*, 52, 105-124.
- Hosking, J. R. M., y Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29, 271–281.
- Hosking, J. R. M., y Wallis, J. R. (1997). *Regional frequency analysis: an approach based on L-moments*. Cambridge University Press.
- Hosking, J. R. M. (2019). L-Moments. R package, version 2.8. URL: <https://CRAN.R-project.org/package=lmom>.
- Hosking, J. R. M. (2019). Regional Frequency Analysis using L-Moments. R package, version 3.2. URL: <https://CRAN.R-project.org/package=lmomRFA>.
- IPCC (2014). Cambio climático 2014: *Informe de síntesis. Contribución de los Grupos de trabajo I, II y III al Quinto Informe de Evaluación del Grupo Intergubernamental de Expertos sobre el Cambio Climático* [Equipo principal de redacción, R.K. Pachauri y L.A. Meyer (eds.)]. IPCC, Ginebra, Suiza.
- Jackson, I. J., y Weinand, H. (1995). Classification of tropical rainfall stations: A comparison of clustering techniques. *International Journal of Climatology*, 15(9), 985–994. <https://doi.org/10.1002/joc.3370150905>
- Jarvis, A., Reuter, H.I., Nelson, A., y Guevara, E. (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m
- Kachroo, R. K., Mkhandi, S. H., y Parida, B. P. (2000). Flood frequency analysis of southern Africa: I. Delineation of homogeneous regions. *Hydrological Sciences Journal*, 45(3), 437–447. <https://doi.org/10.1080/02626660009492340>
- Khan, S. A., Hussain, I., Hussain, T., Faisal, M., Muhammad, Y. S., y Mohamd Shoukry, A. (2017). Regional Frequency Analysis of Extremes Precipitation Using L-Moments and Partial L-Moments. *Advances in Meteorology*, 2017, 1–20. <https://doi.org/10.1155/2017/6954902>
- Laio, F. (2004). Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research*,

- 40(9), 1–10. <https://doi.org/10.1029/2004WR003204>
- Lujano, E., y Obando, O. G. (2015). Análisis de frecuencia regional de las precipitaciones máximas diarias en la región hidrográfica del Titicaca. *Revista de Investigaciones Altoandinas*, 17(1), 53–64.
- Madsen, H., Rasmussen, P. F., y Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. *Water Resources Research*, 33(4), 747–757. <https://doi.org/10.1029/96WR03848>
- Marengo, J.A. y Espinoza, J.C. (2015). Review Article. Extreme Seasonal Droughts and Floods in Amazonia: Causes, Trends and Impacts. *International Journal of Climatology*, 36, 1033–1050.
- Montgomery, D. C., Peck, E. A., y Vining, G. G. . (2012). *Introduction to linear regression analysis*.
- Mujumdar, P., y Nagesh Kumar, D. (2012). References. In *Floods in a Changing Climate: Hydrologic Modeling* (International Hydrology Series, pp. 167-175). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139088428.009
- Naghattini, M., y Pinto, E. J. (2007). *Hidrologia Estatística*. Serviço Geológico do Brasil.
- Nathan, R. J., y McMahon, T. A. (1990). Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology*, 121(1–4), 217–238. [https://doi.org/10.1016/0022-1694\(90\)90233-N](https://doi.org/10.1016/0022-1694(90)90233-N)
- OMM (2011). *Guía de prácticas hidrológicas. Gestión de recursos hídricos y aplicación de prácticas hidrológicas*. Ginebra, Suiza.
- OMM (1994). *Guía de Prácticas Hidrológicas*. OMM N° 168, Ginebra, Suiza.
- OMM (1989) *Statistical Distributions for Flood Frequency Analysis* (Cunnane, C.), Operational Hydrology. Report N° 33, WMO-N° 718, Ginebra, Suiza.
- Parida, B. P., y Moalafhi, D. B. (2008). Regional rainfall frequency analysis for Botswana using L-Moments and radial basis function network. *Physics and Chemistry of the Earth, Parts A/B/C*, 33(8), 614–620. <https://doi.org/10.1016/j.pce.2008.06.011>

- Peel, M. C., Wang, Q. J., Vogel, R. M., y McMahon, T. A. (2001). The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences Journal*, 46(1), 147–155. <https://doi.org/10.1080/02626660109492806>
- Ramachandra Rao, A., y Srinivas, V. V. (2006). Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology*, 318(1), 37–56. <https://doi.org/10.1016/j.jhydrol.2005.06.003>
- R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rodriguez, Y., y Marreno de León, N. (2011). Análisis regional de series de lluvias máximas: consideraciones. *Ingeniería Hidráulica y Ambiental*, XXXII(2), 34–45.
- Saf, B. (2009). Regional Flood Frequency Analysis Using L-Moments for the West Mediterranean Region of Turkey. *Water Resources Management*, 23(3), 531–551. <https://doi.org/10.1007/s11269-008-9287-z>
- Schaefer, M. G., Barker, B. L., Taylor, G. H., y Wallis, J. R. (2006). *Regional precipitation-frequency analysis and spatial mapping of precipitation for 24-hour and 2 hour durations in eastern Washington*. MGS Engineering Consultants, Inc.
- Smith, A., Sampson, C., y Bates, P. (2015). Regional flood frequency analysis at the global scale. *Water Resources Research*, 51, 539–553, <http://dx.doi.org/10.1002/2014WR015814>
- Stedinger, J. R., Vogel, R. M., y Foufoula-Georgiou, E. (1992). Frequency analysis of extreme events. In *Handbook of Hydrology*.
- Tucci, C. (2002) *Regionalização de Vazões*, Brasil, Ed. Universidades/UFRGS, Brasil.
- Veza, P., Comoglio, C., Rosso, M., y Viglione, A. (2010). Low Flows Regionalization in North-Western Italy. *Water Resources Management*, 24(14), 4049–4074. <https://doi.org/10.1007/s11269-010-9647-3>
- Veza, P., Comoglio, C., Viglione, A., y Rosso, M. (2009). The Influence of Soil Characteristics in Low Flows Regionalization. *American Journal of Environmental*

- Sciences*, 5(4), 536–546.
- Viglione, A. (2007). *Metodi statistici non-supervised per la stima di grandezze idrologiche in siti non strumentati*. Politecnico di Torino.
- Viglione, A., Claps, P., y Laio, F. (2007). Mean annual runoff estimation in North-Western Italy. *Water Resources Assessment And Management Under Water Scarcity Scenarios*, edited by G. La Loggia et al., pp. 97 – 121, CSDU, Milan, Italy.
- Viglione, A. (2018). nsRFA: Non-Supervised Regional Frequency Analysis. R package version 0.7-14. <https://CRAN.R-project.org/package=nsRFA>
- Vogel, R. M., y Fennessey, N. M. (1993). L moment diagrams should replace product moment diagrams. *Water Resources Research*, 29(6), 1745–1752. <https://doi.org/10.1029/93WR00341>
- Vogel, R. M., McMahon, T. A., y Chiew, F. H. S. (1993). Floodflow frequency model selection in Australia. *Journal of Hydrology*, 146, 421–449. [https://doi.org/10.1016/0022-1694\(93\)90288-K](https://doi.org/10.1016/0022-1694(93)90288-K)
- Vogel, R. M., Thomas, W. O., y McMahon, T. A. (1993). Flood-Flow Frequency Model Selection in Southwestern United States. *Journal of Water Resources Planning and Management*, 119(3), 353–366. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1993\)119:3\(353\)](https://doi.org/10.1061/(ASCE)0733-9496(1993)119:3(353))
- Wallis, J. R., Schaefer, M. G., Barker, B. L., y Taylor, G. H. (2007). Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrology and Earth System Sciences*, 11(1), 415–442. <https://doi.org/10.5194/hess-11-415-2007>
- Ward, J. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>
- Zaman, M. A., Rahman, A., y Haddad, K. (2012). Regional flood frequency analysis in arid regions: A case study for Australia. *Journal of Hydrology*, 475, 74–83. <https://doi.org/10.1016/j.jhydrol.2012.08.054>