



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA CIVIL – HIDRÁULICA

ANÁLISIS REGIONAL DE FRECUENCIA DE CAUDALES MÁXIMOS EN LA
CUENCA ALTA Y MEDIA DEL RÍO MAGDALENA - COLOMBIA

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA

PRESENTA:
WILSON ALEXANDER HURTADO PÉREZ

TUTOR PRINCIPAL
DR. CARLOS AGUSTÍN ESCALANTE SANDOVAL

COMITÉ TUTOR

DR. JESÚS GRACIA SÁNCHEZ, INSTITUTO DE INGENIERÍA
DRA. LILIA REYES CHÁVEZ, FACULTAD DE INGENIERÍA
DR. RODOLFO SILVA CASARÍN, INSTITUTO DE INGENIERÍA
DR. EDGAR GERARDO MENDOZA BALDWIN, INSTITUTO DE INGENIERÍA

MÉXICO, D. F. NOVIEMBRE DE 2015

JURADO ASIGNADO:

Presidente:	Dr. Jesús Gracia Sánchez
Secretario:	Dra. Lilia Reyes Chávez
1 ^{er} Vocal:	Dr. Carlos A. Escalante Sandoval
2 ^{do} Vocal:	Dr. Rodolfo Silva Casarín
3 ^{er} Vocal:	Dr. Edgar G. Mendoza Baldwin

MÉXICO, D.F.

TUTOR DE TESIS

DR. CARLOS AGUSTÍN ESCALANTE SANDOVAL

FIRMA

AGRADECIMIENTOS

A Dios por las bendiciones recibidas. Al doctor Carlos A. Escalante por su apoyo y recomendaciones para la elaboración de este trabajo. Al comité tutor por sus sugerencias y recomendaciones para mejorarlo.

A Juliana por su apoyo y acompañamiento. A mis compañeros de estudio de quien aprendí mucho y me hicieron sentir uno más de ellos, en especial a Érika, Saúl, Lalo, Regina, Rodolfo, Israel por siempre estar ahí. Los agradezco un buen, a todos.

A Patricia por su disposición y apoyo.

Al Consejo Nacional de Ciencia y Tecnología – CONACYT por la beca otorgada, que sin ella no hubiese sido posible realizar este posgrado.

Un agradecimiento especial al doctor Óscar Fuentes Mariles por el apoyo que me brindó durante el primer semestre.

TABLA DE CONTENIDO

1	INTRODUCCIÓN	11
1.1	MOTIVACIÓN	11
1.2	LONGITUD DEL REGISTRO: FACTOR DETERMINANTE EN EL ANÁLISIS DE FRECUENCIA DE GASTOS MÁXIMOS.....	11
1.2.1	Análisis de homogeneidad	11
1.2.2	Análisis en el sitio	12
1.2.3	Análisis regional.....	12
1.3	OBJETIVOS.....	12
1.4	ESTRUCTURA DE LA TESIS	13
2	ZONA DE ESTUDIO E INFORMACIÓN HIDROLÓGICA	14
2.1	INTRODUCCIÓN	14
2.2	CARACTERIZACIÓN DE LA CUENCA ALTA Y MEDIA	14
2.2.1	Clima	15
2.2.2	Cambio del uso del suelo y de la cobertura.....	16
2.2.3	Caracterización de la cuenca alta	16
2.2.4	Caracterización de la cuenca media	17
2.3	INFORMACIÓN HIDRO-CLIMATOLÓGICA.....	17
2.3.1	Información geográfica.....	17
2.3.2	Información hidrológica	17
2.3.3	Información climatológica.....	23
3	ANÁLISIS DE HOMOGENEIDAD DE SERIES HIDROLÓGICAS.....	24
3.1	Introducción	24
3.2	Pruebas de homogeneidad	24
3.2.1	Análisis de independencia.....	24
3.2.2	Pruebas de cambio en la media	25
3.2.3	Pruebas de cambio en la varianza	25
3.2.4	Análisis de tendencia	25
3.3	No estacionariedad de series hidrológicas	25
3.3.1	Causas de no estacionariedad.....	26
3.3.2	Tratamiento de la no estacionariedad en las series de gastos máximos.....	26
3.3.3	Análisis de frecuencia no estacionario considerando covariables	27

3.4	CASO DE ESTUDIO: CUENCA ALTA Y MEDIA DEL RIO MAGDALENA-COLOMBIA 28	
3.4.1	Caracterización	28
3.4.2	Análisis de homogeneidad	29
3.4.3	Análisis de las posibles causas de no estacionariedad de las estaciones40	
4	DISTRIBUCIONES DE PROBABILIDAD.....	46
4.1	Introducción	46
4.2	FACTORES QUE AFECTAN LA ELECCIÓN DE LA DISTRIBUCIÓN EN GASTOS MÁXIMOS	46
4.3	MÉTODOS DE ESTIMACIÓN DE PARÁMETROS	46
4.4	DISTRIBUCIONES UNIMODALES	47
4.4.1	Distribución Gamma con dos parámetros	47
4.4.2	Distribución de valores extremos tipo I (Gumbel)	49
4.4.3	Distribución general de valores extremos (GVE).....	51
4.5	DISTRIBUCIONES MEZCLADAS	54
4.5.1	Distribución Gumbel – Gumbel (G-G)	55
4.6	COMENTARIO SOBRE LA PARSIMONIA	56
4.7	PRUEBAS DE BONDAD DE AJUSTE	56
4.7.1	Error estándar de ajuste.....	57
4.7.2	Índice Medio de la Desviación Absoluta (MADI).....	57
4.7.3	Índice Desviación Media al Cuadrado (MSDI)	57
4.7.4	Criterio de la información de AKAIKE (CIA)	58
4.7.5	Criterio de Información Bayesiana (BIC)	58
5	ANÁLISIS DE FRECUENCIA EN EL SITIO	59
5.1	INTRODUCCIÓN	59
5.2	ANÁLISIS DE FRECUENCIA DE GASTOS MÁXIMOS ANUALES	60
5.3	CASO DE ESTUDIO	61
5.3.1	Análisis de sensibilidad de las distribuciones mezcladas a la longitud del registro	66
6	ANÁLISIS REGIONAL DE FRECUENCIA DE GASTOS MÁXIMOS	70
6.1	INTRODUCCIÓN	70
6.2	REGIÓN HOMOGÉNEA Y SU DELIMITACIÓN	70
6.2.1	Problemas relacionados con la formación de regiones	71
6.2.2	Técnicas de formación de grupos	72
6.3	VARIABLES EMPLEADAS PARA LA FORMACIÓN DE GRUPOS.....	74

6.4	MEDIDA DE LA DISCORDANCIA.....	75
6.4.1	Medida de la discordancia clásica.....	75
6.4.2	Medida de la discordancia robusta.....	76
6.4.3	Valores críticos	76
6.5	EVALUACIÓN DE LA HOMOGENEIDAD HIDROLÓGICA.....	77
6.6	TÉCNICAS DE ESTIMACIÓN REGIONAL.....	78
6.6.1	Dependencia entre sitios de medición de los gastos máximos anuales	78
6.6.2	Técnica de las estaciones – año	79
6.6.3	Técnica de correlación y regresión múltiple	80
6.6.4	Método de Dalrymple (Avenida índice).....	81
6.6.5	Método de los momentos estandarizados de probabilidad pesada	81
6.6.6	Análisis multivariado como método regional	81
6.7	CASO DE ESTUDIO	81
6.7.1	Selección de variables para la formación de grupos	82
6.7.2	Formación de grupos	85
6.7.3	Medida de la discordancia.....	90
6.7.4	Evaluación de la homogeneidad hidrológica	92
6.7.5	Prueba para evaluar si en las regiones los eventos extremos provienen de más de una población	96
6.7.6	Análisis regional de frecuencia.....	98
7	CONCLUSIONES Y RECOMENDACIONES.....	113
7.1	EL ¿POR QUÉ? DEL ÁREA DE ESTUDIO	113
7.2	CONCLUSIONES	113
7.3	RECOMENDACIONES	114
8	BIBLIOGRAFÍA	115

LISTA DE FIGURAS

Figura 2.1	Zona de estudio y localización de las estaciones hidrométricas.....	22
Figura 2.2	Variación de la precipitación media anual (a) y de los gastos máximos mensuales (b) en la zona de estudio.	23
Figura 3.1	Localización de embalses y estaciones (a) y de estaciones dependientes (b).....	30
Figura 3.2	Localización de las estaciones con cambio en la media	31
Figura 3.3	Estaciones de gastos máximos ubicados aguas abajo de los embalses La Copa, La Playa, El Hato y Betania.....	32
Figura 3.4	Localización de las estaciones que presentan cambio en la varianza .	33
Figura 3.5	Localización de las estaciones que presentan tendencias	35
Figura 3.6	Ejemplo de estaciones que presentaron tendencias	36
Figura 3.7	Análisis de homogeneidad para dos tamaños de muestra	38
Figura 3.8	Análisis de homogeneidad en estaciones con al menos 40 años de registro	40
Figura 3.9	Localización de estaciones no homogéneas. Con relleno rojo las estaciones en las cuales el modelo no estacionario empleando la covariable tiempo es más adecuado.....	43
Figura 3.10	Localización de estaciones no homogéneas. Con relleno azul las estaciones en las cuales el modelo no estacionario empleando la covariable ENSO es más adecuado.	44
Figura 3.11	Localización de las estaciones donde tanto la covariable índices del ENSO y la covariable tiempo describen mejor el comportamiento de los registros.	45
Figura 5.1	Localización de las estaciones en las cuales es adecuado el uso del modelo doble Gumbel	66
Figura 5.2	Función de densidad no paramétrica y función de densidad estimada para el modelo doble Gumbel (24037070).....	68
Figura 5.3	RECM para diferentes periodos de retorno y tamaños de muestra (24037070).....	68
Figura 5.4	Función de densidad no paramétrica y función de densidad estimada para el modelo doble Gumbel (24037110).....	69
Figura 5.5	RECM para diferentes periodos de retorno y tamaños de muestra (24037110).....	69
Figura 6.1	Variación del L-CV con el área de la cuenca.....	85

Figura 6.2	Variación del L-CV con el área de la cuenca. Con relleno rojo las estaciones en las cuales el análisis de frecuencia no estacionario empleando como covariables índices del ENSO es más adecuado.	85
Figura 6.3	Curvas medias f_t para cada grupo formado.....	87
Figura 6.4	Regiones formadas con el método de los trazos multidimensionales..	87
Figura 6.5	Grupos formados con el método de la región de influencia.....	89
Figura 6.6	Grupos formados con el método de Ward.....	90
Figura 6.7	Evaluación de los sitios discordantes para los diferentes grupos empleando la discordancia clásica y la discordancia robusta.	92
Figura 6.8	L-CV estimado para el periodo común de los sitios que conforman cada uno de los grupos.....	94
Figura 6.9	Localización de los sitios que conforman los grupos que son homogéneos tanto física como hidrológicamente.	94
Figura 6.10	Regiones homogéneas de la zona de estudio.....	96
Figura 6.11	Curva de frecuencia empírica conformada por dos curvas (24037110)97	
Figura 6.12	curvas regionales de frecuencia adimensionales estimadas para las regiones de la zona de estudio.....	99
Figura 6.13	Regresiones entre el gasto medio de los gastos máximos anuales y el área de las distintas regiones homogéneas.....	101
Figura 6.14	Curvas regionales de frecuencia empleando la técnica de momentos estandarizados de probabilidad pesada	102

LISTA DE TABLAS

Tabla 2.1	Estaciones hidrométricas en la zona de estudio y sus parámetros estadísticos	19
Tabla 5.1	Eventos de diseño (en m ³ /s) para diferentes periodos de retorno	64
Tabla 6.1	Valores críticos de la discordancia D_i	77
Tabla 6.2	Variables físicas, meteorológicas e hidrológicas de las estaciones	82
Tabla 6.3	Grupos formados mediante el método de los trazos multidimensionales	86
Tabla 6.4	Transformación de las características de los sitios	88
Tabla 6.5	Grupos formados mediante el método de la región de influencia	88
Tabla 6.6	Grupos formados mediante el método de Ward	89
Tabla 6.7	Evaluación de la medida de la discordancia para los diferentes grupos	93
Tabla 6.8	Regiones homogéneas definidas para la zona de estudio	95
Tabla 6.9	Evaluación de la medida de la heterogeneidad para las regiones de la zona de estudio	95
Tabla 6.10	Prueba para evaluar si los eventos extremos provienen de dos poblaciones	98
Tabla 6.11	Valores de las curvas regionales de frecuencia adimensionales para los diferentes periodos de retorno	100
Tabla 6.12	Valores de las curvas regionales de frecuencia empleando la técnica de momentos estandarizados de probabilidad pesada	102
Tabla 6.13	Eventos de diseño (en m ³ /s) para diferentes periodos de retorno para las estaciones de la región 1 estimados por los diferentes casos	104
Tabla 6.14	Eventos de diseño (en m ³ /s) para diferentes periodos de retorno para las estaciones de la región 2 estimados por los diferentes casos	105
Tabla 6.15	Eventos de diseño (en m ³ /s) para diferentes periodos de retorno para las estaciones de la región 3 estimados por los diferentes casos	108
Tabla 6.16	Eventos de diseño (en m ³ /s) para diferentes periodos de retorno para las estaciones de la región 4 estimados por los diferentes casos	109
Tabla 6.17	Relaciones para la estimación de eventos de diseño en sitios no muestreados de la región 1	111
Tabla 6.18	Relaciones para la estimación de eventos de diseño en sitios no muestreados de la región 2	111
Tabla 6.19	Relaciones para la estimación de eventos de diseño en sitios no muestreados de la región 3	111
Tabla 6.20	Relaciones para la estimación de eventos de diseño en sitios no muestreados de la región 4	112

RESUMEN

El conocimiento de la magnitud y frecuencia de los eventos de diseño asociados a los gastos máximos es necesario en el diseño de obras hidráulicas tales como puentes, obras de desvío de ríos, delimitación de zonas de inundación, obras de protección contra inundaciones, entre otras. Sin embargo, en muchos sitios no se cuenta con la información adecuada o no existen registros, lo que plantea un problema para la realización del análisis de frecuencia en el sitio. Sin embargo, existen metodologías como las técnicas regionales y los métodos multivariados que permiten estimar eventos de diseño en sitios no aforados o en aquellos donde la longitud de los registros es corta.

El objetivo de este trabajo fue la estimación de los eventos de diseño para diferentes periodos de retorno en la cuenca alta y media del río Magdalena en Colombia mediante el análisis regional de frecuencia de gastos máximos, empleando 62 estaciones hidrométricas de gastos máximos instantáneos con mínimo 40 años de registros.

El análisis de las causas de no estacionariedad de las estaciones permitió concluir que en al menos, el 42% de las estaciones no homogéneas, la variabilidad climática puede ser la causa. Estos resultados se utilizaron para emplear en el análisis de frecuencia de gastos máximos en el sitio distribuciones mezcladas de dos poblaciones encontrando que en 43 estaciones hidrométricas es más adecuado el uso de este tipo de distribuciones.

Para el análisis regional de frecuencia primero se identificaron las regiones homogéneas. Se formaron los grupos o regiones desde el punto de vista físico empleando el método de la región de influencia, las curvas de Andrews y el algoritmo de Ward y se evaluó la homogeneidad hidrológica de las mismas empleando la técnica de los momentos. - L. Se identificaron para la zona de estudio 4 regiones homogéneas. En general, para todas las regiones homogéneas identificadas, se recomienda realizar el análisis de frecuencia empleando distribuciones mezcladas especialmente para la región compuesta por la zona del altiplano cundi-boyacense, la cuenca del río Chicamocha y el río Lebrija.

La estimación de los eventos de diseño en cada una de las regiones se hizo empleando la técnica de las estaciones – año, que permite el uso de distribuciones mezcladas, la técnica de los momentos - L y el análisis multivariado. Si bien se encontró que el análisis multivariado da el menor error estándar de ajuste, se recomienda que la estimación de eventos de diseño tanto en sitios muestreados como en no muestreados se haga a partir del promedio de las estimaciones obtenidas con las diferentes técnicas.

Para cada una de las regiones homogéneas se obtuvieron curvas adimensionales de frecuencia y relaciones entre los eventos de diseño y parámetros morfométricos de las cuencas para obtener los eventos de diseño en sitios no muestreados.

1 INTRODUCCIÓN

1.1 MOTIVACIÓN

En muchos países de América Latina, sobre todo en los de Centro y Sur América, la información hidrológica es escasa, con pocos registros y las estaciones de medición se ubican principalmente sobre los ríos que cruzan los poblados más importantes. Por lo tanto, muchos de los ríos y cuencas que se ubican lejos de los centros poblados se encuentran sin registros, lo cual plantea un problema serio cuando se requiere conocer los gastos ya sea para el dimensionamiento de una obra hidráulica, para caracterizar los periodos de estiaje o realizar estudios de calidad de agua. En Colombia, las inundaciones pueden ser uno de los fenómenos naturales que más costos, tanto en pérdidas de vidas humanas como económicos genera.

La motivación de este estudio radica principalmente en la falta de información para el diseño de obras hidráulicas en zonas de Colombia donde actualmente se están desarrollando proyectos de infraestructura como el estudio y construcción de centrales hidroeléctricas, presas, construcción de puentes, delimitación de zonas de inundación, construcción de diques de protección contra inundación, obras de desvío, entre otras. Además, la climatología de la zona de estudio es afectada por mecanismos de baja frecuencia como El Niño-Oscilación del Sur (ENSO) lo que provoca que la respuesta hidrológica de la cuenca se origine de más de un mecanismo y por lo tanto, el análisis de frecuencia de gastos máximos empleando análisis univariado puede ser incorrecto y los eventos de diseño estimados no sean los apropiados para el dimensionamiento de las obras hidráulicas.

1.2 LONGITUD DEL REGISTRO: FACTOR DETERMINANTE EN EL ANÁLISIS DE FRECUENCIA DE GASTOS MÁXIMOS

El análisis de frecuencia de caudales máximos desempeña un papel integral en la determinación del riesgo en las cuencas propensas a inundaciones. Un registro largo de gastos máximos anuales es vital para la estimación de los eventos de diseño asociados a grandes periodos de retorno. Sin embargo, en muchas regiones del mundo, el análisis de frecuencia de gastos máximos se realiza con registros cortos lo cual conduce inevitablemente a que las estimaciones sean poco precisas (Senarath, 2013). La longitud de los registros en el análisis de frecuencia define cuando es o no adecuado realizar el análisis en el sitio o si lo más conveniente es realizar un análisis regional de frecuencia.

1.2.1 Análisis de homogeneidad

El análisis de frecuencia de eventos extremos se fundamenta en la hipótesis de que los eventos son independientes, no presentan cambios en la media, ni en la varianza y además no presentan tendencias, es decir, la estimación de los eventos de diseño tradicionalmente se basa en la consideración que las muestras de gastos máximos son estacionarios (Salas y Obeysekera, 2014). Es bien conocido, que esta consideración generalmente no es cierta y puede cambiar con la longitud de registros considerada (Klemes, 1989).

La variabilidad climática puede tener mayor influencia en el comportamiento de las series de gastos máximos con registros menores a 30 años, las cuales en algunos casos pueden mostrar tendencias evidentes (IH, 1999). Sin embargo, en zonas influenciadas por el ENSO, se debe analizar si aún en registros mayores de 30 años las series de gastos máximos pueden mostrar no estacionariedad por efecto de la variabilidad climática.

1.2.2 Análisis en el sitio

Es generalmente aceptable que el análisis en el sitio se realice cuando, al menos, se tienen 30 años de registros (NERC, 1975). Varios autores (Victorov, 1971; Finfrock y Rao, 2003; McCuen y Galloway, 2010) han señalado la importancia de contar con una mayor longitud de registros para poder realizar una adecuada selección de la función de distribución de probabilidades que mejor describe el comportamiento de los registros en el sitio y para obtener una mayor confiabilidad en los eventos estimados, sobre todo en los asociados a grandes periodos de retorno.

1.2.3 Análisis regional

El análisis regional de frecuencia es la herramienta a emplear cuando en el sitio se cuenta con registros cortos o no existen. No obstante, aún para algunos procesos necesarios para la aplicación del análisis regional de frecuencia, como el de la verificación de la homogeneidad hidrológica de las regiones, es aconsejable realizarlo con aquellas estaciones que tengan las longitudes de registros más largas (Dales y Reed, 1989).

1.3 OBJETIVOS

- Identificar para la zona de estudio cuál es la longitud de registros adecuada para realizar el análisis de homogeneidad y por lo tanto el análisis de frecuencia.
- Identificar, entre otras posibles causas, si la variabilidad climática puede explicar la no estacionariedad de las estaciones de gastos máximos de la zona de estudio.
- Evaluar si estadísticamente es mejor el uso de la distribución mezclada doble Gumbel para la estimación de los eventos de diseño para la zona de estudio.
- Evaluar la incertidumbre asociada, mediante simulaciones de Monte Carlo, en el uso de las distribuciones mezcladas en muestras con altos L-coeficientes de asimetría o con bajos L-coeficientes de asimetría.
- Mediante diferentes técnicas identificar las regiones homogéneas desde el punto de vista físico de la zona de estudio.
- Identificar mediante el uso de la medida de la discordancia robusta los sitios discordantes en cada una de las regiones homogéneas identificadas.
- Mediante el uso de la técnica de los Momentos-L evaluar la homogeneidad hidrológica de cada una de las regiones homogéneas identificadas desde el punto de vista físico.
- Estimar los eventos de diseño para las regiones homogéneas de la zona de estudio empleando diferentes técnicas regionales e identificar, a partir del error estándar de ajuste, cuál es la más adecuada para cada una de las regiones.
- Proporcionar relaciones para cada una de las regiones homogéneas para la estimación de los eventos de diseño para diferentes periodos de retorno en sitios no muestreados.

1.4 ESTRUCTURA DE LA TESIS

A continuación se presenta una breve descripción del contenido de los capítulos que componen esta tesis.

El capítulo 2 se describe la zona de estudio y la información hidro-climatológica empleada. Se presentan algunos aspectos para su caracterización como los mecanismos que intervienen en la respuesta hidrológica, brevemente se presenta un análisis de la deforestación sufrida en la cuenca. Las estaciones hidrométricas empleadas en el estudio son caracterizadas de acuerdo a su longitud de registros y algunos parámetros estadísticos, entre otras características.

El capítulo 3 presenta el análisis de homogeneidad de las estaciones de gastos máximos. Se define cuál es la longitud adecuada para realizar los análisis en la zona de estudio y a partir del análisis no estacionario considerando covariables se investiga si la variabilidad climática explica la no estacionariedad de las estaciones que la presentan. También se analizan otras causas que pueden explicar la no estacionariedad de los gastos máximos.

En el capítulo 4 se describen las distribuciones de probabilidad empleadas en el análisis de frecuencia. Se presentan las funciones de distribuciones de probabilidad univariada y las distribuciones de probabilidad mezclada que se emplearán para investigar el comportamiento de los gastos máximos en la zona de estudio.

En el capítulo 5 se presenta un análisis de frecuencia en el sitio. Se hace el análisis para determinar cuál tipo de función de distribución (univariada o mezclada) es más adecuado emplear en la zona de estudio y se hace un análisis de la incertidumbre en las estimaciones a partir de simulaciones de Monte Carlo.

En el capítulo 6 se presenta el análisis regional de frecuencia. Se identifican las regiones homogéneas tanto física como estadísticamente de la zona de estudio y se realiza la estimación de los eventos de diseño empleando varias técnicas regionales. Se dan algunas relaciones para la estimación de los eventos de diseño asociados a diferentes periodos de retorno en sitios no muestreados.

En el capítulo 7 se presentan las principales conclusiones obtenidas del estudio y se dan algunas recomendaciones para abordar futuras investigaciones relacionadas con el análisis realizado en este trabajo.

2 ZONA DE ESTUDIO E INFORMACIÓN HIDROLÓGICA

2.1 INTRODUCCIÓN

Los ríos se constituyen en una de las principales fuentes de desarrollo de un territorio, proporcionando agua, alimentación, transporte, energía hidroeléctrica, recreación, biodiversidad, etc. En el caso colombiano el río Magdalena, que hermanado con el río Cauca, ha sido eje del desarrollo económico del país y recurso natural forjador de la dinámica nacional. El río constituye el eje de desarrollo nacional más importante, desde el punto de vista geográfico-espacial, ambiental, cultural, social, económico, demográfico, urbano, histórico y, por supuesto, hidrográfico (Bernal, 2013).

El río Magdalena nace en el extremo suroccidental del país, a 3685 msnm, en la laguna de la Magdalena ubicada en una pequeña planicie del Páramo de las Papas, correspondiente al Macizo colombiano, en el Departamento del Huila. Su longitud es de aproximadamente 1600 km, de los cuales 886 son navegables. El río es utilizado principalmente para el transporte de hidrocarburos, materias primas y de los pobladores ubicados en las zonas ribereñas. En el Estrecho, el lugar donde el río es más angosto, mide 2.2 m de ancho y en el municipio de Plato, Magdalena, tiene un ancho de 1073 m. Vierte sus aguas en el mar Caribe, en el sitio conocido como Bocas de Ceniza. Después de su nacimiento, el río Magdalena fluye de sur a norte por un valle interandino ubicado entre las cordilleras Central y Oriental de los Andes (Bernal, 2013). A lo largo de su curso recibe unos 500 ríos, numerosas quebradas y registra en la estación Calamar, la más cercana a la desembocadura, un caudal promedio de 8224 m³/s para el periodo de registro entre 1969 y 1999 (Velasco y Granados, 2006).

El Magdalena sirve para muchos propósitos, tanto para extraer agua como para verter residuos de los sistemas de alcantarillado en la mayoría de ciudades colombianas. En menor escala es aprovechable para la navegación fluvial, la pesca y la generación de energía hidroeléctrica (Sarmiento, 2004). La cuenca de drenaje del río Magdalena tiene un área de 257000 km² y representa cerca del 25% del área continental de Colombia. Su red hidrográfica abastece las necesidades hídricas de más de la mitad de la población colombiana asentada en la cuenca, donde se genera aproximadamente el 85% del PIB nacional. Su densidad de población es equivalente a 120 habitantes/km² (Velasco y Granados, 2006). Por sus características morfológicas y de navegabilidad el río ha sido subdividido en tres cuencas, a saber: Alta, media y baja. En este estudio se trabajará con información de las cuenca alta y media.

2.2 CARACTERIZACIÓN DE LA CUENCA ALTA Y MEDIA

La cuenca alta y media del río Magdalena, con un área de 137440 km² (12% del territorio continental colombiano), se ubica entre el flanco oriental de la cordillera Central y el flanco occidental de la cordillera Oriental del sistema montañoso de los Andes. La cuenca está conformada por áreas de los departamentos de Cauca, Huila, Tolima, Cundinamarca, Caldas, Antioquia, Boyacá, Santander, Norte de Santander, Bolívar y Cesar donde se ubican ciudades principales como Bogotá, Bucaramanga, Neiva, Ibagué y Tunja. Se concentra cerca del 40% de la población de Colombia (DANE, 2006) y su economía representa cerca del 40% del PIB nacional (DANE, 2012). Los principales ríos que

drenan al río Magdalena en este tramo son: Suaza, La Plata, Páez, Saldaña, Sumapaz, Bogotá, La Miel, Nare, Carare, Sogamoso y Lebrija.

En la cuenca se encuentran dos de los sistemas hidroeléctricos más importantes de Colombia que son el sistema Nare-Guatapé y el sistema del río Bogotá, además de las centrales hidroeléctricas Miel I, Betania, Hidroprado y Sogamoso y más de 30 centrales hidroeléctricas a filo de agua. Actualmente se está construyendo la central hidroeléctrica de El Quimbo sobre el río Magdalena, además de varios proyectos a filo de agua ubicados en la cuenca del río Nare. La cuenca es una de las zonas con mayor potencial para el desarrollo de proyectos hidroeléctricos (UPME, 2011).

En la parte alta de la cuenca del río Chicamocha se ubica el distrito de riego del Alto Chicamocha-Firavitoba, distrito que cuenta con los embalses La Copa y La Playa que además ayudan a regular las inundaciones en la parte alta de la cuenca del río y el distrito de riego del valle del río Ubaté que cuenta con el embalse El Hato que además es utilizado para el control de inundaciones en la zona.

2.2.1 Clima

La variabilidad de la hidro-climatología de Colombia en diferentes escalas temporales se resume en (Poveda, 2004):

2.2.1.1 Escala Inter-Decadal

La variabilidad climática global a escala de varias décadas está dominada por el cambio climático y el cambio ambiental de largo plazo, así como por fenómenos macroclimáticos que operan en tal escala. Las evidencias de cambio climático en Colombia son múltiples: retroceso de glaciares tropicales, aumento de enfermedades de transmisión vectorial asociadas con el clima, perturbaciones de largo plazo y cambios en la fase y la amplitud de los regímenes hidrológicos. Hay varios fenómenos macro-climáticos que operan a escala de tiempo decadales tales como la Oscilación Decadal del Pacífico (PDO) y la Oscilación del Atlántico Norte (NAO) que tienen una influencia significativa sobre el clima de Colombia.

2.2.1.2 Escala interanual (ENSO)

Los efectos hidro-climáticos de El Niño/Oscilación del Sur (ENSO, por sus siglas en inglés) se pueden resumir en los siguientes aspectos principales: En general durante El Niño (Fase cálida del ENSO) se presenta una disminución en la precipitación y en los caudales medios mensuales de los ríos de Colombia, así como una disminución en la humedad del suelo y en la actividad vegetal. Consistentemente se presenta un incremento de la temperatura media del aire. Todas estas anomalías hidro-climáticas presentan alguna variabilidad regional y temporal. En general, los efectos del Niño son más fuertes y más inmediatos sobre el occidente que sobre el oriente de Colombia (Poveda y Mesa, 1997). Durante La Niña (Fase fría del ENSO) ocurren anomalías contrarias, con eventos de precipitación muy intensos, crecidas de ríos, avalanchas, e inundaciones de planicies aluviales, con las consecuentes pérdidas de vidas humanas, infraestructura y cosechas agrícolas. Desde el punto de vista físico se deduce que, además de los caudales medios mensuales, las dos fases del ENSO afectan los gastos extremos (máximos y mínimos) de distintos periodos de retorno.

2.2.1.3 Escala anual

El ciclo anual de la hidro-climatología de Colombia está dominado por fenómenos físicos como la migración latitudinal de la Zona de Convergencia Intertropical (ZCIT). La

oscilación meridional de la ZCIT que se da como resultado del ciclo anual de temperaturas superficiales por efecto de la insolación, constituye el fenómeno físico de mayor importancia para explicar el ciclo anual (o semi-anual) de la hidro-climatología de Colombia. Sobre el centro de Colombia se presentan dos temporadas de lluvias (abril-mayo y octubre-noviembre) y dos temporadas secas (diciembre-febrero y junio-agosto), como resultado del doble paso de la ZCIT sobre el territorio colombiano. En lugares extremos de la oscilación de la ZCIT, el ciclo anual es unimodal. Poveda y Álvarez, (2011) encontraron que varias series de gastos máximos anuales en Colombia presentan tendencias; que en su mayoría son ocasionadas por la fase cálida del fenómeno del ENSO (Fenómeno de El Niño) que presenta una mayor frecuencia que la fase fría (Fenómeno de La Niña). Así mismo, han encontrado que en la región oriental de Colombia los gastos máximos anuales aumentan con la fase fría del fenómeno del ENSO.

2.2.2 Cambio del uso del suelo y de la cobertura

Para el año de 1985 los Andes colombianos ya presentaban un elevado índice en transformación, existiendo para entonces un remanente de ecosistemas naturales del 44.3% (Armenteras y Rodríguez, 2007). La tasa anual de deforestación de los Andes colombianos para el periodo 1985-2005 fue de aproximadamente 0.67% (Rodríguez, 2011; Armenteras y Rodríguez, 2007). Rodríguez (2011) comparó imágenes de 1985 y 2005 para evaluar la tasa de deforestación en 627 municipios ubicados en zonas de bosque de tierras bajas (<1000 msnm) y en bosques de montaña (>1000 msnm) de los Andes colombianos. El total del área de bosques de la zona de estudio se redujo de 11006893 ha en 1985 a 9528962 ha en 2005 (0.67%), lo que representó una disminución de 7335125 ha a 6405591 ha (0.63%) en bosques de montaña y de 3671768 ha a 3123369 ha (0.75%) en bosques de tierras bajas. En las zonas de bosque de montaña los factores que han tenido mayor influencia son los relacionados con la adecuación de tierras para cultivo, construcción de vías, etc. En los bosques de tierras bajas la deforestación se debió a factores como expansión de la frontera agrícola, incremento de áreas destinadas para la ganadería, colonización, etc. La zona del bosque de tierra baja del valle del río Magdalena presenta la mayor superficie deforestada para el periodo 1985-2005 con 213550 ha aproximadamente, asociada a la vertiente occidental de la cordillera Oriental en las riberas de los ríos Opón, Cascajales, Honduras, Chucurí, Lebrija principalmente; en la vertiente oriental de la cordillera Central en el sector de San Diego en Samaná, (Caldas), riberas del río Samaná, sectores de los municipios de San Francisco, Sonsón, San Carlos, Segovia, y Remedios en el departamento de Antioquia (Armenteras y Rodríguez, 2007).

El ecosistema natural de la zona del altiplano cundiboyacense respecto a 1985 ha presentado una disminución del 75.2% (Armenteras y Rodríguez, 2007). Los bosques y las zonas de pasto eran el principal tipo de cobertura (65.2%) de los Andes colombianos en el año 2000 (Rodríguez, 2011).

A continuación se presenta una caracterización de la cuenca alta y media del río Magdalena.

2.2.3 Caracterización de la cuenca alta

La cuenca alta comprende desde el nacimiento del río en el Páramo de las Papas, hasta Honda en sentido latitudinal, y en sentido longitudinal, de este a oeste, desde las cimas de la cordillera Central hasta las de la Cordillera Oriental, recorre desde su nacimiento 565 km hasta el salto de Honda. Geomorfológicamente este tramo del río es el más joven

con mayores pendientes y donde el río es más estrecho. De acuerdo a Velasco y Granados (2006) la región del valle del Alto Magdalena posee todos los pisos térmicos. En la región plana la temperatura aumenta hacia el norte y en la medida en que disminuye la altitud. La temperatura regional varía entre un poco menos de siete grados hasta valores superiores a 23 grados, en promedio (Velasco y Granados, 2006). El régimen de lluvias desde el nacimiento del río hasta San Agustín es monomodal con periodos lluviosos entre los meses de marzo y septiembre alcanzando su valor máximo en el mes de junio. Las estaciones de gastos máximos de esta región, empleadas en el estudio, presentan un régimen monomodal hasta la confluencia del río Páez con el Magdalena con máximos que se presentan en los meses de junio o julio. Desde este punto hacia aguas abajo el régimen es bimodal, con dos temporadas lluviosas (abril-mayo y octubre-noviembre) y dos temporadas secas (diciembre-febrero y junio-agosto), como consecuencia del doble paso de la ZCIT sobre el territorio (Poveda, 2004). Los suelos de las márgenes del río Magdalena en los departamentos del Huila y Tolima son utilizadas principalmente para los cultivos de cereales.

2.2.4 Caracterización de la cuenca media

Se denomina valle Intermedio del río Magdalena al área comprendida entre la ciudad de Honda y el municipio de Morales, en el departamento de Bolívar. El río en esta región recorre 260 km, y alcanza una diferencia en alturas entre los dos puntos de 145 m (Velasco y Granados, 2006). En este estudio solo se considerará el análisis hasta el sector La Gloria. Las lluvias y gastos presentan un régimen bimodal, con dos temporadas lluviosas (abril-mayo y octubre-noviembre) y dos temporadas secas (diciembre-febrero y junio-agosto), como consecuencia del doble paso de la ZCIT sobre el territorio (Poveda, 2004). Comparada con la cuenca alta esta zona presentan una mayor precipitación media anual.

2.3 INFORMACIÓN HIDRO-CLIMATOLÓGICA

A continuación se presenta la información utilizada para la realización de este estudio, así como las fuentes de las cuales fue obtenida.

2.3.1 Información geográfica

Para obtener las propiedades morfométricas de las cuencas estudiadas, como el área, el perímetro, la longitud del cauce principal, entre otras, se emplearon modelos de elevación digital (DEM) de 30x30 m de resolución, obtenidos del sitio ASTER GDEM (<http://gdem.ersdac.jpacesystems.or.jp/feature.jsp>). La información fue procesada con el software ArcGis 10.0.

2.3.2 Información hidrológica

Para la realización de este estudio se contó con información de 135 estaciones hidrométricas de gastos máximos instantáneos ubicadas en la cuenca alta y media del río Magdalena. Se adquirió a través del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) la información disponible hasta el año 2012. La información faltante fue completada a partir de la técnica de correlación lineal simple con las estaciones vecinas (Salas, 1980). Por razones que se explicarán en el capítulo 3 solo se trabajará con las estaciones que cuenten con 40 años o más de registros. En la Tabla 2.1 se presentan las estaciones de la zona de estudio que tienen 40 o más años de registros y sus principales parámetros estadísticos. También se presentan las relaciones L-

coeficiente de variación (L-CV), L-coeficiente de asimetría (L-g) y L-coeficiente de curtosis (L-k) obtenido mediante la técnica de momentos-L. En la Figura 2.1 se presenta la localización de la zona de estudio y la ubicación de las estaciones hidrométricas.

Tabla 2.1 Estaciones hidrométricas en la zona de estudio y sus parámetros estadísticos

Estación	ID	Latitud (°)	Longitud (°)	Corriente	X (m ³ /s)	S (m ³ /s)	CV	g	k	n	Periodo de registro	% faltantes	L-CV	L-g	L-k
21087050	1	2.71	75.62	Iquira	24.0	17.9	0.75	2.11	7.65	42	1971-2012	0.0	0.36	0.40	0.31
21017050	2	1.85	76.05	Guarapas	78.7	20.3	0.26	0.65	5.14	42	1971-2012	0.0	0.14	0.05	0.22
21017020	3	1.87	76.23	Sombrerillos	153.3	77.4	0.50	1.76	7.11	42	1971-2012	0.0	0.26	0.27	0.21
21017040	4	1.99	76.01	Magdalena	1048.0	271.9	0.26	0.63	3.25	42	1971-2012	0.0	0.15	0.12	0.13
21037010	5	2.03	75.78	Suaza	385.5	118.9	0.31	0.95	3.54	49	1964-2012	0.0	0.17	0.20	0.13
21047010	6	2.23	75.65	Magdalena	1410.1	365.9	0.26	0.98	4.67	42	1971-2012	0.0	0.14	0.15	0.18
21057050	7	2.33	75.94	La plata	278.1	66.4	0.24	-0.04	3.39	40	1973-2012	2.5	0.13	-0.01	0.20
21107030	8	2.54	75.30	Neiva	31.5	22.5	0.71	2.00	7.50	42	1971-2012	0.0	0.35	0.39	0.23
21107020	9	2.57	75.37	Neiva	139.5	82.8	0.59	1.16	4.09	46	1967-2012	0.0	0.32	0.24	0.17
21127020	10	2.95	75.52	Bache	82.7	78.0	0.94	3.83	19.78	42	1971-2012	0.0	0.36	0.50	0.46
21127030	11	2.97	75.59	Bache	42.4	31.1	0.73	3.08	14.44	42	1971-2012	0.0	0.31	0.52	0.35
21137030	12	3.26	75.25	Aipe	401.9	154.5	0.38	0.70	3.02	41	1972-2012	0.0	0.22	0.15	0.14
21147030	13	3.33	74.88	Cabrera	487.0	158.9	0.33	1.09	4.18	41	1972-2012	0.0	0.18	0.22	0.19
21147010	14	3.36	75.12	Cabrera	644.0	494.4	0.77	4.09	22.91	47	1966-2012	0.0	0.30	0.45	0.43
21057030	15	2.46	75.91	Páez	640.8	192.2	0.30	0.65	4.16	42	1971-2012	9.5	0.17	0.09	0.19
22027020	16	3.13	75.66	Ata	128.7	33.8	0.26	0.70	2.66	41	1972-2012	0.0	0.15	0.18	0.09
22027010	17	3.33	75.62	Ata	224.4	80.3	0.36	1.11	3.40	42	1971-2012	0.0	0.19	0.26	0.14
22057060	18	3.48	75.47	Saldaña	922.4	198.3	0.22	0.28	2.87	49	1964-2012	0.0	0.12	0.05	0.12
22057040	19	3.67	75.31	Saldaña	1235.4	411.1	0.33	1.42	5.03	47	1966-2012	0.0	0.18	0.30	0.18
22067010	20	3.92	75.18	Ortega	219.8	163.2	0.74	2.03	8.15	42	1971-2012	7.1	0.37	0.32	0.24
22077060	21	4.17	75.27	Tuamo	76.8	50.6	0.66	1.78	5.44	41	1972-2012	0.0	0.32	0.41	0.23
21217120	22	4.40	75.29	Coello	84.6	40.6	0.48	1.46	4.83	41	1972-2012	9.8	0.25	0.32	0.20
23017030	23	5.20	74.74	Gualí	387.0	278.9	0.72	1.52	4.75	42	1971-2012	16.7	0.37	0.36	0.19
23067020	24	5.47	74.59	Negro	986.7	583.9	0.59	1.58	6.37	61	1952-2012	0.0	0.31	0.24	0.18
21137020	25	3.86	74.96	Chenche	45.1	9.1	0.20	0.77	3.86	42	1971-2012	0.0	0.11	0.14	0.16
21227010	26	4.54	74.88	Opia	198.7	84.3	0.42	0.27	3.79	40	1973-2012	0.0	0.23	0.03	0.20
21257070	27	5.02	74.89	Sabandija	286.9	157.7	0.55	0.98	4.62	41	1972-2012	2.4	0.30	0.15	0.16

Estación	ID	Latitud (°)	Longitud (°)	Corriente	X (m ³ /s)	S (m ³ /s)	CV	g	k	n	Periodo de registro	% faltantes	L-CV	L-g	L-k
22077030	28	4.02	75.57	Cucuaña	63.4	44.5	0.70	3.84	22.84	52	1961-2012	1.9	0.30	0.42	0.33
21187020	29	4.22	75.20	Luisa	162.8	136.8	0.84	3.47	19.62	52	1961-2012	0.0	0.37	0.33	0.27
21197030	30	4.19	74.50	Sumapaz	308.5	148.9	0.48	1.28	5.10	54	1959-2012	0.0	0.26	0.24	0.22
21197110	31	4.40	74.38	Subia	104.4	47.8	0.46	0.98	4.70	41	1972-2012	0.0	0.25	0.12	0.16
23037010	32	5.47	74.66	Magdalena	4261.2	661.7	0.16	0.95	4.64	67	1946-2012	3.0	0.08	0.14	0.15
21167050	33	4.06	74.69	Prado	97.5	38.2	0.39	0.83	3.14	42	1971-2012	0.0	0.22	0.19	0.11
21187030	34	4.23	75.13	Luisa	260.3	162.7	0.63	1.04	3.87	42	1971-2012	0.0	0.34	0.21	0.14
21237020	35	5.20	74.71	Magdalena	3522.9	536.4	0.15	0.45	2.98	79	1934-2012	1.3	0.09	0.09	0.12
21237040	36	4.65	74.66	Seco	227.4	178.6	0.79	1.39	5.34	41	1972-2012	4.9	0.42	0.26	0.15
24017610	37	5.33	73.70	Lenguazaque	26.1	15.6	0.60	1.22	4.09	52	1961-2012	1.9	0.32	0.30	0.11
21087040	38	2.71	75.57	Pedernal	100.8	93.1	0.92	2.04	7.23	40	1973-2012	30.0	0.45	0.36	0.25
23197270	39	7.34	72.99	Vetas	35.2	30.0	0.85	1.44	4.89	42	1971-2012	21.4	0.45	0.31	0.16
23197290	40	7.17	73.15	Lebrija	131.0	60.6	0.46	1.18	3.72	44	1969-2012	0.0	0.25	0.28	0.15
24027050	41	6.29	73.13	Taquiza	276.0	80.3	0.29	0.25	2.08	42	1971-2012	0.0	0.17	0.07	0.05
24037080	42	5.57	73.05	Pesca	18.7	18.0	0.96	2.42	8.89	45	1968-2012	0.0	0.44	0.47	0.30
24037040	43	6.46	72.42	Nevado	29.8	12.6	0.42	1.76	7.49	57	1956-2012	0.0	0.22	0.26	0.19
24017580	44	6.23	73.31	Oibita	169.6	39.4	0.23	1.04	4.88	48	1965-2012	0.0	0.12	0.17	0.27
24037070	45	5.74	73.20	Sotaquirá	14.6	8.4	0.57	0.90	2.93	49	1964-2012	0.0	0.32	0.24	0.08
24037450	46	5.72	73.12	Salitre	6.7	9.4	1.40	4.03	23.35	48	1965-2012	0.0	0.55	0.51	0.39
24037120	47	5.59	72.98	Iza	16.1	12.3	0.76	2.45	9.34	48	1965-2012	0.0	0.35	0.39	0.37
24037300	48	5.72	72.85	Monguí	17.5	9.7	0.55	2.27	10.95	41	1972-2012	0.0	0.27	0.27	0.26
24027040	49	6.53	73.01	Mogoticos	57.7	28.8	0.50	1.07	3.06	41	1972-2012	0.0	0.27	0.29	0.11
24037110	50	5.68	72.99	Chiquito	21.1	14.2	0.67	2.65	11.28	49	1964-2012	0.0	0.30	0.41	0.35
23107130	51	7.29	73.14	Santa cruz	68.2	42.1	0.62	0.83	3.29	44	1969-2012	0.0	0.35	0.18	0.08
24067030	52	7.11	73.42	Sogamoso	2412.8	555.1	0.23	0.85	5.17	41	1972-2012	0.0	0.13	0.07	0.13
24047020	53	6.61	73.28	Suarez	1785.1	371.7	0.21	-0.03	2.15	40	1973-2012	0.0	0.12	-0.02	0.06
24027010	54	6.55	73.13	Fonce	593.8	133.6	0.23	0.38	3.55	58	1955-2012	0.0	0.13	0.05	0.17
24027030	55	6.23	73.17	Pienta	256.9	86.0	0.33	0.86	3.27	42	1971-2012	0.0	0.19	0.19	0.13

Estación	ID	Latitud (°)	Longitud (°)	Corriente	X (m ³ /s)	S (m ³ /s)	CV	g	k	n	Periodo de registro	% faltantes	L-CV	L-g	L-k
23097030	56	6.49	74.40	Magdalena	5584.7	1106.4	0.20	0.40	3.18	77	1936-2012	0.0	0.11	0.06	0.13
23087150	57	6.14	75.38	Negro	48.7	23.7	0.49	1.44	4.98	40	1973-2012	0.0	0.26	0.28	0.19
23087200	58	5.97	75.12	Sto domingo	205.8	84.5	0.41	0.63	3.36	40	1973-2012	0.0	0.23	0.12	0.18
23087180	59	6.38	75.16	Qda la concepcion	69.6	24.5	0.35	-0.23	2.65	40	1973-2012	2.5	0.20	-0.06	0.10
21057060	60	2.46	75.76	Paez	1055.3	302.3	0.29	1.36	7.19	42	1971-2012	4.8	0.15	0.16	0.20
23087190	61	6.00	74.94	Samana norte	612.9	90.7	0.15	0.40	3.35	40	1973-2012	0.0	0.08	0.07	0.18
23127020	62	6.53	74.09	Carare	1425.1	376.6	0.26	0.83	3.37	48	1965-2012	0.0	0.15	0.18	0.18
24037290	63	5.76	72.91	Chicamocha	66.8	38.9	0.58	1.72	8.11	42	1971-2012	0.0	0.30	0.23	0.16
24037320	64	5.77	73.11	Chicamocha	42.8	43.4	1.02	3.21	15.18	41	1972-2012	41.5	0.43	0.48	0.37
24037090	65	5.71	73.24	Chulo	9.7	5.5	0.57	1.91	7.93	49	1964-2012	0.0	0.29	0.26	0.26
24037030	66	5.67	73.22	Tuta	30.3	20.6	0.68	1.72	7.17	58	1955-2012	0.0	0.35	0.26	0.18
24037130	67	5.75	73.19	Chicamocha	37.0	21.3	0.58	1.43	4.89	47	1966-2012	0.0	0.30	0.29	0.24
24017150	68	5.30	73.85	Ubate	25.2	19.6	0.78	1.49	6.89	53	1960-2012	0.0	0.41	0.21	0.14
21097070	69	2.94	75.31	Magdalena	2023.9	558.6	0.28	0.51	2.70	53	1960-2012	0.0	0.16	0.12	0.10

X media de los gastos máximos; S desviación estándar; CV coeficiente de variación; g coeficiente de asimetría; k coeficiente de curtosis; n número de registros; L-CV L-coeficiente de variación; L-g L-coeficiente de asimetría; L-k L-coeficiente de curtosis.

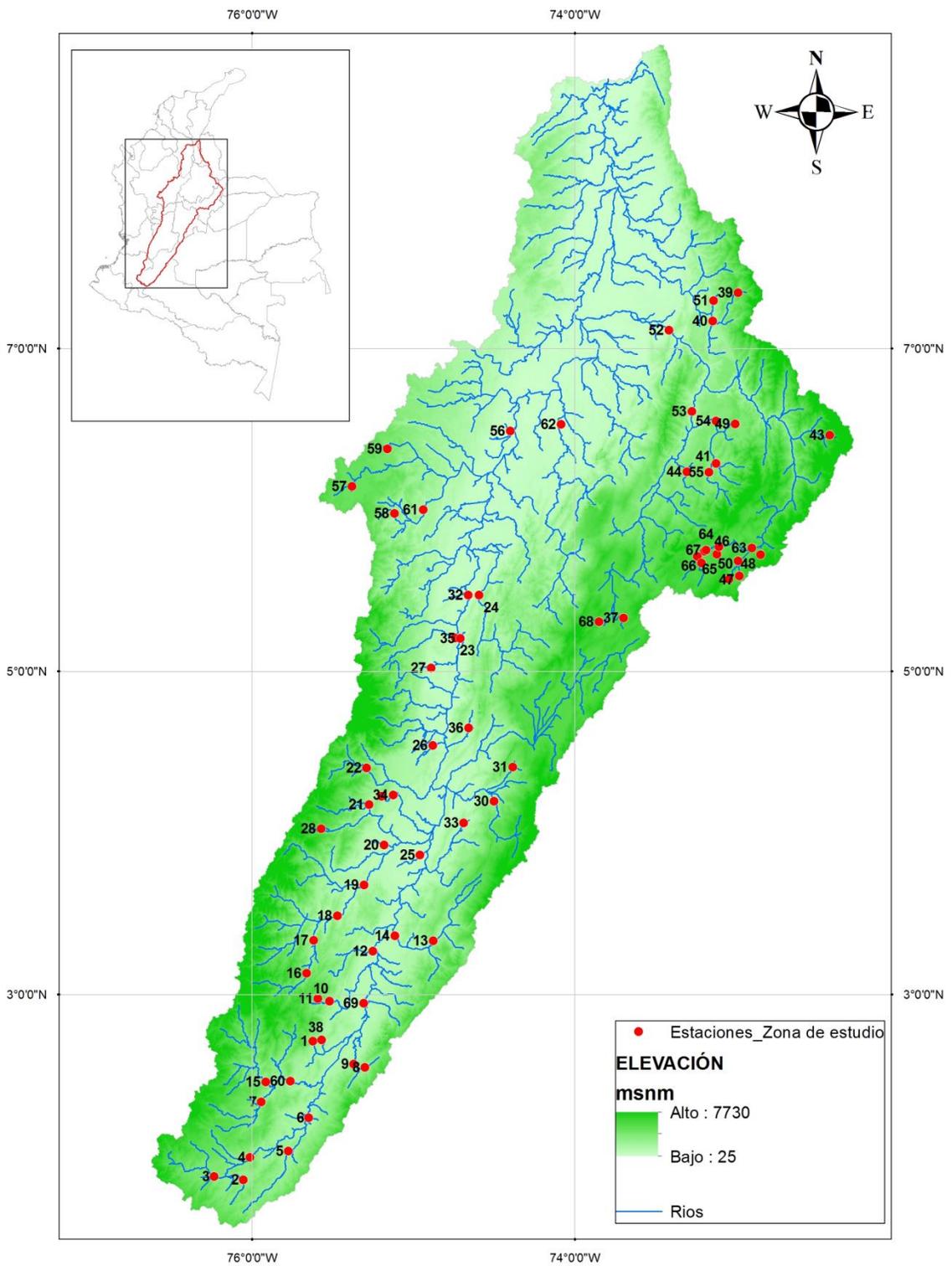


Figura 2.1 Zona de estudio y localización de las estaciones hidrométricas

2.3.3 Información climatológica

A partir de la información de precipitación media anual de 432 estaciones climatológicas operadas por el IDEAM se construyó un mapa de precipitación media anual para la zona de estudio empleando el método de interpolación inversa utilizando un factor de 2. Se emplearon para la construcción del mapa los registros de precipitaciones medias mensuales para el periodo 1981-2010. En la Figura 2.2 se presenta el mapa de precipitación media anual de la zona de estudio. Así mismo se presenta el comportamiento mensual de los gastos máximos en la zona de estudio.

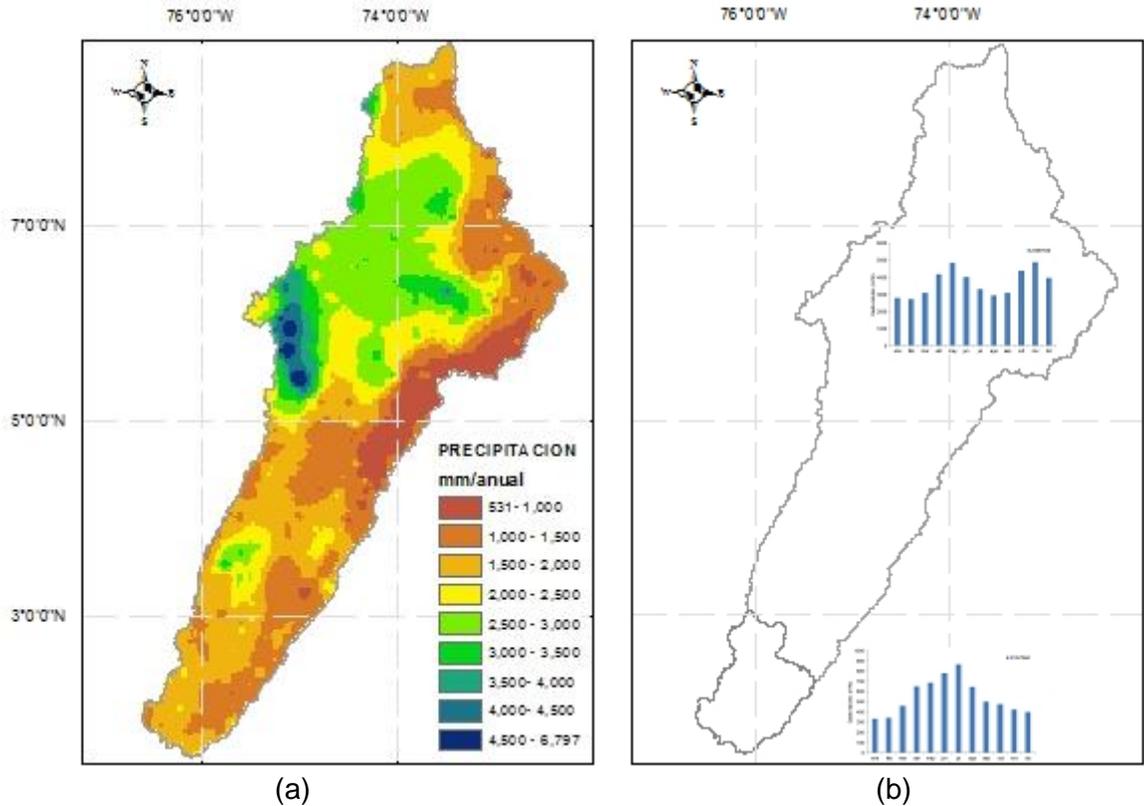


Figura 2.2 Variación de la precipitación media anual (a) y de los gastos máximos mensuales (b) en la zona de estudio.

3 ANÁLISIS DE HOMOGENEIDAD DE SERIES HIDROLÓGICAS

3.1 INTRODUCCIÓN

Los análisis de homogeneidad de series hidrológicas son un aspecto fundamental en los análisis hidrológicos y deben realizarse previamente a cualquier otro análisis, con el objetivo de determinar la calidad de la información hidrológica que se está utilizando. Los análisis hidrológicos serán tan buenos como la información hidrológica que se use (Smith y Campuzano, 2000).

El análisis de frecuencia de eventos extremos se fundamenta en la hipótesis de que los eventos son independientes, no presentan cambios en la media, ni en la varianza y además no presentan tendencias, es decir, la estimación de los eventos de diseño tradicionalmente se basa en la consideración que las muestras de gastos máximos son estacionarios (Petrow y Merz, 2009; Salas y Obeysekera, 2013; Salas y Obeysekera, 2014). Generalmente, las series de gastos máximos en una escala de tiempo anual son estacionarias, sin embargo, esta consideración puede ser incorrecta como resultado de la variabilidad climática de gran escala, desastres naturales como erupciones volcánicas, o cambios inducidos por el hombre como la construcción de embalses, la deforestación, la urbanización, los cambios de usos del suelo. (Salas et al, 1980; Villarini et al, 2009a; Vogel et al, 2011). Adicionalmente, componentes de baja frecuencia de la variabilidad climática, tales como el ENSO (El Niño-Oscilación del Sur) y la PDO (Oscilación Decadal del Pacífico) afectan la ocurrencia y magnitud de los eventos hidrológicos extremos como los niveles del mar y los gastos máximos instantáneos (Mantua et al, 1997; Escalante-Sandoval, 2007; Marengo, 2006).

3.2 PRUEBAS DE HOMOGENEIDAD

Existen una gran variedad de pruebas para el análisis de la homogeneidad de series de gastos máximos. A continuación se presentan las características que se evaluarán en las series de gastos máximos anuales y las pruebas que se emplearán.

3.2.1 Análisis de independencia

Usualmente la autocorrelación o dependencia de series de tiempo de gastos resulta del efecto de la regulación de embalses, lagos o acuíferos, que hacen que el agua permanezca en el sistema un tiempo adicional (Salas, 1991). Generalmente las series de gastos máximos anuales son independientes, sin embargo, series de tiempo no-homogéneas pueden presentar una significativa correlación serial (Potter, 1979). Sveinsson et al, (2005) señalan que las series de gastos máximos pueden ser dependientes no necesariamente por la presencia de almacenamientos sino también por la presencia de mecanismos climáticos de baja frecuencia de gran escala que exhiben patrones de cambios súbitos. Para la evaluación de la independencia serial de los gastos máximos anuales se empleará la prueba de independencia de Anderson (Anderson, 1942).

3.2.2 Pruebas de cambio en la media

Una importante fuente de cambio en la media de series de gastos máximos resulta de cambios en el usos del suelo y la construcción de embalses y estructuras de desvío (Salas, 1991). Otra fuente resulta de la forma de obtención de los registros de niveles, los cambios en las curvas de calibración de las estaciones, cambios de los equipos de medición, erosión de las secciones de medición. También se ha encontrado que los mecanismos de la variabilidad climática de baja frecuencia como el ENSO, la PDO ocasionan cambios en las series de gastos (Sveinsson et al, 2003; Sveinsson et al, 2005, Salas y Obeysekera, 2014).

Las pruebas estadísticas que se aplicaran en este trabajo para la evaluación del cambio en la media son: Prueba de Helmer (Konrad, 1944), Prueba estadística t de Student (Snedecor y Cochran, 1980), Prueba estadística de Cramer, Prueba estadística de Pettit (Pettit, 1979), Prueba Normal Estándar (SNHT) (Alexandersson y Moberg, 1997), Prueba de Buishand (Buishand, 1982), y la Prueba de Von Newman (Buishand, 1982).

3.2.3 Pruebas de cambio en la varianza

Para el análisis de cambio en la varianza se utilizará la prueba de Fisher (Snedecor y Cochran, 1980).

3.2.4 Análisis de tendencia

Una serie presenta tendencia en la media, cuando se presenta un cambio progresivo y gradual en la magnitud o el nivel de la media. Esta tendencia puede ser positiva si hay un incremento gradual de la magnitud de la media o negativa si ocurre lo contrario. Los factores más importantes que ocasionan la presencia de tendencias en las series hidrológicas son los cambios del uso del suelo, la deforestación, y la variabilidad climática (Villarini et al, 2009; Salas y Obeysekera, 2014).

Las pruebas estadísticas que se emplearán para evaluar la tendencia de las series de gastos máximos son la prueba de Mann-Kendall (Kendall, 1975) y la prueba de Spearman. Según Yue et al, (2002), la potencia (definida como la probabilidad de rechazar correctamente la hipótesis nula cuando esta es falsa) en detectar tendencias de las pruebas de Mann-Kendall y la de Spearman es muy similar.

3.3 NO ESTACIONARIEDAD DE SERIES HIDROLÓGICAS

Un registro de gastos es no estacionario si algunas de las propiedades subyacentes (media, varianza) cambian en el tiempo (IH, 1999). Una serie con tendencia es un ejemplo de una serie no estacionaria. La no estacionariedad surge además de cambios repentinos, cambios escalonados en las series de los registros o si hay una marcada fluctuación en los registros. Una serie de registros exhibe fluctuaciones si el promedio de ventanas de tiempo, por ejemplo de 10 años, de la serie de registros va cambiando notablemente pero no en una dirección consistente. Las fluctuaciones están relacionadas con la variabilidad natural (variabilidad climática), especialmente en registros cortos (IH, 1999).

Es importante resaltar que la variabilidad climática no es lo mismo que el cambio climático. Considerando cambio climático las alteraciones se producen a largo plazo. Considerando la variabilidad climática, el clima difiere de un periodo al siguiente pero en promedio mantiene una condición estacionaria, a menos que también haya evidencia de cambio climático (IH, 1999; Kundzewicz y Robson, 2004). La variabilidad climática puede

tener la mayor influencia en el comportamiento de las series de gastos máximos con registros cortos, las cuales en algunos casos pueden mostrar tendencias evidentes que podrían desaparecer cuando la longitud de los registros se incremente y las variaciones en el clima se mantengan en promedio (Salas et al, 1980; IH, 1999; Kundzewicz y Robson, 2004). Debido a que la variabilidad climática es típicamente mayor, esta puede efectivamente ocultar los cambios subyacentes, ya sea los debidos a cambios antropogénicos o al cambio climático (Kundzewicz y Robson, 2004). Las tendencias o cambios en las series de registros cortos deben considerarse principalmente que son producto de la variabilidad climática antes que de factores antropogénicos o el cambio climático (IH, 1999).

3.3.1 Causas de no estacionariedad

Es importante entender las causas de la no estacionariedad de las series de registros toda vez que las implicaciones para el análisis de frecuencia pueden ser distintas. Algunas de las causas de no estacionariedad son (Salas et al, 1980; Salas, 1991; Hanssen-Bauer and Førland, 1994; IH, 1999; Coles, 2001; Sveinsson et al, 2005; Villarini et al, 2009):

- Problemas con los registros de los datos: Por ejemplo errores de transcripción de los datos, errores topográficos, cambios abruptos en las curvas de calibración, relocalización y/o reconstrucción de los sitios de medición y de los equipos de medición.
- Cambios en la cuenca: Por ejemplo cambios en el uso del suelo (notable urbanización, deforestación, incrementos de áreas de cultivo), construcción de estructuras de desvío, construcción de embalses, construcción de obras para el alivio de inundaciones.
- Variaciones en el clima: Por ejemplo la variabilidad climática y el cambio climático.

La identificación de la causa más probable de no estacionariedad requiere de una detallada investigación de los registros de gastos y de la información histórica relacionada con la cuenca. Algunos requerimientos son:

- Visualizar del comportamiento respecto al tiempo de los registros históricos y compararlo con las estaciones vecinas.
- Evaluar pruebas que permitan verificar si las causas de no estacionariedad es por la variabilidad climática.
- Verificar la calidad de los datos, investigar posibles errores de transcripción de los datos, errores topográficos, cambios en las curvas de calibración y sitios de medición, etc.
- Investigar información de la cuenca relacionada con cambios de usos del suelo, urbanización, embalses, sistemas de desvío, etc.

3.3.2 Tratamiento de la no estacionariedad en las series de gastos máximos

Si los registros muestran fuertes cambios (media y varianza), dependencias, y tendencias y son usados para el análisis de frecuencia, los resultados obtenidos, en el mejor de los casos, pueden representar la respuesta media durante el periodo de registros (IH, 1999). Si no se abordan de manera adecuada los resultados pueden ser inconsistentes. Dependiendo de las causas de no estacionariedad las siguientes acciones pueden ser consideradas.

- No estacionariedad ocasionada por problemas con los registros: La acción recomendada es corregir los datos, pero si esto no es posible, puede que sea necesario solo emplear una parte de los registros.
- No estacionariedad ocasionada por cambios en la cuenca: En este caso es preferible emplear solo la parte de los registros después del cambio. Sin embargo, se debe tener cuidado con el tipo de cambio y evaluar su variación en el tiempo. Alternativamente se puede considerar el registro completo pero incorporando en los análisis la no estacionariedad, que puede ser tenida en cuenta empleando covariables en el análisis de frecuencia (Coles, 2001; Katz et al, 2002; Leclerc y Ouarda, 2007; Salas y Obeysekera, 2014).
- No estacionariedad debida a registros cortos: Como se ha comentado anteriormente, en los registros cortos la causa más probable de no estacionariedad es la variabilidad climática. Los registros cortos son muestras con menos de 30 años (Guttman, 1989; IH, 1999; Kundzewicz y Robson, 2000; Kundzewicz y Robson, 2004; Dixon et al, 2006). En el caso de confirmar la variabilidad climática como la causa de no estacionariedad se recomienda realizar el análisis considerando covariables relacionadas con dichos mecanismos climáticos.

Para el análisis del cambio climático se sugiere que al menos se tengan 50 años de registros, (Kundzewicz y Robson, 2000), pero incluso estos no puedan ser suficientes (Chiew y McMahon, 1993). Otra de las consideraciones que debe tener en cuenta para el análisis de cambio climático es que se deben elegir zonas de estudio que hayan sido poco intervenidas. Los cambios en el clima detectados en cuencas que han sufrido cambios antropogénicos no necesariamente pueden relacionarse con el cambio climático (IH, 1999; Kundzewicz y Robson, 2004; Kundzewicz et al, 2005)

- No estacionariedad ocasionada por causas no obvias: Si no es posible identificar las causas que ocasionan la no estacionariedad, los registros pueden ser empleados en el análisis de frecuencia, pero haciendo las respectivas salvedades en los resultados obtenidos.

3.3.3 Análisis de frecuencia no estacionario considerando covariables

Aun cuando es difícil detectar la influencia del cambio climático global en los eventos extremos hidrológicos tales como los gastos máximos, es común encontrar no estacionariedad en las series de gastos máximos ocasionadas por la variabilidad climática, cambios antropogénicos, entre otros factores (Villarini et al, 2009).

Varios autores han propuesto el análisis de frecuencia incorporando covariables para realizar el análisis de frecuencia en series no estacionarias (Smith, 1989; Coles, 2001; Katz et al, 2002; El Adlouni et al, (2007). Coles, (2001) proporciona una descripción general del uso de las covariables en la modelación de eventos extremos. El Adlouni et al, (2007) muestran el uso de las covariables en la modelación de una serie de precipitaciones máximas anuales en la estación Randsburg en California, dicha estación es fuertemente afectada por el fenómeno de El Niño. Ellos emplean el índice de oscilación del sur (SOI) como covariable. Katz et al, (2002) presentan una aplicación en el análisis de gastos máximos empleando como covariable un índice del fenómeno del ENSO.

El objetivo del análisis de frecuencia es identificar, a partir de varias pruebas de bondad de ajuste, la función de distribución de probabilidad que describe los registros históricos y obtener sus parámetros a partir de la información de la muestra por métodos como el de

momentos o el de máxima verosimilitud. Por ejemplo, la función de distribución de probabilidad Gumbel en su condición estacionaria tiene la forma:

$$F(x) = \exp \left[-\exp \left(-\frac{(x - \mu)}{\sigma} \right) \right] \quad (3-1)$$

Donde $\mu \in \mathbb{R}$ y $\sigma > 0$. En el caso no estacionario los parámetros de la distribución Gumbel son obtenidos como una función de las covariables ya sea el tiempo o un índice del ENSO: (μ_t, σ_t) . Para asegurar un valor positivo para el parámetro de escala, se usa una transformación tal que $\sigma_t = \ln \sigma_t$ cuando estimamos los parámetros (Coles, 2001). Así, por ejemplo para el tiempo $\mu_t = \mu_0 + \mu_1 t$ y $\ln \sigma_t = \sigma_0 + \sigma_1 t$.

Generalmente para la estimación de los parámetros se emplea el método de máxima verosimilitud (Coles, 2001; Strupczewski et al, 2001; Katz et al, 2002; Zhang, 2005; El Adlouni et al, 2007, Katz, 2013). Un método simple para comparar la validez de un modelo M_1 contra otro modelo M_0 , tal que $M_0 \subset M_1$, es usar la desviación estadística definida por (Coles, 2001):

$$D = 2\{l_1(M_1) - l_0(M_0)\} \quad (3-2)$$

Donde $l_1(M_1)$ es la máxima verosimilitud del modelo M_1 y $l_0(M_0)$ es la máxima verosimilitud del modelo M_0 . Grandes valores de D indican que el modelo M_1 es más adecuado y explica mejor la variación de los datos que el modelo M_0 . El estadístico D se distribuye de acuerdo a la distribución Chi-cuadrado (χ^2). Por lo tanto el modelo M_1 será aceptado si $D > C_\alpha$, donde C_α es el cuantil $(1 - \alpha)$ de la distribución χ_k^2 (α nivel de significancia), y k es la diferencia entre el número de parámetros de los M_1 y M_0 .

3.4 CASO DE ESTUDIO: CUENCA ALTA Y MEDIA DEL RIO MAGDALENA-COLOMBIA

La cuenca alta y media del río Magdalena se ubica entre el flanco oriental de la cordillera Central y el flanco occidental de la cordillera Oriental del sistema montañoso de los Andes. La cuenca está conformada por áreas de los departamentos de Cauca, Huila, Tolima, Cundinamarca, Caldas, Antioquia, Boyacá, Santander y Cesar donde se ubican ciudades principales como Bogotá, Bucaramanga, Neiva, Ibagué, Tunja. Se concentra cerca del 40% de la población de Colombia (DANE, 2006) y su economía puede representar cerca del 40% del PIB nacional (DANE, 2012). Los principales ríos que drenan al río Magdalena en este tramo son: Suaza, la Plata, Páez, Saldaña, Sumapaz, Bogotá, La Miel, Nare, Carare, Sogamoso y Lebrija.

3.4.1 Caracterización

En la cuenca se encuentran dos de los sistemas hidroeléctricos más importantes de Colombia que son el sistema de Nare-Guatapé y el sistema del río Bogotá, además de las centrales hidroeléctricas Miel I, Betania Sogamoso e Hidroprado y varias centrales hidroeléctricas a filo de agua. Actualmente se encuentra en construcción la central hidroeléctrica de El Quimbo, además de varios proyectos a filo de agua ubicados en la cuenca del río Nare. La cuenca es una de las zonas con mayor potencial para el desarrollo de proyectos hidroeléctricos (UPME, 2011).

En la parte alta de la cuenca del río Chicamocha se ubica el distrito de riego del Alto Chicamocha-Firavitoba, distrito que cuenta con los embalses La Copa y La Playa que además ayudan a regular las inundaciones en la parte alta de la cuenca del río

Chicamocha y el distrito de riego del valle del río Ubaté que cuenta con el embalse El Hato que además es utilizado para el control de inundaciones en la zona. El ecosistema natural de la zona del altiplano cundiboyacense respecto a 1985 ha presentado una disminución del 75.2% (Armenteras y Rodríguez, 2007).

3.4.1.1 Cuenca alta

El régimen de lluvias desde el nacimiento del río hasta San Agustín es monomodal con periodos lluviosos entre los meses de marzo y septiembre alcanzando su valor máximo en el mes de junio. Las estaciones de gastos máximos de esta región, empleadas en el estudio, presentan un régimen monomodal hasta la confluencia del río Páez con el Magdalena con máximos que se presentan en los meses de junio o julio. Desde este punto hacia aguas abajo el régimen es bimodal, con dos temporadas lluviosas (abril-mayo y octubre-noviembre) y dos temporadas secas (diciembre-febrero y junio-agosto), como consecuencia del doble paso de la ZCIT sobre el territorio (Poveda, 2004).

3.4.1.2 Cuenca media

En este estudio solo se considerará el análisis hasta el sector La Gloria. Las lluvias y gastos presentan un régimen bimodal, con dos temporadas lluviosas (abril-mayo y octubre-noviembre) y dos temporadas secas (diciembre-febrero y junio-agosto), como consecuencia del doble paso de la ZCIT sobre el territorio (Poveda, 2004). Comparada con la cuenca alta, esta zona presenta una mayor precipitación media anual.

La zona de estudio es afectada por mecanismos de la variabilidad climática como el ENSO (Mesa y Poveda, 1996; Poveda, 2004; Poveda et al, 2011) y presenta una precipitación media anual de aproximadamente 1980 mm.

3.4.2 Análisis de homogeneidad

Antes de realizar el análisis de homogeneidad al conjunto de las series de registros de gastos máximos, se realizará un análisis a las estaciones que se ubican aguas abajo de embalses y/o cadenas de embalses y posteriormente, se verificará si para la zona de estudio es suficiente para los análisis una longitud mínima de los registros de 30 años, como lo propone la literatura citada. Todas las pruebas se aplicarán para un nivel de significancia del 5%.

3.4.2.1 Análisis de homogeneidad de estaciones ubicadas aguas abajo de embalses

Los análisis de homogeneidad realizados para las estaciones que se ubican aguas abajo de embalses empleando las pruebas señaladas en el numeral 3.2 se presentan a continuación. Se analizarán las series con mínimo 39 años de registro y se considerarán las estaciones para las cuales el tiempo de viaje desde el sitio de la presa hasta la estación es menor a 24 horas y aquellas en las cuales los embalses aguas arriba regulan al menos el 30% del caudal medio multianual.

- Análisis de independencia

Uno de los principales factores que ocasiona autocorrelación en las series de gastos máximos es la regulación ocasionada por embalses (Salas, 1991). Los resultados empleando la prueba de Anderson (Anderson, 1942) se presentan en la Figura 3.1.

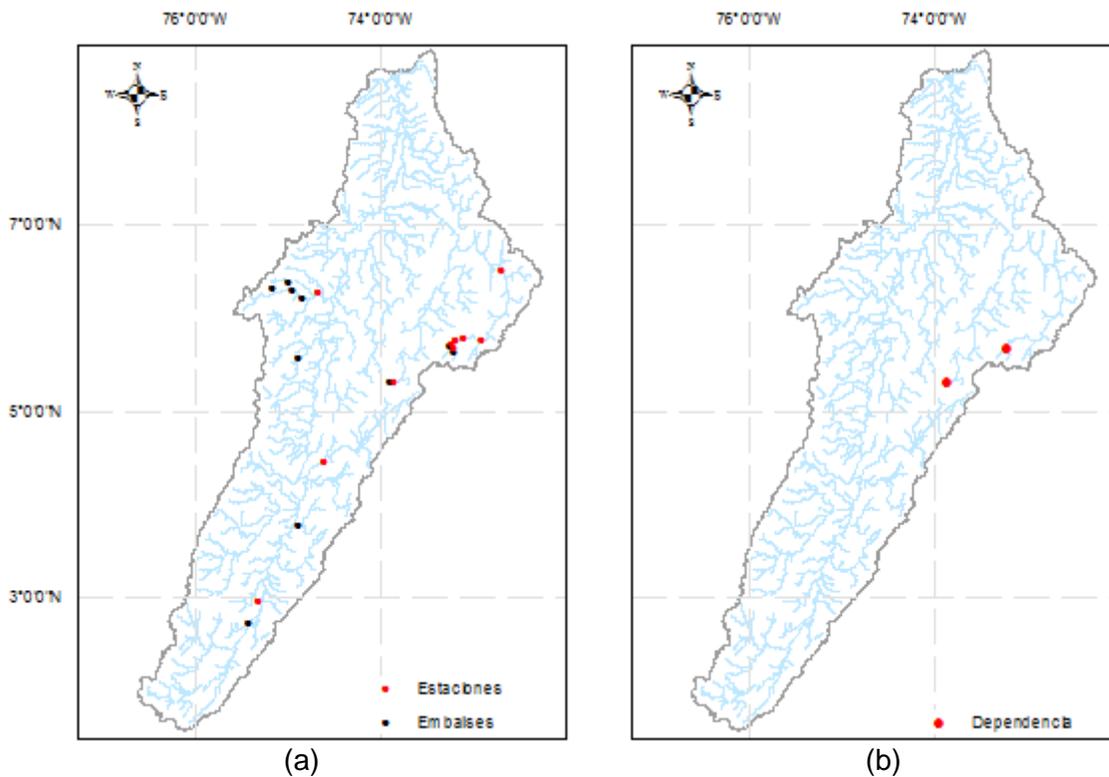


Figura 3.1 Localización de embalses y estaciones (a) y de estaciones dependientes (b).

En Colombia son pocos los embalses que tienen como propósito el control de inundaciones. Entre los que cumplen esta función que se encuentran dentro de la región de estudio están el embalse La Copa y La Playa del distrito de riego del alto Chicamocha (Incoder, 2013), el embalse El Hato del distrito de riego del valle del río Ubaté (CAR, 2005). Así, es probable que la estación 24037030 que se ubica aguas abajo del embalse La Copa y la estación 24017150 que se encuentra aguas abajo del embalse El Hato sean dependientes por efecto de la regulación de estos embalses.

- Análisis de cambio en la media

Los cambios en la media y/o la varianza en las series de gastos pueden ser ocasionados por efectos de regulación de embalses, modificaciones de los sitios de medición, cambios en las curvas de calibración, la variabilidad climática, entre otros factores (Salas, 1991; IH, 1999; Sveinsson et al, 2005; Salas y Obeysekera, 2014). En la Figura 3.2 se presentan los resultados del análisis del cambio en la media. Se aplicaron las 7 pruebas descritas en el numeral 3.2.2. La serie presenta un cambio en la media si al menos 4 pruebas indican que se presenta un cambio, igualmente se ha utilizado la prueba de Pettit para identificar el año donde se produjo el cambio.

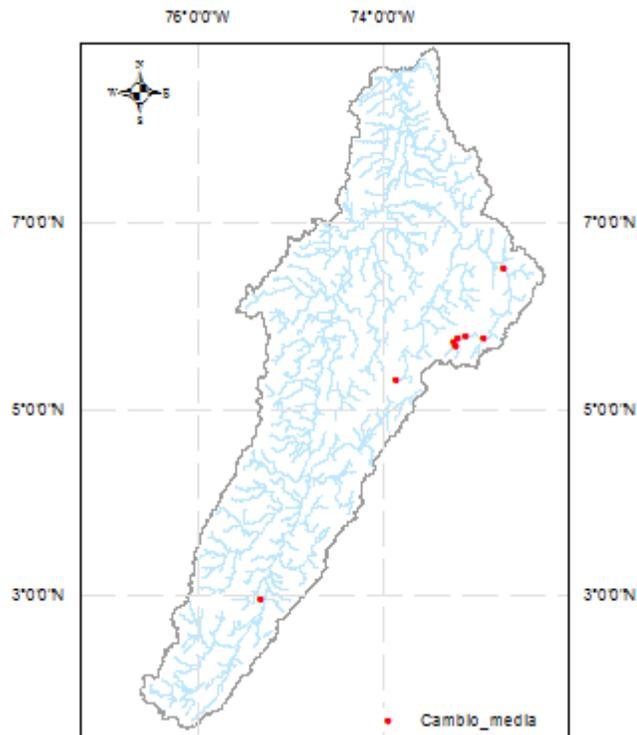


Figura 3.2 Localización de las estaciones con cambio en la media

Los resultados indican que en algunas de las zonas donde se ubican embalses se presentan cambios en la media, principalmente disminuciones en la media después del punto de cambio como se verá más adelante. Para verificar si el cambio en la media es posiblemente por la regulación de los embalses es necesario conocer el inicio de operación de los mismos o las fechas donde se han realizado realces de la cortina para aumentar su capacidad si es el caso. En la Figura 3.3 se presentan las estaciones que se ubican aguas abajo de los embalses La Copa, La Playa, El Hato y Betania. En cada una de las estaciones se señala (línea roja) el año de cambio en la media identificado con la prueba de Pettit.

La estación 24037030 se ubica aguas abajo del embalse La Copa y la disminución en la media a partir de la entrada en operación del embalse (1991) es evidente. Las estaciones 24037130, 24037320, 24037290 y 24037390 localizadas sobre el río Chicamocha se ubican aguas abajo de los embalses La Copa y La Playa del distrito de riego del Alto Chicamocha. En todas las estaciones se presenta una disminución en la media de los gastos máximos desde aproximadamente inicios de 1990, fecha que coincide con el inicio de la operación del embalse La Copa.

La estación 24037090 se ubica aguas abajo del embalse La Playa, en esta estación se presenta una disminución en el valor medio de los gastos máximos a partir del año 1982. Sin embargo el embalse la Playa inició operación a finales de la década de los 60. La estación 24017150 se encuentra aguas abajo del embalse El Hato, esta presenta una disminución en la media a partir de 1988, fecha que es muy próxima a la entrada en operación del embalse. Finalmente, la estación 21097070 presenta una disminución en la media a partir de 1989, fecha que coincide con la entrada en operación del embalse de la central hidroeléctrica de Betania que se ubica aguas arriba de la estación.

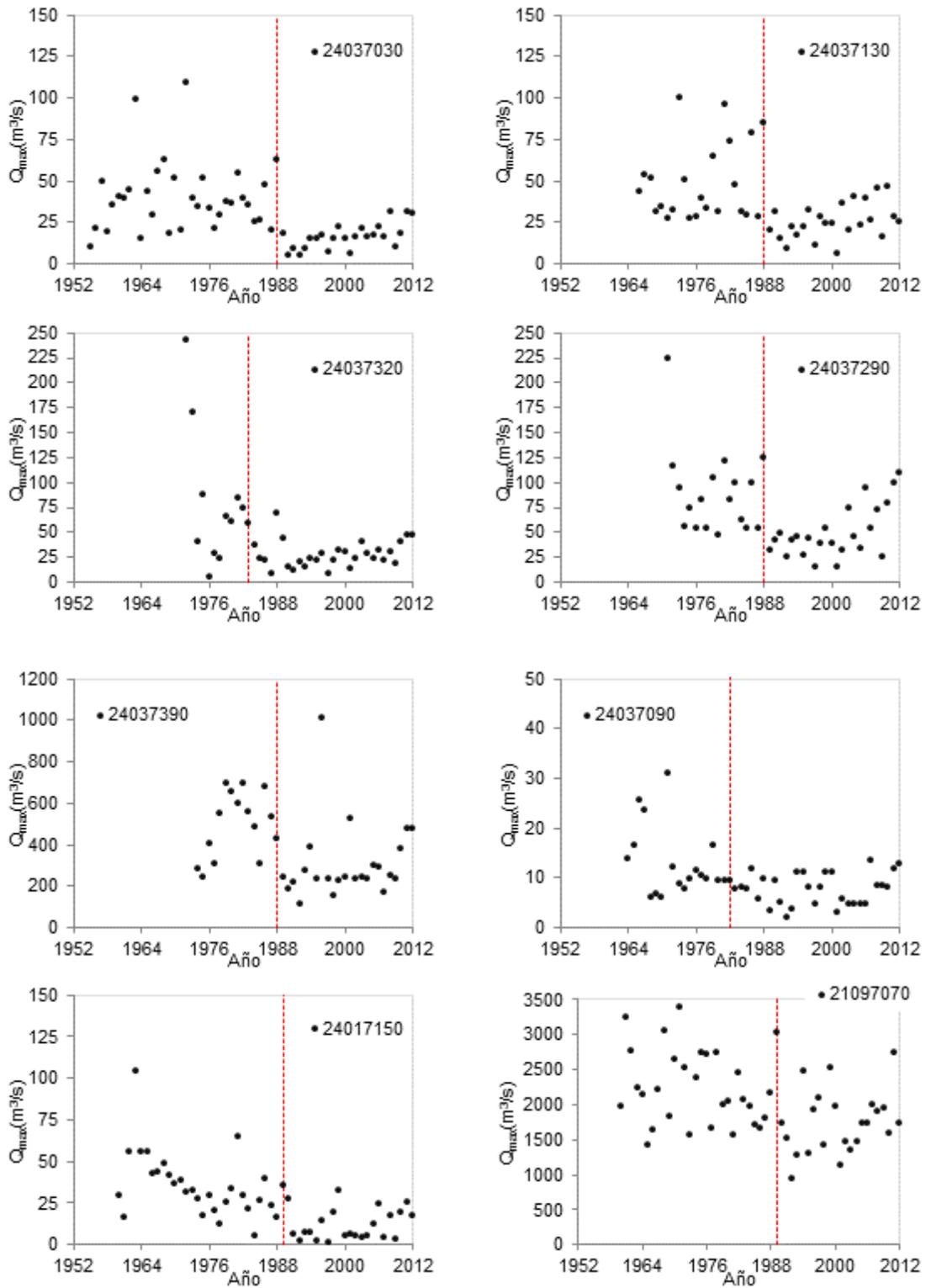


Figura 3.3 Estaciones de gastos máximos ubicados aguas abajo de los embalses La Copa, La Playa, El Hato y Betania

Si bien se ha encontrado que es posible que las disminuciones en la media de las estaciones presentadas en la Figura 3.3 sean por efecto de la regulación de embalses, se deben analizar otros factores como la variabilidad climática, sobre todo en aquellas estaciones ubicadas aguas abajo de embalses utilizados solo para la generación hidroeléctrica.

- Cambio en la varianza

Al igual que el cambio en la media, los cambios en la varianza pueden presentarse por efecto de la regulación de embalses, la variabilidad climática y cambios antropogénicos. En la Figura 3.4 se presentan ubicadas las estaciones que presentan cambios en la varianza, dicho cambio ha sido estimado empleando la prueba de Fisher.

De las estaciones ubicadas aguas abajo de los embalses La Copa, La Playa, El Hato y Betania, las estaciones 24037390, y la 21097070 no presentan cambio en la varianza. Se encuentra que la estación 21207960 que se ubica en el río Bogotá, aguas abajo de la ciudad de Bogotá y el sistema de embalses del río Bogotá, presenta cambio en la varianza. Igualmente la estación 23087210 que se ubica sobre el río Nare, aguas abajo de la cadena de embalses Nare-Guatapé, presenta cambio en la varianza.

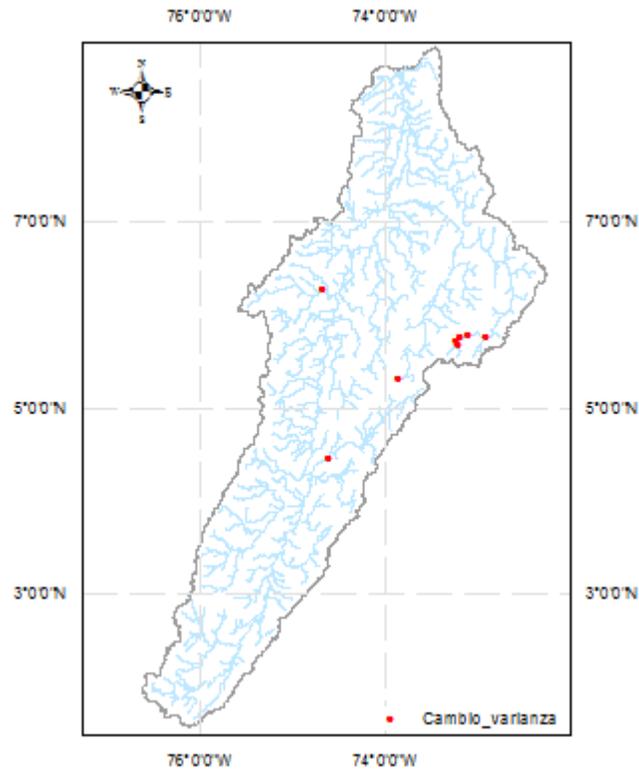


Figura 3.4 Localización de las estaciones que presentan cambio en la varianza

- Análisis de tendencias

Las tendencias en las series de registros de gastos máximos pueden ser ocasionadas por cambios antropogénicos, errores en la medición y sistematización de los datos, y en registros cortos se pueden generar tendencias aparentes producto de la variabilidad climática (Salas et al, 1980; IH, 1999; Kundzewicz y Robson, 2004). La práctica común

para la realización del análisis de tendencias es hacerlo separadamente y no considerar los resultados obtenidos en el análisis de cambio en la media (separar la muestra en dos sub-series antes y después del cambio en la media) (Villarini et al, 2009). Este enfoque puede conducir a resultados engañosos (Villarini et al, 2009; Villarini et al, 2010). Si por ejemplo, se realizará un análisis de tendencias a la serie de registros completa de la estación 24037030 (Véase Figura 3.3) se encontraría que la estación presenta una tendencia decreciente, es decir, que los gastos máximos tenderían a disminuir en el tiempo. Esta situación puede ser errónea, ya que la disminución en los gastos máximos probablemente ha sido generada por la regulación del embalse La Copa, y este con el tiempo tenderá a disminuir su capacidad de regulación y por lo tanto, se irán incrementando los gastos máximos.

Por lo tanto, el análisis de tendencia en este estudio se realizará considerando el punto de cambio en la media, estimado con la prueba de Pettit, en las series que presentan cambio en la media y para los registros completos en aquellas estaciones que no presentan cambio en la media. El análisis se llevará a cabo empleando las pruebas no paramétricas de Mann-Kendall y la de Spearman, aunque sus potencias en detectar tendencias es similar (Yue et al, 2002).

En la Figura 3.5 se presenta la localización de las estaciones que presentan tendencias antes o después del cambio en la media o en todo su registro. Una de las limitaciones del estudio es que al dividir la muestra, en la mayoría de los casos se cuentan con muestras cortas para realizar los análisis, situación que puede afectar la confiabilidad de los resultados.



Figura 3.5 Localización de las estaciones que presentan tendencias

Las estaciones 24017150 y 24037030 presentan correlación serial, sin embargo, al realizar el análisis de tendencias para los periodos antes y después del cambio en la media no presentaron dependencias.

Las estaciones 24037030, 24037130, 24037290 y 24037390 que se ubican aguas abajo de los embalses La Copa y La Playa en la parte alta del río Chicamocha presentan tendencias crecientes después del cambio en la media. Esta situación reafirma la necesidad de realizar los análisis de tendencias y cambios en la media conjuntamente. Si se hubieran analizado los registros completos probablemente se habría encontrado que las series presentaban tendencias decrecientes, cuando la situación es totalmente contraria. Probablemente la tendencia creciente después del cambio en la media en estas estaciones se presente por la pérdida de la capacidad de regulación de los embalses, sin embargo, de acuerdo al Incoder, (2013), se espera que con la compra de unos terrenos adicionales se puede elevar el nivel de operación del embalse La Copa y por lo tanto se aumente la capacidad de regulación del sistema.

La estación 24017150 que se ubica aguas abajo del embalse El Hato, presenta una tendencia decreciente antes del cambio en la media. En la Figura 3.6 se presentan algunas de las estaciones que presentan tendencias.

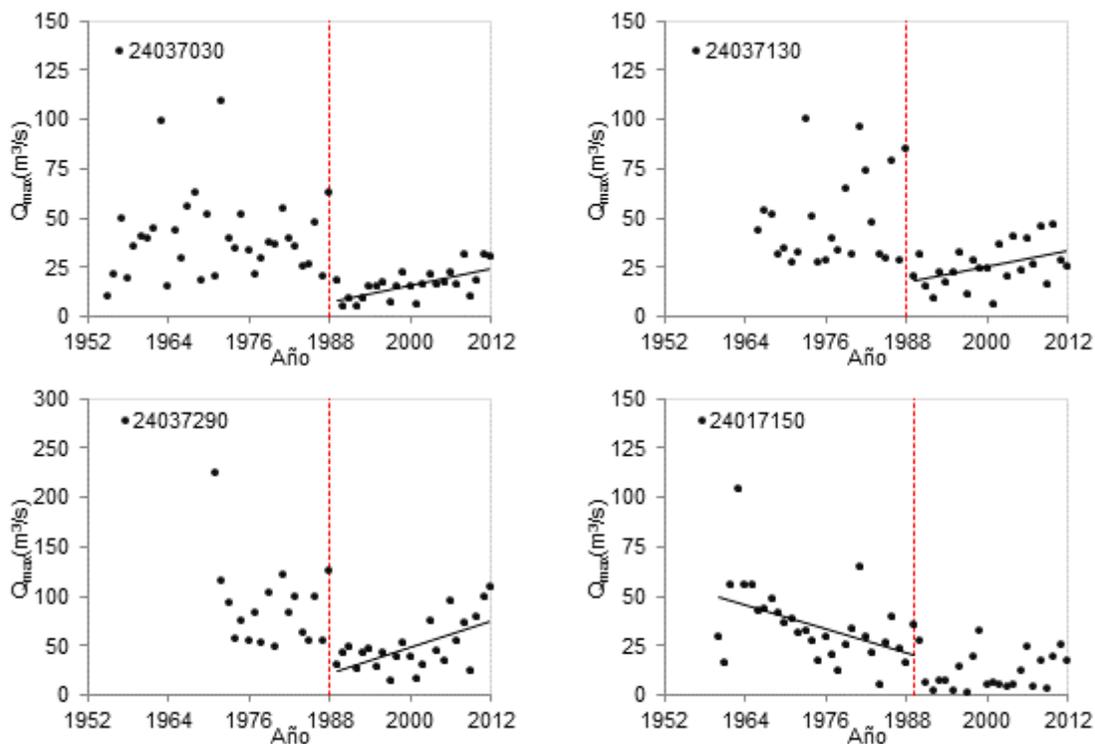


Figura 3.6 Ejemplo de estaciones que presentaron tendencias

Así, las estaciones analizadas que se ubican aguas abajo de embalses no son homogéneas y no se considerarán en los análisis. Sin embargo, cuando se realiza análisis regional de frecuencia es posible utilizar la información de las estaciones hasta antes de la entrada en operación de los embalses, si se cuenta con ella.

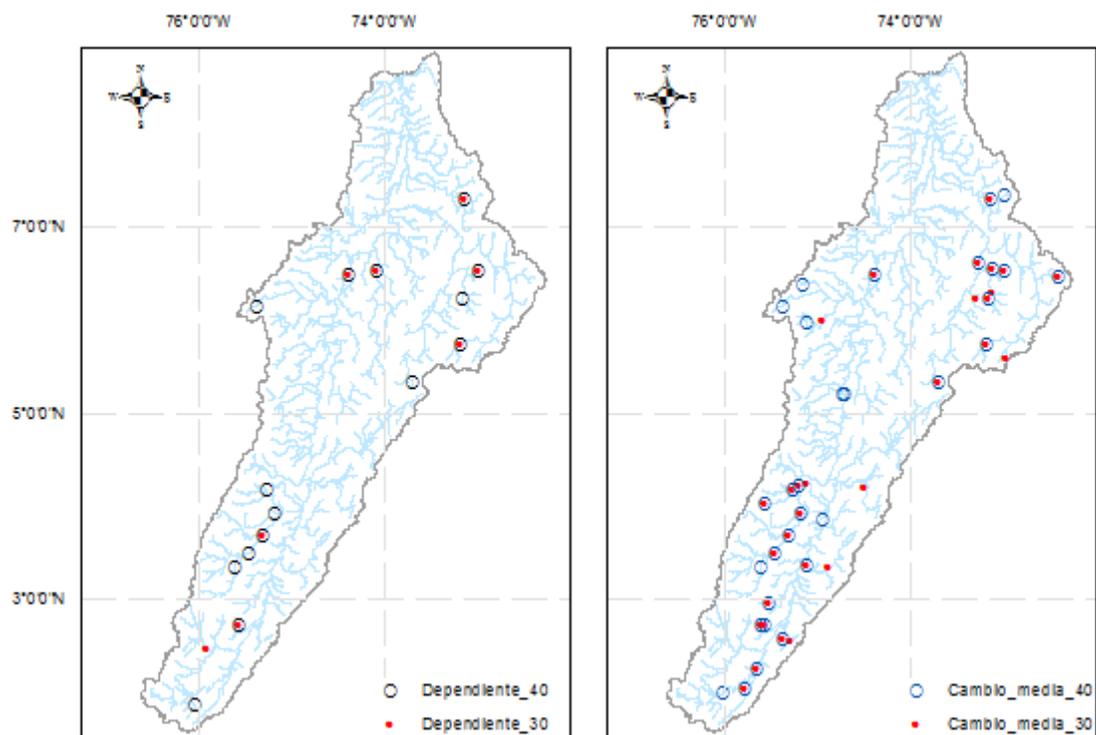
3.4.2.2 Longitud mínima de los registros de gastos máximos

Los registros cortos son muestras con menos de 30 años (Guttman, 1989; IH, 1999; Kundzewicz y Robson, 2000; Kundzewicz y Robson, 2004; Dixon et al, 2006), sin embargo, el análisis de la longitud mínima de los registros aún se encuentra en estudio. Por ejemplo, Das, (2011) y Das y Cunnane, (2012) recomiendan que el análisis de frecuencia en el sitio en zonas con una moderada o alta heterogeneidad se haga solo si se cuentan con al menos 50 años de registros. Por lo tanto, en este trabajo se evaluará si 30 años de registros son recomendables para realizar el análisis de homogeneidad y por ende para los análisis de frecuencia, que es el objetivo de este trabajo.

De las 135 estaciones ubicadas en la región de estudio 69 cuentan con 40 o más años de registros, 7 de estas estaciones se ubican aguas abajo de embalses, así, se cuentan con 62 estaciones con 40 o más años de registros para el análisis. Se plantea realizar el análisis de homogeneidad (siguiendo la metodología del numeral 3.4.2.1) considerando 30 años de registros (1973-2002) y se comparará con el análisis con 40 años de registros (1973-2012). Si 30 años de registro son suficientes no deben presentarse cambios en los análisis, es decir, que si una estación, por ejemplo, presenta cambios (en la media o la varianza) con los primeros 30 años del registro, la adición de información no deberá modificar dicho resultado. En la Figura 3.7 se presentan los resultados del análisis de

homogeneidad considerando los dos tamaños de muestra. Solo se presenta la localización de las estaciones que presentan alguna no homogeneidad.

Tanto para el análisis de independencia como para el análisis de cambio en la media, en la varianza y el análisis de tendencia pareciera que 30 años de registros no son suficientes para el análisis toda vez que al incrementar la longitud de los registros algunas estaciones ya no presentan el cambio o evidencian el mismo (véase Figura 3.7). Así, para la zona de estudio el análisis de homogeneidad con series de mínimo 30 años de registros no es adecuado toda vez que se puede sobre-estimar o subestimar el número de estaciones no homogéneas. Por lo tanto, se recomienda para la zona de estudio que el análisis de homogeneidad y de frecuencia se realicen con estaciones que tengan mínimo 40 años de registros.



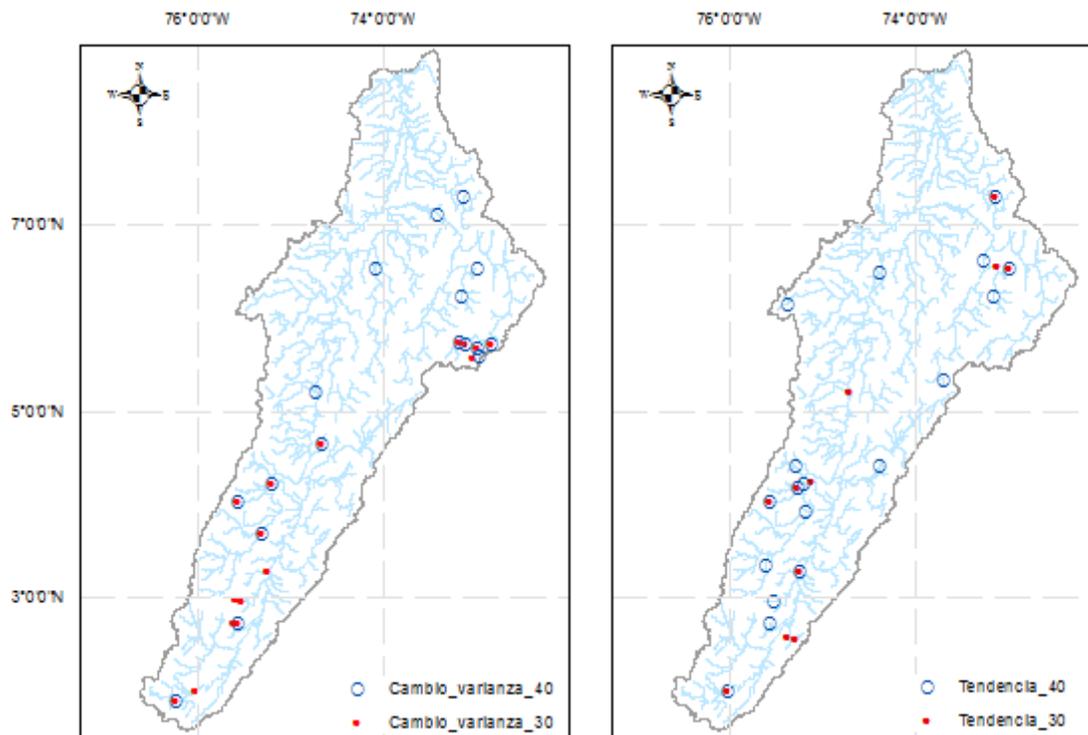
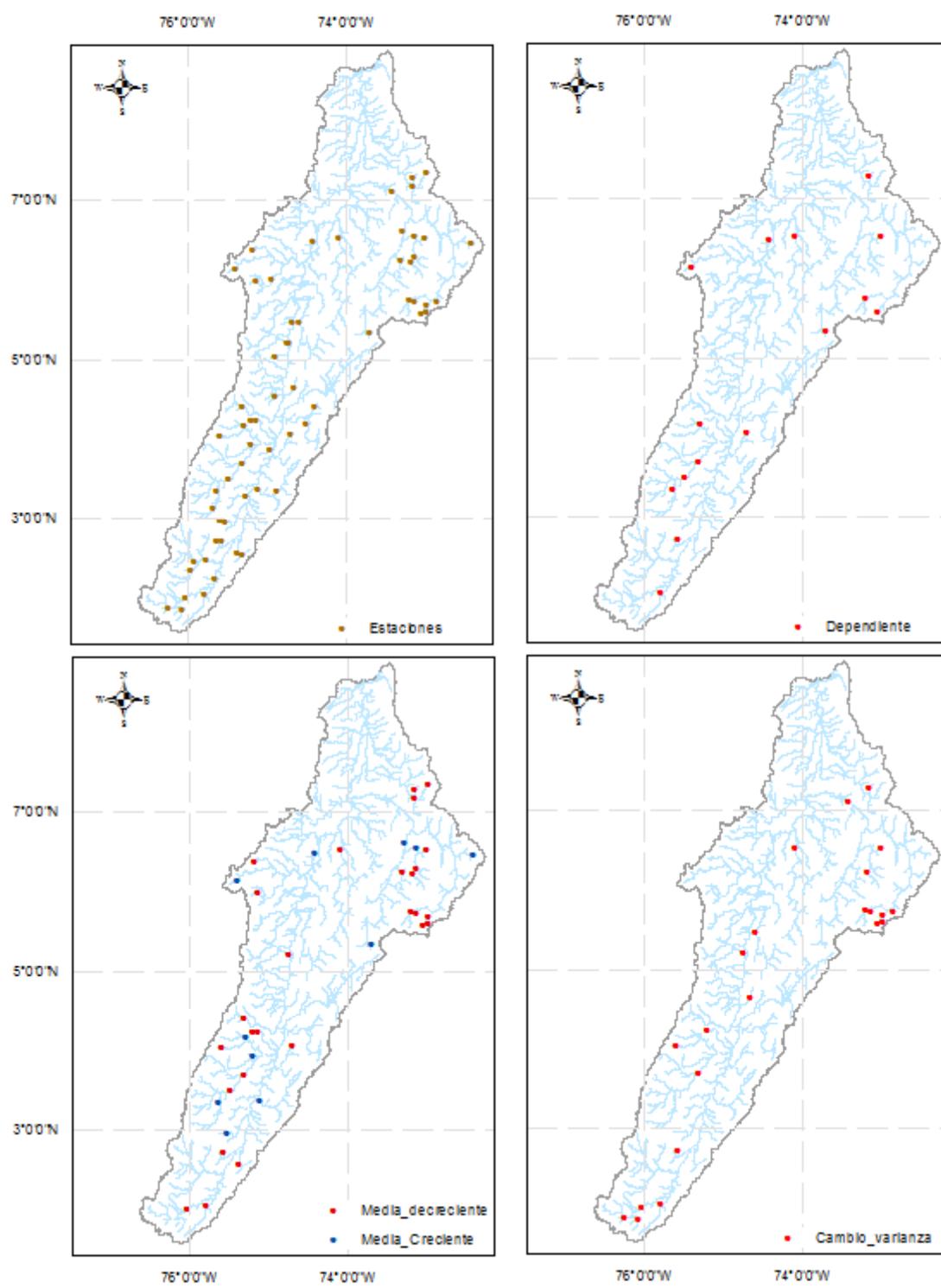


Figura 3.7 Análisis de homogeneidad para dos tamaños de muestra

3.4.2.3 Análisis de homogeneidad de las estaciones con mínimo 40 años de registro

En la Figura 3.8 se presentan los resultados del análisis de homogeneidad de las estaciones que tienen al menos 40 años de registro y que no se encuentran afectadas por la regulación de embalses. Las estaciones 21017020 y la 21197110 no presentaron cambios en la media y mostraron tendencias. Las estaciones 22027010, 23017030 y 24017610 presentan cambio en la media y muestran tendencia antes y después del cambio en la media. Las demás estaciones que muestran tendencias solo la presentan para el periodo antes o después del cambio en la media. Por regiones, las zonas con mayor número de estaciones no homogéneas son las ubicadas en el departamento de Santander y la zona del altiplano cundiboyacense. 45 (72.6%) estaciones de la zona de estudio son no homogéneas.

De las estaciones no estacionarias un 33% presentan correlación serial, un 84% de las estaciones no homogéneas presentan cambio en la media, de éstas el 71% presentan una media decreciente, un 50% presentan cambio en la varianza. Si bien un 51% de las estaciones no estacionarias presentan tendencia, solo 2 estaciones presentan tendencia sin presentar cambio en la media, 3 estaciones presentan tendencia antes y después del cambio en la media y el resto solo presentan tendencia antes o después del punto de cambio en la media.



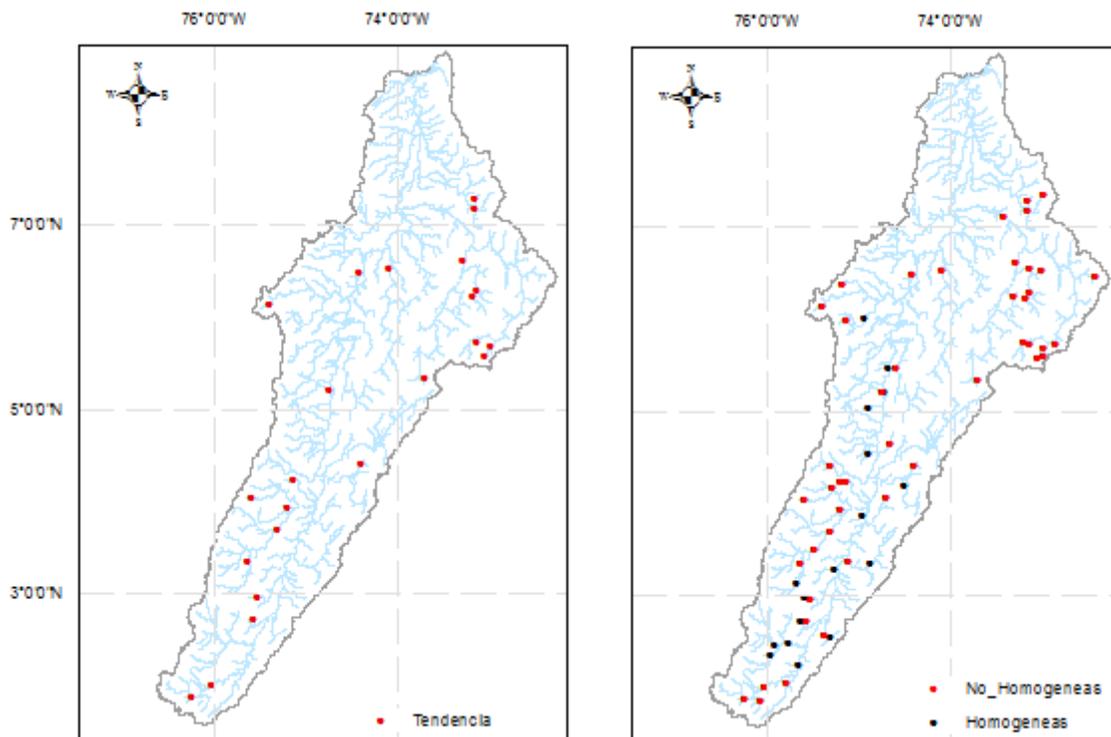


Figura 3.8 Análisis de homogeneidad en estaciones con al menos 40 años de registro

3.4.3 Análisis de las posibles causas de no estacionariedad de las estaciones

Es importante entender las causas de la no estacionariedad de las series de registros toda vez que las implicaciones para el análisis de frecuencia pueden ser distintas. La identificación de dichas causas permitirá definir si es adecuado realizar un análisis estacionario, o si la mejor decisión es realizar el análisis de frecuencia no estacionario.

Las principales causas de no estacionariedad en los registros de gastos se pueden originar por: 1) Problemas en los registros, 2) cambios en la cuenca y 3) la variación del clima.

3.4.3.1 Problemas relacionados con los registros de los gastos

Esta se puede considerar como la principal fuente de incertidumbre de los análisis realizados en este estudio, toda vez que no fue posible contar con la hoja de vida de las estaciones y poder detectar posibles errores en los registros. Un estudio contratado por el IDEAM, (2011) señala algunos inconvenientes que se presentan con la información generada por el IDEAM:

- No hay disponibilidad de las hojas de vida de las estaciones. Los acontecimientos que explican los datos reportados por una estación, como por ejemplo un daño en un instrumento, el robo de algún componente, la ocurrencia de un evento inusitado, etc., son el soporte de interpretación de los datos generados por una estación. La falta de acceso a esta información genera desconfianza en los datos suministrados.
- En términos de la calidad de los datos se observan problemas que empiezan desde su medición. El nivel de obsolescencia de la red es alto y el mantenimiento que se hace de las estaciones no es el adecuado debido a la baja frecuencia del mismo.

Los registros tomados en campo algunas veces no se entienden. Adicionalmente, problemas de orden público generan dificultades en la labor de las personas que operan las estaciones, muchas veces se presenta imposibilidad de tomar el dato por períodos de tiempo considerables.

3.4.3.2 Variabilidad climática

El efecto del ENSO (El Niño-oscilación del sur) en las variaciones interanuales de la precipitación y los gastos en una región ha sido el foco de una variedad de investigaciones. Sin embargo, la conexión potencial entre los efectos del ENSO sobre la variabilidad interanual del clima y la frecuencia o severidad de los eventos de gastos extremos es difícil de documentar (Hirschboeck et al, 2000). Hay solo unas pocas áreas del mundo donde los gastos máximos y la precipitación se pueden relacionar con una sola variable climática o un solo mecanismo de generación, donde puede aparecer una correlación a escala interanual. En muchas regiones, la generación de los gastos máximos es afectada por múltiples mecanismos hidroclimáticos, situación que complica evidenciar la relación entre los gastos máximos y la variabilidad interanual generada por una sola componente climática regional. La relación entre el ENSO y los gastos máximos es geográficamente variable, y en aquellas áreas donde puede ser detectada, puede que no sea consistente o estadísticamente significativa (Hirschboeck et al, 2000). Ishak et al (2013) evaluaron la no estacionariedad de los gastos máximos anuales en Australia, encontrando que el 26% de las estaciones analizadas presentan tendencias decrecientes y solo un 6% presentan tendencias crecientes. Ellos sugirieron que muchas de estas tendencias podrían ser explicadas por la variabilidad climática.

Mecanismos de baja frecuencia como el ENSO tienen una fuerte influencia sobre la hidrología de Colombia (Poveda y Mesa, 1996; Poveda y Mesa; 1997; Poveda, 2004; Poveda et al, 2011). Para la mayor parte del territorio colombiano se ha encontrado que existe una diferencia relevante entre los caudales de La Niña respecto a El Niño y al No ENSO (Álvarez y Poveda, 2006, Poveda et al, 2011), además la zona central del país es afectada por el fenómeno ENSO en su fase fría (La Niña), aumentando los caudales máximos. Los cambios detectados en la media para las series anuales de caudales máximos, en su mayoría están asociados con la fase cálida del ENSO (El Niño). Esto se debe al gran impacto que tiene en las series colombianas y a que su frecuencia es mayor que la fase fría (La Niña) (Poveda y Álvarez, 2012).

Para evaluar si en la zona hay un efecto de la variabilidad climática en el comportamiento de los gastos máximos se emplearán para su evaluación el ajuste de distribuciones de probabilidad no estacionarias considerando covariables como el tiempo y varios índices del ENSO.

- Análisis considerando covariables

Para evaluar si la no estacionariedad de las estaciones de gastos máximos de la zona de estudio es ocasionada por la variabilidad climática se propone realizar un análisis considerando covariables. Dichos análisis generalmente se han empleado para el análisis de frecuencia de eventos extremos (Coles, 2001; Katz et al, 2002; El Adlouni et al, 2007), sin embargo, en este estudio se propone para identificar la posible causa de no estacionariedad de las series de gastos máximos.

Se propone emplear la distribución Gumbel, la distribución general de valor extremo (GEV) y la distribución Gamma de dos parámetros, que en su condición estacionaria tienen la forma:

Para la Gumbel:

$$F(x) = \exp \left[-\exp \left(-\frac{(x - \mu)}{\sigma} \right) \right] \quad (3-3)$$

Para la GEV:

$$F(x) = \exp \left\{ - \left[1 - \left(\frac{x - \nu}{\alpha} \right) \beta \right]^{1/\beta} \right\} \quad (3-4)$$

Para la Gamma tipo II:

$$F(x) = \int_0^x \frac{x^{\beta-1} e^{-x/\alpha}}{\alpha^\beta \Gamma(\beta)} dx \quad (3-5)$$

De acuerdo a Poveda y Mesa, (1996), el ENSO no solo afecta la media de los gastos medios mensuales sino también su desviación típica, además como lo cambios sufridos en la cuenca causan variaciones en estos dos parámetros, se considerarán modelos no estacionarios tanto en el parámetro de escala como en el de localización. Así, asumiendo que los parámetros de las distribuciones son funciones lineales de la covariable tiempo o el ENSO, la función de distribución de probabilidad no estacionaria será:

Para la Gumbel:

$$F(x) = \exp \left[-\exp \left(-\frac{(x - (\mu_0 + \mu_1 Cov(t)))}{\exp(\sigma_0 + \sigma_1 Cov(t))} \right) \right] \quad (3-6)$$

Para la GEV:

$$F(x) = \exp \left\{ - \left[1 - \left(\frac{x - (\nu_0 + \nu_1 Cov(t))}{\exp(\alpha_0 + \alpha_1 Cov(t))} \right) \beta \right]^{1/\beta} \right\} \quad (3-7)$$

Para la Gamma de dos parámetros:

$$F(x) = \int_0^x \frac{x^{(\exp(\beta_0 + \beta_1 Cov(t)) - 1) e^{-x/\exp(\alpha_0 + \alpha_1 Cov(t))}}}{\exp(\alpha_0 + \alpha_1 Cov(t))^{\exp(\beta_0 + \beta_1 Cov(t))} \Gamma(\exp(\beta_0 + \beta_1 Cov(t)))} dx \quad (3-8)$$

Donde $Cov(t)$ representa la covariable ya sea el tiempo o el ENSO. Para evaluar la influencia del ENSO en la serie de gastos máximos se emplearon el índice multivariado de el Niño (MEI), el índice El Niño Oscilacion de Sur (SOI) y el índice Oceánico de El Niño (ONI) de los cuales se tomaron los registros promedios del año hidrológico (Poveda y Álvarez, 2012), para los años 1951 a 2012. Los parámetros de las funciones de distribución de probabilidad fueron obtenidos empleando el método de máxima verosimilitud utilizando el algoritmo de optimización no lineal Gradiente Reducido Generalizado - GRG (Lasdon et al, 1975).

Para evaluar si es más adecuado el análisis no estacionario, es decir, si la covariable tiene una influencia en la serie de registros de la estación o describe mejor su comportamiento, se empleará la desviación estadística D (Coles, 2001) definida en el

numeral 3.3.3. Se considerará como valor crítico el valor de la función Chi-cuadrado para un nivel de significancia del 10% y dos grados de libertad, es decir 4.60. Las estaciones en las que se considera que el análisis estacionario es más adecuado serán aquellas en las cuales tanto la covariable del ENSO como la covariable el tiempo presentan una desviación estadística mayor al valor crítico.

En la Figura 3.9 se presenta la localización de las estaciones no homogéneas. Con relleno de color rojo se presentan las estaciones en las cuales los modelos no estacionarios empleando la covariable tiempo, para las tres distribuciones, es más adecuado que el modelo estacionario, es decir, aquellas en las cuales la desviación estadística es mayor al valor crítico (4.60). De las 45 estaciones que presentan no estacionariedad en 36 (80%) se encuentra que ésta puede ser explicada por la covariable tiempo. Si bien se podría concluir que la estimación de los eventos de diseño en la zona de estudio es recomendable hacerla empleando análisis no estacionario con la covariable el tiempo, se estaría incurriendo en un error toda vez que este análisis asume que la tendencia se presenta en todo el registro y no se está teniendo en cuenta el cambio en la media, por lo que los resultados obtenidos podrían ser engañosos.

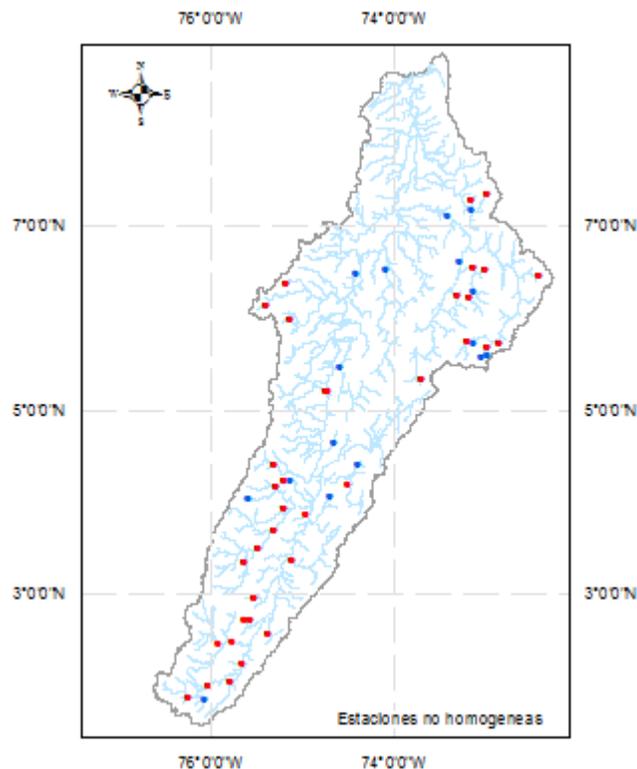


Figura 3.9 Localización de estaciones no homogéneas. Con relleno rojo las estaciones en las cuales el modelo no estacionario empleando la covariable tiempo es más adecuado.

En la Figura 3.10 se presenta la localización de las estaciones no homogéneas, se presenta además, con relleno azul, aquellas en las cuales los modelos no estacionarios empleando como covariable los índices del ENSO para las tres distribuciones es más adecuado, es decir, en las que la desviación estadística es mayor que el valor crítico (4.60). De las 45 estaciones no estacionarias en 23 (51%) es más adecuado emplear los

modelos no estacionarios empleando como covariables índices del ENSO. Cuando se emplea el índice MEI como covariable se obtiene un mayor número de estaciones en las cuales la desviación estadística es superior al valor crítico.

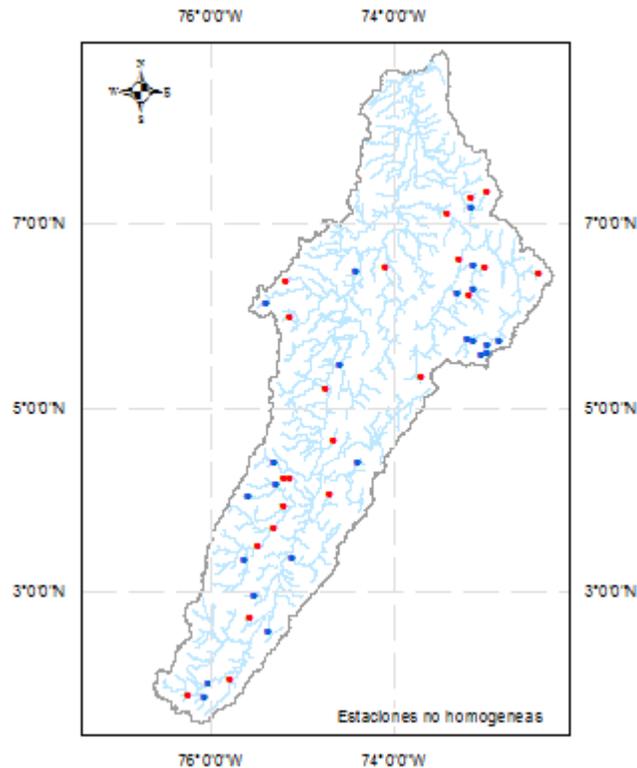


Figura 3.10 Localización de estaciones no homogéneas. Con relleno azul las estaciones en las cuales el modelo no estacionario empleando la covariable ENSO es más adecuado.

Sin embargo, como se comentó anteriormente, se considerará que la variabilidad climática puede explicar la no estacionariedad de las estaciones solo en aquellas en las cuales tanto los modelos no estacionarios con covariable tiempo y con covariable índices del ENSO es más adecuado, donde la desviación estadística es mayor que el valor crítico. En la Figura 3.11 se presenta la localización de las estaciones no homogéneas para las cuales la desviación estadística de los modelos estacionario y no estacionario tanto para la covariable tiempo como para la covariable índices del ENSO dio mayor que el valor crítico (4.60).

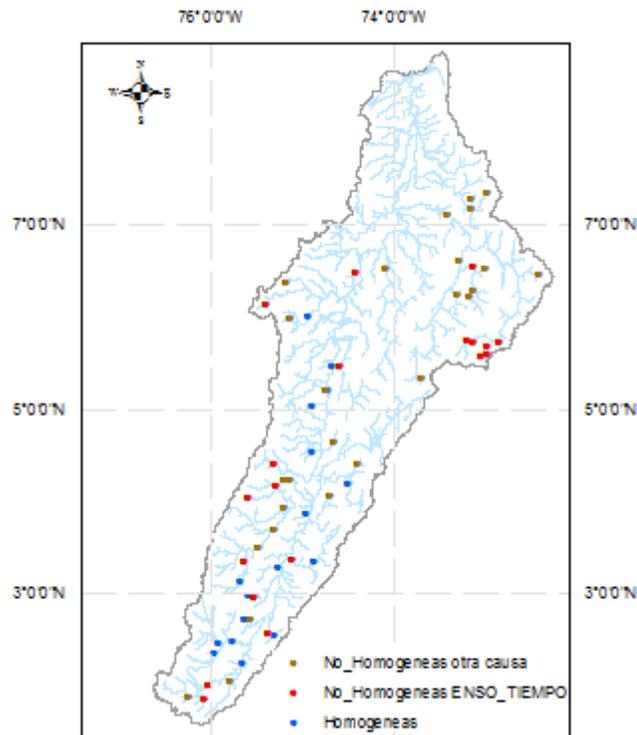


Figura 3.11 Localización de las estaciones donde tanto la covariable índices del ENSO y la covariable tiempo describen mejor el comportamiento de los registros.

De las 44 estaciones que son no homogéneas, en 19 (42%) la desviación estadística de los modelos con covariables (MEI y tiempo) indican que son más adecuados, es decir, que la covariable describe mejor el comportamiento de los datos y por lo tanto, la causa de la no estacionariedad de estas estaciones puede ser la variabilidad climática.

La causa de no estacionariedad en las demás estaciones se puede presentar por la misma variabilidad climática toda vez que los modelos propuestos no sean los más apropiados, además se debe recordar que se están empleando valores medios de los índices del ENSO, análisis que si bien permite obtener resultados robustos puede no ser el más adecuado. Otras causas que pueden originar la no estacionariedad de las estaciones pueden ser problemas en los registros u otras causas como cambios antropogénicos. Así, como no se cuenta con información de las cuencas (cambios de usos del suelo, deforestación, etc) y las hojas de vida de las estaciones no es posible establecer la causa de no estacionariedad de estas estaciones.

En vista de que en el 42% de las estaciones analizadas los gastos máximos provienen de más de un mecanismo, se puede afirmar que el análisis de frecuencia unimodal no es adecuado, ya que pueden existir algunos valores de las muestras que no pueden ser descritos por las funciones de probabilidad de una sola población (Gingras y Adamowski, 1992). El desconocimiento de estos procesos físicos en el análisis de frecuencia convencional es probablemente la principal razón para que las distribuciones de frecuencia comúnmente usadas a menudo no logran dar un ajuste aceptable a los datos históricos (Mtiraoui, 2004). Esta situación implica, además, que en la zona de estudio no solo en muestras de registros cortos (menos de 30 años) se puede presentar no estacionariedad por efecto de la variabilidad climática.

4 DISTRIBUCIONES DE PROBABILIDAD

4.1 INTRODUCCIÓN

El análisis de frecuencia ha sido ampliamente empleado en el diseño de distintas obras hidráulicas y en la estimación de gastos de estiaje asociados a diferentes periodos de retorno para la modelación de calidad de agua en ríos. Un gran número de distribuciones de probabilidad han sido empleadas basados en las diferentes características de los datos hidrológicos a ser analizados, (Zhang, 2005). Los métodos comúnmente usados para la estimación de parámetros son: el método de los momentos, los momentos de probabilidad pesada (Hosking y Wallis, 1997), la técnica de los mínimos cuadrados y el método de máxima verosimilitud. Este último, ofrece una gran versatilidad para la estimación de parámetros en registros que presentan un comportamiento no estacionario (Katz et al, 2002).

4.2 FACTORES QUE AFECTAN LA ELECCIÓN DE LA DISTRIBUCIÓN EN GASTOS MÁXIMOS

Algunos de los principales factores que intervienen y afectan la elección de la distribución son (Cunnane, 1985):

- Método de estimación de los parámetros
- Valores atípicos o outliers presentes en los datos. La presencia de estos debe ser analizada con juicio ya que pueden ser originados por problemas en los registros o dependiendo de la zona, por mecanismos como tormentas tropicales, el ENSO u otro.
- Efecto de los datos históricos. La inclusión de eventos de inundación históricos y la exclusión de eventos de inundación históricos de menor magnitud puede inducir sesgo en la forma de la distribución elegida.
- Efecto de la transformación. Para el uso de la distribución normal y otras, es necesario aplicar transformaciones a los datos para que los mismos tengan un bajo coeficiente de asimetría y satisfagan otras propiedades. Las transformaciones normalmente aplicadas son la de Box-Cox (Box y Cox, 1964) y la logarítmica.
- Efecto de la composición de la población. Las poblaciones de los gastos máximos pueden consistir en dos o más sub-poblaciones clasificables de acuerdo a sus causas definitivas. Estas podrán ser el resultados de: 1) Lluvias convectivas, 2) Huracanes, 3) Tormentas tropicales, 4) Deshielo, y 5) Mecanismo de baja frecuencia como el ENSO.

4.3 MÉTODOS DE ESTIMACIÓN DE PARÁMETROS

Los problemas hidrológicos se analizan a través del uso de un modelo. Los modelos pueden ser del tipo determinístico, paramétrico o estocástico.

Un modelo completamente determinístico sería aquel que se obtiene a través de relaciones físicas y no requiere de datos experimentales para su aplicación.

Un modelo paramétrico puede ser considerado como un determinístico en el sentido de que una vez que se estiman los parámetros del modelo, este siempre genera la misma

salida a partir de la información de entrada. Por otro lado, un modelo paramétrico es estocástico en el sentido de que los parámetros estimados dependen de los datos observados y aquellos cambiarán cuando los datos observados también lo hagan.

Un modelo estocástico es aquel en el cual las salidas son predecibles solo en un sentido estadístico. En un modelo estocástico, el empleo repetido de un grupo dado de entradas del modelo genera salidas que no son las mismas pero siguen cierto modelo estadístico.

Los métodos que se aplicarán en este trabajo para la estimación de los parámetros de las funciones de distribución de probabilidad serán el de los Momentos, la técnica de Máxima Verosimilitud y la técnica de los Momentos-L.

4.4 DISTRIBUCIONES UNIMODALES

En este numeral se describen algunas familias de distribuciones unimodales que comúnmente se emplean en el análisis de frecuencia de gastos máximos y que serán las utilizadas en este estudio.

4.4.1 Distribución Gamma con dos parámetros

$$F(x) = \int_0^x \frac{x^{\beta-1} e^{-x/\alpha}}{\alpha^{\beta} \Gamma(\beta)} dx \quad (4-1)$$

$$f(x) = \frac{x^{\beta-1} e^{-x/\alpha}}{\alpha^{\beta} \Gamma(\beta)} \quad (4-2)$$

$$0 \leq X < \infty$$

$$Si \alpha > 0 \rightarrow \gamma > 0$$

$$\gamma > 0$$

Donde

α parámetro de escala

β parámetro de forma

$\Gamma(\beta)$ función Gamma completa

Ademas,

$$\mu = \beta\alpha \quad (4-3)$$

$$\sigma^2 = \alpha^2\beta \quad (4-4)$$

$$\gamma = \frac{2}{\sqrt{\beta}} \quad (4-5)$$

Estimadores por momentos

$$\hat{\alpha} = \frac{S^2}{\bar{X}} \quad (4-6)$$

$$\hat{\beta} = \left(\frac{\bar{X}}{S}\right)^2 \quad (4-7)$$

Estimadores por máxima verosimilitud

$$\hat{\alpha} = \frac{\bar{X}}{\hat{\beta}} \quad (4-8)$$

El estimadro $\hat{\beta}$ se obtiene al resolver

$$F(\hat{\beta}) = \hat{\mu}_y - \ln(\hat{\mu}_x) + \ln(\hat{\beta}) - \psi(\hat{\beta}) = 0 \quad (4-9)$$

Empleando como valor inicial del parámetro $\hat{\beta}$

$$\hat{\beta} = \frac{1 + \sqrt{1 + \frac{4}{3}C}}{4C} \quad (4-10)$$

$$C = \ln(\bar{x}) - \bar{y} \quad (4-11)$$

La derivada de la ecuación (4-9) es

$$F'(\hat{\beta}) = \frac{1}{\hat{\beta}} - \psi'(\hat{\beta}) \quad (4-12)$$

La aproximación de la función digamma de $\hat{\beta}$ es

$$\psi(\hat{\beta}) = \ln(\hat{\beta}) - \frac{1}{2\hat{\beta}} - \frac{1}{12\hat{\beta}^2} \quad (4-13)$$

La aproximación de la función trigamma de $\hat{\beta}$ es

$$\psi'(\hat{\beta}) = \frac{1}{(\hat{\beta} + 2)} + \frac{1}{2(\hat{\beta} + 2)^2} + \frac{1}{6(\hat{\beta} + 2)^3} - \frac{1}{30(\hat{\beta} + 2)^5} + \frac{1}{42(\hat{\beta} + 2)^7} - \frac{1}{30(\hat{\beta} + 2)^9} + \frac{1}{(\hat{\beta} + 1)^2} + \frac{1}{\hat{\beta}^2} \quad (4-14)$$

Sin embargo, si utilizamos la aproximación de Thom, (1958) para la función digamma ecuación (4-13) y la reemplazamos en la ecuación (4-9) encontramos:

$$12(\ln \widehat{\mu}_x - \mu_y)\beta^2 - 6\beta - 1 = 0 \quad (4-15)$$

Llamando $C = \ln \widehat{\mu}_x - \mu_y$ encontramos que

$$\beta = \frac{1 + \sqrt{1 + \frac{4}{3}C}}{4C} \quad (4-16)$$

Que es idéntica a la ecuación (4-10), por lo tanto, si empleamos la aproximación de Thom, (1958) para estimar la función digamma no es necesario realizar el proceso de iteración, ya que el valor β estimado con la ecuación (4-10) es la solución de la ecuación (4-9).

Estimadores por momentos-L

$$\tau_2 = \frac{\lambda_2}{\lambda_1} \quad (4-17)$$

Para $0 \leq \tau_2 < 0.5$

$$\hat{\beta} = (1 - 0.308 z)/(z - 0.05812 z^2 + 0.01765 z^3) \quad (4-18)$$

$$Z = \pi \tau_2 = 3.1416 \tau_2 \quad (4-19)$$

Para $0.5 \leq \tau_2 < 1$

$$\hat{\beta} = (0.7213 z - 0.5947 z^2)/(1 - 2.1817 z + 1.2113 z^2) \quad (4-20)$$

$$z = 1 - \tau_2 \quad (4-21)$$

El estimador $\hat{\alpha}$ se obtiene con la expresión (4-8) y el valor de (4-18) o (4-20).

4.4.2 Distribución de valores extremos tipo I (Gumbel)

$$F(x) = e^{-e^{\frac{x-v}{\alpha}}} \quad (4-22)$$

$$f(x) = \frac{1}{\alpha} - e^{-\left(\frac{x-v}{\alpha}\right)} e^{-e^{\left(\frac{x-v}{\alpha}\right)}} \quad (4-23)$$

$$-\infty < x < \infty, \quad \alpha > 0$$

Donde

v parámetro de ubicación

α parámetro de escala

$$E(x) = \hat{v} - 0.5772 \hat{\alpha} \quad (4-24)$$

$$\sigma^2 = \frac{\pi^2 \hat{\alpha}^2}{6} \quad (4-25)$$

$$\gamma = 1.1396 \quad (4-26)$$

$$\kappa = 5.4002 \quad (4-27)$$

La variable reducida Gumbel es

$$y_i = \frac{x_i - v}{\alpha} \quad (4-28)$$

Estimadores por momentos

$$\hat{v} = \bar{x} - 0.45 S \quad (4-29)$$

$$\hat{\alpha} = \frac{\sqrt{6}}{\pi} S = 0.78 S \quad (4-30)$$

Estimadores por máxima verosimilitud

Considerando la variable reducida (4-28) se tiene el siguiente proceso iterativo

$$P = n - \sum_{i=1}^n e^{-y_i} \quad (4-31)$$

$$R = n - \sum_{i=1}^n y_i + \sum_{i=1}^n y_i e^{-y_i} \quad (4-32)$$

El criterio de convergencia es:

$$\frac{P}{\hat{\alpha}} \approx 0 \quad y \quad \frac{-R}{\hat{\alpha}} \approx 0$$

Incrementos

$$\begin{aligned} \delta_{v_j} &= (1.11P_j - 0.26R_j) \frac{\alpha_j}{n} \\ \delta_{\alpha_j} &= (0.26P_j - 0.61R_j) \frac{\alpha_j}{n} \end{aligned} \quad (4-33)$$

Nuevos valores

$$\hat{v}_{j+1} = \hat{v}_j + \delta_{v_j} \quad (4-34)$$

$$\hat{\alpha}_{j+1} = \hat{\alpha}_j + \delta_{\alpha_j} \quad (4-35)$$

Estimadores por momentos-L

$$\hat{v} = \lambda_1 - 0.577216\hat{\alpha} \quad (4-36)$$

$$\hat{\alpha} = \frac{\lambda_2}{\ln(2)} \quad (4-37)$$

$$\lambda_1 = \hat{M}_{(0)} \quad (4-38)$$

$$\lambda_2 = 2\hat{M}_{(1)} - \hat{M}_{(0)} \quad (4-39)$$

$$\lambda_3 = 6\hat{M}_{(2)} - 6\hat{M}_{(1)} + \hat{M}_{(0)} \quad (4-40)$$

$$\lambda_4 = 20\hat{M}_{(3)} - 30\hat{M}_{(2)} + 12\hat{M}_{(1)} - \hat{M}_{(0)} \quad (4-41)$$

Una vez conocidas $\lambda_1, \lambda_2, \lambda_3$ y λ_4 se pueden obtener las relaciones de los momentos-L

$$\text{Coeficiente de variación} - L(L - CV) = \tau_2 = \lambda_2/\lambda_1 \quad (4-42)$$

$$\text{Coeficiente de sesgo} - L(L - g) = \tau_3 = \lambda_3/\lambda_2 \quad (4-43)$$

$$\text{Coeficiente de curtosis} - L(L - k) = \tau_4 = \lambda_4/\lambda_2 \quad (4-44)$$

4.4.3 Distribución general de valores extremos (GVE)

$$F(x) = e^{-\left[1 - \left(\frac{x-v}{\alpha}\right)\beta\right]^{1/\beta}} \quad (4-45)$$

$$f(x) = \frac{1}{\alpha} e^{-\left[1 - \left(\frac{x-v}{\alpha}\right)\beta\right]^{1/\beta}} \left(1 - \left(\frac{x-v}{\alpha}\right)\beta\right)^{\frac{1}{\beta}-1} \quad (4-46)$$

Si $\beta = 0$ $-\infty < x < \infty$, entonces la distribución es tipo I (Gumbel)

Si $\beta < 0$ $v + \frac{\alpha}{\beta} \leq x < \infty$, entonces la distribución es tipo II (Fréchet)

Si $\beta > 0$ $-\infty < x \leq v + \frac{\alpha}{\beta}$, entonces la distribución es tipo III (Weibull)

Donde

v parámetro de ubicación

α parámetro de escala, $\alpha > 0$

β parámetro de forma

La variable reducida GVE es:

$$y = -\frac{1}{\beta} \ln \left(1 - \left(\frac{x-v}{\alpha}\right)\beta\right) \quad (4-47)$$

Estimadores por momentos

Para $-11.35 < g < 1.1396$

$$\hat{\beta} = 0.279434 - 0.333535g + 0.048306g^2 + 0.023314g^3 + 0.00376g^4 + 0.000263g^5 \quad (4-48)$$

Para $1.14 < g < 18.95$

$$\hat{\beta} = 0.25031 - 0.29219g + 0.075357g^2 + 0.010883g^3 + 0.000904g^4 + 0.000043g^5 \quad (4-49)$$

$$\hat{A} = \mu_x + \hat{B}E[y] = \bar{x} + \hat{B}E[y] \quad (4-50)$$

$$\hat{B} = \left[\frac{\text{Var}(x)}{\text{Var}(y)}\right]^{(1/2)} \quad (4-51)$$

$$\text{Var}(x) = \sigma_x^2 = S_x^2 \quad (4-52)$$

$$E[y] = \Gamma(1 + \hat{\beta}) \quad (4-53)$$

$$\text{Var}(y) = \Gamma(1 + 2\hat{\beta}) + \Gamma^2(1 + \hat{\beta}) \quad (4-54)$$

Para $\hat{\beta} < 0$

$$\hat{\alpha} = -\hat{\beta}\hat{B} \quad (4-55)$$

$$\hat{v} = \hat{A} + \hat{B} \quad (4-56)$$

Para $\hat{\beta} > 0$

$$\hat{\alpha} = -\hat{\beta}\hat{B} \quad (4-57)$$

$$\hat{v} = \hat{A} - \hat{B} \quad (4-58)$$

Para $\hat{\beta} = 0$

$$\hat{v} = x - 0.45 S \quad (4-59)$$

$$\hat{\alpha} = \frac{\sqrt{6}}{\pi} S = 0.78 S \quad (4-60)$$

Estimadores por máxima verosimilitud

El procedimiento empleado para estimar los parámetros por máxima verosimilitud es el descrito por Jenkinson, (1969). Sin embargo, se han encontrado problemas en la convergencia debido al tamaño del incremento para estimar los parámetros. Esta situación ha sido remediada para los casos analizados, multiplicando los incrementos por 0.1, es decir, solo considerando el 10% del incremento dado.

Considerando la variable reducida (4-47) se tiene el siguiente proceso iterativo

$$P = n - \sum_{i=1}^n e^{-y_i} \quad (4-61)$$

$$Q = \sum_{i=1}^n e^{-y_i + (\beta-1)y_i} - (1-\beta) \sum_{i=1}^n e^{\beta y_i} \quad (4-62)$$

$$R = n - \sum_{i=1}^n y_i + \sum_{i=1}^n y_i e^{-y_i} \quad (4-63)$$

El criterio de convergencia es

$$\begin{aligned} \frac{Q}{\hat{\alpha}} &\approx 0 \\ \frac{1}{\hat{\alpha}} \left(\frac{P+Q}{\hat{\beta}} \right) &\approx 0 \\ \frac{1}{\hat{\beta}} \left[R - \left(\frac{P+Q}{\hat{\alpha}} \right) \right] &\approx 0 \end{aligned} \quad (4-64)$$

Incrementos

$$\delta_{\hat{a}_j} = -\frac{\hat{a}_j}{10n} \left\{ bQ_j + \frac{h(P_j + Q_j)}{\hat{\beta}_j} + \frac{f}{\hat{\beta}_j} \left[R_j - \frac{(P_j + Q_j)}{\hat{\beta}_j} \right] \right\} \quad (4-65)$$

$$\delta_{\hat{\alpha}_j} = -\frac{\hat{\alpha}_j}{10n} \left\{ hQ_j + a \frac{(P_j + Q_j)}{\hat{\beta}_j} + \frac{gs}{\hat{\beta}_j} \left[R_j - \frac{(P_j + Q_j)}{\hat{\beta}_j} \right] \right\}$$

$$\delta_{\hat{\beta}_j} = -\frac{1}{10n} \left\{ fQ_j + gs \left(\frac{(P_j + Q_j)}{\hat{\beta}_j} \right) + \frac{c}{\hat{\beta}_j} \left[R_j - \left(\frac{(P_j + Q_j)}{\hat{\beta}_j} \right) \right] \right\}$$

Donde a , b , c , f , gs y h son los elementos de la matriz de varianza- covarianza de los parámetros de la distribución GVE para $-1.5 < \hat{\beta} < 1.0$ (Raynal, 1982)

$$\begin{aligned} a &= 0.661437 - 0.562798\hat{\beta} + 0.985803\hat{\beta}^2 - 0.059011\hat{\beta}^3 \\ b &= 1.235356 - 0.162161\hat{\beta} - 0.115137\hat{\beta}^2 + 0.009577\hat{\beta}^3 \\ c &= 0.4711 - 0.77627\hat{\beta} + 0.295825\hat{\beta}^2 - 0.009645\hat{\beta}^3 \\ f &= 0.244435 - 0.10287\hat{\beta} - 0.19583\hat{\beta}^2 - 0.016837\hat{\beta}^3 \\ gs &= 0.15373 - 0.411923\hat{\beta} - 0.479209\hat{\beta}^2 - 0.075004\hat{\beta}^3 \\ h &= 0.338937 - 1.209555\hat{\beta} - 0.109822\hat{\beta}^2 - 0.019801\hat{\beta}^3 \end{aligned} \quad (4-66)$$

Nuevos valores

$$\begin{aligned} \hat{v}_{j+1} &= \hat{v}_j + \delta_{v_j} \\ \hat{\alpha}_{j+1} &= \hat{\alpha}_j + \delta_{\alpha_j} \\ \hat{\beta}_{j+1} &= \hat{\beta}_j + \delta_{\beta_j} \end{aligned} \quad (4-67)$$

Estimadores por momentos-L

$$E = \frac{\left\{ 2\hat{M}_{(1)} - \hat{M}_{(0)} \right\} [\ln(2)]}{\left\{ 3\hat{M}_{(2)} - \hat{M}_{(0)} \right\} [\ln(3)]} \quad (4-68)$$

$$\hat{\beta} = 7.859E + 1.9554E^2 \quad (4-69)$$

$$A = \Gamma(1 + \hat{\beta}) \quad (4-70)$$

$$B = 1 - 2^{-\hat{\beta}} \quad (4-71)$$

$$C = (2\hat{M}_{(1)} - \hat{M}_{(0)})\hat{\beta} \quad (4-72)$$

$$D = (A - 1)/\hat{\beta} \quad (4-73)$$

$$\hat{\alpha} = \frac{C}{AB} \quad (4-74)$$

$$\hat{v} = \hat{M}_{(0)} + D\hat{\alpha} \quad (4-75)$$

$$\hat{M}_{(k)} = \frac{1}{n} \sum_{i=1}^{n-k} x_i \binom{n-i}{k} / \binom{n-1}{k} \quad (4-76)$$

$$\hat{M}_{(0)} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4-77)$$

$$\hat{M}_{(1)} = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} x_i(n-i) \quad (4-78)$$

$$\hat{M}_{(2)} = \frac{1}{n(n-1)(n-2)} \sum_{i=1}^{n-2} x_i(n-i)(n-i-1) \quad (4-79)$$

4.5 DISTRIBUCIONES MEZCLADAS

Cada vez son más las evidencias que indican que los gastos provienen de más de una distribución, contrario a como lo asumen los métodos tradicionales (Murphy, 2001; Escalante, 2007; Apipattanavis et al, 2010,). Alila y Mtiraoui (2002) resumen los factores naturales que ocasionan que las muestras sean descritas por más de una distribución: (1) Variación estacional en los mecanismos que producen los gastos máximos, por ejemplo: huracanes, tormentas, inundaciones ocasionadas en primavera por el deshielo, lluvias presentadas en verano, (2) cambios en los patrones climáticos ocasionados por mecanismos de baja frecuencia como el fenómeno del Niño-Oscilación del Sur (ENSO), (3) cambios en el uso del suelo, entre otros. El concepto de mecanismo de generación mixto o mezclado está estrechamente relacionado con la forma de no estacionariedad presente en los registros de gastos máximos (Kidson y Richards, 2005).

En México, los registros hidrométricos de gastos máximos anuales de las zonas afectadas esporádicamente por ciclones tropicales muestran la presencia de dos tipos de crecientes, las precipitaciones comunes o normales de la región y las ciclónicas (Campos, 1989). El U.S. Army Corps of Engineers indica que cuando la cuenca bajo estudio tiene un área inferior a 1300 km², generalmente no requieren un análisis especial de huracanes (Morris, 1982). En los Andes de Colombia (Poveda et al, 2011), en la zona norte de Perú (Waylen y Caviedes, 1986) y en la zona noroeste de México (Escalante, 2007) los gastos máximos son influenciados por el ENSO.

Una técnica que no requiere la separación *a priori* de los mecanismos de generación de los gastos máximos denominada “distribuciones heterogéneas sin clasificación”, considera que los gastos máximos pertenecen a varias poblaciones con distribuciones de probabilidad homogénea específica. De acuerdo a Moran (1959), una distribución heterogénea es una mezcla de m distribuciones homogéneas componentes dada por:

$$F(x) = p_1 F_1(x) + p_2 F_2(x) + \dots + p_m F_m(x) \quad (4-80)$$

Donde $F_1(x), \dots, F_m(x)$ son las distribuciones de probabilidad acumulada de las m distribuciones componentes y p_1, \dots, p_m son parámetros que definen la proporción relativa y satisfacen la condición $p_1 + p_2 + \dots + p_m = 1$.

La aplicación de la ecuación (4-80) requiere del conocimiento *a priori* acerca de las distribuciones homogéneas “madre” para cada componente. También se requiere del conocimiento *a priori* del número de mecanismos que intervienen en la generación de los gastos máximos. Este análisis puede realizarse empleando las técnicas descritas en el capítulo 3.

Un evento extremo es una observación que es muy diferente, ya sea más grande o más pequeña del resto de los datos. Un evento extremo que es improbable que ocurra dentro de la muestra es llamado “punto extraño”.

En el modelado de las series de eventos hidrológicos extremos, por ejemplo, los gastos máximos anuales en una corriente, se puede considerar que éstos son generados por dos diferentes procesos o tipos de tormenta. Es concebible que cierta proporción “ p ” de la población tiene valores que provienen de una distribución $F_1(x)$, mientras que la restante “ $1 - p$ ” tiene una diferente distribución $F_2(x)$, así (Singh y Sinclair, 1972),

$$P(X \leq x) = F(x) = pF_1(x) + (1 - p)F_2(x) \quad (4-81)$$

La estimación de los parámetros de la función de distribución acumulada de la ecuación (4-81) se puede hacer mediante el método de máxima verosimilitud o empleando algoritmos de optimización no lineal maximizando la función de máxima verosimilitud sujeto a unas restricciones. Así mismo, los eventos estimados pueden ser obtenidos de la ecuación (4-81) empleando el método de Newton-Rapson. A continuación se presentan algunos modelos de distribuciones mezcladas de dos componentes que pueden ser empleadas en el análisis de frecuencia de gastos máximos.

4.5.1 Distribución Gumbel – Gumbel (G-G)

$$F(x) = pe^{-e^{-\left(\frac{x-v_1}{\alpha_1}\right)}} + (1 - p)e^{-e^{-\left(\frac{x-v_2}{\alpha_2}\right)}} \quad (4-82)$$

$$f(x) = pe^{-e^{-\left(\frac{x-v_1}{\alpha_1}\right)}} e^{-\left(\frac{x-v_1}{\alpha_1}\right)} + (1 - p)e^{-e^{-\left(\frac{x-v_2}{\alpha_2}\right)}} e^{-\left(\frac{x-v_2}{\alpha_2}\right)} \quad (4-83)$$

$$x > 0 \quad \alpha_i > 0, \quad 0 < p < 1 \quad (4-84)$$

Donde

v_1 parámetro de ubicación de la primera población

α_1 parámetro de escala de la primera población

v_2 parámetro de ubicación de la segunda población

α_2 parámetro de escala de la segunda población

p parámetro de asociación

Estimadores por máxima verosimilitud

Los parámetros se calculan por la maximización de la función logarítmica de verosimilitud sujeta a (4-84).

$$\ln L = \ln \prod_{i=1}^n f(x_i; v_1, \alpha_1, v_2, \alpha_2, p) \quad (4-85)$$

Para el caso de la distribución Gumbel Mixta (G – G) la función es:

$$\begin{aligned}
Ln l(x, \theta) &= \sum_{l=1}^n Ln \left\{ \frac{p}{\alpha_1} \exp \left\{ -\exp \left[-\frac{x - \mu_1}{\alpha_1} \right] \right\} \exp \left[-\frac{x - \mu_1}{\alpha_1} \right] + \frac{(1-p)}{\alpha_2} \exp \left\{ -\exp \left[-\frac{x - \mu_2}{\alpha_2} \right] \right\} \exp \left[-\frac{x - \mu_2}{\alpha_2} \right] \right\} \frac{\partial LnL}{\partial v_1} \\
&= \frac{p}{\alpha_1^2} \sum_{l=1}^n \frac{1}{f(x_i, \theta)} \left\{ \frac{\exp \left\{ -\exp \left[-\frac{x_i - v_1}{\alpha_1} \right] \right\} \exp \left[-\frac{x_i - v_1}{\alpha_1} \right]}{1 - \exp \left[-\frac{x_i - v_1}{\alpha_1} \right]} \right\} = 0 \\
\frac{\partial LnL}{\partial \alpha_1} &= \frac{p}{\alpha_1^2} \sum_{l=1}^n \frac{1}{f(x_i, \theta)} \left\{ \frac{\exp \left\{ -\exp \left[-\frac{x_i - v_1}{\alpha_1} \right] \right\} \exp \left[-\frac{x_i - v_1}{\alpha_1} \right]}{\left[\frac{x_i - v_1}{\alpha_1} \right] \left[1 - \exp \left[-\frac{x_i - v_1}{\alpha_1} \right] \right]} - 1 \right\} = 0 \\
\frac{\partial LnL}{\partial v_2} &= \frac{1-p}{\alpha_2^2} \sum_{l=1}^n \frac{1}{f(x_i, \theta)} \left\{ \frac{\exp \left\{ -\exp \left[-\frac{x_i - v_2}{\alpha_2} \right] \right\} \exp \left[-\frac{x_i - v_2}{\alpha_2} \right]}{1 - \exp \left[-\frac{x_i - v_2}{\alpha_2} \right]} \right\} = 0 \\
\frac{\partial LnL}{\partial \alpha_2} &= \frac{1-p}{\alpha_2^2} \sum_{l=1}^n \frac{1}{f(x_i, \theta)} \left\{ \frac{\exp \left\{ -\exp \left[-\frac{x_i - v_2}{\alpha_2} \right] \right\} \exp \left[-\frac{x_i - v_2}{\alpha_2} \right]}{\left[\frac{x_i - v_2}{\alpha_2} \right] \left[1 - \exp \left[-\frac{x_i - v_2}{\alpha_2} \right] \right]} - 1 \right\} = 0 \\
\frac{\partial LnL}{\partial p} &= \sum_{l=1}^n \frac{1}{f(x_i, \theta)} \left\{ \frac{1}{\alpha_1} \exp \left\{ -\exp \left[-\frac{x_i - v_1}{\alpha_1} \right] \right\} \exp \left[-\frac{x_i - v_1}{\alpha_1} \right] - \frac{1}{\alpha_2} \exp \left\{ -\exp \left[-\frac{x_i - v_2}{\alpha_2} \right] \right\} \exp \left[-\frac{x_i - v_2}{\alpha_2} \right] \right\} = 0
\end{aligned} \tag{4-86}$$

4.6 COMENTARIO SOBRE LA PARSIMONIA

De acuerdo a Salas et al, (1980) el principio de la parsimonia busca encontrar aquella distribución (para nuestro caso) que con el mínimo número de parámetros se ajuste a los registros históricos. Considere que N es el tamaño de la muestra y K es el número de parámetros de la distribución a ajustar, la relación:

$$\delta = \frac{N}{K} \tag{4-87}$$

Es un índice general de la parsimonia de los parámetros. Valores de $\delta = 4 - 6$ podrán indicar que la base del principio de la parsimonia en los parámetros no es respetado. Estrictamente, el valor mínimo de δ debe ser de 15.0. Así, por ejemplo, si trabajamos con una distribución de 3 parámetros deberíamos contar con una serie de al menos 45 años de registro para respetar el principio de la parsimonia. Por lo tanto, a la hora de emplear las distribuciones mezcladas en muestras de pocos registros se debe tener cuidado por el número de parámetros que involucran estas.

4.7 PRUEBAS DE BONDAD DE AJUSTE

A continuación se describen algunos métodos aplicados para elegir la mejor distribución que se describe la muestra de gastos máximos anuales.

4.7.1 Error estándar de ajuste

Kite, (1988) propuso un estadístico que permite seleccionar la mejor opción, entre diferentes modelos en competencia, para el ajuste de una muestra de datos Q_i^j para $i = 1, 2, \dots, n_j$, de un sitio j .

Este estadístico es conocido como el error estándar de ajuste, y tiene la forma

$$EE = \left[\frac{\sum_{i=1}^{n_j} (\hat{Q}_T^j - Q_T^j)^{1/2}}{n_j - mp} \right] \quad (4-88)$$

Donde

Q_T^j son los eventos Q_i^j ordenados de mayor a menor con un periodo de retorno asignado $T = \frac{n_j+1}{m}$ y una probabilidad de no excedencia $P = 1 - \frac{1}{T}$.

n_j longitud en años del registro analizado.

m número de orden del registro.

$m = 1$ para el evento más grande.

$m = n_j$ para el evento más chico.

\hat{Q}_T^j eventos estimadores por cierta distribución de probabilidad para cada periodo de retorno T asignado a la muestra ordenada Q_i^j .

mp número de parámetros de la distribución ajustada, por ejemplo

$mp = 1$ para la exponencial-1

$mp = 2$ para la Normal, Log Normal-2, exponencial-2, Gamma-2, Gumbel

$mp = 3$ para la Log Normal-3, Gamma-3, Log Pearson III, GVE

$mp = 4$ para la de Valores extremos de dos componentes TCEV

$mp = 5$ para la Gumbel de dos poblaciones, Wakeby.

La distribución de mejor ajuste será aquella que proporcione el mínimo valor del estadístico EE . Si una o más distribuciones tienen valores similares del EE , entonces, se deberá optar por aquella distribución que tenga el menor número de parámetros.

4.7.2 Índice Medio de la Desviación Absoluta (MADI)

$$MADI = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - z_i}{x_i} \right| \quad (4-89)$$

4.7.3 Índice Desviación Media al Cuadrado (MSDI)

$$MSDI = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - z_i}{x_i} \right)^2 \quad (4-90)$$

Los índices MADI y MSDI fueron propuestos por Jain y Singh (1987).

4.7.4 Criterio de la información de AKAIKE (CIA)

El criterio de información de Akaike se define como (Akaike, 1974):

$$CIA = -2l + 2k \quad (4-91)$$

Se interpreta como la cantidad de información que se perdió en el proceso, por eso se selecciona aquel con el valor más pequeño.

4.7.5 Criterio de Información Bayesiana (BIC)

El criterio de información Bayesiano se define como (Schwarz, 1978)

$$BIC = -2l + k \ln(N) \quad (4-92)$$

Dadas iguales probabilidades *a priori* para los modelos en competencia, se selecciona aquel con el menor valor de BIC, la cual es equivalente a obtener el modelo con la mayor probabilidad *a posteriori*.

x_i := son los valores reales (del registro)

z_i := son los valores estimados de diferentes periodos de retorno

k := es el número de parámetros del modelo

l := es el logaritmo de máxima verosimilitud $\ln(L)$

N := es el tamaño de la muestra

5 ANÁLISIS DE FRECUENCIA EN EL SITIO

5.1 INTRODUCCIÓN

De acuerdo a las recomendaciones de la NERC, (1975), el procedimiento a llevar a cabo para el análisis de frecuencia de acuerdo al número de registros que se tenga en el sitio es el siguiente:

- En **sitios con menos de 10 años de registros** definitivamente se recomienda que se estimen los gastos de diseño a partir de técnicas de regionalización.
- En **sitios con registros entre 10 y 25 años** se recomienda realizar una estimación en el sitio aplicando las distribuciones de probabilidad vistas en el capítulo 4, sin embargo, se recomienda que se verifique y estimen los gastos de diseño además empleando técnicas de regionalización.
- En **sitios con más de 25 años de registros** se pueden aplicar las metodologías para estimar los gastos de diseño vistas en el capítulo 4 y se recomienda, aunque no es necesario, estimar los gastos empleando técnicas regionales.

Sin embargo, varios estudios han encontrado que es necesario contar con una longitud de registros mayor para tener una adecuada confiabilidad en la estimación de los eventos de diseño. Por ejemplo, Victorov, (1971) analizó el efecto de la longitud de registros en la estimación de gastos máximos en la estación Mankato ubicada sobre el río Minnesota, empleando la distribución Log-Pearson Tipo tres o Log-Gamma de tres parámetros encontró que en el análisis en el sitio el uso de registros cortos de 10, 20, 30 e incluso de 40 años para la estimación de la probabilidad de ocurrencia de gastos máximos puede conducir a errores de cálculo importante.

McCuen y Galloway, (2010) estimaron los errores relativos, para diferentes tamaños de registro y coeficientes de asimetría, en las estimaciones de eventos de diseño de 2, 10 y 100 años empleando la distribución Log-Pearson tipo III. Encontraron que el error relativo incrementa con el periodo de retorno y es mayor para muestras pequeñas. Por ejemplo, para un nivel de confianza del 95% para una muestra con 10 años de registro y coeficiente de asimetría de 1, la estimación del periodo de retorno de 10 años tiene un error relativo de 1.02 y el de 100 años un error relativo de 2.9. Sin embargo, cuando se tiene una longitud de registros de 100 años, los errores relativos disminuyen 27% para el evento de diseño 10 años de periodo de retorno y 63% para el de 100 años de periodo de retorno.

Finfrock y Rao, (2003) realizaron un estudio para definir la longitud de registros necesaria para identificar una adecuada distribución para el análisis de frecuencia. Empleando registros del río Muscatatuck ubicado al sur de Indiana, generaron series sintéticas de 20, 40, 60 y 100 años utilizando las distribuciones General de valores extremos, Generalizada de Pareto y Generalizada Logística. La selección de la distribución la realizaron a partir del diagrama de Momentos-L de las distribuciones General de valores extremos, Generalizada Logística, Generalizada de Pareto, Lognormal de tres parámetros y Gamma de tres parámetros. Para muestras de 20 años de registros encontraron que es casi imposible determinar la distribución a partir de la cual los datos fueron generados. La

correspondencia a la definición de una adecuada distribución aumentó conforme se incrementó el tamaño de la muestra, sin embargo, aún para muestras de 100 años encontraron dificultades para definir la adecuada distribución. Los autores recomiendan emplear muestras con registros largos para realizar los análisis de frecuencia, ya que en muestras de registros cortos es difícil identificar la adecuada distribución de probabilidad.

Das, (2011) y Das y Cunnane, (2011) realizaron un estudio con más de 100 estaciones hidrométricas en Irlanda para evaluar la sensibilidad de la estimación regional Avenida Índice al número de estaciones - año del grupo. En sus análisis encontraron que en regiones con moderada o alta heterogeneidad la estimación del evento de diseño de 100 años de periodo de retorno es recomendable solo si en los sitios el tamaño de la muestra excede los 50 años.

- Periodo de Retorno

Se define como el número de años que transcurren en promedio para que un evento de magnitud dado x sea igualado por lo menos una vez en ese periodo de tiempo.

$$T = \frac{1}{p(x > x)} = \frac{1}{1 - p(x \leq x)} \quad (5-1)$$

Cada espacio muestral tiene su propia función de distribución o de densidad de probabilidad, que normalmente no se conoce a *priori*. Cuando de ese espacio se extrae un grupo de datos (muestra) al azar, se puede esperar que su función de distribución de probabilidad sea similar a la del espacio completo, en particular si la muestra es grande.

En hidrología, normalmente se prefiere trabajar con periodos de retorno en lugar de probabilidades, pues es un concepto que resulta más claro ya que tiene las mismas unidades (tiempo) que la vida útil de las obras.

- Probabilidad de no excedencia

$$F(x \leq x) = 1 - \frac{1}{T} \quad (5-2)$$

- Probabilidad de excedencia

$$F(x) = \frac{1}{T} \quad (5-3)$$

El análisis de frecuencia en el sitio se realizará teniendo en cuenta los resultados obtenidos en el análisis de homogeneidad en el cual se recomienda emplear distribuciones mezcladas para la obtención de los eventos de diseño. Sin embargo, teniendo en cuenta el número de parámetros que involucran estas distribuciones se hará un análisis de sensibilidad para los eventos estimados con estas distribuciones a la longitud de registros de la muestra.

5.2 ANÁLISIS DE FRECUENCIA DE GASTOS MÁXIMOS ANUALES

El análisis de frecuencias de los gastos máximos anuales de una muestra $Q_i, i = 1, 2, \dots, n$, se emplea para proveer la magnitud de un evento \hat{Q}_T , de cierto periodo de retorno T , por medio del ajuste de una distribución de probabilidad, la cual es seleccionada como la mejor entre un grupo de ellas.

La secuencia de análisis es la siguiente:

Paso 1. Recabar la información de los eventos $Q_i, i = 1, 2, \dots, n$. En este punto se debe verificar la calidad y cantidad de la información.

Paso 2. Verificar la homogeneidad de la serie mediante las pruebas de homogeneidad vistas en el capítulo 3.

Paso 3. Obtener los estadísticos muestrales como la media, la desviación estándar, el coeficiente de asimetría, el coeficiente de variación, el coeficiente de curtosis.

Paso 4. Verificar con la prueba de Anderson la independencia de eventos de la serie Q_i .

Paso 5. La serie Q_i se ordena de mayor a menor, se le asigna un periodo de retorno T y una probabilidad de no excedencia.

$$T = \frac{n + 1}{m} \quad (\text{Ley empírica de Weibull})$$
$$P(X \leq x) = 1 - \frac{1}{T} \quad (5-4)$$

Donde

m tamaño de la muestra analizada en años.

n número de orden del registro años.

$P(X \leq x)$ probabilidad de no excedencia.

Paso 6. A la serie Q_i se le ajustan las diferentes distribuciones de probabilidad para el análisis de máximos unimodales (Normal, Log Normal con 2 y 3 parámetros, Gamma con 2 y 3 parámetros, Gamma con 2 y 3 parámetros, Log Pearson tipo III, Gumbel, General de Valores Extremos), y de más de una población (Valores Extremos de dos Componentes. Doble Gumbel, Doble Gamma, etc.), y se selecciona aquella que proporcione el mínimo ajuste de los distintos métodos empleados.

Paso 7. Una vez que se obtiene la distribución de mejor ajuste del registro Q_i , es posible calcular los eventos \hat{Q}_T y sus límites de confianza para los periodos de retorno $T=2, 5, 10, 20, 50, 100, 500, 1000, 5000$ y 10000 años.

5.3 CASO DE ESTUDIO

En el análisis de homogeneidad realizado en el capítulo 3 se encontró que más del 70% de las estaciones de la zona de estudio son no estacionarias, el 98% de las estaciones que son no estacionarias presentan ya sea correlación serial, cambios en la media, cambios en la varianza o alguna combinación de las tres causas de no estacionariedad. Se encuentra además que para la zona de estudio es recomendable realizar el análisis de frecuencia en el sitio en aquellas estaciones que tengan al menos 40 años de registro. Así mismo, se recomienda no realizar el análisis de frecuencia empleando distribuciones unimodales.

Mecanismos de baja frecuencia como el ENSO (El Niño-oscilación del sur) tienen una fuerte influencia sobre la hidrología de Colombia (Poveda y Mesa, 1996; Poveda y Mesa, 1997; Poveda, 2004; Poveda et al, 2011). Empleando análisis con covariables se encontró que en un 42% de las estaciones no estacionarias, (véase capítulo 3) el análisis no estacionario empleando como covariables índices del ENSO es más adecuado para la zona de estudio y por lo tanto, la variabilidad climática puede explicar la no estacionariedad de estas estaciones. Como lo ha sugerido Alila y Mtiraoui (2002) si estos

procesos físicos no son considerados en el análisis de frecuencia convencional, los eventos de diseño estimados pueden ser ineficientes para propósitos de diseño. Por esta razón, el uso de distribuciones mezcladas es una buena opción para mejorar la estimación de gastos máximos (Escalante-Sandoval, 2007).

La correlación serial, los cambios tanto en la media y la varianza de las series de gastos máximos pueden ser provocados por la variabilidad climática (Salas, 1991; IH, 1999; Sveinsson et al, 2005). El concepto de mecanismo de generación mixto o mezclado está estrechamente relacionado con la forma de no estacionariedad presente en los registros de gastos máximos (Kidson y Richards, 2005). Varios estudios (Waylen y Caviedes, 1986; Alila y Mtiraoui, 2002; Mtiraoui, 2004; Escalante-Sandoval, 2007; Apipattanavis et al, 2010; Poveda et al 2011, Salas y Obeysekera, 2014) han recomendado el uso de distribuciones mezcladas en aquellas estaciones que presentan cambios que pueden ser originados por mecanismos de baja frecuencia como el ENSO.

Para un modelo de dos poblaciones se tiene (Singh y Sinclair, 1972):

$$F(x) = pF_1(x) + (1 - p)F_2(x) \quad (5-5)$$

Donde $F_1(x)$ y $F_2(x)$ son las distribuciones de probabilidad acumulada para los regímenes climáticos 1 y 2. p y $(1 - p)$ son los pesos relativos de las distribuciones. En este estudio se empleará la distribución Gumbel para describir el comportamiento de las distintas poblaciones. Así, tenemos:

$$F(x) = pe^{-e^{-\left(\frac{x-\mu_1}{\alpha_1}\right)}} + (1 - p)e^{-e^{-\left(\frac{x-\mu_2}{\alpha_2}\right)}} \quad (5-6)$$

$$f(x) = \frac{p}{\alpha_1} e^{-e^{-\left(\frac{x-\mu_1}{\alpha_1}\right)}} e^{-\left(\frac{x-\mu_1}{\alpha_1}\right)} + \frac{(1 - p)}{\alpha_2} e^{-e^{-\left(\frac{x-\mu_2}{\alpha_2}\right)}} e^{-\left(\frac{x-\mu_2}{\alpha_2}\right)} \quad (5-7)$$

Para $x > 0$, $\alpha_i > 0$ y $0 \leq p \leq 1$.

Donde:

μ_1 : = Parámetro de ubicación de la primera población

α_1 : = Parámetro de escala de la primera población

μ_2 : = Parámetro de ubicación de la segunda población

α_2 : = Parámetro de escala de la segunda población

p : = Peso relativo de la primera población.

Para la obtención de los parámetros de la función doble Gumbel se optimizará la función de máxima verosimilitud empleando el método de optimización no lineal Gradiente Reducido Generalizado (Lasdon et al, 1975) Una de las desventajas de la optimización no lineal es que el valor óptimo obtenido puede ser un máximo local y no el global, por lo tanto, es indispensable el uso de apropiados valores iniciales para obtener una solución adecuada. Para la obtención de los eventos de la muestra y de los de diseño se expresa la ecuación (5-5) de la siguiente manera:

$$F(x) - [pF_1(x) + (1 - p)F_2(x)] = 0 \quad (5-8)$$

Para una dada probabilidad de no excedencia $F(x)$ conocida, la ecuación (5-8) es solo función de x y el problema de la estimación del evento x para una probabilidad de no excedencia $F(x)$ se reduce a encontrar la raíces de la ecuación (5-8). Para la estimación de las raíces se puede emplear el método de Newton-Raphson o el método de bisección.

El método de Newton-Raphson requiere un menor tiempo computacional para encontrar las raíces pero es muy sensible al valor inicial empleado para la solución, por lo que requiere de una mayor experiencia para su empleo. En este estudio se empleó el método de bisección ya que garantiza siempre la respuesta, siempre y cuando la raíz se encuentra entre los valores iniciales de partida que requiere el algoritmo de resolución numérica.

El uso de la distribución Gumbel en las distribuciones mezcladas de dos componentes en muestras heterogéneas de gastos máximos ha sido estudiado por Raynal-Villasenor y Guevara-Miranda, (1997); Escalante-Sandoval, (1998); Escalante-Sandoval, (2007); Evin et al, (2011).

Los resultados obtenidos del ajuste de la distribución mezclada doble Gumbel a las 62 estaciones del estudio se presentan en la Tabla 5.1 Se presentan los eventos de diseño (en m^3/s) obtenidos para periodos de retorno de 2, 5, 10, 25, 50 y 100 años, los parámetros de la distribución, los resultados de las distintas pruebas de bondad de ajuste, así como la desviación estadística.

La desviación estadística se calculó considerando como modelo M_0 el modelo Gumbel unimodal y como modelo M_1 el modelo mezclado, se consideró un nivel de significancia del 10% y por lo tanto un valor crítico de 6.25. Para la distribución Gumbel unimodal se obtuvieron los parámetros por los distintos métodos considerados en el capítulo 3 y se tomó el mejor según los distintos criterios de ajuste.

De acuerdo al criterio de la desviación estadística, en 43 estaciones (69%) de la zona de estudio es adecuado emplear el modelo mezclado Gumbel de dos componentes o doble Gumbel. En las otras estaciones donde no es adecuado el modelo doble Gumbel es posible que otros modelos mezclados arrojen mejores resultados. En la Figura 5.1 se presenta la distribución espacial de las estaciones en las cuales el modelo doble Gumbel, según la desviación estadística, es recomendable su uso. Aunque en 19 estaciones no es adecuado el uso del modelo doble Gumbel, los eventos de diseño del análisis en el sitio serán estimados empleando este modelo.

Tabla 5.1 Eventos de diseño (en m³/s) para diferentes periodos de retorno

ESTACIÓN	Tr(años)						Parámetros					PRUEBAS DE AJUSTE					Desviación Estadística D
	2	5	10	25	50	100	P	μ_1 (m ³ /s)	α_1 (m ³ /s)	μ_2 (m ³ /s)	α_2 (m ³ /s)	EEA (m ³ /s)	MADI	MSDI	CIA	BIC	
21087050	19	30	46	71	81	90	0.91	15.3	7.3	65.0	11.7	3.23	0.07	0.01	335.6	535.6	16.23
21017050	77	94	105	118	129	139	0.05	36.2	1.0	72.6	14.5	4.20	0.03	0.00	371.5	571.5	14.47
21017020	135	192	241	391	409	421	0.95	116.0	44.1	399.5	14.2	7.44	0.04	0.00	470.6	670.6	8.62
21017040	1041	1244	1369	1524	1639	1752	0.23	688.2	49.6	1048.7	162.2	52.63	0.03	0.00	588.5	788.5	7.35
21037010	369	466	529	609	668	726	0.09	234.2	3.1	349.9	83.5	24.93	0.03	0.00	597.1	832.1	9.54
21047010	1363	1673	1876	2133	2325	2517	0.73	1198.9	290.3	1392.5	234.7	66.54	0.02	0.00	618.4	818.4	0.28
21057050	277	329	362	403	433	464	0.21	177.2	49.9	273.1	43.0	7.18	0.02	0.00	457.3	647.3	13.28
21107030	24	45	62	81	95	108	0.66	17.0	5.9	41.4	19.0	4.83	0.05	0.01	355.8	555.8	16.22
21107020	120	190	263	336	360	381	0.93	96.0	50.4	325.6	24.8	7.62	0.05	0.01	529.9	749.9	4.43
21127020	64	94	121	305	391	451	0.95	53.9	23.2	348.1	71.1	19.40	0.05	0.01	427.3	627.3	40.45
21127030	32	52	80	110	131	151	0.74	26.7	5.7	60.3	28.1	10.80	0.04	0.01	355.6	555.6	37.75
21137030	378	512	599	710	791	872	0.05	155.4	6.8	344.5	116.0	28.15	0.04	0.00	532.6	727.6	1.59
21147030	459	598	694	822	920	1021	0.82	406.0	108.3	483.0	168.6	25.99	0.03	0.00	532.4	727.4	0.17
21147010	532	740	912	1833	2619	3075	0.96	462.2	161.8	2383.5	528.1	128.40	0.04	0.00	656.3	881.3	39.43
21057030	645	768	874	1033	1153	1272	0.79	527.4	170.4	684.2	44.2	28.44	0.03	0.00	566.2	766.2	3.40
22027020	121	158	177	198	212	227	0.75	104.3	18.5	158.7	20.7	4.20	0.02	0.00	406.2	601.2	2.20
22027010	200	281	364	401	424	446	0.86	175.7	39.3	365.5	29.7	9.17	0.03	0.00	476.0	676.0	9.33
22057060	897	1087	1210	1366	1482	1597	0.05	559.6	3.1	849.6	164.3	22.51	0.02	0.00	661.7	896.7	7.55
22057040	1108	1526	1884	2121	2273	2418	0.84	989.2	184.3	1866.1	200.6	67.83	0.03	0.00	684.2	909.2	9.03
22067010	177	295	406	592	735	874	0.88	135.0	75.0	400.8	191.1	33.86	0.07	0.01	531.1	731.1	8.16
22077060	62	94	168	189	190	190	0.90	51.6	20.1	188.8	0.6	13.02	0.05	0.01	390.1	585.1	39.67
21217120	71	115	148	174	192	209	0.82	59.7	16.0	140.5	23.6	5.83	0.04	0.00	404.8	599.8	10.26
23017030	290	561	784	1039	1219	1396	0.69	201.5	86.2	546.9	248.3	45.76	0.07	0.01	573.1	773.1	12.77
23067020	859	1366	1725	2188	2534	2879	0.61	595.4	289.6	1040.4	500.1	129.47	0.05	0.01	937.6	1232.6	2.99
21137020	44	51	56	62	67	71	0.07	29.8	0.8	42.4	6.3	1.53	0.02	0.00	305.8	505.8	4.01
21227010	211	251	305	383	440	496	0.78	145.0	80.8	230.4	7.6	13.03	0.08	0.02	469.9	659.9	15.64
21257070	267	396	480	586	665	742	0.07	50.8	4.4	238.2	111.3	30.11	0.05	0.00	527.6	722.6	6.89
22077030	52	77	104	153	189	224	0.85	43.5	14.7	94.1	48.3	18.23	0.04	0.00	494.5	744.5	26.71
21187020	133	209	269	408	661	820	0.97	107.2	61.4	642.9	170.3	38.07	0.10	0.02	620.2	870.2	20.09
21197030	285	412	497	603	683	761	1.00	243.6	112.5	316.0	115.8	31.25	0.07	0.01	690.4	950.4	0.14
21197110	104	140	161	187	207	226	0.29	47.4	9.6	109.4	27.4	12.03	0.04	0.00	428.7	623.7	8.07
23037010	4251	4650	5028	5545	5928	6308	0.85	3901.8	543.0	4458.3	75.6	105.04	0.01	0.00	1054.8	1379.8	7.20

ESTACIÓN	Tr(años)						Parámetros					PRUEBAS DE AJUSTE					Desviación Estadística D
	2	5	10	25	50	100	P	μ_1 (m ³ /s)	α_1 (m ³ /s)	μ_2 (m ³ /s)	α_2 (m ³ /s)	EEA (m ³ /s)	MADI	MSDI	CIA	BIC	
21167050	89	131	149	170	185	200	0.68	67.7	18.1	126.7	21.1	6.07	0.04	0.00	424.2	624.2	3.03
21187030	228	380	478	598	687	775	0.39	110.7	63.2	261.1	125.3	23.17	0.05	0.01	545.6	745.6	1.62
21237020	3455	3938	4254	4652	4947	5240	0.05	2551.5	89.4	3333.2	419.1	51.66	0.01	0.00	1222.5	1607.5	3.57
21237040	183	358	463	590	684	776	0.43	70.7	48.9	246.0	131.2	35.99	0.10	0.02	533.3	728.3	5.34
24017610	20	40	49	59	66	72	0.73	15.1	5.4	42.4	9.1	2.05	0.04	0.00	412.9	662.9	13.89
21087040	75	131	197	377	385	392	0.93	55.4	39.5	374.4	9.2	7.82	0.12	0.03	442.8	632.8	25.15
23197270	27	56	74	96	111	127	0.42	9.8	6.7	37.3	22.1	6.26	0.10	0.03	391.5	591.5	8.75
23197290	112	175	227	266	292	317	0.84	94.0	28.1	222.3	34.3	5.57	0.02	0.00	475.1	685.1	7.41
24027050	260	366	390	418	440	467	0.79	217.5	54.6	370.0	20.6	7.22	0.02	0.00	489.4	689.4	7.31
24037080	13	22	34	69	75	80	0.91	10.4	5.5	65.4	7.0	2.73	0.09	0.01	337.5	552.5	36.29
24037040	27	38	46	55	62	69	0.48	20.2	5.0	30.3	9.8	3.05	0.03	0.00	436.9	711.9	4.23
24017580	163	191	216	256	286	317	0.47	149.4	43.8	157.1	14.6	7.17	0.02	0.00	485.1	715.1	7.94
24037070	12	22	27	32	35	39	0.65	7.9	3.0	21.0	5.1	0.96	0.05	0.01	337.0	572.0	8.66
24037450	4	8	17	27	34	40	0.81	2.6	1.9	14.1	9.0	3.99	0.12	0.03	284.5	514.5	46.11
24037120	13	20	27	54	58	62	0.94	10.6	5.0	54.6	4.1	1.63	0.08	0.01	335.7	565.7	26.49
24037300	15	22	27	36	59	60	0.98	13.2	5.6	59.4	0.3	1.01	0.04	0.00	282.7	477.7	12.80
24027040	47	89	104	117	126	135	0.77	38.0	10.4	96.5	12.3	3.45	0.04	0.00	374.9	569.9	14.68
24037110	17	25	34	54	69	83	0.80	15.0	4.3	24.2	19.9	3.46	0.06	0.01	359.4	594.4	24.44
23107130	63	99	122	151	172	193	0.15	17.4	2.0	59.3	30.1	6.27	0.05	0.00	444.2	654.2	11.64
24067030	2391	2829	3103	3445	3697	3946	0.20	1647.3	110.0	2384.8	357.0	130.15	0.02	0.00	635.0	830.0	5.65
24047020	1845	2085	2223	2392	2516	2638	0.36	1300.2	133.3	1912.6	174.7	39.08	0.01	0.00	586.0	776.0	14.38
24027010	582	695	768	860	929	996	0.10	357.0	59.7	561.0	96.8	13.87	0.02	0.00	739.7	1019.7	7.71
24027030	248	321	365	421	461	502	0.25	157.5	14.3	253.4	57.6	15.83	0.04	0.00	489.6	689.6	5.59
23097030	5535	6396	6982	7896	8580	9259	0.86	4940.6	970.8	6127.8	177.9	91.20	0.01	0.00	1302.8	1677.8	7.54
23087150	42	64	80	99	113	127	0.58	31.5	9.1	53.2	19.8	4.70	0.04	0.00	358.7	548.7	4.44
23087200	201	265	305	356	394	431	0.14	78.0	9.1	193.5	53.5	15.98	0.04	0.00	466.9	656.9	8.37
23087180	73	90	99	114	128	143	0.71	50.0	21.9	84.9	6.0	3.00	0.05	0.01	376.8	566.8	12.40
21057060	1051	1250	1399	1620	1791	1963	0.79	886.5	246.8	1112.8	85.1	84.97	0.04	0.00	599.5	799.5	2.46
23087190	607	681	729	790	834	878	0.10	459.2	2.5	594.1	63.3	10.48	0.01	0.00	466.2	656.2	18.06
23127020	1363	1700	1922	2204	2413	2620	1.00	1254.1	297.0	1350.3	422.9	54.52	0.03	0.00	707.9	937.9	0.02

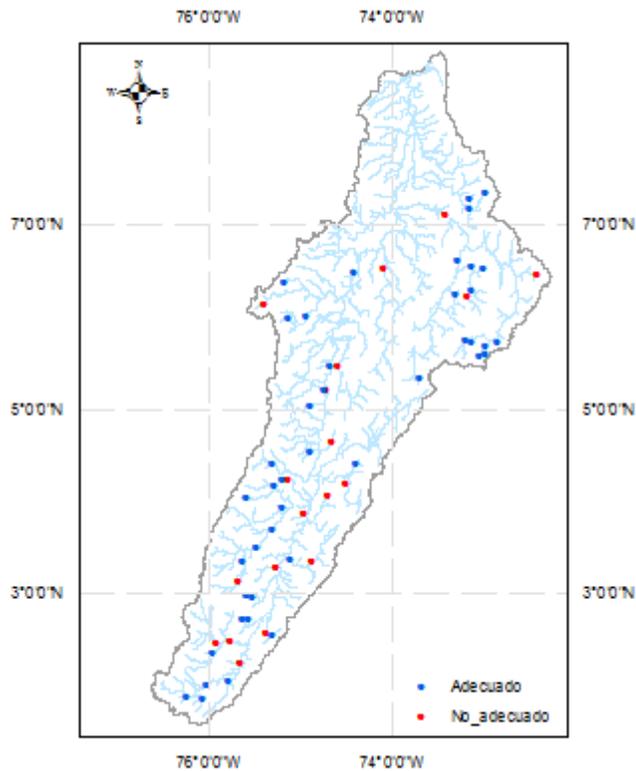


Figura 5.1 Localización de las estaciones en la cuales es adecuado el uso del modelo doble Gumbel

5.3.1 Análisis de sensibilidad de las distribuciones mezcladas a la longitud del registro

El análisis de frecuencia de gastos máximos tiene como objetivo la estimación de los eventos de diseño asociados a diferentes periodos de retorno para el dimensionamiento de obras hidráulicas tales como vertederos, obras de desvío, obras de protección contra inundaciones, etc. Aun cuando varias funciones de probabilidad han sido sugeridas, no hay una función que sea universalmente aceptada (Mtiraoui, 2004).

Las simulaciones de Monte Carlo han sido empleadas en varios estudios para evaluar la eficiencia de un modelo bajo ciertos índices estadísticos como el error cuadrático medio, la varianza, el sesgo, entre otros.

Las simulaciones realizadas permitirán evaluar la eficacia del modelo doble Gumbel en:

- Las estaciones que presentan altos valores en el L-Asimetría ($L - Asimetría \geq 0.3$) y cuya función de densidad no paramétrica presenta una forma de distribución de cola pesada.
- Las estaciones con bajos valores en el L-Asimetría y cuya forma de la función de densidad no paramétrica es multimodal.
- El efecto de la longitud del registro en la estimación de eventos de diseño contenidos en la muestra.

En muestras heterogéneas, las simulaciones de Monte Carlo han sido utilizadas para evaluar la eficiencia de un modelo mezclado de dos componentes empleando la distribución Normal (Leytham, 1984) y para evaluar la eficiencia de un modelo mezclado de dos componentes empleando la distribución Log-normal de dos parámetros (Mtiraoui, 2004).

5.3.1.1 Diseño de la simulación

La simulación empleada en este estudio será del tipo paramétrica en la cual se le ajusta a los datos de la estación de estudio una determinada función de distribución de probabilidad y a partir de esta se generan muestras asociadas a diferentes longitudes y periodos de retorno.

Los pasos para la simulación son los siguientes:

- Se elige una distribución de probabilidad para la generación de las muestras sintéticas. En este estudio se empleará la distribución de Valores Extremos de Dos Componentes (TCEV) la cual ha sido concebida para ser utilizada en muestras heterogéneas (Rossi, 1984).
- Para generar un gasto de determinado periodo de retorno se elige un número aleatorio entre 0 y 1. Este número aleatorio será la probabilidad de no excedencia conocida o el valor de $F(x)$ de la ecuación (5-8), a partir de la cual se estima el gasto.
- Se define el tamaño de la muestra, que para este estudio serán de 20, 30, 40, 50, 60, 70, 80, 90 y 100 años y se generan 1000 series para cada tamaño de muestra.
- A cada muestra generada se le ajusta la distribución doble Gumbel y se estiman algunos eventos de diseño los cuales están contenidos en los datos de la estación de estudio.
- Finalmente se calcula la raíz del error cuadrático medio en porcentaje mediante la siguiente ecuación:

$$RECM(\%) = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} \left[\frac{(Q_i - Q_T)}{Q_T} \right]^2} \cdot 100 \quad (5-9)$$

Donde Q_i son los eventos estimados para diferentes periodos de retorno de las muestras sintéticas y Q_T son los eventos de diseño de diferentes periodos de retorno de la estación de estudio.

5.3.1.2 Resultados

Para las estaciones que tiene un bajo valor en el L-Asimetría se toma la estación 24037070 ($L - Asimetría = 0.24$) para el análisis. En la Figura 5.2 se presentan la función de densidad no paramétrica y la función de densidad obtenida con el modelo doble Gumbel para la estación 24037070, en la Figura 5.3 se presenta el RECM (%) para los diferentes tamaños de muestra y los diferentes periodos de retorno evaluados.

Para esta estación los RECM son menores al 30% para todos los tamaños de muestra, sin embargo, para tamaños de muestra mayores a 60 años los RECM son menores al 18%.

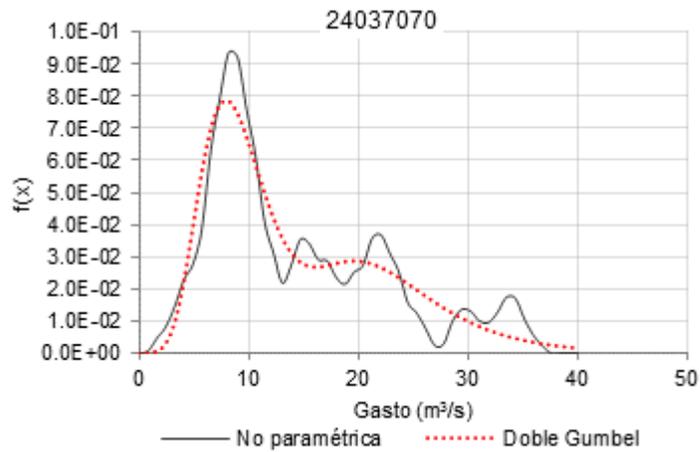


Figura 5.2 Función de densidad no paramétrica y función de densidad estimada para el modelo doble Gumbel (24037070)

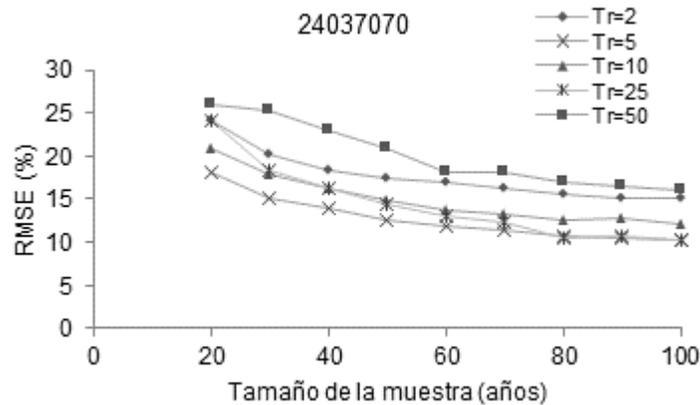


Figura 5.3 RECM para diferentes periodos de retorno y tamaños de muestra (24037070)

Para las estaciones que tienen un alto valor en el L-Asimetría se toma como referencia la estación 24037110 ($L - Asimetría = 0.41$) para el análisis. En la Figura 5.4 se presentan la función de densidad no paramétrica y la función de densidad obtenida con el modelo doble Gumbel para la estación 24037110, en la Figura 5.5 se presenta el RECM (%) para los diferentes tamaños de muestra y los diferentes periodos de retorno evaluados. Para periodos de retorno mayores a 20 años, en general, para todos los tamaños de muestra se presentan RECM mayores al 30%, lo que puede indicar que hay una sub-estimación en los eventos de diseño para grandes periodos de retorno. Para eventos de diseño asociados a periodos de retorno menores a 20 años y tamaños de muestra mayores a 60 años el RECM es menor al 15%. Por lo tanto, para muestras con altos coeficientes de asimetría (L-Asimetría) y registros menores de 60 años, lo recomendable es realizar la estimación de los eventos de diseño a partir de métodos regionales. Si bien, las distribuciones mezcladas son las más adecuadas para la estimación de los eventos en sitios donde los eventos extremos provienen de más de un mecanismo, en sitios con pocos registros, lo más adecuado es usarla a partir de técnicas regionales como la técnica de las estaciones-año y de esta manera se ataca el problema de la parsimonia.

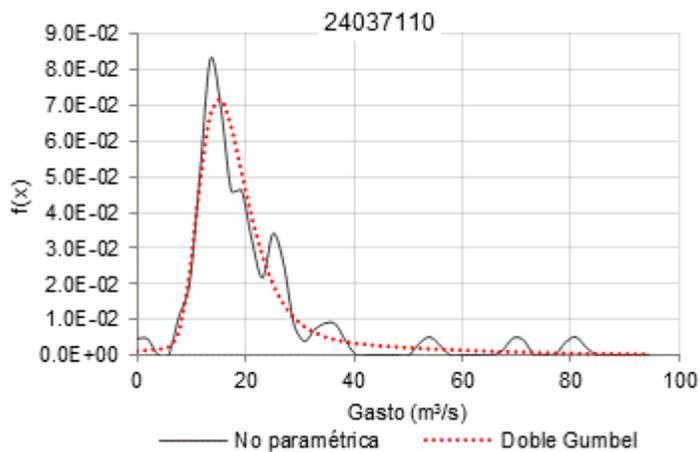


Figura 5.4 Función de densidad no paramétrica y función de densidad estimada para el modelo doble Gumbel (24037110)

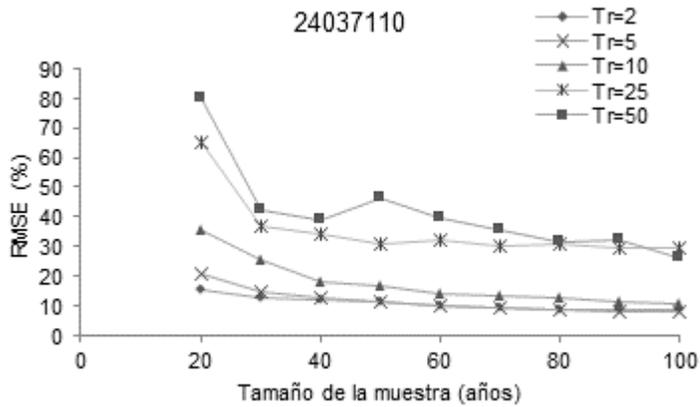


Figura 5.5 RECM para diferentes periodos de retorno y tamaños de muestra (24037110)

El análisis realizado permite concluir que en muestras que exhiben una función de distribución de probabilidad no paramétrica de más de un modo y que tienen un L-Asimetría bajo, la distribución doble Gumbel para muestras de más de 60 años puede dar errores en la estimación menores al 20%. En muestras con menos de 60 años de registro con altos L-Asimetría y con forma de la distribución de densidad no paramétrica de cola pesada, la estimación de los eventos de diseño, para periodos de retorno mayores a 20, con la distribución doble Gumbel puede tener una alta incertidumbre asociada y por lo tanto, lo recomendable en estos sitios es estimar los eventos a partir de métodos regionales.

6 ANÁLISIS REGIONAL DE FRECUENCIA DE GASTOS MÁXIMOS

6.1 INTRODUCCIÓN

El análisis de frecuencia es un problema en hidrología porque la información suficiente muy pocas veces está disponible en un sitio para determinar adecuadamente la frecuencia de los eventos extremos (Stedinger et al, 1991). En muchos sitios, sobre todo en cuencas pequeñas, la información no está disponible. Porque generalmente los registros hidrológicos disponibles son cortos y presentan no estacionariedad o no existen, se hace necesario el empleo de técnicas de regionalización para una adecuada estimación de los gastos de diseño (NERC, 1975; Hosking y Wallis, 1997; IH, 1999; Ilorme y Griffis, 2013). De esta manera, la información de las estaciones vecinas compensa, al menos en una parte, la poca cantidad de registros de un sitio en particular (NRC, 1988). El análisis regional de frecuencia permite estimar los eventos de diseño de un sitio a partir de los gastos de diseño de sitios vecinos a la cuenca que cuenten con información e incluso con información del mismo sitio (Cunnane, 1988; Saf, 2009). La estimación de eventos de diseño mayores a la longitud del registro disponible, resultan en una extrapolación en la cual el error estándar incrementa rápidamente con el incremento en el periodo de retorno (Rao y Hamed, 1997). Varios estudios han demostrado que el uso adecuado de las técnicas de regionalización permite obtener una mejor estimación de los gastos de diseño, en algunos casos con hasta la mitad del error estándar (Stedinger et al, 1991).

La mayor parte de los modelos regionales propuestos en la literatura requieren que la región considerada en el estudio sea homogénea de alguna manera cuantificable. Tanto las técnicas de delimitación de regiones homogéneas como los modelos regionales hacen uso de las características fisiográficas de las cuencas.

6.2 REGIÓN HOMOGÉNEA Y SU DELIMITACIÓN

De todas las etapas del análisis regional de frecuencias, la identificación de regiones homogéneas es a menudo la más complicada y crítica, al requerir de la toma de decisiones subjetivas (Hosking y Wallis, 1997). La formación de grupos es una de las tareas más importantes en las técnicas de regionalización (Burn y Goel, 2000; Ilorme y Griffis, 2009), sin embargo, es altamente dependiente de las variables empleadas para su delimitación (Burn, 1990; Burn et al, 1997; Castellarin et al, 2001).

El objetivo que se plantea es formar grupos de estaciones que satisfagan aproximadamente la condición de homogeneidad, esto es, que compartan la misma distribución de frecuencias excepto por un factor de escala (Hosking y Wallis, 1997). En general, la delimitación regional se ha sustentado en considerar áreas geográficamente continuas, límites políticos o administrativos o límites fisiográficos (Nathan y McMahon, 1990; Hosking y Wallis, 1997; Burn y Goel, 2000). Sin embargo, las regiones formadas de esta manera no necesariamente serán homogéneas en términos de la respuesta hidrológica debido a posibilidad de que exista una gran variabilidad espacial de las características fisiográficas e hidrológicas en este tipo de regiones (Burn y Goel, 2000).

Los requerimientos para asegurar una efectiva transferencia de información de una región a otra son los siguientes (Burn y Goel, 2000):

- Que la región sea hidrológicamente homogénea. Este requerimiento se deriva de la necesidad de asegurar que los eventos extremos que son transferidos al sitio objetivo son similares a los que se han presentado en el sitio.
- La región debe ser identificable, esto implica que se pueda definir una región para una nueva cuenca, la cual puede no estar aforada.
- La región puede ser suficientemente grande. Sin embargo, grandes regiones implican que una mayor cantidad de información de gastos extremos sea incorporada al análisis. Se ha sugerido (Jakob et al, 1999) que una región idealmente podrá contener $5T$ estaciones-año de registros en orden de proporcionar una efectiva estimación del evento de diseño de un periodo de retorno de T años. A medida que el tamaño de la región se incrementa, la homogeneidad tiende a decrecer. Por lo tanto, hay una compensación entre el primer y el tercer requerimiento.

6.2.1 Problemas relacionados con la formación de regiones

Algunos de los problemas relacionados con la formación de grupos o regiones son:

- Un problema común en las técnicas de regionalización, es el relacionado con la selección de las variables para evaluar el grado de similitud entre las diferentes cuencas (Nathan y McMahon, 1990; Ilorme y Griffis, 2009) y por la necesidad de emplear las características que sean indicativas de la respuesta hidrológica y que permitan incluir un sitio no aforado en una región homogénea ya delimitada (Burn, 1997; Burn et al, 1997). Algunas cuencas son similares con respecto a cierta variable, como la cubierta forestal o la lluvia anual, pero no a otras, como las características fisiográficas. Además, con el análisis por grupos, literalmente cualquier conjunto de variables es capaz de generar grupos. Lo cual es muy útil ya que de esta manera se podrá seleccionar las variables de acuerdo con la importancia del problema (Nathan y McMahon, 1990). Medidas de similitud derivadas de las series de gastos máximos como los parámetros estadísticos de los gastos máximos y propiedades relacionadas con la estacionalidad de los gastos no deben ser empleadas (Burn, 1990; Hosking y Wallis, 1997).
- La medida de la similitud de los grupos es altamente dependiente de la escala de las variables empleadas para su formación (Nathan y McMahon, 1990). Así, por ejemplo, es posible que variables como el área tengan una mayor importancia en la medida de la similitud del grupo que variables como la precipitación media anual o la pendiente del cauce principal. Una solución que se ha propuesto a este problema es estandarizar o re-escalar las variables empleadas (Nathan y McMahon, 1990; Zrinji y Burn, 1996; Hosking y Wallis, 1997; Kar et al, 2012). Sin embargo, este procedimiento puede no ser apropiado ya que algunas variables pueden tener un mayor efecto en la forma de la distribución de la región (Hosking y Wallis, 1997) o se puede perder cualquier información (posiblemente útil) de la variabilidad de una variable determinada (Nathan y McMahon, 1990).
- El mapa del coeficiente de variación (CV) o el coeficiente de variación obtenido con los momentos-L (L-CV) se puede emplear para darse una idea inicial en la formación de grupos (Dales y Reed, 1989). Sin embargo, la variación de estos parámetros generalmente está asociada a la variación de los registros; altos valores en muestras

cortas tienen una influencia apreciable en los valores del CV y el L-CV (Dales y Reed, 1989; Hosking y Wallis, 1997). Como la tendencia es agrupar los sitios con altos valores de CV o L-CV el problema surge porque estos valores pueden presentarse en muestras cortas con valores extremos atípicos donde estos tienen una gran influencia en el valor de estos parámetros o en muestra con altos valores atípicos y por lo tanto, altos valores de L-CV, que pueden presentarse por fluctuaciones aleatorias que afectaron el sitio y no a sus vecinos (Hosking y Wallis, 1997).

- Las regiones que son formadas empleando técnicas de formación de grupos generalmente no son estadísticamente homogéneas (Rao y Srinivas, 2006) y las regiones que son estadísticamente homogéneas no lo son físicamente (Ilorme y Griffis, 2009).

Puede pensarse que la selección de la región es un problema trivial, sin embargo, la inclusión o exclusión de información dentro del modelo regional puede producir estimadores de los eventos de diseño poco confiables.

6.2.2 Técnicas de formación de grupos

A continuación se presentan cuatro criterios que pueden emplearse para la formación de regiones. Cabe mencionar que estas técnicas no son excluyentes entre sí, sino que manejadas adecuadamente pueden ser complementarias y definir de manera más confiable la región.

6.2.2.1 Método geográfico

Esta técnica se basa en suponer que una región es homogénea, dada la proximidad de sus cuencas, donde la delimitación se puede realizar trazando un radio de 80 km a partir del centro de gravedad del sitio de proyecto y tomando toda la información proveniente de los sitios que quedan dentro de este radio (Escalante y Reyes, 2008). Esta técnica es recomendable cuando se están realizando estudios de pequeña escala.

6.2.2.1.1 Método de los trazos multidimensionales

Para resolver los problemas de la selección de una adecuada técnica de formación de grupos, de la definición de regiones homogéneas, y de la selección de la región a la cual pertenecen las cuencas bajo estudio, Nathan y McMahon, (1990) propusieron una técnica que se basa en los trazos multidimensionales propuestos por Andrews, (1972).

La heterogeneidad de los grupos que se forman de manera preliminar se evalúa mediante la técnica de los trazos multidimensionales, en la cual un punto en el espacio multidimensional se representa por una curva en dos dimensiones vía la función:

$$f(t) = \frac{X_1}{\sqrt{2}} + X_2 \sin(t) + X_3 \cos(t) + X_4 \sin(2t) + X_5 \cos(2t) + \dots \quad (6-1)$$

Donde X_1, X_2, \dots son las variables fisiográficas y/o climáticas definidas para el análisis de formación de grupos. La función se evalúa en el rango $-\pi \leq t \leq \pi$.

El hecho de que esta función preserve las distancias la hace una técnica ideal de comparación visual para la formación de grupos homogéneos (Escalante y Reyes, 2008). Los grupos de cuencas con comportamientos similares aparecen como una banda de curvas muy próximas unas de otras.

Una característica del método que no se identifica de forma inmediata, consiste en que los resultados que se obtiene dependen del tipo y del ordenamiento de las variables seleccionadas. Las primeras variables son asociadas con componentes cíclicas de baja frecuencia y las últimas de alta frecuencia. Las bajas frecuencias son más fáciles de observar, de esta manera, X_1 representará a aquella variable que, por ejemplo, en el análisis de regresión resultó ser más significativa desde el punto de vista estadístico, X_2 a la segunda y así sucesivamente.

Una vez que se han desarrollado las funciones $f(t)$ para cada una de los sitios involucrados, se generan envolventes que formen conjuntos de curvas de comportamiento similar y se deriva una curva media $\bar{f}(t)$ que represente los atributos físicos de cada grupo formado. Para las curvas donde se tenga duda sobre si pertenece o no a un grupo particular se hará la siguiente prueba:

$$SS = \sum_{t=-\pi}^{\pi} \{ [f^j(t)] - [\bar{f}(t)]_k \}^2 \quad (6-2)$$

Donde SS es la suma de las diferencias al cuadrado entre los puntos que forman la curva en el sitio j , y la curva característica $\bar{f}(t)$ de cada uno de los k grupos ya formados. La cuenca se localizará en la región k donde se haya obtenido el menor valor de SS .

6.2.2.2 Método de la región de influencia

En esta técnica cada sitio se considera el centro de su propia región (Burn, 1988, 1990; Zrinji y Burn, 1996). La identificación de una región de influencia se apoya en la medición de la distancia Euclidiana en un espacio de atributos multidimensionales. El conjunto de atributos (características fisiográficas y/o climatológicas) se relaciona a las características de los eventos extremos. También se define una función de peso que refleja la importancia relativa de cada sitio dentro de la región.

La distancia Euclidiana D_{jb} entre una estación cualquiera j y la estación base b , para la cual se genera la región de influencia, se expresa como

$$D_{jb} = \left[\sum_{i=1}^p W_i (C_j^i - C_b^i)^2 \right]^{1/2} \quad (6-3)$$

$$C_j^i = \frac{X_j^i}{S(X^i)}$$

$$C_b^i = \frac{X_b^i}{S(X^i)}$$

Donde

p número de atributos considerados

C_j^i y C_b^i valores estandarizados del atributo i para los sitios j y la estación base b

X_j^i valores del atributo i del sitio j

X_b^i valores del atributo i de la estación base b

$S(X^i)$ desviación estándar del atributo i considerando todos los sitios j y la estación b .

W_i peso asociado con el atributo i .

El proceso de estandarización elimina las unidades de cada atributo y reduce cualquier diferencia de escala del rango de valores.

El grupo de atributos seleccionados son función de la información disponible en la red de estaciones. La elección de los atributos requiere cierto juicio ingenieril, sin embargo, pueden obtenerse como en la técnica de delimitación multidimensional al examinar la correlación entre las características estadísticas de los eventos extremos y las características fisiográficas y/o climatológicas.

El próximo paso dentro del proceso de identificación de la región de influencia implica proponer un valor de tolerancia que acote la media de la distancia. Así, se pudieran excluir todas aquellas estaciones que tengan una distancia más grande que el valor promedio \bar{D}_{jb} desde el sitio base.

6.2.2.3 Método de Ward

Es una de las técnicas más utilizadas, ya que tiende a formar grupos más o menos equivalentes en tamaño (Hosking y Wallis, 1997; Ilorme y Griffis, 2013), y es considerada adecuada para la formación de grupos en estudios de gastos máximos (Hosking y Wallis, 1997, pp. 58–59).

El método de Ward (Ward, 1963; Johnson and Wichern 2002; Ilorme y Griffis, 2013), es un algoritmo de aglomeración jerárquica el cual inicialmente considera que el sitio sirve en su propio grupo o región. El algoritmo combina grupos mediante un análisis enfocado en la varianza en la cual la similitud entre los miembros de una región es medida en términos de la suma de los cuadrados del error (ESS). Para una región k que contiene N_k sitios, el ESS es calculado como:

$$ESS_k = \sum_{j=1}^{N_k} (X_j - \bar{X})^T (X_j - \bar{X}) \quad (6-4)$$

Donde $X_j = [x_1, x_2, \dots, x_p]^T$ es un vector de p atributos medidos en el sitio j , y $\bar{X} = [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p]^T$ en el que cada elemento representa el valor medio de los atributos de los N_k sitios de una región. En cada paso del análisis, se considera la unión de todos los pares posibles de agrupaciones, y los dos grupos cuya combinación da como resultado el menor incremento en ESS_k (mínima pérdida de información) se unen.

6.3 VARIABLES EMPLEADAS PARA LA FORMACIÓN DE GRUPOS

Para la aplicación de los métodos de formación de grupos, cada sitio debe ser presentado con sus características individuales (variables). Algunos atributos comúnmente empleados son (Burn, 1990; Hosking y Wallis, 1997; Escalante y Reyes, 2008; Kar et al, 2012):

- Variables físicas: área de la cuenca, perímetro de la cuenca, pendiente media del cauce principal, pendiente media de la cuenca, latitud, longitud, longitud del cauce principal, tiempo de concentración, orden de la cuenca, coordenadas del centroide de la cuenca, elevación máxima, elevación mínima, variables relacionadas con el tipo de suelo de la cuenca.

- Variables meteorológicas: Precipitación media anual, precipitación media mensual multianual, días de precipitación apreciable, precipitación media de las máximas de 24 horas, entre otras.

Para la selección de las variables que mejor describen la respuesta hidrológica de la zona de estudio se puede emplear el análisis de correlación múltiple entre los eventos de diseño para diferentes periodos de retorno obtenidos del análisis en el sitio y las variables físicas y/o meteorológicas obtenidas para cada uno de los sitios. También se puede emplear para definir las variables para la formación de grupos el análisis por componentes principales.

6.4 MEDIDA DE LA DISCORDANCIA

Si un sitio es discordante con otro aparentemente similar, entonces puede existir un problema con los datos del sitio discordante. Para la comparación de los datos de diferentes sitios, algunas técnicas, tales como la curva de doble masa o los gráficos cuantil-cuantil son bien conocidos (Hosking y Wallis, 1997). En el contexto del análisis regional se evaluarán la medida de la discordancia clásica (Hosking y Wallis, 1997) y la medida de la discordancia robusta (Neykov et al, 2007) para detectar los sitios discordantes dentro de una región.

A nivel práctico, se recomienda el empleo de la medida de discordancia en dos ocasiones. En primer lugar, al comienzo del análisis, procurando abarcar la totalidad de las estaciones o regiones suficientemente amplias como para que su número sea elevado. Con esta primera aproximación se trata de detectar aquellos sitios que difieren significativamente de la mayoría, sea por la presencia de valores atípicos o por la existencia de errores de transcripción de los datos, cambios de los sitios de medición o cambios inducidos por el hombre. (Hosking y Wallis, 1997).

Posteriormente, una vez identificadas las diferentes regiones homogéneas, es aconsejable recalcular la medida de la discordancia de cada estación en su región asignada. En aquellos sitios que aquí resulten discordantes habrá que considerar la posibilidad de trasladarlos a otra región, si bien esta opción no prevalecerá sobre las similitudes físicas con el resto de sitios de su región asignada. De hecho, son habituales los eventos que afectan únicamente a algunas estaciones de una determinada región homogénea desde el punto de vista físico, y ello puede dar pie a pensar erróneamente que las estaciones afectadas son discordantes con el resto (Hosking y Wallis, 1997).

6.4.1 Medida de la discordancia clásica

La medida de la discordancia clásica se establece en términos de las relaciones τ_2, τ_3 y τ_4 de los momentos-L. Si se considera (Hosking y Wallis, 1997):

$$U_i = [\tau_2^i, \tau_3^i, \tau_4^i]^T \quad (6-5)$$

$$\bar{U} = \frac{1}{ms} \sum_{i=1}^{ms} U_i \quad (6-6)$$

$$SC = \frac{1}{ms} \sum_{i=1}^{ms} (U_i - \bar{U})(U_i - \bar{U})^T \quad (6-7)$$

Donde

U_i vector que contiene las relaciones de los momentos-L de cada sitio i en la región

\bar{U} promedio no pesado de los valores U_i

SC matriz de covarianza muestral de los valores U_i

m número total de estaciones en la región

T indica transpuesta

Entonces la medida de discordancia D_i de un sitio i se define como

$$D_i = \frac{1}{3} (U_i - \bar{U})^T (SC)^{-1} (U_i - \bar{U}) \quad (6-8)$$

6.4.2 Medida de la discordancia robusta

La medida de la discordancia clásica es estimada en términos de la media muestral y de la matriz de covarianza muestral de relaciones τ_2, τ_3 y τ_4 de los momentos-L. Es bien sabido, sin embargo, que la media muestral y la matriz de covarianza muestral, así como cualquier análisis basado en estas, puede ser fuertemente afectado por incluso algunos pocos datos atípicos (Rousseeuw y Leroy, 1987; Saf, 2010). Para resolver los problemas derivados de la media muestral y de la matriz de covarianza, Neykov et al, (2007) proponen el reemplazo de estos estimadores por el uso del estimador del determinante de la covarianza mínima (MCD). La medida de la discordancia robusta se estima como (Neykov et al, 2007):

$$RD^2(U_i) = (U_i - \bar{U}_h)^T C^{-1} (U_i - \bar{U}_h) \quad (6-9)$$

Donde el vector $U_i = [\tau_2^i, \tau_3^i, \tau_4^i]^T$ contiene las relaciones de los momentos L del sitio i , el vector $\bar{U}_h = [\bar{\tau}_2, \bar{\tau}_3, \bar{\tau}_4]^T$ contiene el valor medio de las relaciones de los momentos L para todos los sitios h dentro del subconjunto, y la matriz de covarianza del subconjunto se estima como sigue:

$$C = C_m \frac{1}{h} \sum_{j=1}^h (U_j - \bar{U}_h)(U_j - \bar{U}_h)^T \quad (6-10)$$

El factor de C_m se elige de tal manera que sea consistente con el modelo normal multivariado y sea insesgado con muestras pequeñas (Pison et al, 2002). De acuerdo a Neykov et al, (2007) la medida de la discordancia robusta es mejor para detectar sitios discordantes comparada con la medida de la discordancia clásica y por lo tanto será la utilizada en este estudio. La estimación de la discordancia robusta se hará empleando el paquete rrcov disponible el software R (Todorov, 2009).

6.4.3 Valores críticos

Los valores críticos para la discordancia D_i dependiendo del número de estaciones en la región se presentan en la Tabla 6.1 (Hosking y Wallis, 1997).

Tabla 6.1 Valores críticos de la discordancia D_i

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

Una vez que se ha delimitado la región de influencia para un sitio base, se debe determinar su homogeneidad.

6.5 EVALUACIÓN DE LA HOMOGENEIDAD HIDROLÓGICA

La confiabilidad del análisis regional de frecuencia es altamente dependiente del grado de homogeneidad hidrológica (Saf, 2009). Para determinar la homogeneidad regional o para evaluar si una región propuesta es homogénea desde el punto de vista hidrológico, la prueba estadística H , llamada medida de la heterogeneidad, fue propuesta por Hosking y Wallis, (1993). Se comparan las variaciones entre sitios en los momentos L de la muestra para un grupo de sitios con los cuales se espere se forme una región completamente homogénea. La variación entre sitios de la relación de los momentos L es medida como la desviación estándar (V) del $L - CV$ en el sitio, ponderado proporcionalmente con la longitud de los registros del sitio. Los promedios regionales de las relaciones de los momentos L son calculadas por las siguientes expresiones (Hosking y Wallis, 1997):

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

$$t_3^R = \frac{\sum_{i=1}^N n_i t_3^{(i)}}{\sum_{i=1}^N n_i} \quad (6-11)$$

$$t_4^R = \frac{\sum_{i=1}^N n_i t_4^{(i)}}{\sum_{i=1}^N n_i}$$

Donde N es el número de estaciones en la región, para la i -ésima estación con longitud de muestra n_i y relaciones de momentos L ($t^{(i)}, t_3^{(i)}, t_4^{(i)}, i = 1, 2, \dots, N$). La desviación estándar ponderada de la relación de momentos L en una región es calculada mediante la siguiente expresión:

$$V = \left\{ \frac{\sum_{i=1}^N n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^N n_i} \right\}^{1/2} \quad (6-12)$$

El cálculo de la desviación estándar es llevado a cabo mediante un proceso de simulación ajustando la distribución Kappa de cuatro parámetros al conjunto de estaciones de la región. La variación entre sitios de cada región es calculada y la media (μ_v) y la

desviación estándar (σ_v) de la variación entre sitios es obtenida. Entonces la medida de la heterogeneidad H es calculada como:

$$H = (V_{obs} - \mu_v) / \sigma_v \quad (6-13)$$

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

6.6 TÉCNICAS DE ESTIMACIÓN REGIONAL

El análisis regional hidrológico suministra la forma de inferir eventos para ciertos periodos de retorno en sitios con escasa o nula información, a través del empleo de información de los sitios de la región e incluso información del sitio de estudio (Cunnane, 1988). Algunos métodos empleados para el análisis regional de frecuencia, especialmente el de la *avenida índice* asumen que la región es un conjunto de sitios de medición cuyo comportamiento de la frecuencia de los gastos máximos es homogéneo de alguna manera cuantificable. El análisis de frecuencia regional explota esta homogeneidad para estimar los eventos de diseño que, en la mayoría de los casos, son más confiables que los estimados a partir del análisis en el sitio (Cunnane, 1988).

6.6.1 Dependencia entre sitios de medición de los gastos máximos anuales

La magnitud de los gastos máximos anuales en ríos de una misma región geográfica pueden mostrar alguna asociación o dependencia por alguna o todas de las siguientes razones (Cunnane, 1988):

- En algunos años, los gastos máximos anuales en todos los sitios pueden ser ocasionados por un evento meteorológico ocurrido en la región el cual afecta todas las cuencas cualitativamente de la misma manera, es decir, que se producen eventos de caudales máximos que son mayores que los promedios registrados en todos los sitios. Esta es una fuente inicial de dependencia.
- Otra fuente de dependencia es originada por la ocurrencia de años relativamente secos en los cuales los gastos máximos son bajos en toda la región, en cuyo caso todos los gastos máximos anuales serán bajos, aunque no necesariamente causados por el mismo evento meteorológico.
- Una tercera causa de dependencia débil puede ser ocasionada por la ausencia persistente de precipitaciones extremas en eventos meteorológicos extensos originando gastos máximos de moderada magnitud en todos los sitios, causados generalmente por diferentes eventos meteorológicos.

Algunos métodos del análisis regional de frecuencia hacen uso de la dependencia entre sitios de forma explícita y no podrían funcionar sin esta, por ejemplo, los métodos tipo extensión de registros, mientras otros métodos dependen de su completa (estaciones-año) o parcial (avenida índice) ausencia. La dependencia entre sitios es amiga o enemiga en el análisis regional de frecuencia dependiendo de las circunstancias, una paradoja que ha sido poco investigada (Cunnane, 1988). Una consecuencia de la dependencia entre sitios es la pérdida de información (Reed, 1994). Por ejemplo, cuando una tormenta afecta varios sitios, no hay redundancia de información porque los eventos extremos observados provienen del mismo evento meteorológico.

Rosbjerg y Madsen, (1995) analizaron el efecto de la homogeneidad y la dependencia entre sitios para cinco métodos regionales entre ellos el método de la avenida índice y el método de las estaciones-año empleando la distribución general de valores extremos para este último. En su estudio encontraron que las varianzas más pequeñas empleando la información regional se encuentran para los métodos de la avenida índice y el de las estaciones-año, además de que son métodos que dan resultados muy similares.

Hoskin y Wallis, (1988) evaluaron el efecto de la dependencia entre sitios en el método de los momentos estandarizados de probabilidad pesada, sus principales conclusiones fueron: Cualquier sesgo en la estimación de los eventos de diseño no se modifica por la presencia de la dependencia entre sitios. La precisión de la estimación de los eventos de diseño decrece cuando la dependencia entre sitios está presente, pero este efecto es menos importante para efectos prácticos que el sesgo ocasionado en la estimación de eventos de diseño por la heterogeneidad (diferentes distribuciones de frecuencia de gastos máximos en la región). Aun cuando tanto la heterogeneidad como la dependencia entre sitios están presentes y la forma de la distribución de frecuencias de gastos máximos no sea conocida, el análisis regional de frecuencias es más exacto que el análisis en el sitio.

Stedinger, (1983) analizó el efecto de la dependencia entre sitios, y concluyó que la principal consecuencia es la de incrementar la varianza de la media de los estimadores insesgados, recomendando el uso de registros largos cuando se tienen estaciones con registros de periodo común y que presentan correlación.

6.6.2 Técnica de las estaciones – año

El método de las estaciones-año para estimar eventos de largos periodos de retorno busca evitar o reducir la necesidad de realizar la extrapolación de la distribución ajustada. En lugar de ello, el método de las estaciones-año combina los registros de sitios individuales, para formar un solo registro de longitud igual a la suma de la longitud de los registros individuales. Al igual que otros métodos de regionalización, el método asume que la distribución de frecuencia de los gastos máximos anuales es idéntica en cada sitio una vez que los gastos máximos anuales han sido estandarizados por la división de una variable índice, generalmente el gasto medio de los gastos máximos anuales (Buishand, 1991; Stewart et al, 1999). Además, el método de las estaciones-año asume que los registros en los sitios son mutuamente independientes. Esto sólo es válido si los sitios están muy dispersos o si sus periodos de registro son mutuamente excluyentes. Esta última condición no se cumple en la práctica, mientras que, si la primera condición se cumple, el supuesto subyacente de que las distribuciones de frecuencia sean idénticas en todos los sitios es poco probable que se cumpla. Por lo tanto, en la práctica, el enfoque carece de justificación (Stewart et al, 1999). A pesar de estas debilidades, el método de las estaciones-año continúa encontrando aplicación para cumplir el requisito ingenieril de tener suficientes registros para realizar el análisis de frecuencia en la estimación de eventos de largos períodos de retorno.

Un procedimiento típico cuando se utiliza el método de las estaciones-año es evitar el uso de información de las estaciones cercanas al sitio de estudio, para evitar la dependencia excesiva en los datos de los gastos máximos anuales cuando son agrupados (Stewart et al, 1999). Sin embargo, Reed y Stewart (1989) muestran que el método de las estaciones-año, en el cual el registro total combinado tiene una longitud de registro efectiva igual a la suma de la longitud de los registros individuales, introduce menos sesgo

de lo que parece a primera vista, esperando que las estimaciones tengan una mayor variabilidad en lugar de un mayor sesgo.

Buishand (1984) empleó el método de las estaciones-año para evaluar la distribución conjunta de eventos máximos diarios de precipitación en dos sitios, es decir, para el análisis bivariado en estaciones de precipitación máxima diaria. Los registros de las estaciones-año fueron elegidos de tal manera que la distancia entre estaciones fuera de al menos 30 km, considerando a esta distancia entre estaciones como la segura para preservar la independencia.

Este método considera el tratamiento de una sola muestra de datos conformada por un registro estandarizado de eventos, el cual una vez que se construye se ajusta a un conjunto de distribuciones de probabilidad. Esta técnica regional a diferencia de otras no requiere que las muestras tengan una longitud de registro común (Cunnane, 1988; Garros-Berthet, 1994). Con referencia al tamaño de muestra aceptable en cada muestra participante en el análisis, se recomienda que al menos sea de 10 años. Esta recomendación no es del todo estricta, ya que si se tiene un sitio con, por ejemplo, 9 años, pero dentro de este están eventos extremos que puedan afectar el comportamiento del fenómeno en la región, entonces se deberán incluir en el estudio.

Existe otra recomendación en cuanto el número de estaciones-año que se requieren para una correcta estimación del evento \hat{Q}_T , y es que $L \geq 5T$ donde T es el periodo de retorno en años.

6.6.3 Técnica de correlación y regresión múltiple

Las características en la frecuencia de los gastos máximos anuales pueden expresarse en términos de algunos parámetros geomorfológicos y/o factores climatológicos de la región homogénea previamente definida (Cruff y Rantz, 1965). La dependencia o independencia de las variables se estima para cada sitio aforado mediante un proceso de correlación y regresión discriminante, donde la aportación de cada variable que en él intervengan sean estadísticamente significativas para la estimación de la variable dependiente analizada. En la práctica, el número de variables independientes no debe exceder aproximadamente un tercio del número de observaciones. Por lo tanto se debe tener cuidado de no arriesgar la confiabilidad de la estimación de los coeficientes por incluir variables de más (Escalante y Reyes, 2008).

El modelo de regresión puede ser del tipo

$$Y = a + b_1X_1 + \dots + b_sX_s \quad (6-14)$$

O también

$$Y = aX_1^{b_1} X_2^{b_2} \dots X_s^{b_s} \quad (6-15)$$

Donde

Y variable dependiente

X_k características fisiográficas y/o climatológicas ($k=1, \dots, s$)

a, b_k conjunto de parámetros a estimar de los datos

6.6.4 Método de Dalrymple (Avenida índice)

Este método (Dalrymple, 1960) es del tipo de la avenida índice y considera que las muestras se ajustan a la distribución de valores extremos tipo I (Gumbel). Un requisito para esta técnica es que las muestras tengan igual longitud de registro. En esta técnica los gastos máximos instantáneos son estandarizados con el gasto de diseño de periodo de retorno de 2.33 años.

6.6.5 Método de los momentos estandarizados de probabilidad pesada

Este método regional fue propuesto por Wallis (1980) y es similar a la técnica de la avenida índice propuesta por Dalrymple (1960), pues en esta técnica el método de la avenida índice se basa en los momentos estandarizados de probabilidad pesada (Madsen y Rosbjerg, 1997). Se recomienda para su aplicación trabajar con un periodo común.

6.6.6 Análisis multivariado como método regional

En el análisis de frecuencia cuando se emplea el análisis multivariado en la estimación de los parámetros se produce una mejora significativa comparada a cuando se estiman de forma univariada para el mismo tipo de distribución (Rueda, 1981; Raynal, 1985; Raynal y Salas, 1987; Escalante y Raynal, 1994; Escalante-Sandoval, 1998; Raynal-Villaseñor y Salas, 2008). Cuando se aplican los modelos multivariados, en general se obtiene una disminución en el error estándar de ajuste (Escalante-Sandoval, 1998; Raynal-Villaseñor y Salas, 2008). Las estaciones a emplear en el análisis multivariado son las pertenecientes a la misma región homogénea (Escalante-Sandoval, 1998; Escalante y Reyes, 2008).

El modelo a emplear en este estudio es el modelo logístico bivariado el cual tiene la forma (Gumbel, 1960):

$$F(x, y, m) = \exp \left\{ - \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{1/m} \right\} \quad (6-16)$$

$$\begin{aligned} f(x, y, m) = & \frac{f(x)f(y)}{F(x)F(y)} [-\ln F(x)]^{m-1} [-\ln F(y)]^{m-1} \\ & \cdot \exp \left\{ - \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{1/m} \right\} \\ & \cdot \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{\frac{1}{m}-2} \left\{ (m-1) \right. \\ & \left. + \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{\frac{1}{m}} \right\} \end{aligned} \quad (6-17)$$

Donde $F(x, y, m)$ es la función de distribución acumulada bivariada, $F(x)$ y $F(y)$ distribuciones marginales y m es el parámetro de asociación, mayor a 1. $f(x, y)$ es la función de densidad del modelo bivariado.

6.7 CASO DE ESTUDIO

La zona de estudio comprende la cuenca alta y media del río Magdalena que se ubica entre el flanco oriental de la cordillera central y el flanco occidental de la cordillera oriental del sistema montañoso de los Andes. La cuenca está conformada por áreas de los

departamentos de Cauca, Huila, Tolima, Cundinamarca, Caldas, Antioquia, Boyacá, Santander y Cesar donde se ubican ciudades principales como Bogotá, Bucaramanga, Neiva, Ibagué, Tunja. Se concentra cerca del 40% de la población de Colombia (DANE, 2006) y su economía puede representar cerca del 40% del PIB nacional (DANE, 2012). Los principales río que drenan al río Magdalena en este tramo son: Suaza, la Plata, Páez, Saldaña, Sumapaz, Bogotá, La Miel, Nare, Carare, Sogamoso y Lebrija.

A continuación se presenta la aplicación de la metodología definida para el análisis regional de frecuencia en la zona de estudio.

6.7.1 Selección de variables para la formación de grupos

Para cada una de las 62 estaciones ubicadas en la zona de estudio y que cuentan con al menos 40 años de registro, empleando el software ArcGis 10.0, se obtuvieron las siguientes variables:

- Físicas: Latitud, longitud, área de la cuenca, perímetro, coordenadas del centroide, pendiente media de la cuenca, longitud del cauce principal, pendiente media del cauce (Taylor y Schwarz, 1952), elevación máxima, elevación mínima.
- Meteorológicas: Precipitación media anual en la cuenca.

La selección de las variables principales se hizo a partir del análisis de correlación múltiple entre los eventos de diseño obtenidos para el análisis en el sitio para periodos de retorno de 2, 5, 10, 25, 50 y 100 años y las variables físicas y la meteorológica. Este análisis se hizo empleando el paquete estadístico XLSTAT mediante la herramienta mínimos cuadrados ordinarios (OLS).

Hasta periodos de retorno de 50 años, las variables que mejor describen la respuesta hidrológica de la cuenca fueron: el área de la cuenca, la precipitación media anual, la pendiente media del cauce, la pendiente media de la cuenca, la elevación máxima, la longitud del cauce principal, el perímetro y la elevación mínima. Siendo el área de la cuenca y la precipitación media anual las variables más significativas.

Para el periodo de retorno 100 años las variables más representativas fueron el área de la cuenca, la pendiente del cauce principal, la elevación mínima y el perímetro de la cuenca.

Por lo tanto, las variables que mejor describen la respuesta hidrológica de las cuencas y que se emplearan para la formación de grupo serán: el área de la cuenca, la precipitación media anual, la elevación máxima del río, la elevación mínima del río, la pendiente media del cauce principal, la pendiente media de la cuenca, la longitud del cauce principal y el perímetro. En la Tabla 6.2 se presentan las variables seleccionadas para cada una de las estaciones, también se presenta la latitud y L-CV.

Tabla 6.2 Variables físicas, meteorológicas e hidrológicas de las estaciones

Estación	ID	L-CV	Latitud(°)	Área(Km ²)	PMA(mm)	P(km)	Sc(%)	Lc(km)	Sr(%)	C _{max} (msnm)	C _{min} (msnm)
21087050	1	0.356	2.71	124.4	1854.7	53.0	50.2	17.9	6.89	2466	1015
21017050	2	0.140	1.85	256.1	1825.5	105.0	30.7	42.0	1.54	2342	1253
21017020	3	0.260	1.87	325.6	1784.8	113.6	34.2	35.3	3.36	2903	1295
21017040	4	0.146	1.99	3000.8	1677.0	327.1	40.3	103.7	1.47	3123	996
21037010	5	0.170	2.03	994.4	1615.5	202.3	35.9	82.3	1.07	2281	888
21047010	6	0.142	2.23	5650.3	1555.0	467.9	37.0	175.0	0.63	3123	691
21057050	7	0.134	2.33	1227.0	1743.7	204.1	33.6	68.3	2.36	3315	1142
21107030	8	0.353	2.54	214.8	1393.6	81.6	40.2	28.6	3.97	2277	978

Estación	ID	L-CV	Latitud(°)	Área(Km ²)	PMA(mm)	P(km)	Sc(%)	Lc(km)	Sr(%)	C _{max} (msnm)	C _{min} (msnm)
21107020	9	0.321	2.57	617.8	1403.0	152.0	48.7	34.1	3.10	2083	679
21127020	10	0.362	2.95	302.1	1762.0	97.2	45.3	30.4	4.19	2500	866
21127030	11	0.313	2.97	222.2	1758.0	77.0	44.7	19.7	4.35	2500	1301
21137030	12	0.216	3.26	612.9	1900.1	160.0	38.4	69.1	1.80	1872	369
21147030	13	0.177	3.33	1667.2	1582.2	256.4	35.7	83.9	2.70	3023	677
21147010	14	0.296	3.36	2696.4	1416.8	384.7	36.9	118.7	1.91	3022	363
21057030	15	0.166	2.46	2471.3	1609.0	315.1	49.9	85.0	1.52	3322	981
22027020	16	0.148	3.13	926.0	1587.2	191.4	51.2	64.5	2.68	3497	1224
22027010	17	0.194	3.33	1381.8	1730.2	247.8	47.8	92.8	2.61	3497	748
22057060	18	0.123	3.48	4409.3	1918.4	376.2	53.9	97.2	2.28	3277	534
22057040	19	0.176	3.67	6226.9	1997.6	496.2	50.8	133.1	1.36	3277	444
22067010	20	0.371	3.92	290.3	1780.8	104.9	30.2	36.9	2.82	1951	358
22077060	21	0.324	4.17	274.4	1574.7	95.4	38.4	37.9	3.54	3275	668
21217120	22	0.250	4.40	797.4	1444.5	173.3	54.7	38.8	3.36	3072	1160
23017030	23	0.371	5.20	796.9	2498.4	233.3	35.6	101.0	1.63	3453	206
23067020	24	0.309	5.47	3117.9	1725.4	327.5	37.2	123.4	0.57	2044	305
21137020	25	0.112	3.86	266.9	1566.0	137.9	10.9	55.5	0.60	760	294
21227010	26	0.234	4.54	288.5	1446.8	113.7	13.5	32.3	0.71	634	354
21257070	27	0.303	5.02	252.3	2050.1	78.6	27.0	25.3	2.90	1290	268
22077030	28	0.296	4.02	255.3	1496.4	94.6	43.6	29.3	4.14	3268	1527
21187020	29	0.370	4.22	141.2	1737.4	69.5	40.3	28.1	3.97	2538	708
21197030	30	0.255	4.19	1165.0	1251.3	217.8	31.5	78.6	1.81	3524	716
21197110	31	0.250	4.40	159.6	896.6	58.6	36.2	20.4	4.27	2554	1459
23037010	32	0.085	5.47	56507.2	1581.8	1917.1	34.5	673.6	0.08	3114	169
21167050	33	0.219	4.06	144.8	1806.8	61.0	28.9	19.9	0.43	1846	466
21187030	34	0.344	4.23	232.4	1729.5	93.3	34.7	42.5	2.14	2538	588
21237020	35	0.086	5.20	54344.1	1545.1	1839.7	34.4	636.8	0.08	3112	206
21237040	36	0.420	4.65	248.3	1222.7	83.2	52.4	33.6	2.10	1458	365
24017610	37	0.320	5.33	274.8	800.3	110.6	24.1	28.8	0.71	3297	2572
21087040	38	0.450	2.71	102.5	1894.8	50.1	50.3	18.5	6.28	2230	718
23197270	39	0.451	7.34	245.9	1107.2	72.6	47.2	23.5	6.77	3604	1623
23197290	40	0.250	7.17	571.6	1315.4	127.4	32.8	68.0	1.71	3328	628
24027050	41	0.168	6.29	605.0	1784.0	127.7	47.2	29.7	2.42	2230	1259
24037080	42	0.435	5.57	106.9	710.8	50.5	26.4	15.8	2.57	3195	2592
24037040	43	0.219	6.46	157.8	1044.0	63.9	40.5	20.8	8.97	4524	2574
24017580	44	0.122	6.23	325.2	2840.6	131.4	28.3	45.8	0.95	2609	1362
24037070	45	0.317	5.74	140.1	1133.2	65.4	26.8	19.5	1.10	3018	2559
24037450	46	0.551	5.72	53.0	811.6	41.7	21.6	13.8	1.06	3133	2561
24037120	47	0.346	5.59	100.3	812.2	57.5	26.3	23.0	3.17	3580	2543
24037300	48	0.275	5.72	31.3	867.7	27.9	30.1	9.6	4.71	3599	2867
24027040	49	0.271	6.53	183.3	2014.5	73.4	29.0	18.5	0.52	1872	1592
24037110	50	0.301	5.68	418.9	739.5	125.1	24.2	32.8	1.58	3195	2501
23107130	51	0.346	7.29	127.2	1414.6	57.1	44.7	22.3	4.10	1940	738
24067030	52	0.127	7.11	21239.1	1487.7	1011.9	33.9	339.5	0.41	3382	161
24047020	53	0.120	6.61	9762.3	1872.6	664.9	29.6	227.0	0.62	2609	542
24027010	54	0.126	6.55	2079.7	2003.2	273.7	38.9	79.4	0.54	2552	1101
24027030	55	0.186	6.23	690.0	2092.8	168.9	39.7	44.7	1.73	2552	1279
23097030	56	0.112	6.49	73926.5	1886.3	2309.9	33.5	838.9	0.07	3115	109

Estación	ID	L-CV	Latitud(°)	Área(Km ²)	PMA(mm)	P(km)	Sc(%)	Lc(km)	Sr(%)	C _{max} (msnm)	C _{min} (msnm)
23087150	57	0.257	6.14	295.9	2270.0	98.3	23.3	22.2	0.06	2164	2097
23087200	58	0.230	5.97	196.2	4078.3	87.5	48.5	37.7	2.29	2240	814
23087180	59	0.201	6.38	119.5	3474.3	60.0	27.8	23.0	1.66	2288	1642
21057060	60	0.151	2.46	4845.1	1666.7	436.6	43.9	113.0	1.38	3322	789
23087190	61	0.083	6.00	1509.5	4172.7	222.6	41.6	52.0	0.75	1696	389
23127020	62	0.146	6.53	5424.0	2450.9	547.5	34.4	198.0	0.31	1755	106

L-CV L-Coeficiente de variación, PMA Precipitación media anual, P Perímetro de la cuenca, Sc Pendiente media de la cuenca, Lc Longitud del cauce principal, Sr Pendiente media del cauce principal, C_{max} Elevación máxima del río, C_{min} Elevación mínima del río.

6.7.1.1 Breve comentario sobre el efecto de la escala espacial en el L-CV

El L-CV, el cual es la relación entre la desviación estándar y la media de las series de gastos máximos anuales, representa el grado de variabilidad en la distribución de frecuencia de los gastos máximos (Mtiraoui, 2004). Por lo tanto, es uno de los parámetros clave en el análisis regional de frecuencia (Blöschl y Sivapalan, 1997). Dawdy (1961) notó que en algunas regiones de Estados Unidos la pendiente de la curva de frecuencia de los gastos máximos, y por lo tanto el L-CV, tiende a decrecer con el incremento del tamaño de la cuenca. Este comportamiento planteó un problema para el análisis regional de frecuencia ya que algunas metodologías como la de la *Avenida índice* asumen que el L-CV es constante en toda la región (Mtiraoui, 2004).

La variación del L-CV con el tamaño de la cuenca tiene importantes implicaciones en la teoría de escalamiento aplicada a los gastos máximos. De interés primordial es el cambio que ocurre en el L-CV como los cambios en el tamaño de la cuenca y los procesos que controlan este patrón de escalamiento (Mtiraoui, 2004). Sin embargo, Hosking y Wallis (1997), han encontrado que, en algunas regiones, solo el área de drenaje es incapaz de explicar la variación en el sitio de la distribución de frecuencias de los gastos máximos anuales.

En la Figura 6.1 se presenta la variación del L-CV con el tamaño de la cuenca. Se puede observar que se presenta un comportamiento decreciente en L-CV a medida que el tamaño de la cuenca incrementa, sin embargo se presenta una considerable dispersión de los puntos al alrededor de la línea de regresión del L-CV versus el área de la cuenca. Esta dispersión puede ser explicada por la influencia de mecanismos como el ENSO. Por ejemplo, la estación 23067020 y la estación 21147010 que tienen áreas de 3118 y 2696 km² respectivamente presentan valores en el L-CV cercanos a 0.3, estos altos valores en el L-CV pueden ser originados por la presencia de valores atípicos en las muestras que son mayores al 250% del valor medio de los gastos máximos anuales y que se presentaron en años La Niña. De acuerdo al análisis con covariables realizado en el capítulo 3 se encontró que en estas estaciones es más adecuado realizar el análisis de frecuencia no estacionario empleando como covariables índices del ENSO. Una situación similar se presenta para estaciones como la 24037450 (53, 0.55), 24037120 (100, 0.35), 24037070 (140, 0.32), 21127020 (302, 0.36), 24037080 (107, 0.44). En paréntesis el área y el L-CV. En la Figura 6.2 se presentan (con relleno rojo) los puntos en los cuales el análisis de frecuencia no estacionario empleando como covariables índices del ENSO es más adecuado.

Otra causa que puede explicar la dispersión es que algunas estaciones que tienen altos valores en el L-CV presentan precipitaciones medias anuales menores a 1000 mm y algunas con valores menores a 0.23 presentan precipitaciones medias anuales mayores a 3000 mm.

Aunque se está realizando el análisis con estaciones que tienen 40 o más años de registro, es probable que valores atípicos ocasionados por la variabilidad de la muestra o por el ENSO, como en la estación 21087040 que tiene 40 años de registro y tiene tres eventos que son un 200% mayor a la media de los gastos máximos anuales (registrados en años La Niña), expliquen también la dispersión en los valores del L-CV.

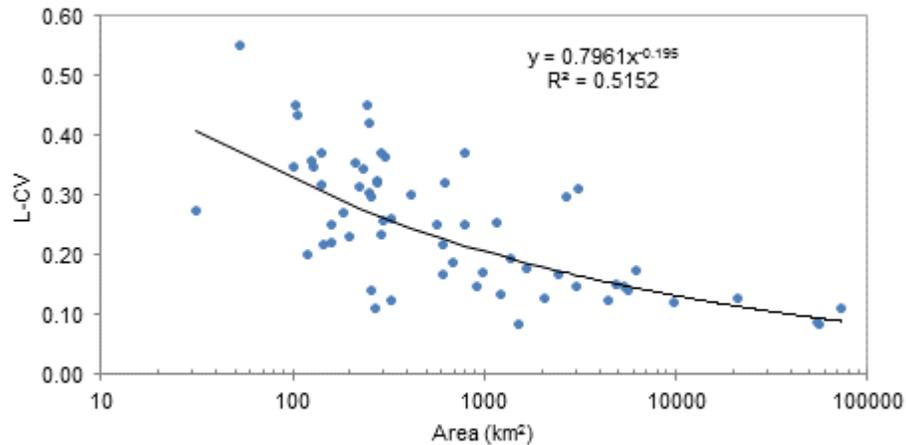


Figura 6.1 Variación del L-CV con el área de la cuenca

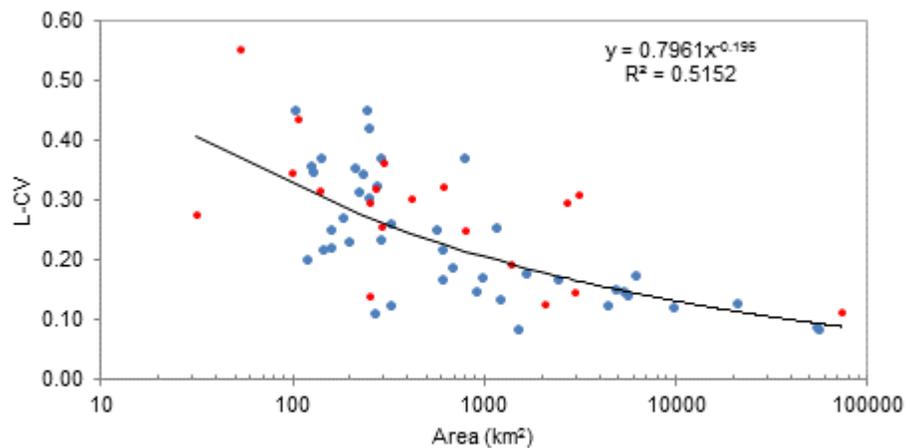


Figura 6.2 Variación del L-CV con el área de la cuenca. Con relleno rojo las estaciones en las cuales el análisis de frecuencia no estacionario empleando como covariables índices del ENSO es más adecuado.

Del breve análisis realizado al comportamiento del L-CV con el tamaño de la cuenca se puede concluir que no solo el área afecta dicho comportamiento, mecanismos como el ENSO o la precipitación también tienen una influencia en la variación de este parámetro.

6.7.2 Formación de grupos

Una vez se han definido las variables que mejor describen la respuesta hidrológica de las cuencas, se procede a la formación de grupos o regiones físicamente homogéneas

aplicando el método de los trazos multidimensionales, el método de la región de influencia y el método de Ward. Para la aplicación de las diferentes técnicas se han descartado 4 estaciones cuyas áreas de la cuenca son mayores a 10000 km². Sin embargo se incluyen una vez formado los grupos.

6.7.2.1 Método de los trazos multidimensionales

Las variables empleadas para la aplicación de este método fueron el área de la cuenca (A), la pendiente media de la cuenca (Sc), la precipitación media anual (PMA), la longitud del cauce principal (Lc), la pendiente media del cauce principal (Sr), la elevación máxima (Cmax) y la elevación mínima (Cmin). El método de estandarización empleado en esta técnica es (Kar et al, 2012):

$$x_{kn} = \frac{Y_{n(\text{MAX})} - Y_{kn}}{Y_{n(\text{MAX})} - Y_{n(\text{MIN})}} \quad (6-18)$$

Donde Y_{kn} es la n-ésima característica en la variable k ; y $Y_{n(\text{MAX})}$ y $Y_{n(\text{MIN})}$ son el valor máximo y mínimo de la variable k . La función para hallar los trazos multidimensionales, aplicada a las variables previamente estandarizadas, es:

$$f(t) = \frac{A}{\sqrt{2}} + PMA \sin(t) + C_{max} \cos(t) + S_c \sin(2t) + S_r \cos(2t) + C_{min} \sin(3t) + L_c \cos(3t) \quad (6-19)$$

Las curvas medias $\bar{f}(t)$ para cada grupo formado se presentan en la Figura 6.3. Los grupos 4 y 5 presentan una mayor diferencia en las curvas medias respecto a los otros grupos. Las curvas de los tres primeros grupos presentan un comportamiento similar, las mayores diferencias se presentan para los intervalos $-\pi \leq t \leq -2\pi/3$ y $2\pi/3 \leq t \leq \pi$. A partir de este método la zona de estudio se divide en 5 grupos o regiones homogéneas de acuerdo sus características fisiográficas. En la Tabla 6.3 se presentan las estaciones que conforman cada grupo y en la Figura 6.4 se presenta su ubicación.

Tabla 6.3 Grupos formados mediante el método de los trazos multidimensionales

Región	Estaciones	Nro. Estaciones	Estaciones-año
1	2,3,4,5,6,7,15,60	8	341
2	1,8,...,14,16, ...,31,32,33,34,35,36,38	30	1382
3	41,44,52,54,55,56,57,59	8	388
4	37,39,42,43,45,46,47,48,50	9	431
5	40,49,51,53,58,61,62	7	297

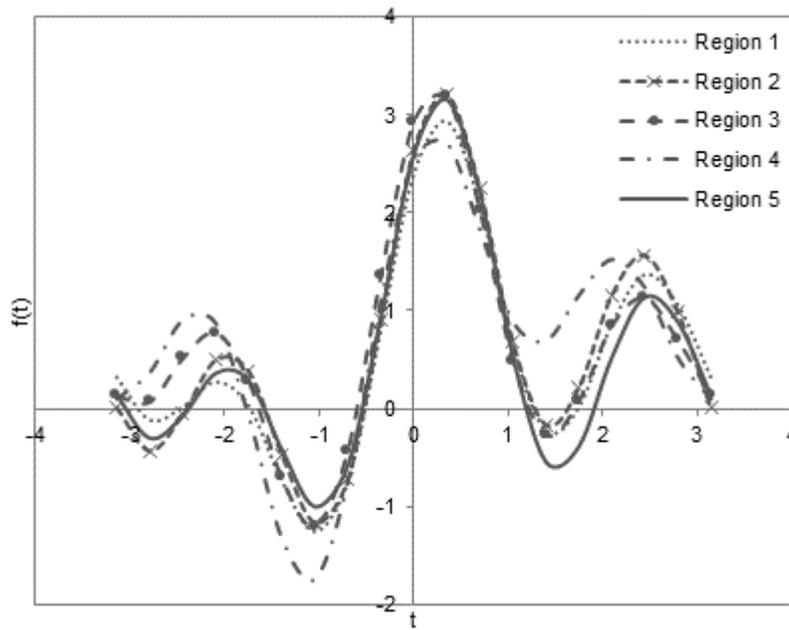


Figura 6.3 Curvas medias $\bar{f}(t)$ para cada grupo formado

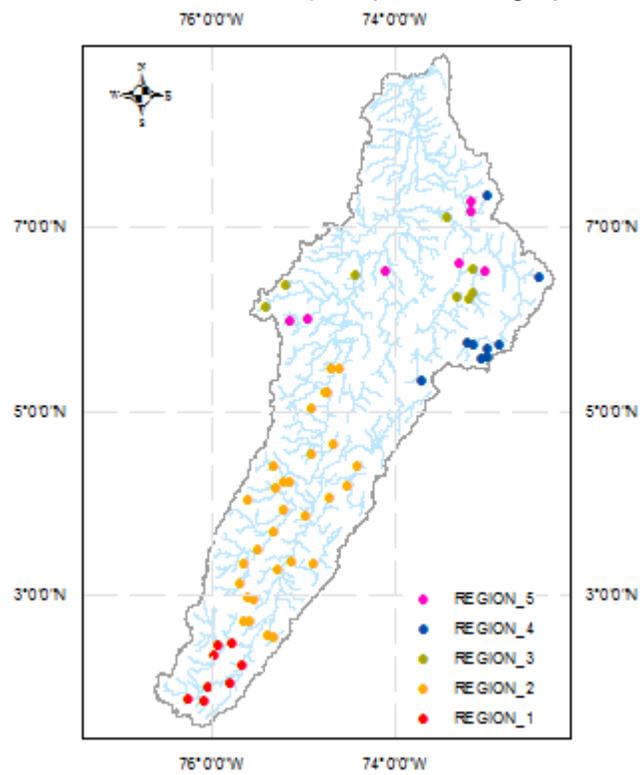


Figura 6.4 Regiones formadas con el método de los trazos multidimensionales

6.7.2.2 Método de la región de influencia

Las variables empleadas para la aplicación de este método fueron: La latitud, el área de la cuenca (A), la pendiente media de la cuenca (Sc), la precipitación media anual (PMA), el perímetro (P), la longitud del cauce principal (Lc), la pendiente media del cauce principal (Sr), la elevación máxima (Cmax) y la elevación mínima (Cmin). A algunas de las variables se les aplicaron transformaciones no-lineales para reducir efectos de escala (Hosking y Wallis, 1997). Todas las variables para cada uno de los sitios fueron estandarizadas empleando la desviación estándar de cada variable. En la Tabla 6.4 se presentan las transformaciones aplicadas a cada una de las variables y el proceso de estandarización empleado. A las variables área de la cuenca, precipitación media anual, elevación máxima y latitud se multiplicaron por 3 para darles una mayor importancia en el análisis.

Tabla 6.4 Transformación de las características de los sitios

Característica del sitio, X	Variable para la formación del grupo, Y
Latitud (grados), X	$Y=X/\sigma(X)$
Area (Km ²), X	$Y=\log(X)/\sigma(X)$
PMA (mm), X	$Y=\log(X)/\sigma(X)$
P(km), X	$Y=\log(X)/\sigma(X)$
Sc (%), X	$Y=X/\sigma(X)$
Lc (km), X	$Y=\log(X)/\sigma(X)$
Sr (%), X	$Y=X/\sigma(X)$
Cmax (msnm), X	$Y=(X)^{0.5}/\sigma(X)$
Cmin (msnm), X	$Y=(X)^{0.5}/\sigma(X)$

$\sigma(X)$ denota la desviación estándar de la variable X medida en los 58 sitios.

Una vez se estima la distancia Euclidiana, el siguiente paso es definir el valor límite para la formación de los grupos. La metodología descrita en el numeral 6.2.2.2 propone como valor límite el valor medio de la distancia Euclidiana. Sin embargo, para la zona de estudio se ha encontrado que este valor excluye muchas estaciones y por lo tanto se forman demasiados grupos con pocas estaciones, lo que implica que se puedan tener problemas para la transferencia de información. Para este estudio se propone trabajar con un valor límite igual al 60% del valor medio de la distancia Euclidiana, valor que permite la formación de pocos grupos con un mayor número de sitios incluidos. Se debe tener en cuenta que regiones con un gran número de estaciones incluidos permiten una mejor transferencia de la información pero la homogeneidad estadística pueda decrecer (Burn, 1990).

En la Tabla 6.5 se presentan los cuatro grupos formados empleando el método de la región de influencia, las estaciones que los conforman y el número de estaciones-año de cada uno. En la Figura 6.5 se presenta su localización.

Tabla 6.5 Grupos formados mediante el método de la región de influencia

Región	Estaciones	Nro. Estaciones	Estaciones-año
1	2,3,4,5,6,7,12,13,14,15,16,17,18,19,25,60	16	691
2	1,8,9,10,11,20,21,22,23,24,26,27,28,29,30,31,32,33,34,35,36,38	22	1032
3	41,44,52,53,54,55,56,57,58,59,61,62	12	556
4	37,39,40,42,43,45,46,47,48,49,50,51	12	560

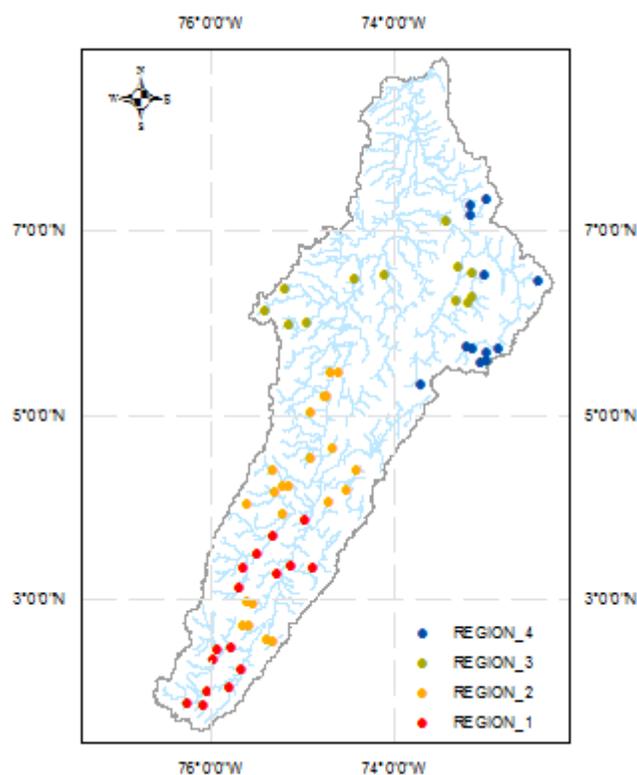


Figura 6.5 Grupos formados con el método de la región de influencia

6.7.2.3 Método de Ward

Para este método se utilizan las mismas variables y el mismo proceso de estandarización aplicado al método de la región de influencia. Para la aplicación del método se empleó la herramienta LmomRFA (Hosking, 2009) disponible en el paquete estadístico R. Para la aplicación del método se requiere la definición inicial del número de grupos, por lo que se proponen 4. En la Tabla 6.6 se presentan los 4 grupos formados empleando el método de Ward, así como las estaciones que componen cada grupo y el número de estaciones-año de cada grupo. En la Figura 6.6 se presenta la localización de las regiones.

Tabla 6.6 Grupos formados mediante el método de Ward

Región	Estaciones	Nro. Estaciones	Estaciones-año
1	4,6,7,13,14,15,16,17,18,19,30,60	12	529
2	1,2,3,5,8,9,10,11,12,20,21,22,25,26,27,28,29,32,33,34,35,36,38	23	1050
3	23,24,40,41,44,49,51,52,53,54,55,56,57,58,59,61,62	17	788
4	31,37,39,42,43,45,46,47,48,50	10	472

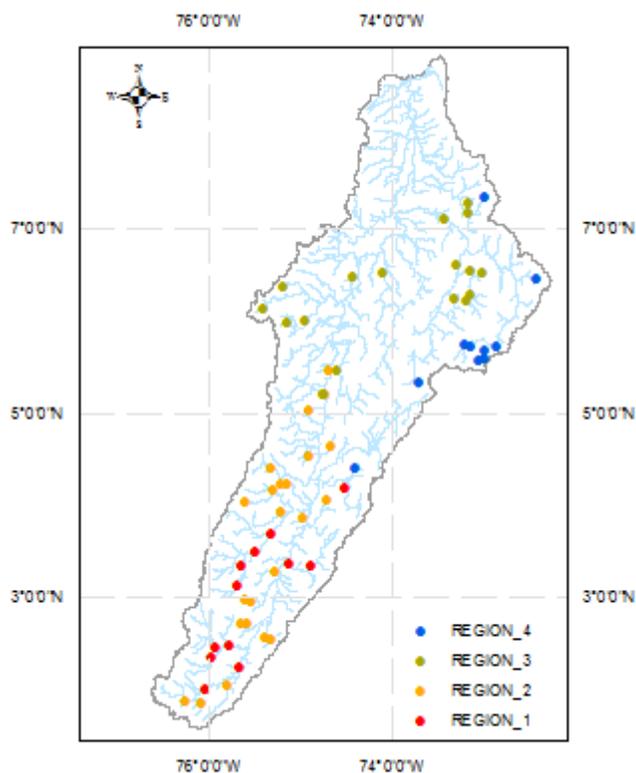


Figura 6.6 Grupos formados con el método de Ward

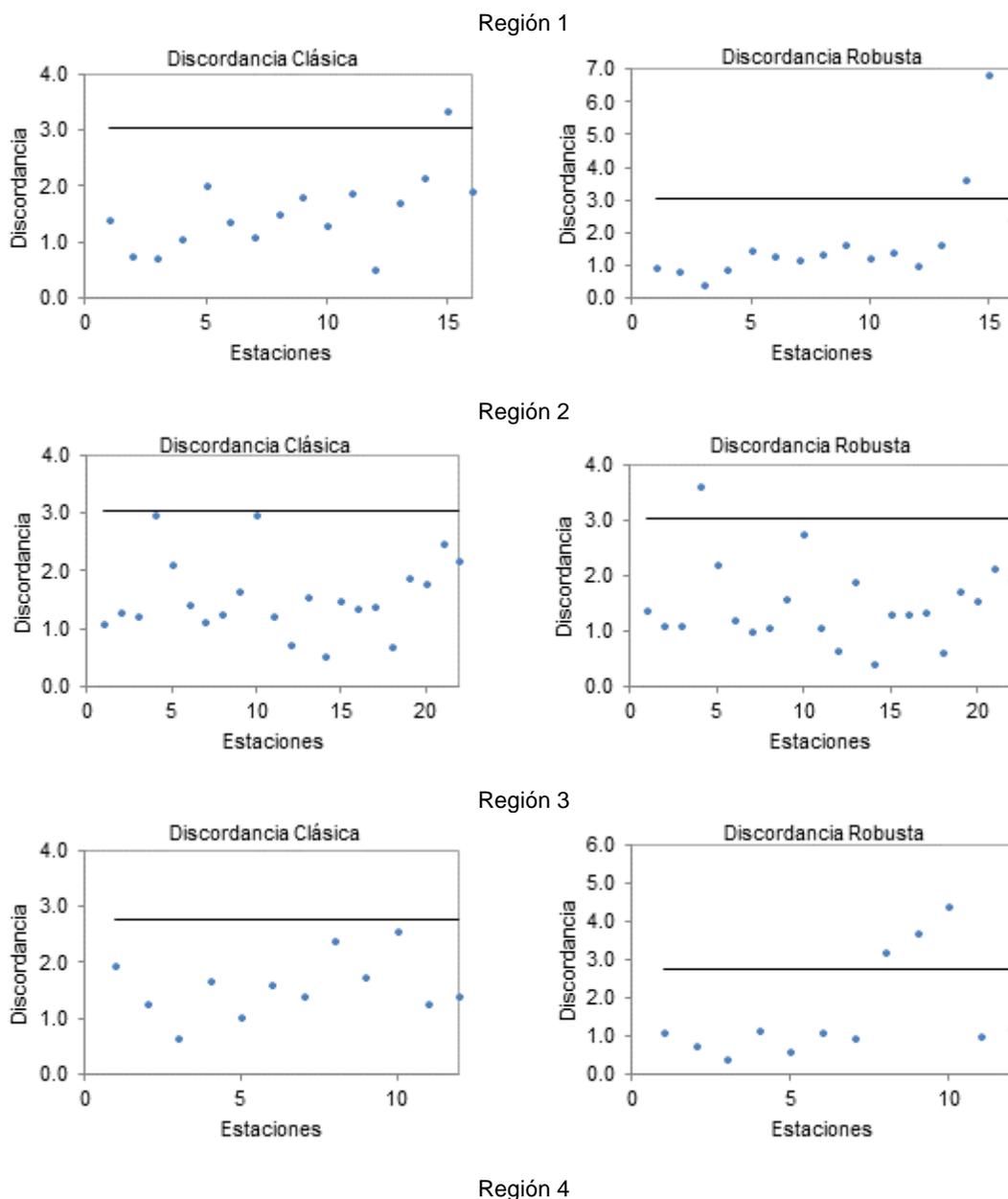
Para los tres métodos empleados en la formación de grupos se presentan diferencias sobre todo en los grupos (regiones) 1 y 2. Los resultados del método de la región de influencia y del método de Ward, salvo por unas estaciones, son muy similares. Por lo tanto, para los análisis de evaluación de las estaciones discordantes y de la evaluación de la homogeneidad estadística de las regiones se emplearán los grupos formados mediante el método de la región de influencia.

6.7.3 Medida de la discordancia

Si bien se recomienda evaluar la medida de la discordancia en todo el conjunto de sitios de la zona de estudio antes de realizar el análisis de formación de grupos, es posible que los sitios que son discordantes respecto al total no lo sean respecto al grupo al que pertenecen, por lo tanto solo se evaluará la medida de la discordancia en los grupos ya definidos. Para este estudio se emplearán los grupos formados mediante el método de la región de influencia.

La evaluación de la medida de la discordancia involucra el cálculo de las relaciones τ_2 , τ_3 y τ_4 de los momentos-L. Dichas relaciones serán estimadas para un periodo común de 40 años (1973-2012), considerando que las condiciones climáticas en cada grupo (región) para el periodo común pueden ser “estables”. Se empleó el paquete estadístico rrcov (Todorov, 2009) para la estimación de la discordancia clásica (Hosking y Wallis, 1997) y para la estimación de la discordancia robusta (Neykov et al, 2007), el paquete se encuentra disponible en el software R.

En la Figura 6.7 se presenta la evaluación de la discordancia para las 4 regiones formadas empleando el método de la región de influencia. La medida de la discordancia robusta (DR) detecta más sitios discordantes que la discordancia clásica (DC). Para la región 1 la DC detecta como sitio discordante la estación 15, mientras que la DR detecta además la estación 3; Para la región 2 la DC no detecta sitios discordantes, la DR detecta la estación 10 como sitio discordante; Para la región 3 la DC no detecta sitios discordantes, la DR detecta las estaciones 57, 58, y 59 como sitios discordantes; Para la región 4 la DC no detecta sitios discordantes, la DR detecta la estación 39 como sitio discordante. Los sitios discordantes incrementan la heterogeneidad de las regiones y por lo tanto tienen un efecto importante en la homogeneidad hidrológica, así, la eficacia de la medida de la discordancia robusta en detectar dichos sitios se evaluará cuando se estime la homogeneidad hidrológica de las regiones.



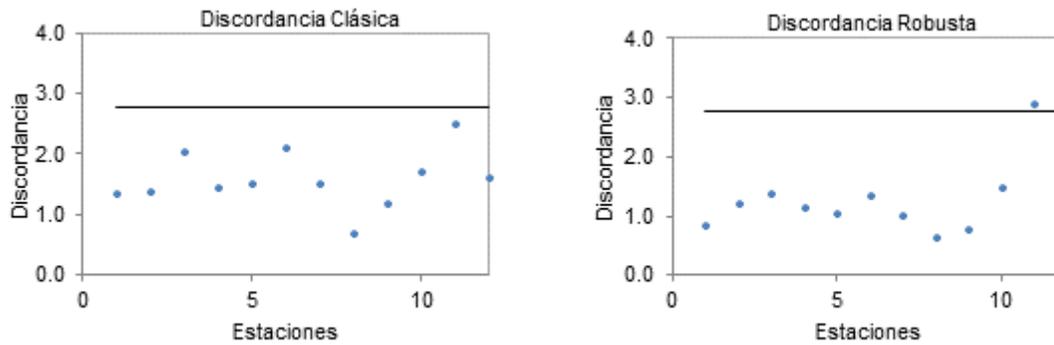


Figura 6.7 Evaluación de los sitios discordantes para los diferentes grupos empleando la discordancia clásica y la discordancia robusta.

6.7.4 Evaluación de la homogeneidad hidrológica

La confiabilidad del análisis regional de frecuencia es altamente dependiente del grado de homogeneidad hidrológica (Saf, 2009). Para determinar la homogeneidad regional o para evaluar si una región propuesta es homogénea desde el punto de vista hidrológico, la prueba estadística H , llamada medida de la heterogeneidad, fue propuesta por Hosking y Wallis, (1993). Un grupo (región) es homogéneo si todos los sitios del grupo tienen la misma curva de frecuencia adimensional, es decir, la misma distribución de probabilidad. Se considera heterogénea si en el grupo existen muchos sitios con diferentes distribuciones de probabilidad. La medida de la heterogeneidad (Hosking y Wallis, 1993) es utilizada para evaluar si un grupo es homogéneo o heterogéneo y se define en términos de las relaciones de los momentos $-L$ y puede ser basada en:

- Solo en el L-CV (Estadístico H_1)
- L-CV y L-Asimetría (Estadístico H_2)
- L-Asimetría y L-Curtosis (Estadístico H_3)

En este estudio la heterogeneidad de los grupos, formados con el método de la región de influencia, se evaluó de acuerdo al estadístico H_1 . Al igual que en la evaluación de los sitios discordantes, para esta prueba solo se consideraron los registros correspondientes al periodo común (1973-2012). Para la estimación de los estadísticos H_1 , H_2 y H_3 se empleó el paquete estadístico LmomRFA (Hosking, 2009) disponible en el software R. En términos del estadístico H_1 se considera la región homogénea si $H_1 < 1$, posiblemente heterogénea si $1 < H_1 < 2$ y heterogénea si $H_1 > 2$ (Hosking y Wallis, 1997). La homogeneidad está estrechamente relacionada con el número de sitios del grupo y por lo tanto, a medida que el número de sitios en el grupo es mayor la homogeneidad tiende a decrecer (Burn, 1990; Rao y Hamed, 2000). Así, en este estudio se consideró la región homogénea si el estadístico $H_1 < 2$. En la Tabla 6.7 se presenta la evaluación de la medida de la heterogeneidad para los diferentes grupos considerando todos los sitios y sin considerar los sitios discordantes.

Todos los grupos formados a partir de las características físicas y climáticas de los sitios son heterogéneos. Retirar solo los sitios discordantes detectados con la discordancia robusta no garantiza que se obtenga el valor en la medida de la heterogeneidad deseado y por lo tanto, fue necesario retirar más sitios (regiones 1, 3, 4) o retirar sitios que no

habían sido detectados como discordantes en el caso de la región 2. La búsqueda de la homogeneidad hidrológica del grupo, en términos del estadístico H_1 , tiende a disminuir la confiabilidad de la estimación regional, toda vez que para obtenerla, la cantidad de los sitios en cada grupo tiende a decrecer.

Tabla 6.7 Evaluación de la medida de la discordancia para los diferentes grupos

Región	Heterogeneidad, todas las estaciones			Heterogeneidad, sin sitios discordantes			Sitios discordantes
	H_1	H_2	H_3	H_1	H_2	H_3	
1	7.78	2.43	1.19	1.34	1.14	0.20	3,12,14,17
2	7.70	5.29	2.46	1.57	2.14	0.71	24,32,33,35
3	8.90	1.26	0.71	1.55	-0.91	0.29	55,57,58,59,61
4	5.74	0.87	0.37	1.92	-0.10	0.76	39,46

En la Figura 6.8 se presenta el L-CV estimado para el periodo común de los grupos analizados, en cada una de las regiones, los valores de las últimas estaciones corresponden a los de los sitios discordantes que ocasionan una medida de la heterogeneidad $H_1 > 2$. En las regiones 1 y 4 son los sitios con mayor valor, respecto al grupo, en el L-CV los que producen la heterogeneidad. En la región 2 son los sitios con menor valor en el L-CV los que producen la heterogeneidad del grupo. Finalmente, en la región 3 son los sitios con altos (3 sitios) y con bajos (1 sitio) valores en L-CV los que hacen que el grupo sea heterogéneo. Por lo tanto, si bien la medida de la discordancia robusta da una idea inicial de los sitios que pueden producir la heterogeneidad en el grupo, se recomienda, en el caso de que el grupo formado sea heterogéneo en términos del estadístico H_1 , retirar aquellos sitios que presenten una alta dispersión respecto al grupo en el L-CV para lograr el valor en el estadístico H_1 deseado.

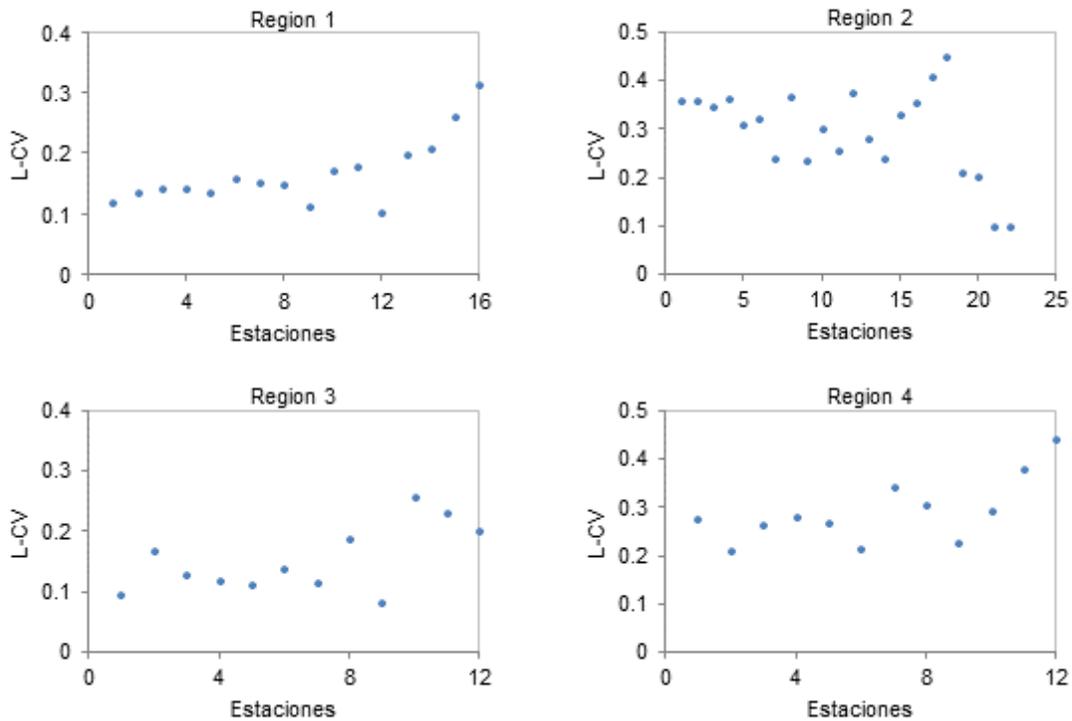


Figura 6.8 L-CV estimado para el periodo común de los sitios que conforman cada uno de los grupos

En la Figura 6.9 se presenta la localización de los sitios que conforman cada uno de los grupos que pueden ser considerados homogéneos tanto física como estadísticamente. De acuerdo a los resultados obtenidos en el capítulo 3 la región 4 es no estacionaria, recomendando el uso de modelos no estacionarios que incorporen índices del ENSO o distribuciones mezcladas para realizar el análisis de frecuencia de gastos máximos.

Aunque se puede “pasar” los sitios discordantes de una región a otra, se debe tratar que este traslado se haga a estaciones vecinas para favorecer la estimación de eventos, sobre todo, en los sitios no aforados.

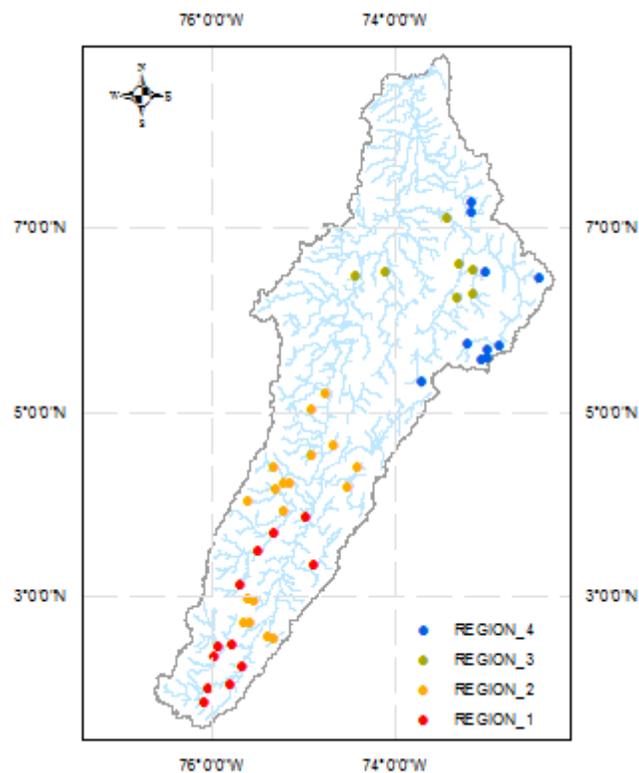


Figura 6.9 Localización de los sitios que conforman los grupos que son homogéneos tanto física como hidrológicamente.

6.7.4.1 Regiones homogéneas de la zona de estudio

Como algunos de los sitios, desde el punto de vista físico, de las regiones 1 y 2 se están cruzando se pasarán estaciones de la región 1 a la región 2 y viceversa con el objetivo de obtener una mejor delimitación de las regiones desde el punto de vista físico, lo que facilita la estimación de eventos en cuencas no aforadas. En la Tabla 6.8 se presenta la nueva delimitación de las regiones y los sitios que las conforman. Estas regiones delimitadas serán las utilizadas para la estimación de los eventos de diseño por los métodos regionales.

Tabla 6.8 Regiones homogéneas definidas para la zona de estudio

Región	Estaciones	Nro. Estaciones	Estaciones-año
1	2,3,4,5,6,7,15,16,17,18,19,60	12	520
2	1,8, ...,14,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,36,38	26	1203
3	41,44,52,53,54,55,56,57,58,59,61,62	12	556
4	37,39,40,42,43,45,46,47,48,49,50,51	12	560

En la Tabla 6.9 se presenta la evaluación de la medida de la heterogeneidad para las nuevas regiones definidas para la zona de estudio. La región 1 presenta un sitio discordante, dicho sitio exhibe, respecto a los otros sitios del grupo, un alto valor en L-CV. Los sitios discordantes en la región 3 son aquellos que presentan, respecto a los sitios del grupo, bajos valores en el L-CV. Por lo tanto, quedan definidas las regiones de la zona de estudio que son homogéneas tanto física como hidrológicamente y que serán empleadas para la realización del análisis regional de frecuencia.

Tabla 6.9 Evaluación de la medida de la heterogeneidad para las regiones de la zona de estudio

Región	Heterogeneidad, todas las estaciones			Heterogeneidad, sin sitios discordantes			Sitios discordantes
	H ₁	H ₂	H ₃	H ₁	H ₂	H ₃	
1	4.82	2.07	1.07	1.41	1.77	1.02	3
2	9.02	6.02	2.83	1.89	2.94	1.63	13,22,25,32,33,35
3	8.90	1.26	0.71	1.55	-0.91	0.29	55,57,58,59,61
4	5.74	0.87	0.37	1.92	-0.10	0.76	39,46

Así, las principales cuencas que conforman las regiones homogéneas de la zona de estudio son:

- Región 1: Parte alta del río Magdalena hasta la confluencia con el río Páez y el río Saldaña hasta la estación 22057040.
- Región 2: Cuenca del río Magdalena aguas abajo de la confluencia del río Páez hasta la estación 23037010. Sin embargo, solo se considera la parte baja del río Bogotá y además toda la cuenca del río Negro.
- Región 3: Comprende la parte baja del río Magdalena desde la estación 23037010. Sin embargo, esta región solo considera en la cuenca del río Sogamoso la correspondiente a la cuenca del río Suárez en su parte baja y media. Esta región no incluye la cuenca del río Chicamocha y el río Lebrija.
- Región 4: Comprende toda la cuenca del río Chicamocha y la del río Lebrija. Además considera la parte alta del río Suárez, específicamente la zona que queda sobre el departamento de Boyacá y la parte alta del río Bogotá.

De acuerdo a los resultados obtenidos del análisis de homogeneidad, se puede considerar a la región 4 como una región no estacionaria y en la cual un gran porcentaje de sus estaciones la variabilidad climática puede explicar la no estacionariedad.

En la Figura 6.10 se presenta la localización de los sitios que conforman cada una de las regiones homogéneas de la zona estudio.

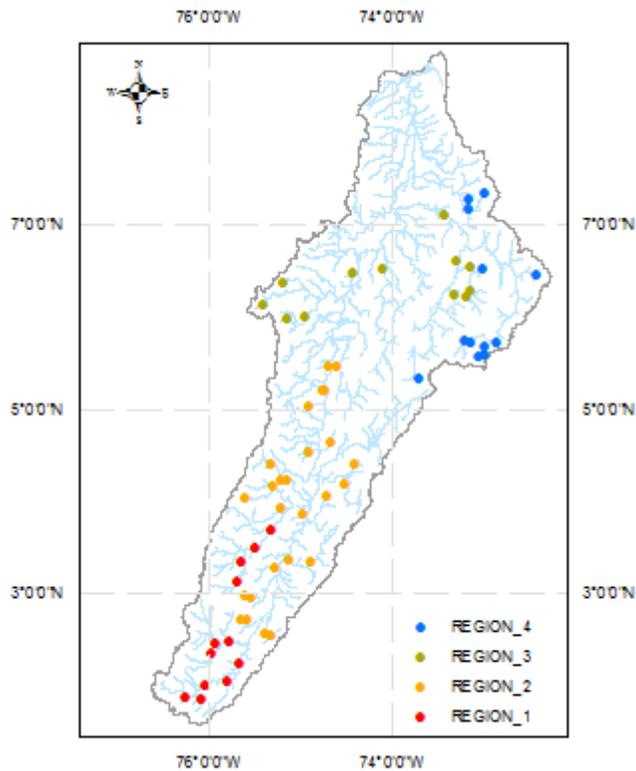


Figura 6.10 Regiones homogéneas de la zona de estudio

6.7.5 Prueba para evaluar si en las regiones los eventos extremos provienen de más de una población

Un cambio en la pendiente de la curva empírica de frecuencia puede ser un indicio de que los eventos extremos en el sitio provienen de más de una población (Potter, 1958; Mtiraoui, 2004). Por lo tanto, la curva empírica de frecuencia se puede formar por dos o más curvas rectas, cada una representando un mecanismo diferente que influye en la generación de los eventos extremos. En la Figura 6.11 muestra que la curva de frecuencia empírica de los registros de la estación 24037110 es formada por dos curvas rectas, lo que puede indicar que los eventos extremos provienen de dos poblaciones, en el capítulo 3 se encontró que en esta estación es más adecuado el análisis no estacionario empleando covariables del ENSO. Sin embargo, si no es acentuado el cambio en la pendiente de la curva de frecuencia empírica o las líneas coinciden, los registros en la estación pueden provenir de la misma población. Y por lo tanto, los eventos estimados para las diferentes poblaciones pueden ser similares (Wang, 2000).

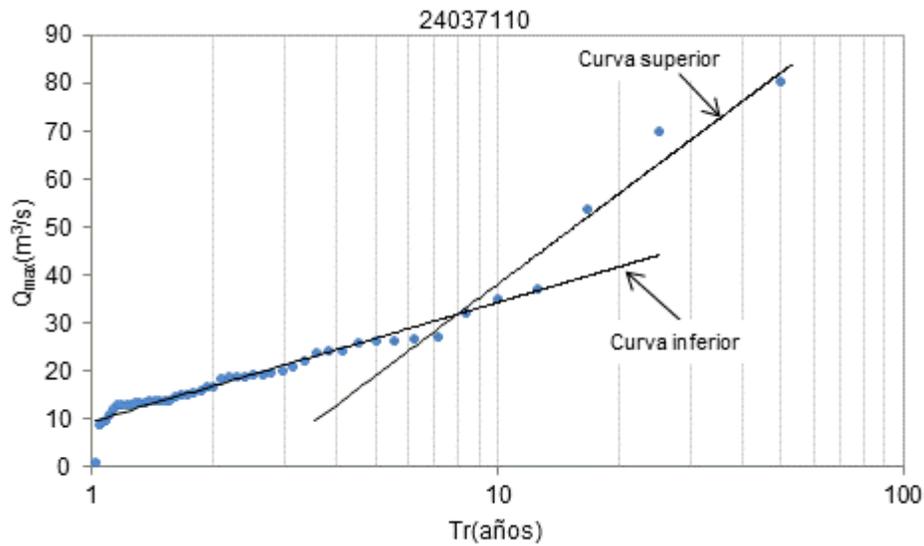


Figura 6.11 Curva de frecuencia empírica conformada por dos curvas (24037110)

Wang (2000) propone una prueba para evaluar si los eventos en una región homogénea, previamente definida, provienen de dos poblaciones. El estadístico R_m propuesto por Wang (2000) se calcula de la siguiente manera:

$$R_m = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \left(\frac{Q_{10I}}{Q_{10S}} + \frac{Q_{50I}}{Q_{50S}} \right) \quad (6-20)$$

Donde Q_{10I} y Q_{50I} son los eventos de diseño de periodos de retorno de 10 y 50 años de la población que conforma la curva inferior (Véase Figura 6.11); Q_{10S} y Q_{50S} son los eventos de diseño de periodos de retorno de 10 y 50 años de la población que conforma la curva superior; N es el número de estaciones de la región homogénea. Teóricamente, si las dos curvas coinciden los eventos de diseño para periodos de retorno de 10 y 50 años serán similares y el estadístico R_m será muy cercano a 1. Por lo tanto, si el estadístico R_m es significativamente diferente de 1, se puede considerar que los eventos extremos en la región provienen de dos poblaciones.

El proceso para evaluar si los eventos extremos en las regiones homogéneas definidas provienen de dos poblaciones es el siguiente:

- Calcular los Q_{10I} , Q_{50I} y Q_{10S} y Q_{50S} para cada estación empleando la distribución Gumbel y estimando los parámetros por Momentos-L. La proporción, de que cantidad de registros en cada estación conforman ya sea la curva inferior o superior, la dará el valor p , donde p es el peso relativo de la primera población (curva inferior) obtenido del ajuste de la distribución Doble Gumbel a cada una de las estaciones de la zona de estudio (Tabla 5.1).
- Calcular el estadístico R_m para cada región.
- Evaluar, con un nivel de confianza del 99% si el estadístico R_m es estadísticamente diferente de 1, usando la prueba t-student con $N - 1$ grados de libertad.

Los resultados del estadístico R_m para las diferentes regiones se presentan en la Tabla 6.10. En aquellas estaciones en las cuales las proporciones son muy pequeñas y la curva superior o inferior está compuesta por menos de tres registros se asumió que el estadístico R_m es igual a 1. De acuerdo a los resultados obtenidos, los eventos extremos en las regiones provienen de por lo menos dos poblaciones.

Tabla 6.10 Prueba para evaluar si los eventos extremos provienen de dos poblaciones

Región	R_m	σ_{Rm}	N	t_d	t	Prueba
1	0.711	0.186	12	5.387	3.106	Difiere de 1
2	0.634	0.296	26	6.287	2.787	Difiere de 1
3	0.642	0.188	12	6.575	3.106	Difiere de 1
4	0.412	0.218	12	9.331	3.106	Difiere de 1

R_m valor del estadístico R_m , σ_{Rm} Desviación estándar del estadístico R_m , N número de estaciones, t_d estadístico de prueba, t valor de la prueba t-student para N-1 grados de libertad y 99% de nivel de confiabilidad.

6.7.6 Análisis regional de frecuencia

Aunque la homogeneidad hidrológica de los grupos definidos para la zona de estudio se logra sin considerar los sitios discordantes, varios autores han señalado la importancia de evaluar dichos sitios porque se puede perder información valiosa del comportamiento de los gastos máximos en la región. Por ejemplo, un evento meteorológico extremo localizado puede afectar solo unos pocos sitios de la región. Si tal evento es igualmente probable que afecte en el futuro a otros sitios, entonces sería correcto tratar todos los sitios como una región homogénea, aun cuando algunos sitios parezcan discordantes respecto al conjunto de sitios de la región (Hosking y Wallis, 1997). Lo ideal de la región es que sea homogénea. Sin embargo, una representativa pero heterogénea región puede dar mejores estimaciones de los eventos máximos que el análisis en el sitio o una región que se ha hecho homogénea por una inapropiada remoción de sitios (IH, 1999). Algunas de las estaciones (3, 39, 46, 57, 58, 59) discordantes lo son porque presentan altos valores atípicos presentados en años La Niña, Klemes (2000) ha argumentado que los altos valores atípicos en los registros de gastos máximos son importantes para la extrapolación de eventos asociados a grandes periodos de retorno, ya que definen el tramo superior de la curva de frecuencia de los gastos máximos.

En el IH (1999) definen la homogeneidad hidrológica en términos del estadístico H_2 considerando la región como fuertemente heterogénea si $H_2 > 4$. Por lo tanto, excepto la región 2, las demás regiones con todos sus sitios serían adecuadas para la estimación de eventos regionales. Para la estimación de los eventos mediante métodos regionales se emplearán todos los sitios de la región. Se considerará el método de las estaciones-año, el método de los momentos estandarizados de probabilidad pesada y el análisis multivariado, este último solo se puede aplicar en sitios que cuenten con al menos 10 años de registro (Escalante-Sandoval, 1994).

6.7.6.1 Método de las estaciones-año

La técnica de las estaciones-año tiene la ventaja de que permite el uso de cualquier tipo de distribución. Mecanismos de baja frecuencia como el ENSO (El Niño-oscilación del sur) tienen una fuerte influencia sobre la hidrología de Colombia (Poveda y Mesa, 1996; Poveda y Mesa; 1997; Poveda, 2004; Poveda et al, 2011). Del análisis realizado en el capítulo 3 se encontró que el 70% de las estaciones de la zona de estudio son no estacionarias y en el 42% de éstas es más adecuado emplear análisis no estacionario

empleando como covariables índices del ENSO. En el capítulo 5 se encontró que comparado con el modelo Gumbel el modelo doble Gumbel es más adecuado en un 69% de las estaciones, y mediante el estadístico R_m , se encontró que los eventos extremos en las regiones provienen de por lo menos dos poblaciones, por lo tanto, en la aplicación de la técnica de las estaciones-año se emplearán las distribuciones mezcladas, específicamente el modelo doble Gumbel que fue descrito en el capítulo 5. En la

Figura 6.12 se presentan las curvas regionales de frecuencia adimensional estimadas para cada una de las regiones de la zona de estudio y en la Tabla 6.11 se presentan los valores de las curvas para los diferentes periodos de retorno. Las curvas regionales de frecuencia adimensional de las regiones 1 y 3 son similares, por lo que inicialmente se podría afirmar que se trata de una misma región. Sin embargo, este argumento debe tratarse con cuidado porque la ausencia de una diferencia estadísticamente significativa entre las curvas regionales de frecuencia adimensionales puede meramente reflejar la carencia de datos (Hosking y Wallis, 1997). La diferencia en la forma de las curvas regionales de frecuencia adimensional de las regiones 1 y 2 o 3 y 4 se puede explicar por la diferencia en el valor del L-CV que se presenta en las distintas regiones, mientras que en las regiones 1 y 3 el valor medio del L-CV es de 0.15, en las regiones 2 y 4 el valor medio es de 0.29. Por lo tanto, en las regiones 2 y 4 se presenta una mayor variabilidad y es más común que se presenten altos valores atípicos. Farquharson et al, (1992) han indicado que el clima parece ser el principal factor que determina la forma de la curva regional de frecuencia.

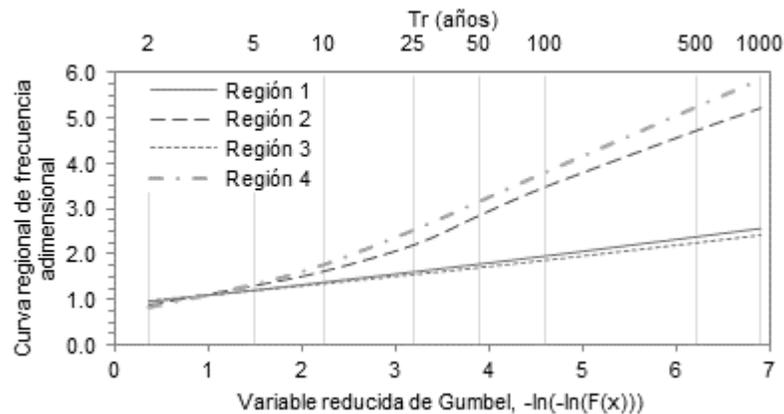


Figura 6.12 curvas regionales de frecuencia adimensionales estimadas para las regiones de la zona de estudio

Tabla 6.11 Valores de las curvas regionales de frecuencia adimensionales para los diferentes periodos de retorno

Tr (años)	Eventos de diseño adimensionales			
	Región 1	Región 2	Región 3	Región 4
2	0.95	0.88	0.97	0.83
5	1.21	1.31	1.20	1.34
10	1.38	1.63	1.36	1.78
25	1.61	2.20	1.56	2.54
50	1.78	2.86	1.71	3.17
100	1.96	3.48	1.87	3.80
500	2.38	4.72	2.25	5.24
1000	2.57	5.23	2.43	5.86

En sitios muestreados, con al menos 30 años de registro, se pueden emplear los valores adimensionales de la Tabla 6.11 para obtener los eventos de diseño asociados a diferentes periodos de retorno al multiplicarlos por el valor medio de los gastos máximos anuales del sitio en estudio. A continuación se dan unas relaciones para obtener el gasto promedio de los gastos máximos anuales en sitios no muestreados o con registros cortos, adicionalmente se puede emplear el software HidroSig (Poveda et al, 2007) para obtener este valor.

Las relaciones que se presentan a continuación fueron obtenidas con las estaciones de cada una de las regiones homogéneas que tuvieran al menos 30 años de registros.

- Región 1

$$\bar{Q} = 0.4189A^{0.9316} \quad (6-21)$$

Donde A es el área de cuenca en km^2 y \bar{Q} es el gasto promedio de los gastos máximos anuales en m^3/s . El coeficiente de determinación de la relación es $R^2 = 0.8828$ y el error estándar de la estimación es de 0.5737.

- Región 2

$$\bar{Q} = 2.5736A^{0.6773} \quad (6-22)$$

Donde A es el área de cuenca en km^2 y \bar{Q} es el gasto promedio de los gastos máximos anuales en m^3/s . El coeficiente de determinación de la relación es $R^2 = 0.7908$ y el error estándar de la estimación es de 0.5885.

- Región 3

$$\bar{Q} = 2.5243A^{0.7076} \quad (6-23)$$

Donde A es el área de cuenca en km^2 y \bar{Q} es el gasto promedio de los gastos máximos anuales en m^3/s . El coeficiente de determinación de la relación es $R^2 = 0.8778$ y el error estándar de la estimación es de 0.4563.

- Región 4

$$\bar{Q} = 0.6822A^{0.6844} \quad (6-24)$$

Donde A es el área de cuenca en km^2 y \bar{Q} es el gasto promedio de los gastos máximos anuales en m^3/s . El coeficiente de determinación de la relación es $R^2 = 0.7512$ y el error estándar de la estimación es de 0.6308.

En la Figura 6.13 se presentan las curvas de regresión entre el gasto medio de los gastos máximos anuales y el área de los sitios de cada una de las regiones homogéneas que tienen 30 o más años de registro. Dichas relaciones, como se comentó anteriormente, pueden ser empleadas para la estimación del gasto medio de los gastos máximos anuales en sitios con registros cortos o no muestreados para obtener los eventos de diseño.

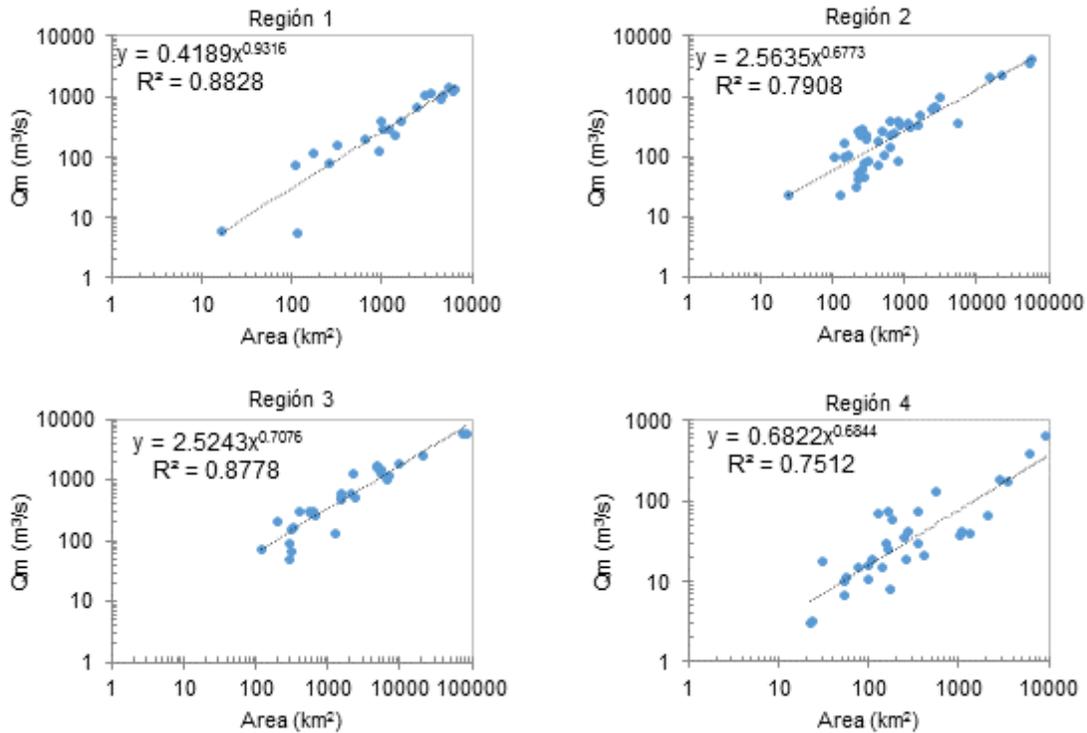


Figura 6.13 Regresiones entre el gasto medio de los gastos máximos anuales y el área de las distintas regiones homogéneas

6.7.6.2 Método de los momentos estandarizados de probabilidad pesada (PMWs)

El método de los momentos estandarizados de probabilidad pesada es un método similar a la técnica de la avenida índice, diferenciándose de esta al emplear los momentos L para su desarrollo. Si bien es una técnica que ha sido empleada ampliamente para análisis de gastos máximos (Hosking y Wallis, 1997; Rao y Hammed, 1997; Kumar y Chatterjee, 2005; Saf, 2009), poco se ha documentado sobre su uso y efectividad en el análisis regional de frecuencia de muestras que provienen de más de un mecanismo de generación como es el caso de zonas donde los gastos máximos están influenciados por tormentas tropicales, el deshielo o mecanismos de baja frecuencia como el ENSO.

Para su aplicación se empleó los registros comunes de cada región y se empleó la distribución general de valores extremos (GEV) para la estimación de los eventos de

diseño. Los parámetros de la distribución se obtuvieron empleando las relaciones dadas en el numeral 4.4.3. En la Figura 6.14 se presentan las curvas regionales de frecuencia adimensionales obtenidas con el método de los momentos estandarizados de probabilidad pesada y empleando la distribución GEV. Para la obtención de los momentos se emplearon un registro común de 40 años para las regiones 1, 2, y 3 y de 41 para la región 4. En la Tabla 6.12 se presentan los valores adimensionales de las curvas regionales de frecuencia. El procedimiento para obtener los eventos de diseño en sitios con registros cortos o no muestreados es igual el empleado en la técnica de las estaciones año.

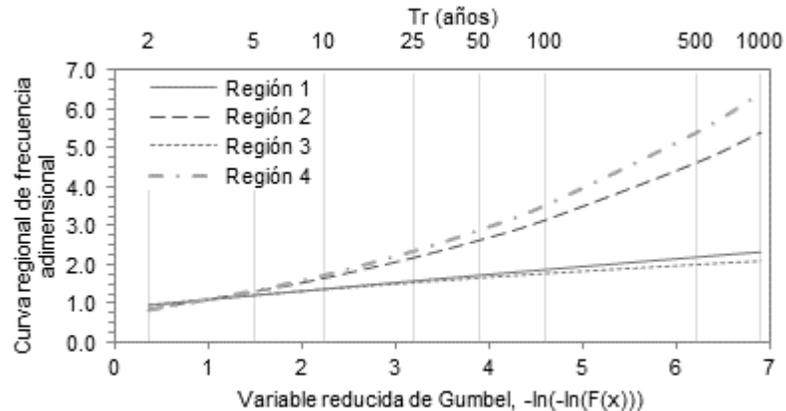


Figura 6.14 Curvas regionales de frecuencia empleando la técnica de momentos estandarizados de probabilidad pesada

Tabla 6.12 Valores de las curvas regionales de frecuencia empleando la técnica de momentos estandarizados de probabilidad pesada

Tr (años)	Eventos de diseño adimensionales			
	Región 1	Región 2	Región 3	Región 4
2	0.96	0.86	0.97	0.83
5	1.21	1.31	1.22	1.33
10	1.38	1.66	1.37	1.74
25	1.58	2.18	1.54	2.35
50	1.73	2.64	1.66	2.90
100	1.87	3.15	1.78	3.53
500	2.19	4.62	2.01	5.40
1000	2.33	5.40	2.10	6.42

6.7.6.3 Análisis multivariado

En general, el análisis multivariado empleado en el análisis regional de frecuencia permite una disminución del error estándar de ajuste (Escalante-Sandoval, 1998; Raynal-Villaseñor y Salas, 2008). El modelo a emplear en este estudio es el modelo logístico bivariado el cual tiene la forma (Gumbel, 1960):

$$F(x, y, m) = \exp \left\{ - \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{1/m} \right\} \quad (6-25)$$

Donde $F(x)$ y $F(y)$ serán el modelo mezclado Gumbel o doble Gumbel (ecuación 5-6) descrito en el capítulo 5. El método de máxima verosimilitud fue propuesto por Raynal (1985) para la estimación de los parámetros del modelo bivariado, y fue seleccionado debido a su amplia aplicabilidad y eficiencia, propiedades que no son fáciles de encontrar en otros métodos de estimación de parámetros.

$$\begin{aligned} F(x) &= p e^{-e^{-\frac{(x-\mu_1)}{\alpha_1}}} + (1-p) e^{-e^{-\frac{(x-\mu_2)}{\alpha_2}}} \\ F(y) &= q e^{-e^{-\frac{(y-\mu_3)}{\alpha_3}}} + (1-q) e^{-e^{-\frac{(y-\mu_4)}{\alpha_4}}} \end{aligned} \quad (6-26)$$

La forma general de la función de verosimilitud del modelo bivariado es (Raynal, 1985):

$$L(x, y, \underline{\theta}) = \left[\prod_{i=1}^{n_1} f(r_i, \underline{\theta}_1) \right]^{I_1} \left[\prod_{i=1}^{n_2} f(x_i, y_i, \underline{\theta}_2) \right]^{I_2} \left[\prod_{i=1}^{n_3} f(s_i, \underline{\theta}_3) \right]^{I_3} \quad (6-27)$$

Donde n_1 es la longitud de los registros antes del periodo común, n_3 es la longitud de los registros después del periodo común, n_2 es la longitud de los registros del periodo común, r es la variable con longitud n_1 , x, y son las variables con longitud n_2 , s es la variable con longitud n_3 , y I_i es un número indicador tal que $I_i = 1$ si $n_i > 0$ o $I_i = 0$ si $n_i = 0$.

El logaritmo de la función de máxima verosimilitud es:

$$\ln L(x, y, \underline{\theta}) = I_1 \sum_{i=1}^{n_1} \ln f(r_i, \underline{\theta}_1) + I_2 \sum_{i=1}^{n_2} \ln f(x_i, y_i, \underline{\theta}_2) + I_3 \sum_{i=1}^{n_3} \ln f(s_i, \underline{\theta}_3) \quad (6-28)$$

Los parámetros del modelo bivariado son obtenidos por la maximización de la función de la ecuación 6-30. Para este estudio se empleó el método de optimización no lineal Gradiente Reducido Generalizado (Lasdon et al, 1975) para maximizar dicha función. Debido a que la optimización no lineal no garantiza el óptimo global se debe emplear unos adecuados valores iniciales para obtener una solución cercana al óptimo global, para este estudio se han empleado como valores iniciales los obtenidos del ajuste del modelo mezclado en el análisis en el sitio.

El procedimiento para estimar los parámetros del análisis en el sitio empleando el análisis multivariado es el siguiente (Escalante-Sandoval, 2007):

- Paso 1: Seleccionar alguna estación para ser analizada (Estación base)
- Paso 2: Identificar el conjunto de estaciones vecinas, las pertenecientes a la región homogénea, que se pueden relacionar con la estación base.
- Paso 3: Para cada estación base, se deben analizar todos los posibles pares de combinaciones. Los valores iniciales requeridos para la optimización de la ecuación 6-30 son los obtenidos del análisis en el sitio (Tabla 5.1) y asumiendo 1.1 como el valor inicial del parámetro de asociación m .

- Paso 4: Para cada combinación bivariada con la estación base, se calcula el error estándar de ajuste (EEA). La mejor combinación es elegida de acuerdo al menor EEA encontrado de todos los pares posibles de combinaciones.
- Paso 5: La estimación regional en el sitio de los eventos de diseño para cada estación base se hace con los parámetros del modelo mezclado obtenido con la mejor combinación bivariada.

6.7.6.4 Comparación del análisis regional y el análisis en el sitio

Para evaluar los resultados de las distintas metodologías se presenta una comparación de las estimaciones obtenidas en el análisis en el sitio (caso 1), con el método regional de las estaciones-año (caso 2), con el análisis multivariado (caso 3) y con el método de los momentos estandarizados de probabilidad pesada (Caso 4). Principalmente se compararán los resultados obtenidos en el error estándar de ajuste (EEA).

En la Tabla 6.13, Tabla 6.14, Tabla 6.15 y Tabla 6.16 se presentan los resultados obtenidos en los eventos de diseño para periodos de retorno de 2, 5, 10, 25, 50 y 100 años para cada una de las estaciones de las distintas regiones obtenidos por los diferentes casos de análisis. Se presenta además el EEA y la estación (mejor estación) de la región con la que obtiene el menor error estándar de ajuste en el análisis bivariado (Caso 3).

Aun cuando algunos autores (Klemes, 2000) han señalado que la técnica de los momentos L es menos sensible a valores atípicos, lo que puede generar que se pierda información útil para la estimación de los eventos y más cuando estos valores atípicos pueden ser generados por diferentes mecanismos como el ENSO, la magnitud del error estándar de ajuste (EEA) del método de los momentos estandarizados de probabilidad pesada y del método de las estaciones-año es similar para los sitios de las regiones analizadas. En general, para todas la regiones, el uso del análisis bivariado para la estimación de los parámetros en el análisis en el sitio da menores EEA que en análisis en el sitio y los métodos regionales.

- Región 1

Tabla 6.13 Eventos de diseño (en m³/s) para diferentes periodos de retorno para las estaciones de la región 1 estimados por los diferentes casos

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
21017050	1	77	94	105	118	129	139	4.20	
	2	76	95	109	128	141	155	5.06	
	3	77	94	105	119	129	139	4.20	4
	4	75	96	109	126	138	150	5.18	
21017040	1	1041	1244	1369	1524	1639	1752	52.63	
	2	1006	1280	1474	1718	1900	2080	46.65	
	3	1046	1275	1417	1594	1725	1854	39.21	60
	4	1001	1272	1448	1667	1826	1983	38.51	
21037010	1	369	466	529	609	668	726	24.93	
	2	368	466	532	619	685	753	24.56	
	3	370	469	533	613	673	732	23.80	60
	4	369	468	532	611	668	724	24.69	
21047010	1	1363	1673	1876	2133	2325	2517	66.54	
	2	1348	1712	1952	2255	2479	2702	57.30	

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
21057050	3	1356	1695	1919	2203	2413	2622	56.11	2
	4	1349	1712	1946	2235	2444	2647	56.63	
	1	277	329	362	403	433	464	7.18	
	2	267	336	387	452	501	549	18.53	
	3	277	331	365	408	440	471	6.78	
21017020	4	265	337	384	444	488	531	17.73	18
	1	135	192	241	391	409	421	7.44	
	2	147	184	209	240	264	286	43.00	
	3	135	192	241	391	409	421	7.44	
21057030	4	148	184	208	235	255	274	43.81	18
	1	645	768	874	1033	1153	1272	28.44	
	2	617	777	894	1042	1151	1259	34.00	
	3	648	775	888	1051	1174	1296	25.96	
22027020	4	612	775	882	1016	1115	1212	39.71	19
	1	121	158	177	198	212	227	4.20	
	2	123	156	178	205	226	246	5.32	
	3	120	159	185	198	208	217	3.44	
22027010	4	123	156	177	204	223	241	4.70	2
	1	200	281	364	401	424	446	9.17	
	2	215	271	309	358	396	434	27.75	
	3	200	285	368	405	429	452	9.00	
22057060	4	216	272	307	350	380	410	28.72	5
	1	897	1087	1210	1366	1482	1597	22.51	
	2	876	1118	1284	1502	1670	1841	70.74	
	3	907	1078	1188	1327	1430	1532	17.62	
22057040	4	879	1121	1279	1477	1623	1767	61.78	18
	1	1108	1526	1884	2121	2273	2418	67.83	
	2	1184	1498	1704	1964	2157	2348	122.08	
	3	1104	1525	1952	2271	2476	2672	53.83	
21057060	4	1188	1500	1696	1934	2103	2264	126.96	15
	1	1051	1250	1399	1620	1791	1963	84.97	
	2	1005	1280	1465	1704	1885	2069	76.54	
	3	993	1239	1447	1804	2127	2464	58.40	
	4	1010	1280	1454	1668	1824	1974	81.23	

- Región 2

Tabla 6.14 Eventos de diseño (en m³/s) para diferentes periodos de retorno para las estaciones de la región 2 estimados por los diferentes casos

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
21087050	1	19	30	46	71	81	90	3.23	24
	2	21	32	39	52	67	83	7.86	
	3	19	30	49	73	83	92	2.79	
	4	21	31	40	52	63	75	7.82	
21107030	1	24	45	62	81	95	108	4.83	
	2	28	41	51	69	89	109	8.51	

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
21107020	3	24	46	63	83	97	111	4.35	13
	4	27	41	52	68	82	98	8.67	
	1	120	190	263	336	360	381	7.62	
	2	123	182	227	306	401	491	20.17	
	3	120	191	267	344	370	393	6.61	
21127020	4	120	182	231	303	366	436	19.72	28
	1	64	94	121	305	391	451	19.40	
	2	73	109	135	182	233	279	46.10	
	3	64	96	129	369	451	518	9.96	
21127030	4	72	108	137	179	214	254	49.22	24
	1	32	52	80	110	131	151	10.80	
	2	37	56	69	93	120	146	14.67	
	3	32	56	88	122	145	168	9.18	
21137030	4	37	56	70	92	111	131	14.97	30
	1	378	512	599	710	791	872	28.15	
	2	353	526	658	897	1165	1411	68.49	
	3	377	526	623	744	834	923	24.14	
21147030	4	344	526	670	884	1070	1282	56.92	26
	1	459	598	694	822	920	1021	25.99	
	2	427	638	798	1087	1410	1710	94.00	
	3	453	598	709	854	962	1071	19.24	
21147010	4	418	638	813	1072	1297	1552	81.60	13
	1	532	740	912	1833	2619	3075	128.40	
	2	569	847	1055	1419	1818	2197	266.14	
	3	528	734	908	2014	2743	3222	99.40	
22067010	4	557	845	1069	1398	1680	1996	278.16	11
	1	177	295	406	592	735	874	33.86	
	2	194	288	358	480	620	759	64.60	
	3	180	289	386	720	784	832	17.89	
22077060	4	189	287	364	477	576	687	66.54	38
	1	62	94	168	189	190	190	13.02	
	2	68	101	125	166	218	270	20.23	
	3	60	91	188	202	210	218	5.81	
21217120	4	66	101	127	167	201	239	19.31	33
	1	71	115	148	174	192	209	5.83	
	2	74	111	138	187	244	298	8.16	
	3	72	120	153	183	203	222	4.62	
23017030	4	73	111	141	185	224	267	6.19	9
	1	290	561	784	1039	1219	1396	45.76	
	2	342	507	629	838	1092	1354	112.92	
	3	302	511	921	1078	1167	1249	31.02	
23067020	4	334	505	640	838	1009	1202	109.61	28
	1	859	1366	1725	2188	2534	2879	129.47	
	2	870	1291	1607	2159	2807	3458	125.83	
	3	855	1311	1692	2664	2895	3060	68.63	
21137020	4	844	1289	1642	2172	2634	3161	133.13	
	1	44	51	56	62	67	71	1.53	

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
	2	40	59	74	101	130	158	12.88	
	3	43	52	58	66	72	78	1.04	20
	4	39	59	76	100	121	145	13.45	
	1	211	251	305	383	440	496	13.03	
21227010	2	174	259	324	443	576	698	31.21	
	3	211	251	306	384	441	498	12.99	34
	4	170	259	330	437	530	636	31.46	
21257070	1	267	396	480	586	665	742	30.11	
	2	252	375	467	634	826	1007	25.09	
	3	263	404	498	617	705	792	23.93	10
	4	246	374	475	627	759	909	33.11	
	1	52	77	104	153	189	224	18.23	
	2	56	83	104	141	181	218	19.48	
	3	54	76	93	122	314	317	6.36	24
	4	55	83	105	139	168	200	20.93	
	1	133	209	269	408	661	820	38.07	
	2	144	213	266	360	461	554	68.85	
	3	133	210	273	499	803	990	19.83	38
	4	141	213	269	352	424	504	73.55	
	1	285	412	497	603	683	761	31.25	
	2	271	404	503	683	893	1088	29.08	
	3	272	401	514	664	776	886	18.14	14
	4	265	403	512	675	816	976	23.55	
	1	104	140	161	187	207	226	12.03	
	2	92	137	171	232	302	367	10.48	
	3	102	141	164	193	214	235	11.01	27
	4	89	136	174	229	278	333	10.82	
	1	4251	4650	5028	5545	5928	6308	105.04	
	2	3725	5584	7007	9608	12426	15040	1347.42	
	3	4259	4670	5065	5591	5980	6365	96.74	28
	4	3648	5603	7148	9440	11426	13676	1546.81	
	1	89	131	149	170	185	200	6.07	
	2	86	128	160	218	283	342	14.73	
	3	88	134	156	180	198	215	4.83	24
	4	84	128	162	214	260	311	11.96	
	1	228	380	478	598	687	775	23.17	
	2	229	340	423	571	747	915	43.96	
	3	219	386	502	644	748	851	16.48	38
	4	224	339	431	567	687	822	45.61	
	1	3455	3938	4254	4652	4947	5240	51.66	
	2	3075	4617	5802	7978	10308	12464	1065.57	
	3	3476	3956	4253	4624	4899	5172	48.10	26
	4	3014	4629	5907	7809	9459	11332	1325.18	
	1	183	358	463	590	684	776	35.99	
	2	201	296	368	494	643	789	70.49	
	3	181	365	475	608	705	800	32.57	13
	4	196	296	375	492	594	709	76.62	

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
21087040	1	75	131	197	377	385	392	7.82	
	2	89	132	163	219	280	341	50.29	
	3	75	133	201	378	388	396	7.41	33
	4	87	131	166	217	261	310	51.81	

- Región 3

Tabla 6.15 Eventos de diseño (en m³/s) para diferentes periodos de retorno para las estaciones de la región 3 estimados por los diferentes casos

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
24027050	1	260	366	390	418	440	467	7.22	
	2	268	331	373	429	473	518	19.40	
	3	261	366	390	418	441	468	7.17	53
	4	267	335	377	425	458	489	16.14	
24017580	1	163	191	216	256	286	317	7.17	
	2	165	205	231	265	291	318	11.06	
	3	163	191	219	262	295	327	6.24	52
	4	164	208	234	264	285	304	12.37	
24067030	1	2391	2829	3103	3445	3697	3946	130.15	
	2	2345	2900	3275	3769	4153	4550	153.49	
	3	2377	2836	3127	3489	3757	4022	125.68	57
	4	2334	2943	3314	3750	4050	4330	155.43	
24047020	1	1845	2085	2223	2392	2516	2638	39.08	
	2	1735	2141	2421	2798	3096	3406	138.32	
	3	1853	2106	2251	2429	2560	2689	33.79	52
	4	1724	2177	2456	2784	3012	3227	130.46	
24027010	1	582	695	768	860	929	996	13.87	
	2	576	716	809	931	1025	1121	29.49	
	3	584	700	776	872	942	1012	13.42	57
	4	574	726	817	925	998	1067	31.34	
24027030	1	248	321	365	421	461	502	15.83	
	2	250	308	346	396	435	474	27.40	
	3	247	326	374	434	478	522	13.00	53
	4	250	312	349	392	421	448	25.60	
23097030	1	5535	6396	6982	7896	8580	9259	91.20	
	2	5418	6753	7648	8817	9717	10640	462.78	
	3	5533	6395	6984	7899	8584	9263	91.27	62
	4	5410	6810	7657	8642	9317	9943	397.28	
23087150	1	42	64	80	99	113	127	4.70	
	2	47	58	66	75	81	88	13.38	
	3	43	61	85	113	120	126	1.92	52
	4	48	59	65	73	78	82	13.34	
23087200	1	201	265	305	356	394	431	15.98	
	2	201	246	276	317	348	381	35.35	
	3	200	279	330	394	442	489	10.37	59
	4	200	249	278	312	336	358	36.54	

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
23087180	1	73	90	99	114	128	143	3.00	56
	2	68	82	93	109	121	133	7.67	
	3	73	90	99	113	128	143	3.00	
	4	67	84	95	107	116	125	8.91	
23087190	1	607	681	729	790	834	878	10.48	59
	2	595	741	839	964	1060	1157	76.07	
	3	607	685	735	798	844	890	9.57	
	4	593	751	848	961	1038	1110	76.26	
23127020	1	1363	1700	1922	2204	2413	2620	54.52	56
	2	1388	1714	1932	2219	2440	2668	60.72	
	3	1323	1692	1936	2242	2469	2693	44.83	
	4	1379	1742	1961	2218	2394	2558	55.14	

Región 4

Tabla 6.16 Eventos de diseño (en m³/s) para diferentes periodos de retorno para las estaciones de la región 4 estimados por los diferentes casos

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
24017610	1	20	40	49	59	66	72	2.05	48
	2	22	35	46	67	85	103	3.67	
	3	20	40	49	59	66	72	2.05	
	4	22	35	45	62	76	93	3.04	
23197270	1	27	56	74	96	111	127	6.26	42
	2	29	47	62	88	110	133	10.62	
	3	27	60	79	103	120	137	4.89	
	4	30	46	60	81	99	121	12.67	
23197290	1	112	175	227	266	292	317	5.57	42
	2	108	175	236	339	426	511	24.36	
	3	112	178	230	272	299	324	4.36	
	4	108	176	230	313	387	473	16.58	
24037080	1	13	22	34	69	75	80	2.73	49
	2	16	25	33	46	57	68	8.80	
	3	13	22	36	69	75	80	2.59	
	4	16	25	33	44	54	66	9.12	
24037040	1	27	38	46	55	62	69	3.05	49
	2	24	40	55	78	97	115	6.75	
	3	27	38	46	55	62	69	3.05	
	4	25	40	52	72	89	109	4.92	
24037070	1	12	22	27	32	35	39	0.96	50
	2	12	19	26	37	48	57	2.52	
	3	12	22	27	33	37	41	0.84	
	4	12	19	25	34	43	52	1.98	
24037450	1	4	8	17	27	34	40	3.99	39
	2	6	9	12	16	20	23	5.09	
	3	4	8	18	29	37	44	3.66	
	4	6	9	11	15	18	22	6.71	

Estación	Caso	Periodo de retorno (años)						EEA (m ³ /s)	Mejor estación
		2	5	10	25	50	100		
24037120	1	13	20	27	54	58	62	1.63	
	2	13	22	29	40	50	60	4.53	
	3	13	20	28	56	60	64	1.35	45
	4	13	21	28	38	47	57	4.95	
24037300	1	15	22	27	36	59	60	1.01	
	2	14	23	31	45	56	67	2.35	
	3	16	23	28	38	59	59	0.91	47
24027040	1	47	89	104	117	126	135	3.45	
	2	48	77	103	149	187	226	11.20	
	3	47	90	105	119	129	138	3.17	42
	4	48	77	101	137	169	206	8.08	
24037110	1	17	25	34	54	69	83	3.46	
	2	17	28	38	54	66	79	4.22	
	3	17	25	38	64	83	100	1.79	45
	4	18	28	37	50	61	74	4.82	
23107130	1	63	99	122	151	172	193	6.27	
	2	56	90	121	175	221	266	10.21	
	3	64	102	126	156	178	201	4.92	42
	4	56	90	118	161	199	244	10.15	

Como se comentó anteriormente, el método multivariado permite obtener una disminución en el error estándar de ajuste con lo cual se disminuye la incertidumbre en las estimaciones de los eventos en los sitios. Sin embargo, no es posible emplear esta técnica en sitios no muestreados, pero si se pueden emplear las estimaciones de los eventos obtenidos con esta técnica para combinarlo con las propiedades morfométricas de la cuenca y mediante la técnica de correlación y regresión múltiple poder obtener relaciones para estimar los eventos de diseño en sitios no muestreados.

Así, se obtuvieron para cada una de las regiones relaciones de máximo dos variables entre algunos parámetros morfométricos de la cuenca y los eventos de diseño para periodos de retorno de 2, 5, 10, 25, 50 y 100 años, estimados mediante la técnica del análisis multivariado para obtener los eventos de diseño en sitios no muestreados. Las relaciones para la obtención de los eventos de diseño en sitios no muestreados se presentan en las siguientes tablas. En las regiones 2 y 4 las variables que mejor describen la respuesta hidrológica de la cuenca son el área y la elevación mínima de la cuenca; en la región 1 son el área, la elevación máxima del río y la pendiente media de la cuenca; en la región 3 son el perímetro de la cuenca, la pendiente media de la cuenca, la longitud del cauce principal y la pendiente media del río.

Tabla 6.17 Relaciones para la estimación de eventos de diseño en sitios no muestreados de la región 1

Tr (años)	Relación	R ²
2	$Q_2 = 496580.0610A^{1.0115}C_{max}^{-1.8224}$	0.932
5	$Q_5 = 192086.2940A^{0.9853}C_{max}^{-1.6496}$	0.931
10	$Q_{10} = 69057.8612A^{0.9720}C_{max}^{-1.4909}$	0.926
25	$Q_{25} = 45.8511A^{0.9571}S_c^{-1.1757}$	0.870
50	$Q_{50} = 53.4261A^{0.9788}S_c^{-1.2375}$	0.874
100	$Q_{100} = 60.6306A^{0.9987}S_c^{-1.2912}$	0.877

Q_i Evento de diseño asociado al periodo de retorno i en m³/s; A : Área de la cuenca en km²; C_{max} : Elevación máxima del río en msnm; S_c : Pendiente media de la cuenca en porcentaje.

Tabla 6.18 Relaciones para la estimación de eventos de diseño en sitios no muestreados de la región 2

Tr (años)	Relación	R ²
2	$Q_2 = 651.2465A^{0.5363}C_{min}^{-0.7578}$	0.825
5	$Q_5 = 1195.9779A^{0.4853}C_{min}^{-0.7381}$	0.796
10	$Q_{10} = 1845.1967A^{0.4509}C_{min}^{-0.7288}$	0.787
25	$Q_{25} = 2329.3053A^{0.4284}C_{min}^{-0.6928}$	0.722
50	$Q_{50} = 1202.9870A^{0.4360}C_{min}^{-0.5693}$	0.675
100	$Q_{100} = 1349.4421A^{0.4311}C_{min}^{-0.5673}$	0.664

Q_i Evento de diseño asociado al periodo de retorno i en m³/s; A : Área de la cuenca en km²; C_{min} : Elevación mínima del río en msnm.

Tabla 6.19 Relaciones para la estimación de eventos de diseño en sitios no muestreados de la región 3

Tr (años)	Relación	R ²
2	$Q_2 = 0.1507P^{1.4793}S_r^{0.3666}$	0.975
5	$Q_5 = 0.2520P^{1.4251}S_r^{0.3487}$	0.974
10	$Q_{10} = 0.0047P^{1.2001}S_c^{1.4581}$	0.980
25	$Q_{25} = 0.0101P^{1.1778}S_c^{1.3158}$	0.983
50	$Q_{50} = 0.0497S_c^{1.3604}L_c^{1.1108}$	0.986
100	$Q_{100} = 0.0538S_c^{1.3640}L_c^{1.1073}$	0.988

Q_i Evento de diseño asociado al periodo de retorno i en m³/s; P : Perímetro de la cuenca en km; S_r : Pendiente del cauce principal m/km; S_c : Pendiente media de la cuenca en porcentaje, L_c : Longitud del cauce principal en km.

Tabla 6.20 Relaciones para la estimación de eventos de diseño en sitios no muestreados de la región 4

Tr (años)	Relación	R^2
2	$Q_2 = 38656.3288A^{0.2988}C_{min}^{-1.1891}$	0.743
5	$Q_5 = 58165.2577A^{0.3067}C_{min}^{-1.1773}$	0.789
10	$Q_{10} = 62322.3221A^{0.2643}C_{min}^{-1.1172}$	0.876
25	$Q_{25} = 33392.3944A^{0.2304}C_{min}^{-0.9678}$	0.869
50	$Q_{50} = 47344.3103A^{0.1678}C_{min}^{-0.9516}$	0.847
100	$Q_{100} = 39622.6235A^{0.1967}C_{min}^{-0.9341}$	0.872

Q_i Evento de diseño asociado al periodo de retorno i en m^3/s ; A : Área de la cuenca en km^2 ; C_{min} Elevación mínima del río en msnm.

7 CONCLUSIONES Y RECOMENDACIONES

7.1 EL ¿POR QUÉ? DEL ÁREA DE ESTUDIO

El área de estudio utilizada en esta tesis se elige por los mecanismos climáticos que pueden afectar su respuesta hidrológica. Varios autores han identificado que en el área de estudio el fenómeno del ENSO afecta, en su fase fría, la magnitud y frecuencia de los caudales máximos. Las estaciones hidrométricas fueron elegidas para:

- Analizar la no estacionariedad de los registros y sus posibles causas
- Evaluar la influencia del fenómeno del ENSO en la magnitud de los caudales máximos.
- Identificar el tipo de distribución más adecuado para la estimación de los eventos de diseño asociados a diferentes periodos de retorno
- Identificación de regiones homogéneas para la estimación de eventos empleando técnicas regionales
- Identificación del mejor método regional para la estimación de los eventos de diseño

7.2 CONCLUSIONES

Los aspectos que pueden tener implicaciones en el análisis de frecuencia de caudales máximos en la zona de estudio son reflejados a partir de las siguientes conclusiones:

- Si bien hay ciertos autores que indican que en registros cortos (menos de 30 años) la causa más probable de no estacionariedad es la variabilidad climática, para la zona de estudio se ha encontrado que la variabilidad climática puede causar no estacionariedad aún en estaciones con más de 40 años de registros y por lo tanto, 30 años de registros no pueden ser suficientes para obtener un análisis de homogeneidad confiable.
- Del análisis de homogeneidad realizado a 62 estaciones de la zona de estudio se ha encontrado que 45 estaciones son no homogéneas. De las estaciones no estacionarias, un 33% presentan correlación serial, un 84% de las estaciones no homogéneas presentan cambio en la media, de éstas el 71% presentan una media decreciente, y un 50% presentan cambio en la varianza. Si bien un 51% de las estaciones no estacionarias presentan tendencia, solo 2 estaciones presentan tendencia sin presentar cambio en la media, 3 estaciones presentan tendencia antes y después del cambio en la media y el resto solo presentan tendencia antes o después del punto de cambio en la media.
- De las 45 estaciones no estacionarias en 23 (51%) es más adecuado emplear los modelos no estacionarios empleando como covariables índices del ENSO. En 19 (42%) estaciones la desviación estadística de los modelos con covariables (MEI y tiempo) indican que son más adecuados, es decir, que la covariable describe mejor el comportamiento de los datos y por lo tanto, la causa de la no estacionariedad de estas estaciones puede ser la variabilidad climática.
- Varios autores han indicado la conveniencia del uso de distribuciones mezcladas en muestras heterogéneas donde los gastos máximos proviene de más de un mecanismo.

A partir del criterio de la desviación estadística entre el modelo Gumbel y el modelo doble Gumbel o modelo mezclado Gumbel de dos componentes se ha encontrado que en 43 estaciones (69%) de la zona de estudio es más adecuado emplear este último para la estimación de los eventos de diseño.

- A partir del empleo del método de la región de influencia, el algoritmo de Ward y el método de los trazos multidimensionales se han identificado 4 regiones homogéneas desde el punto de vista físico, las cuales a partir de la técnica de los momentos-L se ha encontrado que son homogéneas desde el punto de vista hidrológico. Así, se han identificado 4 regiones homogéneas para la zona de estudio tanto física como hidrológicamente. Recomendando el uso de distribuciones mezcladas para la estimación de los eventos de diseño en las distintas regiones homogéneas especialmente en la región 4.
- Mediante el uso del error estándar de ajuste se ha encontrado para la zona de estudio que el método multivariado modelo logístico bivariado con distribuciones marginales doble Gumbel permite una disminución del error estándar de ajuste y es el recomendado para la estimación de los eventos de diseño en sitios muestreados. Para los sitios no muestreados se han proporcionado relaciones empleando la técnica de regresión y correlación múltiple entre los eventos obtenidos con el análisis bivariado y las propiedades morfométricas de las cuencas.

7.3 RECOMENDACIONES

Los siguientes párrafos presentan algunas recomendaciones que pueden ayudar a un mejor entendimiento de los mecanismos que intervienen en la respuesta hidrológica de la cuenca.

- En el uso de modelos no estacionarios con covariables además de emplear modelos con una sola covariable, se pueden emplear modelos con dos covariables como el ENSO y el tiempo, que pueden permitir evidenciar que en la zona además de los mecanismos de baja frecuencia, los cambios antropogénicos modifican o han modificado la respuesta hidrológica de la zona de estudio.
- Para mejorar la estimación en sitios no muestreados, sobre todo en cuencas pequeñas, se recomienda realizar análisis comparativos con información de cuencas pequeñas instrumentadas y que permitan mejorar las estimaciones en este tipo de cuencas.
- En algunas de las estaciones empleadas en el estudio que tienen un área mayor a 10000 km² parece que hay un mayor efecto del área en el L-CV y por ende en la forma de la distribución de frecuencias. Por lo tanto, se debe tener cuidado cuando se emplee el análisis regional en este tipo de cuencas cuando se encuentren insertas en regiones homogéneas con altos L-CV.
- Si bien se ha encontrado que el análisis multivariado disminuye el error estándar de ajuste, se recomienda que la estimación de los eventos tanto en sitios muestreados como en no muestreados se obtenga a partir del promedio de los resultados obtenidos con las diferentes técnicas empleadas en este estudio.

8 BIBLIOGRAFÍA

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716-723.
- Alexandersson H, Moberg A. (1997). Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. *International Journal of Climatology* 17: 25–34.
- Alila, Y., and Mtiraoui, A. (2002). Implications of heterogeneous flood frequency distributions on traditional stream-discharge prediction techniques. *Hydrolog. Process.*, 16, 1065–1084.
- Anderson, R. L. (1942). Distribution of the serial correlation coefficients. *Annals of Mathematical Statistics* 13(1): 1–13.
- Andrews, D. F. (1972). Plots of high-dimensional data. *Biometrics*, 28, 125-136.
- Armenteras, D. and N. Rodríguez. (2007). Monitoreo de los ecosistemas andinos 1985-2005: síntesis. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, pp: 176.
- Apipattanavis, S., Rajagopalan, B., and Lall, U.(2010): Local polynomial based flood frequency estimator for mixed population, *J. Hydrol. Eng.*, 15, 680–691.
- Bernal, D. E., (2013). El río Magdalena: Escenario primordial de la patria. *Revista Credencial*, Banco de la República de Colombia. Credencial historia No. 282.
- Bloschl, G., Sivapalan, M., (1997). Process controls on regional flood frequency: coefficient of variation and basin scale. *Water Resour. Res.* 33, 2967–2980.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society B*, 26, 211–252.
- Buishand TA. 1982. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* 58: 11–27.
- Buishand, T.A., (1984). Bivariate extreme-value data and the station year method. *Journal of Hydrology* 69, 77–95.
- Buishand, T.A., (1991). Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal* 36 (4), 345–365.
- Burn, D.H., (1990). Evaluation of regional flood frequency-analysis with a region of influence approach. *Water Resour. Res.* 26 (10), 2257–2265.
- Burn, D.H., Zrinji, Z., Kowalchuk, M., (1997). Regionalization of catchments for regional flood frequency analysis. *J. Hydrol. Eng.* 2 (2), 76–82.
- Burn, D.H. and Goel, N.K., (2000). The formation of groups for regional flood frequency estimation. *Hydrolog. Sci. J.*, 45, 97–112.
- Campos, A. D. F (1989)., “Estimación de los parámetros óptimos de la distribución Gumbel mixta por medio del algoritmo de Rosenbrok”. *Ingeniería Hidráulica en México*, Vol. V, num. 3, pp. 9 – 18.
- Castellarin, A., Burn, D.H., Brath, A., (2001). Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *J. Hydrol.* 241 (3–4), 270–285.
- Coles, G. S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, 208 pp., Springer, New York.

- Chiew, F. H. S. & McMahon, T. A. (1993) Assessing the adequacy of catchment streamflow yield estimates. *Aust. J. Soil Res.* 31, 665–680.
- Cunnane, C (1985)., Factors affecting choice of distribution for flood series, *Hydrol. Sci. J.*, 30, 25-36.
- Cunnane, C., (1988). Methods and merits of regional flood frequency analysis. *J. Hydrol.* 100, 269–290.
- Cruff, R.W. and Rantz, S.E. (1965) A comparison of methods used in flood frequency studies for coastal basins in California. *U.S. Geol. Surv. Water Supply Pap.*, 1580-E, 56pp.
- Dales M. Y, Reed D. W. (1989). Regional flood and storm hazard assessment. Institute of Hydrology Rep. No. 102. IH, Wallingford, UK.
- Dalrymple T (1960). Flood-frequency analyses, US Geological Survey water-supply paper 1543-A.
- Das, S., (2011). Examination of flood estimation techniques in the Irish context. Thesis (PhD), Department of Engineering Hydrology, NUI Galway, 2010.
- Das S, Cunnane C (2012). Performance of flood frequency pooling analysis in a low CV context. *Hydrological Sciences Journal.*;57(3):433-444.
- Dawdy. D. R. (1961). Variation of flood ratios with size of drainage area, *US Geol. Sun: Res.*, 424-C. paper C36.
- Departamento Administrativo Nacional de Estadística - DANE (2006). Censo general.
- Departamento Administrativo Nacional de Estadística - DANE (2012). Boletines de economía.
- Dixon H, Lawler DM, Shamseldin AY, Webster P, (2006)., The effect of record length on the analysis of river flow trends in Wales and central England, *Climate Change and Variability Hydrological Impacts*, IAHS Publication 308, 490-495.
- El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., and Bobee, B. (2007)., Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, 43, doi:10.1029/2005WR004545.
- Escalante, C. and J. A. Raynal., (1994). A Trivariate Extreme Value Distribution Applied to Flood Frequency Analysis. *Journal of Research of The National Institute of Standards and Technology* 99(4):369-375.
- Escalante-Sandoval, C. (1998) Multivariate extreme value distribution with mixed Gumbel marginals. *J. Am. Water Resour. Assoc.* 3(2), 321–333.
- Escalante-Sandoval. C., (2007). A Mixed Distribution with EV1 and GEV Components for Analyzing Heterogeneous Samples. *Ingeniería, Investigación y Tecnología*, Vol. 8, No. 3, Julio/Septiembre, pp. 123-133.
- Escalante, C., y L. Reyes (2008). *Técnicas Estadísticas en Hidrología*. Facultad de Ingeniería, UNAM. México.
- Evin G, Merleau J, Perreault L (2011) Two-component mixtures of normal, gamma, and Gumbel distributions for hydrological applications. *Water Resour Res* 47. doi:10.1029/2010wr010266.
- Farquharson, F. A. K., Meigh, J. R. and Sutcliffe, J. V. (1992) Regional flood frequency analysis in arid and semi-arid areas. *J. Hydrol.* 138, 487-501.
- Finrock, H. and Rao, A. (2003) Length of Data Necessary to Identify Distributions in Flood Frequency Analysis. *World Water & Environmental Resources Congress 2003*: pp. 1-7.
- Gingras, D. and Adamowski, K. (1992) Coupling of nonparametric frequency and L-moment analyses for mixed distribution identification. *Wat. Resour. Bull.* 28(2), 263-272.

- Gumbel, E. J. (1960) Distributions des valeurs extremes en plusieurs dimensions. Publ. Inst. Stat. Univ. Paris. 9, 171–173.
- Guttman, N. B (1989)., Statistical descriptors of climate ,Bull. Am. Meteorol. Soc., 70, 602-607, 1989.
- Hanssen-Bauer I, Førland EJ. 1994. Homogenizing long Norwegian precipitation series. Journal of Climate 7: 1001–1013.
- Hirschboeck, K.K., L.L. Ely, and R.A. Maddox, (2000). Hydroclimatology of meteorologic floods, Inland Flood Hazards, edited by E.E. Wohl, Cambridge University Press, pp. 39-72.
- Hosking, J.R.M, and J.R.Wallis (1988): The effect of intersite dependence on regional flood frequency analysis, *Water Resour. Res.*, 24(4), 588-600.
- Hosking, J. R. M., and Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29, 271-281. Correction: *Water Resources Research*, 31 (1995), 251.
- Hosking, JRM; Wallis, JR (1997). Regional Frequency Analysis: An Approach Based on L-moments. Cambridge University Press.
- Hosking, J. Package ImomRFA (2009). Regional Frequency Analysis using L moments. CRAN Repository. <http://cran.r-project.org/web/packages/lmomRFA/lmomRFA.pdf>.
- IDEAM (2011). “Evaluación de la Capacidad del IDEAM para Producir la Información Hidrológica, Meteorológica y Ambiental que Necesita el País”. Econometría Consultores – 2011 – Producto 4 Informe Final y Bases de Datos.
- Instituto colombiano de desarrollo rural – Incoder (2013). Proyecto: Análisis, diseño y construcción de distritos de riego y drenaje a nivel nacional. Código BPIN 0049-00004 9999.
- Institute of Hydrology (IH), (1999). Flood Estimation Handbook. Institute of Hydrology, Wallingford.
- Ilorme, F. and Griffis, V. (2009) Providing a Physical Basis for Statistical Homogeneity in Regional Frequency Analyses. World Environmental and Water Resources Congress 2009: pp. 1-10.
- Ilorme, F.; Griffis, V.W (2013). A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation. *Journal of Hydrology*, v. 492, p. 151-162.
- Ishak, E. H., Rahman, A., Westra, S., Sharma, A., and Kuczera, G.(2013): Evaluating the non-stationarity of Australian annual maximum flood, *J. Hydrol.*, 494, 134–145.
- Jain, D. and Singh, V.P., (1987). Comparison of Some Flood Frequency Distributions Using Empirical Data. “ Ed. V.P. Singh. Hydrologic Frequency Modeling”, D. Reidel Publishing Company, p 467- 485, The Netherlands.
- Jakob, D., Reed, D. W. & Robson, A. J. (1999) Choosing a pooling-group, chapter C6. In: Flood Estimation Handbook Vol. 3. Statistical procedures for flood frequency estimation. Institute of Hydrology, Wallingford, UK.
- Jenkinson, A.F., (1969). Estimation of Maximum Floods, chap. 5. World Meteorological Office Technical Note 98.
- Johnson, R. A. and D. W. Wichern, (2002). Applied Multivariate Statistical Analysis. 5th Edition, Prentice Hall, Upper Saddle River, NJ, pp. 767.
- Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, 25, 1287–1304.
- Kar, A., Goel, N., Lohani, A., and Roy, G. (2012). ”Application of Clustering Techniques Using Prioritized Variables in Regional Flood Frequency Analysis—Case Study of Mahanadi Basin.” *J. Hydrol. Eng.*, 17(1), 213–223.

- Katz, R.W. (2013). Statistical methods for nonstationary extremes. Extremes in a Changing Climate, Springer Netherlands.
- Kendall, M. G (1975)., Rank Correlation Methods, 4th ed., Charles Griffin: London.
- Klemes, V. (1989) The improbable probabilities of extreme floods and droughts. In: Hydrology and Disasters, ed. O. Starosolszky & O. M. Melder, James & James, London, 43-51.
- Klemes, V. (2000): Tall tales about tails of hydrological distributions, *J. Hydrol. Eng.*, 5(3), 227–231.
- Kidson R and Richards K (2005) Flood frequency analysis: assumptions and alternatives. *Progress in Physical Geography* 29(3): 392–410.
- Kite, G. W (1988)., "Frequency and risk analysis in hydrology". Water Resources Publication, Colorado State, University Fort Collins.
- Konrad, V. (1944)., *Methods in Climatology*: Harvard University Press, Cambridge, Massachusetts.
- Kundzewicz, Z. W. and Robson, A (2000). Detecting Trend and Other Changes in Hydrological Data .World Climate Programme—Water, World Climate Data and Monitoring Programme, WCDMP-45, WMO/TD no. 1013, World Meteorological Organization, Geneva, Switzerland.
- Kundzewicz, Z.W., Robson, A.J., (2004). Change detection in hydrological records – a review of the methodology. *Hydrol. Sci. J.* 49 (1), 7–19.
- Kundzewicz ZW, Graczyk D, Maurer T, Pinskiwar I, Radziejewski M, Svensson C, Szwed M. (2005). Trend detection in river flow series: 1. Annual maximum flow. *Hydrological Sciences Journal* 50: 797–810.
- Kumar, R. and Chatterjee, C. (2005). "Regional Flood Frequency Analysis Using L-Moments for North Brahmaputra Region of India." *J. Hydrol. Eng.*, 10(1), 1–7.
- Leclerc, M. and Ouarda, T. (2007):, Non-stationary regional flood frequency analysis at ungauged sites, *J. Hydrol.*, 343, 254–265.
- Lasdon, L. S; . Waren, A. D; Jain, A and Ratner, M. W.(1975). Design and testing of a Generalized Reduced Gradient Code for Nonlinear Optimization. NTIS National Technical Information Service U. S. Department of Commerce, Cleveland.
- Leytham, K. M (1984)., Maximum likelihood estimates for the parameters of mixture distributions, *Water Resour. Res.*, 20(7), 896-902.
- Madsen, H. and Rosbjerg, D. (1997). The partial duration series method in regional index flood modelling. *Water Resources Research*, Vol. 33, No. 4, pp. 737–46.
- Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M., and Francis, R. C. (1997). A Pacific Interdecadal Climate Oscillation with impacts on salmon production. *Bull. Am. Meteorol. Soc.*, 78(6), 1069–1079.
- Marengo, H., (2006). Case study: dam safety during construction, lessons of the overtopping diversion works at Aguamilpa dam. *Journal of Hydraulic Engineering*, 132 (11), 1121–1127.
- McCuen, R. and Galloway, K. (2010). Record Length Requirements for Annual Maximum Flood Series. *J. Hydrol. Eng.*, 15(9), 704–707.
- Moran, P.A.P. (1959). *The Theory of Storage*. John Wiley and Sons, New York, NY.
- Mtiraoui A. (2004), *Mixture Distribution and Spatial Scale Effects on Flood Hydrology*, PhD Dissertation, Department of Forest Resources Management, Faculty of Forestry, University of British Columbia, Vancouver, Canada.
- Murphy, P. J. (2001), Evaluation of mixed-population flood-frequency analysis, *J. Hydrol. Eng.*, 6, 62–70.

- Nathan, R. and McMahon, T. (1990): Identification of homogeneous regions for the purposes of regionalization, *J. Hydrol.*, 121(1–4), 217–238.
- National Research Council, (1988). Estimating Probabilities of Extreme Floods, Methods and Recommended Research, Report by the Committee on Techniques for Estimating Probabilities of Extreme Floods, National Academy Press, Washington D.C.
- Neykov, N.M., Neytchev., P. N., Van Gelder, PHAJM., Todorov, V.K (2007). Robust detection of discordant sites in regional frequency analysis. *Water Resour Res*, 43.
- NERC (1975). Flood Studies Report. Natural Environment Research Council, London (five volumes).
- Obeysekera, J. and Salas, J.D., (2013). Quantifying the Uncertainty of Design Floods Under Nonstationary Conditions, *ASCE J. Hydrologic Engineering*, Vol. 18, No. 5, May 2013, pp. 518-526.
- Obeysekera, J. and Salas, J.D., (2014). Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events. *J. Hydrol. Eng.*, 19(3), 554–568.
- Pettitt, A. N. (1979), A non-parametric approach to the change-point problem, *Appl. Stat.*, 28, 126–135.
- Petrow, T., Zimmer, J., Merz, B., (2009). Changes in the flood hazard in Germany through changing frequency and persistence of circulation patterns. *Natural Hazards and Earth System Sciences* 9, 1409–1423.
- Pison, G., Van Aelst, S., and Willems, G. (2002), Small Sample Corrections for LTS and MCD, *Metrika*, 55, 111-123.
- Poveda, G., Mesa O. J. (1996). Las fases extremas del ENSO - El Niño y La Niña - y su influencia sobre la hidrología de Colombia. *Revista de Ingeniería Hidráulica en México*. Vol. XI, No. 1, 21-37.
- Poveda G, Mesa O J (1997). Feedbacks between hydrological processes in tropical South America and large scale oceanic– atmospheric phenomena. *J Clim* ;10:2690–702.
- Poveda, G., (2004). La hidroclimatología de Colombia: una síntesis desde la escala inter-decadal hasta la escala diurna. *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales* 107, 201–222.
- Poveda, G., Vélez, J., Mesa, O., Cuartas, A., Barco, J., Mantilla, R., Mejía, J., Hoyos, C., Ramírez, J., Ceballos, L., Zuluaga, M., Arias, P., Botero, B., Montoya, M., Giraldo, J., and Quevedo, D. (2007). Linking Long-Term Water Balances and Statistical Scaling to Estimate River Flows along the Drainage Network of Colombia. *J. Hydrol. Eng.*, 12(1), 4–13.
- Poveda, G., Alvarez, D. y Rueda, O. (2011). Hydro-climatic variability over the Andes of Colombia associated with enso: a review of climatic processes and their impact on one of the Earth's most important biodiversity hotspots, *Clim Dyn*, vol. 35, No. 6, pp.1-17.
- Potter, W. D. (1958), Upper and lower frequency curves for peak rates of runoff, *Eos Trans. AGU*, 39, 100 – 105.
- Potter, K.W., (1979). Annual precipitation in the Northeast United States: long memory, short memory, or no memory. *Water Resources Research* 15 (2), 340–346.
- Rao, A.R. and K.H. Hamed (1997). Regional Frequency Analysis of Wabash River Flood Data by L-Moments, *ASCE Journal of Hydrologic Engineering*, Vol. 2, No. 4, pp. 169-180.
- Rao, A. Ramachandra and Khaled H. Hamed (2000). *Flood Frequency Analysis*, CRC Press LLC, Boca Raton, FL.
- Rao, A. R and Srinivas, V. V., (2006) Regionalization of Watersheds by Fuzzy Cluster Analysis. *Journal of Hydrology*, Elsevier, Vol.318, Issues 1-4, pp.57-79.

- Raynal, J. A., (1985). Bivariate Extreme Value Distributions Applied to Flood Frequency Analysis. Ph. D. Dissertation, Civil Engineering Department, Colorado State University, 237 pp.
- Raynal J. A. and J. D. Salas, (1987). Multivariate Extreme Value Distributions in Hydrological Analyses. *Water for the Future: Hydrology in Perspective*, Proceedings of the Rome Symposium. IAHS Publications 164:111-119.
- Raynal-Villasenor, J. A. and Guevara-Miranda, J. L. (1997). Maximum Likelihood Estimators for the Two Populations Gumbel Distribution, *Hydrological Science and Technology J.*, Vol. 13, No. 1-4, pp 47-56.
- Raynal-Villasenor, J. and Salas, J. (2008) Using Bivariate Distributions for Flood Frequency Analysis Based on Incomplete Data. *World Environmental and Water Resources Congress 2008*: pp. 1-9.
- Reed D.W. and Stewart E.J. (1989) Focus on rainfall growth estimation. *Proc. 2nd Nat. Hydrol. Symp.*, Sheffield, Sept. 1989: 3.57-3.65.
- Reed, D. W. (1994), Rainfall frequency analysis for flood design, in *Coping with Floods*, NATO ASI Series Volume 257, 59-75.
- Rodríguez E. N (2011). Land-cover and land-use change and deforestation in Colombia: spatial dynamics, drivers and modelling. Thesis (PhD) Universitat Autònoma de Barcelona, Departament de Biologia Animal, de Biologia Vegetal i d'Ecologia.
- Rosbjerg, D., and H. Madsen (1995), Uncertainty measures of regional flood frequency estimators, *J. Hydrol.*, 167, 209 – 224.
- Rossi, F., Fiorentino, M. & Versace, P. (1984) Two-component extreme value distribution for flood frequency analysis. *Water Resour. Res.* 20(7), 847–856.
- Rousseeuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, New York: John Wiley.
- Rueda, E., 1981. Transfer of Information for Flood Related Variables, M. Sc. Thesis, Civil Engineering Department, Colorado State University, 112 pp.
- Saf, B. (2009). "Regional flood frequency analysis using L-moments for the büyük and küçük menderes river basins of Turkey", *Journal of Hydrologic Engineering*, 14 (8), pp. 783–794.
- Saf, B. (2010): Assessment of the effects of discordant sites on regional flood frequency analysis, *J. Hydrol.*, 380, 362–375.
- Salas, J.D. (1980) Transfer of information to improve estimates of flood frequencies. Chapter 17. *Hydrology for Transportation Engineers*. Thomas G. Sanders (editor). Washington, D.C.: Federal Highway Administration, pp. 592-653.
- Salas, J.D., J.W. Delleur, V. Yevjevich and W.L. Lane, (1980). *Applied Modeling of Hydrologic Time Series*. 1st Edn., Water Resources Publication, Littleton, Colorado, ISBN-10: 0918334373, pp: 484.
- Salas, J.D., (1991). Analysis and modeling of hydrologic time series, Chapter 19 in "Handbook of Hydrology", McGraw-Hill, New York.
- Schwarz, G, (1978), Estimating the dimension of a model, *Annals of Statistics* 6, 461-464.
- Senarath, S. U., (2013). The Significance of the Record Length in Flood Frequency Analysis. *American Geophysical Union, Fall Meeting 2013*.
- Sing, K. P. and Sinclair, R. A (1972): Two-distribution method for flood frequency analysis, *J. Hydraul. Div.*, 98, 29–44.
- Snedecor, G.W. and Cochran, W.G. (1980). *Statistical Methods*. Ames Iowa: The Iowa State University Press.

- Smith R.L., (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Stat Sci* 4:367–377.
- Smith Q. R and Campuzano O. C., (2000). Análisis exploratorio para la detección de cambios y tendencias en series hidrológicas. In: XIV Seminario de Hidráulica e Hidrología, 7-9 de septiembre, Villa de Leyva, Boyacá.
- Sveinsson, O. G. B. , Salas, J. D. , Boes, D. C. , and Pielke, R. A. (2003). “Modeling the dynamics of long term variability of hydroclimatic processes.” *J. Hydrometeorol.* , 4 (3) , 489–505.
- Sveinsson, O. G. B., Salas, J. D., and Boes, D. C. (2005). Prediction of extreme events in hydrologic processes that exhibit abrupt shifting patterns. *J. Hydrol. Eng.*, 10.1061/(ASCE)1084-0699(2005)10:4(315), 315–326.
- Stedinger JR (1983). Estimating a regional flood frequency distribution. *Water Resour Res*;19:503–10.
- Stewart, E.J., Reed, D.W., Faulkner, D.S., Reynard, N.S., (1999). The FORGEX method of rainfall growth estimation I: Review of requirement. *Hydrology and Earth System Sciences* 3 (2), 187– 195.
- Stedinger J. R., Vogel R. M., and Efi F., (1991). Frequency analysis of extreme events, Chapter 18 in “Handbook of Hydrology”, McGraw-Hill, New York.
- Strupczewski, W.G., Singh, V.P., Feluch, W., (2001). Non-stationary approach to at site flood frequency modelling. I. Maximum likelihood estimation. *Journal of Hydrology* 248, 123–142.
- Taylor, A. B., & Schwarz, H. E. (1952). United hydrograph lag and peak flow related to basin characteristics. *Trans., Am. Geophys. Union*, 33(2), 235-246.
- Thom, Herbert C. S (1958)., “A Note on the Gamma Distribution,” *Monthly Weather Review*, Vol. 86, No. 4, PP. 117-132.
- Todorov, V. Filzmoser, P, (2009). An Object-Oriented Framework for Robust Multivariate Analysis. *Journal of Statistical Software*, 32(3), 1-47.
- Unidad de Planeación Minero Energética.- UPME (2011). Plan Preliminar de Expansión de Referencia Generación – Transmisión 2011-2025.
- Velasco R. A., y Granados O. M., (2006). Tendencias e incidencia de los fenómenos macroclimáticos en la hidrología del alto y medio magdalena. *Ciencia e ingeniería neogranadina*, Vol. 16. No. 2, pp. 29-44.
- Victorov, P. (1971). “Effect of period of record on flood prediction.” *J. Hydr. Div.*, 97(11), 1853–1964.
- Villarini, G., F. Serinaldi, J. A. Smith, and W. F. Krajewski, 2009: On the stationarity of annual flood peaks in the continental United States during the 20th century. *Water Resour. Res.*, 45, W08417, doi:10.1029/2008WR007645.
- Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., and Krajewski, W. F.(2009a): Flood frequency analysis for nonstationary annual peak records in an urban drainage basin, *Adv. Water Resour.*, 32, 1255–1266.
- Villarini, G., Smith, J.A., 2010. Flood peak distributions for the eastern United States. *Water Resources Research* 46 (W06504). doi:10.1029/2009WR008395.
- Vogel, R. M., Yaindl, C., and Walter, M.: Nonstationary: flood magnification and recurrence reduction factors in the United States, *J. Am. Water Resour. As.*, 47, 464–474, 2011.
- Wallis, J. R (1980)., Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures, in *Piene e Siccit&* edited by E. Guggino, G. Rossi, and E. Todini, pp. 3-36, Fondazione Politecnica del Mediter., Catania, Italy.
- Wang, Y. (2000), Development of methods for regional flood estimates in the province of British Columbia, Canada, Ph.D. Thesis, The University of British Columbia.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236–244.

Waylen, P.R, Caviedes, C.N. (1986). El Nino and floods along the north Peruvian littoral. *Journal of Hydrology* 89: 141–156.

Yue, S., P. Pilon, and G. Cavadias (2002), Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series, *J. Hydrol.*, 259, 254– 271.

Yue, S., P. Pilon, and B. Phinney (2003), Canadian streamflow trend detection: Impacts of serial and cross-correlation, *Hydrol. Sci. J.*, 48(1), 51–63.

Zhang, L. (2005). Multivariate hydrological frequency analysis and risk mapping, Ph.D. thesis, Louisiana State University, Baton Rouge, Louisiana (USA).

Zrinji, Z., Burn, D.H., (1996). Regional flood frequency with hierarchical region of influence. *J. Water Resour. Plann. Mgmt* 122 (4), 245-252.