



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Ciencias (Hidrología Ambiental)

Modelos regionales de transferencia de información hidrológica para el diseño de
minicentrales hidroeléctricas

Opción de titulación
Tesis

Que como parte de los requisitos para obtener el Grado de
Maestría en Ciencias (Hidrología Ambiental)

Presenta:

Ing. Rosa Celic Reséndiz Mendoza

Dirigido por:

Dr. Martín Alfonso Gutiérrez López

Dr. Martín Alfonso Gutiérrez López
Presidente


Firma

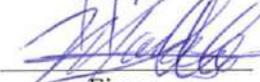
M.C. Felipe Ortiz Arredondo
Secretario


Firma

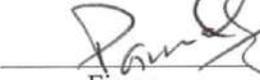
M.I. Pablo Talamantes Contreras
Vocal

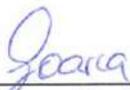

Firma

M.C. Marcela Susana Duhne Ramírez
Suplente


Firma

M.I. Cristóbal Bernabé Pascual
Suplente


Firma

Dr. Manuel Toledano Ayala
Director de la Facultad

Dra. Ma. Guadalupe Flavia Loarca Piña
Director de Investigación y Posgrado

RESUMEN

La implementación de obras hidráulicas en sitios no aforados es un problema común al realizar la estimación del caudal asociado a un evento de diseño, esto repercute en subdimensionar o sobredimensionar la obra en cuestión. Esta investigación tuvo dos enfoques, teórico y práctico. Respecto al primero, fue la elaboración de 6 capítulos relacionados a los pasos fundamentales en el desarrollo de un Análisis Regional de Frecuencias de Crecientes, donde se incluyeron ejemplos de aplicación para mayor difusión y comprensión del tema. El segundo enfoque fue la aplicación de la metodología propuesta a un sitio no aforado con el objetivo de determinar el potencial hidroeléctrico como parte del diseño de una minicentral hidroeléctrica. En la Región Hidrológica 10 (Sinaloa), se agruparon las estaciones aforadas en 4 zonas homogéneas mediante el método de K-medias. La verificación de regiones se realizó mediante el cociente estadístico regional. Se seleccionó la zona 4 para la transferencia de información hidrológica. En la selección de los descriptores más importantes de la cuenca se propuso una metodología de tres etapas mediante un análisis de componentes principales mejorando en un 26.40% la varianza de los descriptores empleados respecto a un proceso tradicional, con los mismos se obtuvieron las ecuaciones para obtener la avenida índice con un error relativo máximo de 0.067 respecto al caudal observado y un análisis de dispersión de la línea de tendencia de un 99.83%. La derivación de las curvas de crecimiento regional se creó a partir de la distribución de frecuencia regional log-normal de tres parámetros, la cual fue empleada para la estimación de los cuantiles regionales. Para el conocimiento de la permanencia del caudal a través del tiempo y para la proyección del comportamiento del mismo para la implementación de la obra hidráulica se estimó la ecuación de la Curva de Duración de Caudal regional, donde se creó una ecuación regional con un ajuste polinómico de grado 10. Los resultados obtenidos hacen viable la aplicación de esta metodología en sitios no aforados por medio de la evaluación de cada etapa del análisis, además con la información proporcionada en la fundamentación teórica, abre campo a futuras investigaciones en el análisis regional hidrológico en sitios no aforados.

(Palabras clave: cuencas no aforadas, potencial hidroeléctrico, agrupación de estaciones aforadas, análisis de componentes principales, avenida índice, caudal observado, curva de crecimiento regional, distribución de frecuencia regional, curva de duración de caudal)

SUMMARY

The design of hydraulic structures in ungauged sites is a common problem in the estimation of the flow associated with a design event, this has repercussions in undersizing or oversizing the hydraulic structure. The ungauged sites are those sites where there is no hydrometric station, so there is an absence of historical records of the variable in question. This research had two approaches, theoretical and practical. In regard to the first one, 6 chapters were written related to the fundamental steps in the development of a Regional Analysis of Flow Frequencies (ARFC), where application examples were included for greater diffusion and it is understood by readers of the present writing. The second approach was the application of the proposed methodology to an ungauged site with the objective of determining the hydroelectric potential as part of the design of a hydroelectric mini-power plant. At the Hydrological Region 10 (Sinaloa), hydrometric stations were grouped into 4 homogeneous zones using the K-means method. The verification of regions was carried out using the regional statistical ratio with a limit value of 0.3. Zone 4 was selected for the transference of hydrological information. In the selection of the most important descriptors of the basin, a three-step methodology was proposed by means of a principal components analysis (PCA), improving by 26.40% the variance of the descriptors used related to a traditional process, with those descriptors were obtained the equations to get the index avenue with a maximum relative error of 0.067 related to the flow observed, with a trend line dispersion analysis of 99.83%. The regional growth curves derivation was created from the regional frequency distribution, log-normal of three parameters (LN3), which was used for the estimation of the regional quantiles. For the knowledge of the permanence of the flow through time and for the projection of the behavior of the same for the hydraulic structure implementation, the equation of the Flow Duration Curve by region was estimated, where the polynomial adjustment of degree 10 obtained was of 99.99%. The results obtained make the application of this methodology feasible in ungauged sites by means of the evaluation of each analysis stage, in addition to the information provided in the theoretical foundation, it opens the field for future research in hydrological regional analysis in ungauged sites.

(Key words: ungauged basins, hydroelectric potential, grouped gauged basins, principal component analysis, index flood, observed flow, regional growth curves, regional frequency distribution, flow duration curve)

DEDICATORIA

A mi familia

AGRADECIMIENTOS

A Dios por darme la fuerza, la paciencia y el conocimiento para perseverar en mis metas.

A mi Familia, a mis papás, José Jaime y Camerina por apoyarme, por motivarme a siempre seguir adelante y por ser un ejemplo a seguir y a mis hermanos, Jaime Emmanuel, Lucía Mariana y Elissa Itzel por todos sus consejos y su apoyo.

Agradezco a mi hermana, la Ing. Mariana Reséndiz, a la M. en C. Marce Duhne y al Dr. Alejandro Ávalos, por todas las revisiones, observaciones y comentarios realizados a este trabajo que ayudaron a enriquecerlo más.

A la Comisión Federal de Electricidad, a la jefatura y delegación sindical del Centro de Anteproyectos del Bajío, al Ing. Germán Gurría, al Ing. Gaspar Parra y al Ing. Víctor Ledesma, por permitirme asistir a mis clases.

Al director de tesis (Dr. Alfonso Gutiérrez) así como a los sinodales (Maestros: Marce Duhne, Felipe Ortiz, Pablo Talamantes, y Cristóbal Pascual) porque me permitieron mostrarles mi trabajo de investigación así como todos sus comentarios en las evaluaciones durante este proceso.

TABLA DE CONTENIDOS

RESUMEN	II
SUMMARY	III
DEDICATORIA	IV
AGRADECIMIENTOS	V
TABLA DE CONTENIDOS	V
ÍNDICE DE TABLAS	XIII
ÍNDICE DE FIGURAS	XV
1. INTRODUCCIÓN	1
2. HIPÓTESIS	4
3. OBJETIVOS	4
3.1. OBJETIVO GENERAL	4
3.2. OBJETIVOS PARTICULARES	4
4. JUSTIFICACIÓN	5
5. FUNDAMENTACIÓN TEÓRICA	6
6. CARACTERIZACIÓN DE LA CUENCA	9
6.1. TIPOS DE VARIABLES	11
6.1.1. Variables geomorfológicas	11
6.1.2. Variables climatológicas	14
6.1.3. Variables hidrométricas	14
6.1.4. Estructuras hidráulicas	15
6.1.5. Propiedades hidrológicas del suelo	16
6.1.6. Uso de suelo	16
6.1.7. Variables físico-químicas y ambientales	17
6.1.8. Variables relacionadas a la geología	18
6.2. ECUACIONES EMPLEADAS EN LA DESCRIPCIÓN DE LA CUENCA	18
6.3. COMENTARIOS Y RECOMENDACIÓN DE VARIABLES	33

7.	ANÁLISIS PREVIO DE DATOS	39
7.1.	ESTADÍSTICA MULTIVARIADA DESCRIPTIVA	40
7.1.1.	<i>Estadísticos básicos de una muestra hidrológica</i>	<i>40</i>
7.1.2.	<i>Estadística descriptiva multivariada.....</i>	<i>42</i>
7.1.3.	<i>Estadística descriptiva gráfica:.....</i>	<i>44</i>
7.2.	REVISIÓN DE LA CONSISTENCIA DE LOS DATOS	44
7.3.	PRUEBAS DE CALIDAD ESTADÍSTICA A LOS DATOS	45
7.3.1.	<i>Etapas de un análisis exploratorio de datos</i>	<i>46</i>
7.3.2.	<i>Valores atípicos, outliers.....</i>	<i>46</i>
7.3.3.	<i>Pruebas de normalidad</i>	<i>47</i>
7.3.4.	<i>Pruebas de homogeneidad</i>	<i>47</i>
7.3.5.	<i>Pruebas de revisión de estacionariedad.....</i>	<i>48</i>
7.3.6.	<i>Métodos para detectar tendencia</i>	<i>49</i>
7.4.	CANTIDAD DE VARIABLES A EMPLEAR EN EL ANÁLISIS REGIONAL	49
7.5.	NORMALIZACIÓN DE MUESTRAS	53
7.6.	ANÁLISIS ESTADÍSTICO MULTIVARIADO.....	53
7.6.1.	<i>Marco teórico de un ACP.....</i>	<i>55</i>
7.7.	COMENTARIOS Y RECOMENDACIONES	58
8.	ANÁLISIS REGIONAL DE FRECUENCIAS ARF.....	60
8.1.	INTRODUCCIÓN A MOMENTOS-L	62
8.2.	ETAPAS DEL ANÁLISIS REGIONAL DE FRECUENCIAS BASADO EN MOMENTOS-L ...	64
8.2.1.	<i>Revisión de datos.....</i>	<i>65</i>
8.3.	FORMACIÓN DE REGIONES HOMOGÉNEAS	68
8.3.1.	<i>Medidas de heterogeneidad propuestas por Hosking y Wallis</i>	<i>70</i>
8.3.2.	<i>Otras medidas de homogeneidad</i>	<i>71</i>
8.4.	SELECCIÓN DE UNA DISTRIBUCIÓN REGIONAL DE FRECUENCIAS	71
8.4.1.	<i>Pruebas de bondad de ajuste.....</i>	<i>72</i>
8.5.	ESTIMACIÓN DE CUANTILES.....	74
8.6.	COMENTARIOS Y RECOMENDACIONES	75
9.	DELIMITACIÓN DE REGIONES HOMOGÉNEAS.....	76
9.1.	SELECCIÓN DE VARIABLES.....	83

9.1.1.	<i>Datos faltantes</i>	85
9.2.	ÍNDICE DE PROXIMIDAD – MEDIDAS DE SIMILITUD	86
9.2.1.	<i>Distancias</i>	87
9.3.	REGLAS DE AGREGACIÓN.....	88
9.3.1.	<i>Método de agrupamiento por Ward</i>	89
9.3.2.	<i>K-Medias</i>	97
9.4.	MÉTODOS GRÁFICOS DE DELIMITACIÓN DE REGIONES HOMOGÉNEAS	101
9.4.1.	<i>Método geográfico</i>	101
9.4.2.	<i>Rostros de Chernoff</i>	101
9.4.3.	<i>Curvas de Andrews</i>	102
9.4.4.	<i>Árbol de clasificación</i>	103
9.5.	MÉTODOS ESTADÍSTICOS DE DELIMITACIÓN DE REGIONES HOMOGÉNEAS	106
9.5.1.	<i>Métodos de agrupación por fecha de ocurrencia: similitud estacional</i>	106
9.5.2.	<i>Análisis de componentes principales (ACP)</i>	113
9.5.3.	<i>Análisis de Correspondencia (CA)</i>	113
9.5.4.	<i>Regiones de influencia</i>	114
9.5.5.	<i>Análisis de Correlación Canónica</i>	116
9.5.6.	<i>Agrupación por residuales</i>	117
9.6.	CANTIDAD ÓPTIMO DE LOS GRUPOS FORMADOS.....	118
9.6.1.	<i>Validez del agrupamiento</i>	120
9.7.	VERIFICACIÓN DE REGIONES HOMOGÉNEAS.....	120
9.7.1.	<i>Métodos gráficos: curvas de Andrews</i>	120
9.7.2.	<i>Medidas de discordancia</i>	121
9.7.3.	<i>Medidas de heterogeneidad con la técnica de Momentos-L</i>	121
9.7.4.	<i>Pruebas estadísticas: cociente estadístico regional</i>	121
9.7.5.	<i>Test de Wiltshire</i>	122
9.7.6.	<i>Test de Langbein</i>	124
9.8.	COMENTARIOS Y RECOMENDACIONES	124
10.	MÉTODOS REGIONALES	127
10.1.	MÉTODOS REGIONALES	133
10.1.1.	<i>Método estaciones-años</i>	133

10.1.2.	<i>Técnica de correlación y regresión múltiple</i>	136
10.1.3.	<i>Método regional de la Transformación de Box-Cox</i>	143
10.1.4.	<i>Avenida índice</i>	144
10.1.5.	<i>Método de la cuadrícula</i>	152
10.1.6.	<i>Método derivado del uso de funciones de distribución de frecuencia estándar</i> 154	
10.1.7.	<i>Análisis regional de máximos</i>	157
10.2.	VALIDACIÓN DEL MÉTODO REGIONAL	159
10.3.	COMENTARIOS Y RECOMENDACIONES	161
11.	FUNDAMENTOS GENERALES EN EL DISEÑO DE UN SISTEMA DE MINIGENERACIÓN	162
11.1.	DISEÑO DE UN SISTEMA	162
11.2.	MICROHIDROGENERACIÓN	162
11.3.	LOS COMPONENTES DE UN SISTEMA DE HIDROGENERACIÓN	164
11.4.	ENERGÍA A PARTIR DEL AGUA Y CÁLCULO DE POTENCIAL.....	165
11.5.	GENERACIÓN MEDIA ANUAL	167
11.6.	FACTOR DE PLANTA	167
11.7.	FACTORES DE CARGA Y DE PLANTA.	168
11.8.	COSTO UNITARIO DE LA ENERGÍA	168
11.9.	DECISIONES DE COSTO-BENEFICIO	168
11.10.	CAPACIDAD Y ESTUDIO DE LA DEMANDA.....	169
	<i>Informes de factibilidad</i>	170
	<i>Estudios generales en presas pequeñas (anteproyecto)</i>	170
11.11.	EVALUACIÓN DEL RECURSO HIDROENERGÉTICO.....	171
	<i>Análisis de la información hidrométrica</i>	171
11.12.	OBRAS CIVILES.....	172
	<i>Introducción</i>	172
	<i>Bocatomas</i>	174
	<i>Aliviaderos</i>	176
	<i>Desarenadores y cámara de carga</i>	176
	<i>Canales</i>	176

<i>Tuberías de presión</i>	177
<i>Turbinas hidráulicas</i>	177
11.13. TIPOS DE TURBINAS HIDRÁULICAS	179
<i>Selección de turbina</i>	181
<i>Instalación de turbinas</i>	182
<i>Curvas características</i>	183
11.14. COMENTARIOS	186
12. METODOLOGÍA.....	187
12.1. REVISIÓN DE LA LITERATURA	187
12.2. SELECCIÓN DE ESTACIONES HIDROMÉTRICAS	187
12.2.1. <i>Selección de estaciones de aforo</i>	187
12.2.2. <i>Selección de estaciones discordantes</i>	187
12.2.3. <i>Calidad de los registros</i>	187
12.3. DELIMITACIÓN DE CUENCAS Y SELECCIÓN DE CARACTERÍSTICAS DE LA CUENCA	188
12.3.1. <i>Delimitación de cuencas</i>	188
12.3.2. <i>Selección de variables descriptivas más representativas de la cuenca</i>	188
12.4. DELIMITACIÓN DE REGIONES HOMOGÉNEAS.	189
12.5. ANÁLISIS REGIONAL DE FRECUENCIAS DE CRECIENTES (ARFC)	189
12.5.1. <i>Selección de una distribución de frecuencias regional</i>	189
12.5.2. <i>Generación de ecuaciones de regresión</i>	190
12.5.3. <i>Derivación de curvas de crecimiento regional</i>	190
12.6. GENERACIÓN DE CURVAS DE DURACIÓN DE CAUDAL (CDC)	190
12.7. APLICACIÓN DE LA METODOLOGÍA PROPUESTA A UN SITIO DE INTERÉS	191
13. ZONA DE ESTUDIO	192
13.1. HIDROLOGÍA	192
13.2. CLIMA	193
13.3. PRECIPITACIÓN	193
13.4. CENTRALES HIDROELÉCTRICAS EXISTENTES EN LA RH10	195
13.5. ESTUDIOS RECIENTES	195
14. RESULTADOS Y DISCUSIÓN	197
14.1. REVISIÓN DE LA LITERATURA	197

14.2.	SELECCIÓN DE ESTACIONES HIDROMÉTRICAS APLICABLES EN EL ANÁLISIS REGIONAL	197
14.3.	DELIMITACIÓN DE CUENCAS Y SELECCIÓN DE CARACTERÍSTICAS DE LA CUENCA	198
14.4.	DELIMITACIÓN DE REGIONES HOMOGÉNEAS	206
14.5.	ANÁLISIS REGIONAL DE FRECUENCIAS DE CRECIENTES (ARFC)	209
14.5.1.	<i>Selección de una distribución de frecuencia regional</i>	209
14.5.2.	<i>Generación de ecuaciones de regresión</i>	211
14.6.	GENERACIÓN DE CURVAS DE DURACIÓN DE CAUDAL (CDC)	214
14.7.	SELECCIÓN DE UN SITIO PARA LA APLICACIÓN DE LA METODOLOGÍA REGIONAL..	216
14.7.1.	<i>Gasto medio anual</i>	217
14.7.2.	<i>Curva regional</i>	218
14.7.3.	<i>Curva de Duración de Caudal en la zona de interés</i>	220
15.	CONCLUSIONES	223
16.	REFERENCIAS	226
ANEXO I	247
	CARACTERÍSTICAS DE LA CUENCA	247
	<i>Geomorfología de la cuenca</i>	247
	<i>Variables climatológicas</i>	254
	<i>Variables hidrométricas</i>	260
	<i>Variables de ingeniería</i>	267
	<i>Propiedades hidrológicas del suelo</i>	268
	<i>Variables uso del suelo</i>	268
	<i>Variables ambientales y físicas</i>	269
	<i>Variables geológicas</i>	269
ANEXO II	270
	APLICACIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES	270
ANEXO III	288
	EJERCICIOS DRH.....	288
	<i>Método de Ward</i>	288
	<i>Método de Ward aplicado a datos de precipitación</i>	295

<i>Ejercicio de Método de K-Medias</i>	301
<i>Método del árbol de regresión</i>	308

ÍNDICE DE TABLAS

Tabla 8-1. Tabla de simbología empleada en Momentos y Momentos-L (Hosking y Wallis, 1997).....	61
Tabla 8-2. Valores críticos para las medidas de discordancia (Hosking y Wallis, 1997) ...	68
Tabla 8-3. Valores para medir el nivel de heterogeneidad (Hosking y Wallis, 1997)	71
Tabla 9-1. Valores críticos de χ^2 (Campos-Aranda, 2010a)	123
Tabla 10-1. Límites de la prueba de homogeneidad propuesta por Dalrymple.....	147
Tabla 10-2. Límites de la prueba de homogeneidad para dos poblaciones.	149
Tabla 11-1 Clasificación de las presas (CFE, 1980)	163
Tabla 11-2. Pérdidas y eficiencias típicas a carga parcial	167
Tabla 11-3. Características de las turbinas (González-García, 2014)	178
Tabla 11-4. Eficiencia del grupo de generación en porcentaje (Coz <i>et al.</i> , 1985)	181
Tabla 13-1 Presas con almacenamiento mayor a 1000 hm ³ (CONAGUA, 2012)	195
Tabla 14-1 Resultados de la prueba de Mann-Kendall para 24 estaciones.	197
Tabla 14-2 Características 24 estaciones.....	198
Tabla 14-3 Primer filtro de variables.....	201
Tabla 14-4 Valores propios para 16 variables.	202
Tabla 14-5 Correlación de los dos primeros componentes principales	204
Tabla 14-6 Valores propios para 32 variables.	204
Tabla 14-7 Resultados obtenidos por región homogénea.....	209
Tabla 14-8 Variables independientes y dependientes de cada estación.	212
Tabla 14-9 Ecuaciones de regresión para cada región homogénea.	212

Tabla 14-10 Comparación de caudal observado y estimado en cada región homogéneo.	213
Tabla 14-11 Valores de los coeficientes de cada ecuación de CDC regional	216
Tabla 14-12 Comparación de predicciones históricas de crecientes	218
Tabla 14-13 Estimación de cuantiles regionales	219
Tabla 14-14 Cálculo de potencia instalable y generación media anual en la RH4.	221
Tabla II-1. Datos estadísticos de 39 estaciones.....	275
Tabla II.2- Estadísticos descriptivos de las variables medidas.....	276
TablaII-3. Valores-propios y porcentaje de la varianza.....	278
Tabla III-1. Datos para ejemplo de Ward.....	293
Tabla III-2. Matriz de pérdida de inercia mínima inicial.....	295
Tabla III-3. Resumen de valores de inercia calculados.....	299
Tabla III-4. Variables empleadas en la delimitación de regiones homogéneas.....	300
Tabla III-5. Estaciones climatológicas de la Región Hidrológica 10 empleadas.....	301
Tabla III-6. Datos estadísticos de las estaciones climatológicas.....	303
Tabla III-7. Resultados de la clasificación por el método de Ward.....	305
Tabla III-8. Coeficientes Momentos-L Datos estadísticos.....	306
Tabla III-9. Coeficientes Momentos-L Datos estadísticos.....	312
Tabla III-10. Verificación de la calidad de partición óptima con otros elementos.....	312

ÍNDICE DE FIGURAS

Figura 7-1. Desviación estándar en una muestra (Chow <i>et al.</i> , 1994).....	41
Figura 7-2. Coeficiente de asimetría en una muestra con datos hidrológicos (Chow <i>et al.</i> , 1994).....	42
Figura 7-3 Esquema de tratamiento de datos en modelos de regresión (Ahn y Palmer, 2016).	53
Figura 8-1. Esquema de medidas de discordancia (Hosking y Wallis, 1997)	67
Figura 11-1. Componentes de un sistema de micro hidrogenación (Coz <i>et al.</i> , 1985). ..	165
Figura 11-2. Pérdidas típicas del sistema para una microcentral operando a plena carga (Coz <i>et al.</i> , 1995).	166
Figura 11-3. Componentes de una microcentral hidráulica (Coz <i>et al.</i> , 1995).....	173
Figura 11-4. Ubicación de la bocatoma y empleo de un barraje de derivación (Coz <i>et al.</i> , 1995).....	175
Figura 11-5. Ubicación de la bocatoma (Coz <i>et al.</i> , 1995).....	176
Figura 11-6. Componentes de una tubería de presión (Coz <i>et al.</i> , 1995).	177
Figura 11-7. Turbinas hidráulicas a) Francis, b) Kaplan, c) Kaplan doble álabe, d) Hélice, e) Kaplan (Dolores-Juárez, 1992).....	178
Figura 11-8. Turbina Turgo	180
Figura 11-9. Turbina Michell-Banki.	180
Figura 11-10. Turbina Francis.	180
Figura 11-11. Turbina Kaplan	181
Figura 11-12. Diagrama de selección de turbinas hidráulicas (Coz <i>et al.</i> , 1985).....	182
Figura 11-13 Diagrama de selección de turbinas (Nava-Mastache y García-Gutiérrez, 2013)	183

Figura 11-14. Curvas de eficiencia potencia 1. Kaplan, 2. Francis, 3. Pelton, 4. Hélice (Dolores-Juárez, 1992).	184
Figura 11-15. Característica universal de turbina Kaplan (Dolores-Juárez, 1992).	185
Figura 11-16. Características de operación de una turbina Kaplan de diámetro de 880 cm y velocidad de 62.5 r.p.m. (Dolores-Juárez, 1992).....	185
Figura 13-1 Mapa de ubicación de la Región Hidrológica 10, Sinaloa, México.	192
Figura 13-2 Mapa que muestra el clima en la Región Hidrológica 10, Sinaloa, México...	194
Figura 13-3 Precipitación Media Anual (PMA) Región Hidrológica 10, Sinaloa, México	194
Figura 13-4 Página de consulta INERE.....	195
Figura 14-1 Delimitación de cuencas en la RH10.....	200
Figura 14-2 Polígonos de Thiessen para el cálculo de la precipitación media anual.	201
Figura 14-3 Círculo de correlación entre los componentes 1 y 2. a) cuatro cuadrantes, b) cuadrante II, c) cuadrante III, d) cuadrante IV.....	203
Figura 14-4 Valores propios de la matriz de correlación de 32 variables.	205
Figura 14-5 Valores propios de la matriz de correlación de 16 variables	206
Figura 14-6 Gráfica de distancia de ligadura a través de los pasos de agrupamiento de elementos.....	207
Figura 14-7 Dendograma de Ward para 25 variables.....	208
Figura 14-8 Agrupación por el método de K-medias.....	208
Figura 14-9 Diagrama de cocientes de Momentos-L para la RH1.	209
Figura 14-10 Diagrama de cocientes de Momentos-L para la RH2.	210
Figura 14-11 Diagrama de cocientes de Momentos-L para la RH3.....	210
Figura 14-12 Diagrama de cocientes de Momentos-L para la RH4.....	211

Figura 14-13 Gráficas de dispersión del caudal observado y estimado en las cuatro regiones (a - d).....	214
Figura 14-14 Curvas de duración de caudal para la región homogénea 1	214
Figura 14-15 Curvas de duración de caudal para la región homogénea 2.....	215
Figura 14-16 Curva de duración de caudal para la región homogénea 3.	215
Figura 14-17 Curva de duración de caudal para la región homogénea 4	216
Figura 14-18 Ubicación del proyecto Río Verde.....	218
Figura 14-19 Curvas de crecimiento regional trazadas con la curva de crecimiento regional en la RH4.....	219
Figura 14-20 Curva del caudal esperado en la RH4.....	220
Figura 14-21 Caudal mensual específico en la región homogénea 4	220
Figura 14-22 CDC en la RH4.....	222

1. INTRODUCCIÓN

Cuando se planea un proyecto de diseño de manejo de recursos hídricos es importante el conocimiento de la ocurrencia de eventos de inundaciones (Gado y Nguyen, 2016), por lo que uno de los principales problemas en el diseño de obras hidráulicas es la recopilación de información hidrológica confiable, el problema se agrava cuando en el sitio de proyecto no hay estaciones de aforo o las estaciones están parcialmente aforadas, es decir, no hay continuidad en los registros históricos de las variables de interés que permitan realizar un análisis de frecuencias en el sitio de proyecto, normalmente esta situación se presenta en países en desarrollo y subdesarrollo (Saf, 2009) donde no hay suficiente cobertura de estaciones de aforo.

La estimación de la magnitud de diseño de la variable en cuestión llega a ser un reto para el profesionalista debido a que en el diseño de las obras hidráulicas es necesario contar con un periodo de registro histórico de datos (Isik *et al.*, 2008), y en muchas de las ocasiones la escasez en el monitoreo de registro repercute en la subestimación y en la sobreestimación de caudales de diseño.

Una opción viable a este problema es la transferencia de información por medio de la agrupación de regiones homogéneas, formadas por las estaciones aforadas y la transferencia de información hacia las zonas no aforadas, es decir, la aplicación de técnicas regionales de transferencia de información hidrológica.

La regionalización hidrológica es un conjunto de ecuaciones, que permiten estimar eventos hidrológicos en sitios con registros escasos o nulos, a sitios que permanecen en condiciones similares, es decir, dentro de una misma región homogénea la respuesta hidrológica es similar. Para comprender qué significa la regionalización hidrológica es necesario identificar todos aquellos procedimientos y técnicas que se ven involucrados en la regionalización y que contribuyen en la estimación precisa de la variable de interés, pero que de manera aislada no cumplen con el propósito de regionalizar un espacio. Las técnicas y procesos antes mencionados constan de calcular isolíneas de alguna variable hidrométrica en cartografía, así como emplear métodos como *kriging* o *splines* en interpolación espacial, realizar análisis múltiple de frecuencias con diversas distribuciones de probabilidad, ponderación espacial

con características fisiográficas de la región y estimación de envolventes de eventos hidrológicos en análisis de extremos.

Mencionado lo anterior, todo proceso de regionalización requiere de seis pasos integrales: 1) caracterizar la variabilidad (espacio-temporal) de las mediciones, 2) identificar las variables que describen el fenómeno a estimar, 3) Seleccionar un índice de proximidad, 4) identificar regiones hidrológicamente homogéneas, 5) construir ecuaciones regionales de transferencia de información, y 6) verificar la validez de las ecuaciones regionales.

La modelización hidrológica no es una metodología nueva, la implementación de métodos multivariados a datos hidro-meteorológicos comienza a finales de la década de 1940 (Baeriswyl y Rebetz, 1997), por lo que los modelos regionales se han empleado en la evaluación de los recursos hídricos, en estudios de proyección del impacto del cambio climático, así como en la evaluación de la respuesta hidrológica ante los agentes antropogénicos que intervienen en el cambio de las condiciones iniciales naturales de las cuencas (Peterson *et al.*, 2011).

En la estimación de eventos hidrológicos existen dos métodos que son aplicables, cuando se cuenta con información de aforos cerca la zona de interés se aplica un Análisis de Frecuencias del sitio (AF), en el caso de ausencia y escasez de registros la metodología a aplicar consiste en un Análisis Regional de Frecuencias (ARF), en este último es necesario estimar las características descriptivas de las cuencas (Mishra *et al.*, 2009). Además, es de gran utilidad emplear un ARFC en la estimación de avenidas cuando los cuantiles de inundación tienen una longitud de registro inferior a $5T$, donde T es el periodo de retorno de interés (Reed *et al.*, 1999; Rao y Srinivas, 2006).

En el procedimiento de la formación de regiones homogéneas no se ha desarrollado una metodología universal ya que existen distintos métodos aplicables, que pueden ser en regiones contiguas geográficamente, regiones no contiguas geográficamente y vecinos hidrológicos (Ouarda *et al.*, 2001; Ouarda *et al.*, 2006; Wazneh *et al.*, 2015).

En el análisis regional la selección de variables es un paso fundamental, sin embargo, en la literatura no existen muchos estudios que muestren a detalle la selección de las variables óptimas (Ilorme y Griffis, 2013).

En la delimitación de las regiones homogéneas es necesario dividir la gran región heterogénea en sub-regiones hidrológicas homogéneas (Dodangeh *et al.*, 2014). El objetivo del agrupamiento de las variables hidrológicas es el análisis de la variabilidad espacial de la similitud de características hidrológicas, como las propiedades geográficas, físicas y estadísticas de las observaciones (Sarhadi y Modarres, 2011) que permitan la agrupación de regiones con la misma respuesta hidrológica, por lo que la extrapolación de los registros para los efectos de cuantiles de diseño es más precisa (Nathan y Mc Mahon, 1990; Yu *et al.*, 2002).

Después de la delimitación de regiones homogéneas, es importante la ubicación de la estación no aforada dentro de alguna de las regiones formadas, para lo cual se recurre a un análisis de técnicas regionales, que mediante la formación de una ecuación en función de las características de la zona de estudio permita transferir información obteniendo de esta forma un caudal índice.

Mediante este trabajo de tesis se busca aportar en el área de investigación hidrológica una metodología que permita el cálculo del caudal en un punto de interés no aforado, es decir, de un sitio donde no se cuente con información hidrométrica y donde se planea la implementación de una minicentral hidroeléctrica, estimando el caudal mediante métodos regionales de transferencia de información hidrológica, por medio de una fundamentación teórica que respalde la elección de la técnica de regionalización empleada.

2. HIPÓTESIS

La transferencia de información hidrológica a sitios con registros escasos o nulos, puede realizarse empleando un análisis multivariado de las variables más representativas ya sean climatológicas, fisiográficas y ambientales entre cuencas hidrológicamente homogéneas, que superen el 80% de la varianza acumulada respecto a la matriz de datos originales.

3. OBJETIVOS

3.1. Objetivo General

Construir modelos regionales de transferencia de información hidrológica en sitios sin registros o con poca información por medio de ecuaciones regionales de transferencia de información hidrológica, midiendo el error de cada ecuación con un caudal que ha sido medido y observado, con un error relativo regional menor a un 10%.

3.2. Objetivos Particulares

- Recopilar y dar recomendaciones sobre las características fisiográficas a utilizar en un análisis regional en el país.
- Recopilar y analizar los diversos tipos de métodos de delimitación de regiones hidrológicamente homogéneas, empleados para la transferencia de lluvias y caudales, así como proporcionar algunos ejemplos de aplicación.
- Recopilar y analizar los diversos tipos de métodos regionales empleados para la transferencia de datos de lluvias y caudales.
- Analizar la metodología referente a un Análisis Regional de Frecuencias de Crecientes (ARFC).
- Recopilar y analizar los tratamientos previos aplicados en un ARFC.
- Analizar las variables más representativas de una región hidrológica, mediante un Análisis Multivariado.

4. JUSTIFICACIÓN

Este tema de investigación surge debido a la relevancia en la generación de energías limpias que ha cobrado interés a través de los años en México (CMM, 2011; Valle-Pereña y Ortega-Navarro, 2012; IMTA, 2017) así como en la generación de investigación científica y tecnológica en la medición del agua (CONAGUA, 2014).

Actualmente la mayor parte de la energía renovable producida en el país es la hidroeléctrica que aporta el 48% del total de energía renovable (IMTA, 2017). Impulsando el desarrollo de esta energía limpia, las Prospectivas de Energías Renovables 2012-2026 plantean un potencial importante en la implementación de centrales de pequeña escala en las que la generación no exceda los 30 MW (Valle-Pereña y Ortega-Navarro, 2012).

Debido a lo anterior esta investigación se enfocó en la generación de energía hidroeléctrica en centrales de pequeña escala en sitios no aforados, por lo que este trabajo de investigación corresponde a un contenido de interés actual en materia del sector hídrico con la aplicación de una regionalización hidrológica debido a escasez de registros,

Actualmente el BANDAS cuenta con datos de 2070 estaciones hidrométricas, aunque existe una gran cobertura de estaciones en la República Mexicana, la necesidad de la implementación de obras hidráulicas en sitios sin aforos es frecuente, lo que implica el empleo de métodos de regionalización hidrológica.

La modelización hidrológica es ampliamente utilizada para la predicción de avenidas, la evaluación de los recursos hídricos, la proyección del impacto del cambio climático, así como en la evaluación de la respuesta hidrológica ante los agentes antropogénicos que intervienen en el cambio de las condiciones iniciales de las cuencas (Peterson *et al.*, 2011).

La metodología desarrollada en este trabajo se aplicó en la Región Hidrológica 10 - Sinaloa, se eligió en primera instancia por la importancia en el tema de regionalización hidrológica siendo uno de los tres sitios con mayor información de estudios en la República Mexicana, por lo que la metodología aplicada en esta investigación permite la comparación de resultados obtenidos en análisis previos en algunas etapas del desarrollo de este trabajo.

5. FUNDAMENTACIÓN TEÓRICA

La fundamentación teórica de este tema de investigación se desarrolló en los siguientes seis capítulos, esto con el fin de proporcionar al lector de este trabajo una serie de capítulos que incluyen el estado del arte bajo el enfoque de la regionalización hidrológica. Un objetivo a desarrollar en esta investigación fue el contar con información bibliográfica y marco teórico de los principales temas que involucran la transferencia de información hidrológica de los sitios aforados al sitio no aforado. Por lo que el orden de presentación de los capítulos está en función de las consideraciones que implica un análisis regional, en algunos capítulos se incluyen ejemplos de aplicación, los cuales se encuentran en el apartado de ANEXO II y ANEXO III.

En el capítulo 6 CARACTERIZACIÓN DE LA CUENCA se introduce a la metodología de las técnicas de regionalización y se expone la importancia de la selección de las variables a emplear en estudios regionales. La cantidad de las variables expuestas en este documento se debe únicamente a aquellas empleadas en las técnicas regionales, clasificando a las variables en ocho tipos: geomorfológicas, climatológicas, hidrométricas, estructuras hidráulicas, propiedades hidrológicas del suelo, uso de suelo, físico-químicas y ambientales, y geológicas. Además, se incluyen recomendaciones en la selección de las variables dependiendo del objeto de estudio, así como del periodo de registro a emplear, y del tratamiento previo a los datos. Se incluyó un ANEXO I para consulta de la descripción de cada tipo de variable, así como de las ecuaciones para obtener las características geomorfológicas.

En el capítulo 7 ANÁLISIS PREVIO DE DATOS se describe como el analista debe recopilar información descriptiva debida a la zona de interés y estadística debida a la hidrometría. La primera se logra mediante los datos de fisiografía y climatología, mientras que para la segunda se toman como fuente de información los registros históricos hidrométricos de los aforos, este capítulo se enfocó en el tratamiento aplicado a datos hidrométricos.

El análisis previo de datos busca brindar al analista un primer conocimiento del comportamiento de las variables, así como la identificación de valores atípicos en las series de registros, ya sea mediante métodos gráficos o métodos estadísticos, para un posterior análisis de consistencia y de calidad de datos, que permitan la identificación de tendencias en las series. Además que este análisis previo permite el conocimiento de las variables más

importantes mediante la normalización de datos y la aplicación de un análisis de componentes principales (ACP).

En el capítulo 8 ANÁLISIS REGIONAL DE FRECUENCIAS ARF se describe cómo se desarrolla el ARF de crecientes mediante la metodología de Momentos L, esta metodología fue planteada por Hosking y Wallis (1997) la cual se ha difundido en los casos de ausencia en cantidad de registros confiables de la zona de estudio.

El capítulo incluye la introducción de la metodología Momentos L, después se describen las etapas del análisis regional basado en L-Momentos, formación de regiones homogéneas, verificación de la homogeneidad, selección de una distribución regional de frecuencias, pruebas de bondad de ajuste y estimación de cuantiles.

En el capítulo 9 DELIMITACIÓN DE REGIONES HOMOGÉNEAS se muestran los distintos métodos para la delimitación de regiones homogéneas, este trabajo inicia con una introducción de la selección de variables necesarias para la delimitación de regiones, además se incluyó una descripción de los índices de proximidad como medida de similitud, así como de las reglas de agregación para la formación de grupos.

En la descripción de los distintos métodos de delimitación se realizó una breve reseña de los distintos estudios que se han desarrollado. Por último, se incluyen algunos ejemplos de agrupación de elementos homogéneos, que permiten al lector conocer ejemplos de aplicación (consultar el ANEXO II).

En el capítulo 10 MÉTODOS REGIONALES se presentan los distintos métodos de transferencia de información regional de un sitio aforado a un sitio no aforado. Se exponen las primeras técnicas que iniciaron con el concepto de regionalización para continuar con los métodos pioneros en el análisis regional, como el método racional, seguido del método de la avenida índice, método de regresión múltiple, entre otros. Para finalizar se detallan las técnicas de validación de transferencia de información hidrológica.

En el capítulo 11 FUNDAMENTOS GENERALES EN EL DISEÑO DE UN SISTEMA DE MINIGENERACIÓN se muestra un extracto de los principales manuales de presas pequeñas y de mini y micro centrales hidráulicas, estos manuales están dirigidos a especialistas y profesionistas encargados del desarrollo de generación de energía. Por lo que en este

documento se muestran los principales lineamientos y metodología básica para la realización y evaluación de proyectos en micro centrales hidroeléctricas.

Para mayor información referirse a los manuales adquisición de datos (Camargo-Hernández y Franco, 1998), en hidrotecnia en escurrimiento dirigirse al manual de diseño de obras civiles (CFE, 1981), manual de presas pequeñas (Marsal, 1974) y dimensionamiento de una presa derivadora (Sánchez-Bribiesca y Fuentes-Mariles, 1997).

6. CARACTERIZACIÓN DE LA CUENCA

El primer aspecto importante en la metodología regional es el análisis de las variables descriptivas de la zona de estudio disponibles, para ello en el análisis regional se emplean dos tipos de datos, características de la zona de estudio, o los valores estadísticos de las series de tiempo de las variables hidro-meteorológicas (D.F. Campos-Aranda, 2014a). Además, es necesario contar con variables no correlacionadas entre sí que repercutan en la precisión del modelo, para ello son aplicables las técnicas multivariadas por medio de un análisis de componentes principales.

En un proceso de regionalización, primero se asume que se tiene una región heterogénea, por lo que el primer paso es la formación de grupos hidrológicamente homogéneos mediante las variables descriptivas de cada cuenca aforada, el siguiente paso en la regionalización consiste en estimar la incertidumbre de los parámetros del modelo y la estimación de cuantiles (Dodangeh *et al.*, 2014).

La formación de regiones homogéneas estima las zonas que comparten la misma respuesta hidrológica en función de las características descriptivas y climatológicas, de forma que se reservan los estadísticos de las series de datos para verificar las condiciones de medidas de heterogeneidad.

Actualmente existen diversas técnicas de regionalización por lo que es factible la transferencia de información, sin embargo, la aplicación de diversas técnicas de regionalización ha llevado a obtener distintos resultados sobre cuál es el enfoque óptimo, debido a ello Heřmanovský y Pech (2013) identificaron cuales son los principales factores que repercuten en la obtención de resultados distintos, siendo que la primera diferencia se debe a la cantidad y tipo de datos, la segunda diferencia en el modelo hidrológico empleado y la tercera causa está relacionada con el hecho de que cada enfoque de regionalización emplea un procedimiento diferente (Gubareva, 2012; Heřmanovský y Pech, 2013).

Este capítulo está enfocado al primer factor que repercute en el análisis regional (Heřmanovský y Pech, 2013), es decir, en la cantidad y tipos de datos empleados. Autores como Nathan y McMahon (1990) y Caratti *et al.*, (2005) proponen considerar al inicio de la investigación qué variables están relacionadas con el fenómeno a estudiar, si lo anterior no se toma en cuenta es posible que se generen grupos homogéneos con poco significado físico.

Al encontrar una relación entre las variables descriptivas de la cuenca, es posible la transferencia de información a un sitio no aforado. Por lo que la importancia en el conocimiento de cuántos descriptores de la cuenca emplear es una pregunta frecuente al iniciar un análisis de modelos regionales. Debido a que no se han establecido teóricamente la cantidad de descriptores de una cuenca que deben ser empleados en un análisis regional (Heřmanovský y Pech 2013). Existen propuestas de la cantidad de variables a emplear algunos de ellos son mediante un análisis de correlación y análisis multivariado para descartar la multicolinealidad de las variables (Malekinezhad *et al.*, 2011).

Los atributos empleados para la delimitación de regiones homogéneas bajo el enfoque del ARFC se clasifican en cinco tipos (Rao y Srinivas, 2006): i) características fisiográficas de la cuenca, ii) ubicación geográfica, iii) medidas de tiempo en la respuesta hidrológica, como el tiempo pico, tiempo de retraso, entre otros, iv) variables meteorológicas y, v) estadísticos descriptivos en el sitio; aunque este último punto está bajo la observación al ser no recomendado por el hecho de emplear los estadísticos exclusivamente para la revisión de la homogeneidad de las regiones formadas (Hosking y Wallis, 1997; Núñez *et al.*, 2016).

Existen diversos estudios que se han realizado con el objetivo de formar grupos homogéneos y de esta forma comprender el comportamiento de la distribución espacial de las variables descriptivas de las cuencas. Se han desarrollado estudios con el objetivo de analizar el impacto hidrológico generado por la implementación de presas (Matteau *et al.*, 2009), estudio de predicción de caudales mínimos para la planeación de recursos hídricos y su manejo (Nathan y McMahon, 1990; Yu *et al.*, 2002), análisis de sequías (Dodangeh *et al.*, 2014), monitoreo de variables con fines ecológicos (Kahya *et al.*, 2008), análisis de calidad de agua (Sanborn y Bledsoe, 2006), entre otros.

En cada uno de los diversos estudios se emplearon distintos tipos de variables, por lo que en este capítulo se desarrolló una clasificación de las mismas además se incluyen las ecuaciones que describen a las variables (consultar el ANEXO I) esta actividad se desarrolló por medio de la revisión del estado del arte de los diversos estudios publicados.

Se clasificaron las variables que intervienen en el análisis regional de la siguiente forma:

- Geomorfológicas: La clasificación de variables geomorfológicas, la define Spring (2006) como el conjunto de variables que reflejan el comportamiento hidrológico y que pueden ser empleadas para evaluar la respuesta hidrológica en la cuenca, estas variables están relacionadas con las características físicas y características de los canales de drenaje en la cuenca. Algunos ejemplos de estas variables son área, perímetro, longitud del cauce principal, pendiente del cauce principal, entre otros.
- Climatológicas: Referentes a todas las variables que están relacionadas con las condiciones atmosféricas de un clima. Ejemplos de estas variables son precipitación, nieve, temperatura, humedad relativa y radiación.
- Hidrométricas: Variables asociadas al caudal natural o regulado
- Estructuras hidráulicas: Geométricas de infraestructura hidráulica, dimensiones de una obra hidráulica.
- Propiedades hidrológicas del suelo: Variables relacionadas con el tipo de suelo
- Uso de suelo: Cobertura porcentual del tipo de uso de suelo en la cuenca.
- Físico-químicas: Todas aquellas variables relacionadas con la calidad y contaminación del agua, es decir, variables relacionadas con la concentración de elementos en el agua, sólidos suspendidos totales, grado de acidez o basicidad en un medio acuoso, conductividad eléctrica, oxígeno disuelto, Demanda Biológica de Oxígeno (DBO), Demanda Química de Oxígeno (DQO), concentración de sedimentos y temperatura del agua.
- Geológicas: Clasificación del tipo de rocas subyacente al área de drenaje.

6.1. Tipos de variables

6.1.1. Variables geomorfológicas

En cuanto a la morfometría de la cuenca las variables empleadas en el análisis regional como el método de la avenida índice, análisis regional de crecientes y en la delimitación de regiones homogéneas por métodos estadísticos multivariados se encuentran las siguientes: área de la cuenca, perímetro de la cuenca, relación de compacidad, relación de elongación, longitud de la cuenca, ancho de la cuenca, tiempo de concentración, orientación y aspecto de la cuenca, en esta última característica se incluyen las variables específicas de aspecto cara al norte, aspecto cara al este y aspecto promedio.

El área de drenaje, es una de las variables más empleadas, Sanborn y Bledsoe (2006) relacionan el área de la cuenca con la magnitud del caudal y como variable más importante

en estudios de avenida índice, por lo que es identificada como la variable de mayor importancia en la delimitación de regiones homogéneas (Kult *et al.*, 2014; Nobert *et al.*, 2014).

El perímetro de la cuenca, se ha empleado en estudios que se relacionan con análisis regional de crecientes, análisis de regresión múltiple, análisis de regresión difusa, análisis de caudales mínimos, y análisis de homogeneidad de regiones con caudal medio.

La relación de compacidad, se ha empleado en el análisis de caudales medios. La relación de elongación ha sido una variable empleada en la regionalización de caudales mínimos y conservación de caudal ecológico, además de ser empleada en el método de la avenida índice.

El parámetro de forma, se ha empleado en estudios que involucran la delimitación de regiones homogéneas, así como en la estimación del método de la avenida índice.

El ancho medio de la cuenca, ha sido una variable poco empleada en los análisis regionales, se cuenta con el registro del empleo de ésta, en un estudio realizado en México con el objetivo de aplicar métodos estadísticos multivariados para predicción de caudales máximos en la Mixteca Oaxaqueña (Álvarez-Olguín *et al.*, 2011).

Se han desarrollado diversos estudios que están relacionados con los cauces en las cuencas aforadas para estimar el caudal máximo, estas variables son las siguientes: la longitud de la corriente principal, la longitud promedio de las corrientes de los primeros tres órdenes, longitud total de la red de drenaje en la cuenca, ancho máximo, mínimo y medio de las corrientes, clasificación del orden de corrientes, cantidad porcentual del orden de corrientes, frecuencia, densidad, orientación, elevación máxima y mínima de la corriente principal, pendiente promedio.

En cuanto a la clasificación del orden de corrientes en una cuenca, la clasificación por Strahler es empleada por los investigadores, inclusive, algunos autores como Singh *et al.* (2014) han seleccionado la cantidad porcentual presente en los primeros cuatro órdenes como variable necesaria en un análisis de transferencia de información de caudales diarios.

Las variables que están relacionadas con la topografía de la cuenca son las siguientes siendo las variables más comúnmente empleadas en los estudios de homogeneidad regional en

caudales máximos y mínimos son la longitud y altitud de la cuenca, mientras que otras variables son menos empleadas, como el desnivel de la cuenca (Álvarez-Olguín *et al.*, 2011), el relieve relativo de la cuenca y el relieve de la cuenca dividido por la longitud de la cuenca (Sanborn y Bledsoe, 2006), la profundidad promedio de la superficie a la roca madre (Sanborn y Bledsoe, 2006; Singh *et al.*, 2014) son algunos de ellos.

En cuanto a las propiedades hidrológicas y con objeto del estudio de caudal mínimo y caudal diario las variables más empleadas son la humedad (Acreman y Sinclair, 1986), tipo de suelo (Acreman y Sinclair, 1986), permeabilidad, infiltración, índice de flujo base y capacidad de agua disponible (Kult *et al.*, 2014).

Las variables relacionadas en estudios de caudal ecológico y caudal mínimo son: pendiente media ponderada de la cuenca, pendiente media de la cuenca, cantidad porcentual del tipo de pendiente moderada y pronunciada, así como el índice topográfico de humedad (Sanborn y Bledsoe, 2006; Kult *et al.*, 2014), y la relación de la pendiente de la cuenca con la elongación de la misma (Sanborn y Bledsoe, 2006).

Para la estimación de caudales medios, se encuentran otras variables fisiográficas propuestas y que sin embargo, han sido poco empleadas y muy específicas en determinados estudios, un ejemplo de ello, fue la investigación realizada por Paris y Zucarelli (2004) por medio del índice de Glushkov para determinar el lecho del río, con el objetivo de establecer puntos de control para el monitoreo de caudales medios.

En cuanto a variables más representativas en zonas montañosas y mesetas orientales boscosas Singh *et al.* (2014) encontraron la variable de geología seguida de drenaje, topografía y clima, como las mediciones de similitud dominante en las cuencas, mientras que para llanuras húmedas los autores encontraron la variable de uso del suelo se relaciona más fuertemente con la transferencia exitosa de parámetros en esas zonas.

Matteau *et al.* (2009) encontraron para las presas en Québec que las variables hidrológicas que son afectadas por la implementación de presas son las corrientes y caudales en verano e invierno, así como el tiempo de registro del caudal medio mensual máximo y la variabilidad estacional de las corrientes, corrientes mensuales y el promedio mínimo mensual de corrientes. En cuanto a los factores que afectan la magnitud del cambio hidrológico son las

siguientes: el régimen regulado y en menor magnitud el tamaño de la cuenca. La innovación de ese estudio radica en el análisis multivariado de la clasificación de regímenes de caudal en ríos regulados, siendo los pioneros en proponer una clasificación de los ríos regulados de acuerdo con los impactos producidos por las presas.

6.1.2. Variables climatológicas

Las variables climatológicas que se emplean en los modelos de análisis regional se pueden clasificar en precipitación, nieve, temperatura, humedad relativa y radiación.

La variable de precipitación media anual se emplea en estudios del proceso lluvia-escurrimiento, para la regionalización en caudales mínimos (Mamun *et al.*, 2010; Nathan y McMahon, 1990), en el análisis de caudales medios y desarrollo del método de la avenida índice (Acreman y Werritty, 1987), en la delimitación de regiones hidrológicas (Ouarda *et al.*, 2008; Álvarez-Olguín *et al.*, 2011; Malekinezhad *et al.*, 2011; Sarhadi y Modarres, 2011; Gubareva, 2012; Arellano-Lara y Escalante-Sandoval, 2014; Singh *et al.*, 2014) en el estudio de la distribución espacial y temporal de la lluvia (Baeriswyl y Rebetez, 1997), en la evaluación de falla en estructuras (Requena *et al.*, 2016) y en estudios de sequías (Núñez *et al.*, 2016).

Existen otras variables derivadas de la precipitación que también se han empleado como descriptoras siendo las siguientes: el valor de la precipitación asociada a un periodo de retorno, intensidad anual, los estadísticos relacionados a la lluvia anual, como la varianza, desviación estándar, promedio máximo anual, coeficiente de variación, cantidad de días de lluvia por año, desviación estándar de la cantidad de días de lluvia por año, y coeficiente de variación de la cantidad de días de lluvia por año.

6.1.3. Variables hidrométricas

La obtención de las variables como el caudal mínimo es importante debido a que la estimación de los estadísticos relacionados a ellos es de gran importancia en el manejo y administración de los recursos hídricos, así como en las políticas de operación de las centrales hidroeléctricas y en presas de derivación (Dodangeh *et al.*, 2014).

Dodangeh *et al.* (2014) establecen que para el cálculo de los caudales mínimos, se debe estudiar las curvas de duración de caudal, las principales características son aquellas curvas que exceden el 95% de la descarga en el tiempo, y la variable que describe la descarga mínima

de 7 días con un intervalo de recurrencia de 10 años. Además los autores Yu *et al.* (2002) proponen el empleo de variables hidrogeológicas debido a que en su investigación los autores recopilaron información sobre estas variables que se ven influenciadas por los caudales mínimos, además de las variables hidrogeológicas identificaron las siguientes variables que intervienen dentro del concepto “caudal mínimo” encontrando para ello diversos índices que ayudan a caracterizar el caudal, siendo los siguientes: (1) tiempo con caudal cero, (2) caudal base, (3) flujos mínimos anuales de 1 día, 10 días y 30 días, (4) la media de los déficits máximos anuales observados durante eventos consecutivos de caudal bajo por debajo de alguna descarga de referencia o umbral de referencia, (5) duración media anual máximas de los acontecimientos durante los cuales los caudales caen por debajo de un umbral de caudal y (6) cuantiles de flujo extraídos de una curva de duración de flujo de flujo (FDC) que se superan con cierta probabilidad.

Siempre que la intención del investigador sea trabajar con caudales mínimos, se deberá estimar el índice de caudal base debido a que esta variable es clave en la comprensión del proceso hidrológico del escurrimiento directo (Lee *et al.*, 2005).

En la predicción de caudales medianos, Paris y Zucarelli (2004) identificaron regiones homogéneas al noroeste de Argentina con el objetivo de ubicar puestos de aforo que permitieran obtener estimaciones representativas del caudal mediano, definido como un caudal asociado con el 50% de probabilidad de ocurrencia. La característica del grupo de régimen de ríos la consideraron necesaria, debido a que ésta provee una clasificación objetiva de la variación de las corrientes dentro de un año, que, en conjunto con la variable del caudal medio anual, permite la estimación del caudal medio mensual.

En la predicción de caudales máximos, los datos hidrométricos de máximos anuales además de los cocientes de caudales estacionales son empleados.

6.1.4. Estructuras hidráulicas

Las variables empleadas en las técnicas regionales son principalmente la capacidad del embalse a analizar, la capacidad almacenada total de todos los embalses aguas arriba del embalse a analizar y el grado de almacenamiento (Matteau *et al.*, 2009).

En estudios referentes al impacto por acciones antrópicas (Matteau *et al.*, 2009) debido a la construcción de obras hidráulicas, se realizó un estudio en Québec donde los autores retomaron la importancia del estudio de la acción de restauración de ríos a razón del daño provocado por acciones antrópicas. El estudio se realizó bajo dos objetivos principales: el primero analizar las variables que son modificadas por presas y la identificación de los factores que influyen en la magnitud de los cambios hidrológicos. El análisis se llevó a cabo mediante dos técnicas, un ACP y un ACC (Análisis de Componentes Principales y Análisis de Correlación Canónica, respectivamente). Mediante ACP se obtuvieron las variables modificadas por presas siendo los siguientes: corrientes y caudales en verano e invierno, tiempo de registro del caudal medio mensual máximo y la variabilidad estacional de las corrientes; corrientes mensuales y el promedio mínimo mensual de corrientes. Para la identificación de factores que afectan en el cambio hidrológico se empleó la técnica de ACC siendo el régimen regulado y el tamaño de la cuenca, siendo un menor efecto este último.

6.1.5. Propiedades hidrológicas del suelo

Las variables de las propiedades hidrológicas del suelo se han empleado en estudios relacionados al caudal ecológico (Sanborn y Bledsoe, 2006), en la estimación de caudales mensuales (Kult *et al.*, 2014), y en el proceso lluvia-escurrimiento (BRATH *et al.*, 2001).

Las variables comúnmente empleadas en estos estudios son la permeabilidad del suelo (Kult *et al.*, 2014; Singh *et al.*, 2014; Ahn y Palmer, 2016; Singh *et al.*, 2016), tasa media de la infiltración de la cuenca (Requena *et al.*, 2016) y el índice topográfico de la humedad (Sanborn y Bledsoe, 2006; Kult *et al.*, 2014; Singh *et al.*, 2014; Saharia *et al.*, 2017).

6.1.6. Uso de suelo

Las variables descriptivas de la cubierta de suelo son diversas y dependen de las condiciones climáticas particulares de cada zona de estudio. Los descriptores de suelo empleados por diversos autores son los siguientes: desde las cantidades porcentuales por tipo de suelo, así como la fracción del tipo de suelo y porcentaje de aridez en estudios de análisis de caudales mínimos, caudal base y caudal ecológico (Sanborn y Bledsoe, 2006; Kult *et al.*, 2014). Este tipo de variables son empleadas para la evaluación del efecto antropogénico (Johnston y Shmagin, 2008; Peterson *et al.*, 2011), caudales máximos (Nathan y McMahon, 1990; Laaha y Blöschl, 2006; Ahn y Palmer, 2016), caudales mínimos y ecológicos (Sanborn y Bledsoe, 2006).

6.1.7. Variables físico-químicas y ambientales

Las variables físico-químicas y ambientales más empleadas son el índice de acidez, conductividad específica, sólidos suspendidos y concentración de sulfato (Hynds *et al.*, 2014). Normalmente estas variables están enfocadas en la determinación de análisis de componentes principales de los elementos más importantes que afectan algún proceso físico, como la evaluación de la calidad del agua.

En la evaluación de la calidad del agua WQE, por sus siglas en inglés, se han desarrollado tres principales métodos, el primero es, evaluación sintética difusa (FSE), análisis de agrupamiento difuso (FCA) y sistema de inferencia difusa (FIS) (Wong y Hu, 2013).

En el área de calidad del agua se han implementado métodos multivariados que han demostrado ser eficaces (Panda *et al.*, 2006; Wong y Hu, 2013). En Quesland, Australia (McNeil *et al.*, 2005) se realizó un estudio enfocado en el análisis de las series históricas de muestras grandes, aproximadamente de 34,000 muestras acumuladas durante 30 años, con métodos de agrupación de K-medias en dos etapas se obtuvo la agrupación preliminar de 347 grupos y la agrupación final de 9 grupos. Encontrando la composición del agua superficial, alta en sodio, cloro y magnesio y con bajas cantidades de calcio y sulfatos.

Se han desarrollado estudios en la evaluación de contaminación del agua por acciones antropogénicas (Panda *et al.*, 2006) en hidrogeoquímica de la hidrogeología volcánica (Erwin *et al.*, 2009). En la evaluación del impacto antropogénico en agua superficial, donde se emplearon las técnicas de agrupamiento difuso, análisis multivariado y técnicas de GIS (Güler *et al.*, 2012) en análisis de formación de grupos de acuerdo a las características del perfil fisicoquímico en muestras con contaminación de agua (Hynds *et al.*, 2014).

Algunos estudios se han enfocado en realizar comparacion de SOM, análisis de agrupamiento y PCA. Astel *et al.*, (2007) y Lachance y Bernard (1979) aplicaron en un conjunto de datos de indicadores químicos, para el análisis de la calidad del agua. Kirk y Williams (1985) desarrollaron una estandarización de variables, por medio de variables químicas los autores desarrollaron un conjunto de técnicas de análisis de aglomerados, MANOVA, análisis canónico y análisis discriminante.

6.1.8. Variables relacionadas a la geología

En cuanto a los descriptores relacionados con la geología, éstos son empleados para la descripción de caudales mínimos (Yu *et al.*, 2002; Laaha y Blöschl, 2006; Sanborn y Bledsoe, 2006).

Yu *et al.* (2002) emplearon métodos multivariados (Método de Ward y Método de *K-means*) para regionalizar caudales mínimos a partir de variables hidrogeológicas, a partir de un análisis de componentes principales, donde además realizan un análisis discriminante y un análisis de agrupamiento, emplearon tanto el enfoque de regresión tradicional así como regresión difusa para la regionalización de curvas de duración de caudal en sitios no aforados, identificando las variables hidrogeológicas que tienen correspondencia con caudales mínimos.

6.2. Ecuaciones empleadas en la descripción de la cuenca

Abreviaciones

A	Área de la cuenca [km ²]
L	Longitud de la cuenca [km]
C	Compacidad [adim]
B	Parámetro de forma [km ⁻¹]
BH	Parámetro de forma de Horton [adim]
W_{max}	Ancho máximo de la cuenca [km]
W_{mean}	Ancho medio de la cuenca [km]
Tc	Tiempo de concentración [h]
S_{len}	Longitud de la corriente principal [km]
\overline{S}_{len}	Longitud promedio de la corriente [km]
L_{TOT}	Longitud total de la red de drenaje [km].
W_L	Ancho de las corrientes [m]
$SFREQ$	Frecuencia de corrientes [uniones* km ⁻²]
D	Densidad de corrientes [km ⁻¹]
E	Relación de elongación [adim]
C_R	Relación circular [km]

a) Factor de forma de Horton

$$BH = \frac{A}{L^2} \quad (6.1)$$

donde BH es el parámetro de forma de Horton [adim], A es el área de la cuenca [km^2], y L es la longitud de la cuenca [km].

b) Parámetro de forma

$$B = \frac{P}{A} \quad (6.2)$$

donde B es el parámetro de forma [km^{-1}], A es el área de la cuenca [km^2], y P es el perímetro de la cuenca [km].

c) Relación de elongación

$$E = \frac{1.128A^{0.5}}{L} \quad (6.3)$$

donde E es la relación de elongación [adimensional], A es el área de la cuenca [km^2], y L es la longitud de la cuenca [km].

d) Relación de circularidad

$$C_r = \frac{4\pi A}{P} \quad (6.4)$$

donde C_r es la relación de circularidad [km], A es el área de la cuenca [km^2], y P es el perímetro de la cuenca [km].

e) Compacidad

$$C = \frac{A/P^2}{100} \quad (6.5)$$

donde C es la compacidad [adim], A es el área de la cuenca [km^2], y P es el perímetro de la cuenca [km].

f) Pendiente del cauce principal

$$S = \frac{h}{L} \quad (6.6)$$

donde S es la pendiente del cauce principal [km/km], h es el desnivel del punto con mayor elevación al punto con menor elevación de la cuenca [km], y L es la longitud de la cuenca [km].

g) Similitud cuenca de drenaje

El valor de la constante SC se emplea para indicar las cuencas que son geomorfológicamente similares (Spring, 2006).

$$SC = \frac{A}{L^2} \quad (6.7)$$

donde SC es la similitud cuenca de drenaje [adimensional], A es el área de la cuenca [km²], y L es la longitud de la cuenca [km].

h) Densidad de corrientes

$$D = \frac{L_{TOT}}{A} \quad (6.8)$$

donde D es la densidad de corrientes [km⁻¹], L_{TOT} es la longitud total de la red de drenaje [km], y A es el área de la cuenca [km²].

i) Pendiente media de la cuenca

Se calcula como el desnivel entre las curvas de nivel en metros (d), la y el área de la cuenca (A)

$$Pm = \frac{dL_C}{A} \quad (6.9)$$

donde Pm es la pendiente media de la cuenca [m²/ m²], A es el área de la cuenca [m²], d es el desnivel entre las curvas de nivel [m], y L_C es la longitud entre las curvas de nivel [m].

j) Densidad hidrográfica

Es la relación que existe entre el número de corrientes y el área de drenaje, donde D_h está en corrientes/km².

$$D_h = \frac{N_c}{A} \quad (6.10)$$

donde D_h es la densidad hidrográfica [corrientes/km²], N_c es el número de corrientes [corrientes], y A es el área de la cuenca [m²].

k) Relación de bifurcación

Es la relación que existe entre el número de cauces de cualquier orden entre el número de cauces del orden inmediato superior.

$$R_b = \frac{N_u}{N_{u+1}} \quad (6.11)$$

donde R_b es la relación de bifurcación [adimensional], N_u es el número de cauces de orden u , y N_{u+1} es el número de cauces de orden $u + 1$.

l) Longitud media de corriente específica

La longitud media de una corriente de determinado orden, se obtiene dividiendo la longitud total de la corriente entre el número de segmentos

$$\bar{L}_i = \frac{\sum_{i=1}^n L_i}{N_i} \quad (6.12)$$

donde \bar{L}_i es la longitud media de corrientes [m], L_i es la longitud de la corriente de orden específico [m], y N_i es el número total de corrientes específica [adimensional].

m) Índice de Glushkov

Medida descriptiva para asociar los ríos por las características de los lechos rocosos (Paris y Zucarelli, 2004):

$$K_{go} = \frac{w^{1/2}}{h} \quad (6.13)$$

donde K_{go} es el índice de Glushkov, siendo rocosos si K_{go} tiene el valor de 1.4, arenosos con K_{go} con valor de 6.50, w es el ancho de la sección del río [m], y h es la profundidad del cauce.

n) Tiempo de concentración

A continuación las siguientes ecuaciones que describen el tiempo de concentración (Cevallos-Flores, 2016), esta una variable empleada en el diseño hidrológico e hidráulico. Se define como el tiempo requerido para que todo el sistema contribuya eficazmente a la generación del caudal en un punto dado, es decir, en el punto de salida (Cevallos-Flores, 2016).

1) ASC Modificada

$$tc = \frac{0.007 \times n^{0.8} \times L^{0.8}}{S^{0.4} \times P_2^{0.5}} \quad (6.14)$$

donde tc es el tiempo de concentración [min], n es el coeficiente de rugosidad de Manning [$m^{-1/3}s$], L es la longitud del curso de agua más largo [ft], S es la pendiente promedio del cauce principal [adimensional], y P_2 es la precipitación con un periodo de retorno de 2 años para una lluvia de 24 horas [in].

2) ASCE

$$tc = \frac{7.2983 \times L^{0.6} \times n^{0.6}}{i^{0.4} \times S^{0.3}} \quad (6.15)$$

donde tc es el tiempo de concentración [h], n es el coeficiente de rugosidad de Manning [$m^{-1/3}s$], L es la longitud del curso de agua más largo [km], i es la intensidad de la lluvia [mm/h], y S es la pendiente promedio del cauce principal [adimensional].

3) Bransby-Williams

$$tc = 14.6 \times L \times A^{-0.1} \times S^{-0.2} \quad (6.16)$$

donde tc es el tiempo de concentración [min], L es la longitud del curso de agua más largo [km], S es la pendiente promedio del cauce principal [adimensional], y A es el área de la cuenca [km²].

4) California Culvert Practice

$$tc = 60 \times \left[\frac{0.87075 \times L^3}{H} \right]^{0.385} \quad (6.17)$$

donde tc es el tiempo de concentración [min], L es la longitud del curso de agua más largo [km], y H es el desnivel entre el punto más alto y más bajo de la cuenca [m].

5) Carter

$$tc = 0.0977 \times \frac{L^{0.6}}{S^{0.3}} \quad (6.18)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km] y S es la pendiente promedio del cauce principal [adimensional].

6) Clark

$$tc = 0.335 \times \left[\frac{A}{S^{0.5}} \right]^{0.593} \quad (6.19)$$

donde tc es el tiempo de concentración [h], S es la pendiente promedio del cauce principal [adimensional], y A es el área de la cuenca [km²].

7) Corps Engineers

$$tc = 0.191 \times L^{0.76} \times S^{-0.19} \quad (6.20)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

8) Dooge

$$tc = 0.365^{0.41} \times S^{-0.17} \quad (6.21)$$

donde tc es el tiempo de concentración [h], y S es la pendiente promedio del cauce principal [adimensional].

9) DNOS

$$tc = 0.419 \times \frac{A^{0.3} \times L^{0.2}}{K \times S^{0.4}} \quad (6.22)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], A es el área de la cuenca [km²], S es la pendiente promedio del cauce principal [adimensional], y K es el coeficiente del tipo de superficie [adimensional].

10) Ecuación del SCS

$$tc = \frac{100 \times L^{0.8} \times \left[\frac{1000}{CN} - 9 \right]^{0.7}}{1900 \times (S \times 100)^{0.5}} \quad (6.23)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [ft], CN es el número de curva de escurrimiento [adimensional], y S es la pendiente promedio del cauce principal [adimensional].

11) Epsey

$$tc = 6.89 \times \left[\frac{L}{\sqrt{S}} \right]^{0.36} \quad (6.24)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

12) Epsey-Winslow

$$tc = 31 \times \Phi \times L^{0.29} \times S^{-0.145} \times I^{-0.6} \quad (6.25)$$

donde tc es el tiempo de concentración [min], Φ es el factor de conducción [adimensional], L es la longitud del curso de agua más largo [ft], I es el porcentaje de área impermeable en la cuenca [%], y S es la pendiente promedio del cauce principal [adimensional].

13) Federal Aviation Agency

$$tc = 3.26(1.1 - C) \frac{(L \times 1000)^{0.5}}{(S \times 100)^{0.333}} \quad (6.26)$$

donde tc es el tiempo de concentración [min], C es el coeficiente de escurrimiento del método racional [adimensional], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

14) George Rivero

$$tc = \frac{16 \times L}{\left[(1.05 - 0.2 \times p) \times (100 \times S)^{0.04} \right]} \quad (6.27)$$

donde tc es el tiempo de concentración [min], L es la longitud del curso de agua más largo [km], p es la relación entre el área cubierta de vegetación y el área de la cuenca [adimensional], y S es la pendiente promedio del cauce principal [adimensional].

15) Giandotti

$$tc = \frac{4 \times \sqrt{A} + 1.5 \times L}{25.3 \times \sqrt{S \times L}} \quad (6.28)$$

donde tc es el tiempo de concentración [h], A es el área de la cuenca [km²], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

16) Haktanir y Sezen

$$tc = 0.7473 \times L^{0.841} \quad (6.29)$$

donde tc es el tiempo de concentración [h], y L es la longitud del curso de agua más largo [km].

17) Henderson y Wooding

$$tc = 0.94 \times i^{-0.4} \times \left[\frac{n \times L}{\sqrt{S}} \right]^{0.6} \quad (6.30)$$

donde tc es el tiempo de concentración [min], n es el coeficiente de rugosidad de Manning del cauce [$m^{-1/3}s$], L es la longitud del curso de agua más largo [ft], i es la intensidad de la lluvia [mm/h], y S es la pendiente promedio del cauce principal [adimensional],

18) Izzard

$$tc = \frac{41.025 \times (0.0007 \times i^{0.8} + c) \times L^{0.333} \times n^{0.605}}{i^{0.667} \times S^{0.333}} \quad (6.31)$$

donde tc es el tiempo de concentración [min], n es el coeficiente de rugosidad de Manning del cauce [$m^{-1/3}s$], L es la longitud del curso de agua más largo [ft], S es la pendiente promedio del cauce principal [adimensional], i es la intensidad de la lluvia [ft/s], y c es el coeficiente de escorrentía del método racional [adimensional].

19) Johnstone Cross

$$tc = 5 \times \left(\frac{L}{\sqrt{S}} \right)^{0.5} \quad (6.32)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [mi], y S es la pendiente promedio del cauce principal [ft/mi].

20) Kerby-Hathaway

$$tc = \left[\frac{0.67 \times n \times L}{\sqrt{S}} \right]^{0.467} \quad (6.33)$$

donde tc es el tiempo de concentración [min], n es el coeficiente de rugosidad de Manning del cauce [$m^{-1/3}s$], L es la longitud del curso de agua más largo [ft], S es la pendiente promedio del cauce principal [adimensional].

21) Kinematic wave

$$tc = 7.35 \times n^{0.6} \times i^{-0.4} \times L^{0.6} \times S^{-0.3} \quad (6.34)$$

donde tc es el tiempo de concentración [h], n es el coeficiente de rugosidad de Manning del cauce [$m^{-1/3}s$], L es la longitud del curso de agua más largo [km], i es la intensidad de la lluvia [mm/h], y S es la pendiente promedio del cauce principal [adimensional].

22) Kirpich

$$tc = 0.0663 \times L^{0.77} \times S^{-0.385} \quad (6.35)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

23) McCuen

$$tc = 2.253 \times i^{-0.7164} \times L^{0.5552} \times S^{-0.2070} \quad (6.36)$$

donde tc es el tiempo de concentración [h], i es la intensidad de la lluvia [mm/h], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

24) Método Racional Generalizado

$$tc = \left[\frac{60 \times n \times L}{H^{0.3}} \right] \quad (6.37)$$

donde tc es el tiempo de concentración [min], n es el coeficiente de rugosidad de Manning del cauce [$m^{-1/3}s$], L es la longitud del curso de agua más largo [km], y H es el desnivel entre el punto más alto y el más bajo de la cuenca [m].

25) Modelo Eagleson

$$tc = 0.0111 \times L \times n \times Rh^{-2/3} \times S^{-1/2} \quad (6.38)$$

donde tc es el tiempo de concentración [min], n es el coeficiente de rugosidad de Manning del cauce [$m^{-1/3}s$], L es la longitud del curso de agua más largo [ft], S es la pendiente promedio del cauce principal [adimensional] y Rh es el radio hidráulico [ft].

26) Morgali y Linsley

$$tc = \frac{0.933 \times L^{0.6} \times n^{0.6}}{i^{0.4} \times S^{0.3}} \quad (6.39)$$

donde tc es el tiempo de concentración [min], n es el coeficiente de rugosidad de Manning del cauce [$m^{-1/3}s$], L es la longitud del curso de agua más largo [ft], i es la intensidad de la lluvia [mm/h], y S es la pendiente promedio del cauce principal [adimensional].

27) Papadakis y Kazan

$$tc = \frac{2.1539 \times n^{0.52} \times L^{0.5}}{i^{0.38} \times S^{0.31}} \quad (6.40)$$

donde tc es el tiempo de concentración [h], n es el coeficiente de rugosidad de Manning del cauce [$m^{-1/3}s$], L es la longitud del curso de agua más largo [km], i es la intensidad de la lluvia [mm/h] y S es la pendiente promedio del cauce principal [adimensional].

28) Passini

$$tc = \frac{[0.108 \times (A \times L)^{1/3}]}{S^{0.5}} \quad (6.41)$$

donde tc es el tiempo de concentración [h], A es el área de la cuenca [km^2], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

29) Pérez

$$tc = \frac{L}{72 \times \left(\frac{H}{L}\right)^{0.6}} \quad (6.42)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], y H es el desnivel entre el punto más alto y el más bajo de la cuenca [m].

30) Pickering

$$tc = \left(\frac{0.871 \times L^3}{H} \right)^{0.385} \quad (6.43)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], y H es el desnivel entre el punto más alto y el más bajo de la cuenca [m].

31) Picking

$$tc = 0.0883 \times L^{0.667} \times S^{-0.333} \quad (6.44)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

32) Pilgrim y McDermott

$$tc = 0.76 \times A^{0.38} \quad (6.45)$$

donde tc es el tiempo de concentración [h], y A es el área de la cuenca [km²].

33) Simas-Hawkins

$$tc = 0.322 \times A^{0.594} \times L^{-0.594} \times S^{-0.150} \times S_{SCS}^{0.31} \quad (6.46)$$

$$S_{SCS} = \frac{25400}{CN} - 254 \quad (6.47)$$

donde tc es el tiempo de concentración [h], A es el área de la cuenca [km²], L es la longitud del curso de agua más largo [km], S es la pendiente promedio del cauce principal [adimensional], S_{SCS} es la máxima capacidad de retención [mm], y CN es el número de la curva de escurrimiento [adimensional].

34) Snyder

$$tc = 1.2 \times (L \times L_{cg})^{0.3} \quad (6.48)$$

donde t_c es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], y L_{cg} es la distancia desde la salida hasta el centro de gravedad de la cuenca [mi].

35) Témez

$$t_c = 0.3 \times \left[\frac{L}{S^{0.25}} \right]^{0.76} \quad (6.49)$$

donde t_c es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

36) Valencia y Zuluaga

$$t_c = 1.7694 \times A^{0.325} \times L^{-0.096} \times S^{-0.290} \quad (6.50)$$

donde t_c es el tiempo de concentración [h], A es el área de la cuenca [km²], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [%].

37) Van Sickle

$$t_c = 0.55 \times \left(\frac{L \times Lcu}{\sqrt{S}} \right) \quad (6.51)$$

donde t_c es el tiempo de concentración [min], L es la longitud del curso de agua más largo [mi], Lcu es el perímetro de la cuenca [mi], y S es la pendiente promedio del cauce principal [adimensional].

38) Vente Chow

$$t_c = 0.1602 \times L^{0.64} \times S^{-0.32} \quad (6.52)$$

donde t_c es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

39) Ventura-Heras

$$tc = \alpha \times \frac{A^{0.5}}{S} \quad 0.04 \leq \alpha \leq 0.13 \quad (6.53)$$

$$\alpha = \frac{L}{\sqrt{S}} \quad (6.54)$$

donde tc es el tiempo de concentración [h], α es el parámetro en función de la pendiente, A es el área de la cuenca [km²], L es la longitud del curso de agua más largo [km], y S es la pendiente promedio del cauce principal [adimensional].

40) Williams

$$tc = \frac{0.272 \times L \times A^{0.4}}{D \times S^{0.2}} \quad (6.55)$$

donde tc es el tiempo de concentración [h], L es la longitud del curso de agua más largo [km], A es el área de la cuenca [km²], D es el diámetro equivalente al área de la cuenca [km], y S es la pendiente promedio del cauce principal [adimensional].

41) Woolhisier y Ligget's

$$tc = 7.3015 \times \left(\frac{n \times L}{S^{0.5}} \right)^{0.6} \times i^{-0.4} \quad (6.56)$$

donde tc es el tiempo de concentración [h], n es el coeficiente de rugosidad de Manning del cauce [m^{-1/3}s], L es la longitud del curso de agua más largo [km], i es la intensidad de la lluvia [mm/h], y S es la pendiente promedio del cauce principal [adimensional].

42) Yen y Chow's

$$tc = 1.2 \times \left(\frac{n \times L}{S^{0.5}} \right)^{0.6} \quad (6.57)$$

donde tc es el tiempo de concentración [h], n es el coeficiente de rugosidad de Manning del cauce [m^{-1/3}s], L es la longitud del curso de agua más largo [km], S es la pendiente promedio del cauce principal [adimensional].

6.3. Comentarios y recomendación de variables

The World Meteorological Organization (WMO) recomienda un periodo de información mínimo de 30 años que se debe recabar en estudios hidrológicos, sin embargo, diversos estudios enfocados en la regionalización marcan la recolección de registros entre un periodo de 10 a 20 años (GREHYS, 1996; Ouarda *et al.*, 2006; Soltani y Reza, 2006; Campos-Aranda, 2014b). Otra recomendación indica la selección de la cantidad de registros dependiendo de los T años de periodo de retorno a simular, la relación es $5T$ la cantidad de registros, es decir, cinco veces el periodo de retorno de interés (Reed *et al.*, 1999; Rao y Srinivas, 2006; Ilorme y Griffis, 2013).

Además de la revisión del periodo de registro observado, algunos autores exhortan la verificación en la homogeneidad de los registros debido a la escasez de datos (Kahya *et al.*, 2008). Aunado a las pruebas de homogeneidad en los registros Machiwal y Jha (2012) recomiendan un análisis exploratorio de datos previos al análisis regional con el propósito de verificar los supuestos del Análisis Regional de Frecuencias (ARF).

Existen diversos problemas en la regionalización, el primero de ellos es la selección de variables, sin embargo, en la literatura no existen muchos estudios que muestren a detalle la selección de las variables óptimas (Bates *et al.*, 1998; Ilorme y Griffis, 2013). Nathan y McMahon (1990) proponen como primer paso el análisis del empleo de las variables que se relacionen con el objeto de estudio, además de la importancia de ponderar las variables; asimismo identificaron los problemas comunes en la regionalización por el método de agrupamiento, como la magnitud en la escala de las variables, los métodos de ligadura y las distancias empleadas.

En el análisis de caudales, es importante la selección de estaciones hidrométricas cuyos escurrimientos aforados no estén siendo afectados por embalses, es decir, que no sean corrientes provenientes de escurrimiento controlado por efecto de diseño de operación de plantas hidroeléctricas (Campos-Aranda, 2014b).

Cuando el hidrólogo trabaja con caudales, debe considerar las tres diferentes formas en las cuáles puede extraer información para desarrollar un ARF (Sarhadi y Modarres, 2011): (1) series de máximos anuales (am, por sus siglas en inglés), (2) series de duración parcial (POT,

por sus siglas en inglés) y (3) datos de series de tiempo (TS, por sus siglas en inglés). En las series de máximos anuales, sólo la observación máxima anual es considerada, por lo que se puede involucrar pérdida de información, es decir, si el segundo o el tercero máximo de algún año superan al primero máximo de los otros años. Por lo que con los datos de series de duración parcial este problema se evita, al establecer un valor base. Por lo que en la selección de una u otra variable se deberán tener en cuenta las ventajas y desventajas de cada método. Autores como Lins (1985) propusieron incluir en el análisis de caudales, aquellas estaciones con desviación inferior al 10% del caudal medio y con capacidad de almacenamiento menor al 10% del escurrimiento anual (Kahya *et al.*, 2008).

Existen críticas en cuanto a la elección de variables de los estadísticos de los registros provenientes de las crecientes y se han señalado como variables no aptas para la delimitación, debido a que se emplearían datos estadísticos tanto para la formación de regiones homogéneas y posteriormente para la verificación de homogeneidad de las regiones formadas (Burn *et al.*, 1997; Campos-Aranda, 2014b). Debido a ello se recomienda el uso de estadísticos únicamente para la verificación de la homogeneidad de las regiones formadas (Hosking y Wallis, 1997).

El gran conjunto de datos implica que la escala de cada variable sea distinta, por lo que el hidrólogo puede trabajar con datos normalizados, para que, en el proceso de regionalización este factor no altere los resultados de selección de variables, así como la formación de grupos homogéneos (Soltani y Reza, 2006), otro procedimiento consiste en realizar una prueba de normalidad por medio de la prueba de gráfica de correlación y los datos se tratan por medio de la transformación de Box-Cox, por lo que las variables pueden escalarse con valores de 0 a 1, evitando de esta forma asignar mayor peso a cualquiera de ellas (Ilorme y Griffis, 2013). Soltani y Reza (2006) aplicaron la normalización de variables con logaritmo natural, seguido de una aplicación de análisis de componentes principales y por último un análisis discriminante con las variables más representativas.

La recomendación de algunos autores (Ilorme y Griffis, 2013) respecto a la selección de variables físicas es por medio de una combinación de técnicas como lo son: análisis de agrupamiento, análisis de componentes principales (ACP), análisis de correlación canónica

(ACC) y análisis múltiple discriminante, por lo que al realizar estas técnicas es más precisa la estimación de cuantiles.

Una técnica principalmente usada es el ACP cuando se emplea un gran número de datos de distintas escalas, como pueden ser variables geomorfológicas, climatológicas, ambientales, físico-químicas, uso de suelo, entre otras (Sanborn y Bledsoe, 2006). La técnica ACP consiste en lo siguiente: a partir de un conjunto de datos, se crean combinaciones lineales de los mismos y se constituyen los vectores no correlacionados, es decir, los componentes principales cuya característica principal es que son ortogonales entre sí, además que en orden descendente conservan de la mayor a la menor varianza de los datos originales, de manera que es posible identificar cuáles son las variables más representativas y descartar la multicolinealidad de las mismas (Sanborn y Bledsoe, 2006).

Las técnicas multivariadas, como el ACP y ACC, han demostrado ser métodos óptimos en el análisis de variables que son modificadas por presas y en la identificación de los factores que influyen en la magnitud de los cambios hidrológicos (Matteau *et al.*, 2009), para este tipo de análisis se deben incluir como características las siguientes: caudales en verano e invierno, tiempo de registro del caudal medio mensual máximo y la variabilidad estacional de las corrientes; corrientes mensuales y el promedio mínimo mensual de corrientes, régimen regulado y el tamaño de la cuenca.

En la creación de grupos homogéneos, se busca obtener las zonas que producen la misma respuesta hidrológica, por ello es necesario contar con las características morfo-litológicas en estudios de caudal mínimo (Meijerink, 1985; Mamun *et al.*, 2010) que permitan al investigador descubrir irregularidades geotectónicas, es decir, irregularidades espaciales e irregularidades en los límites de la zona de estudio, por medio de parámetros geomorfológicos que demuestren la relación entre las cuencas de drenaje de los ríos y las estructuras geotectónicas (Gubareva, 2012).

En la predicción de caudales mínimos, Nathan y McMahon (1990) emplearon características tanto fisiográficas como climatológicas, geomorfológicas, e hidrogeológicas. Por otro lado Yu *et al.* (2002) emplearon variables de tipo hidrogeológicas, para formar mediante un análisis multivariado regiones hidrogeológicamente homogéneas. Nathan y McMahon (1990) proponen no emplear en el análisis multivariado de variables más representativas a la

variable del área de la cuenca de drenaje, ya que ésta representa un factor de escala. Ahn y Palmer (2016) enfatizan que la avenida o también llamada creciente es negativamente afectada por el tamaño de la cuenca en los métodos de Análisis Regional de Frecuencia de Crecientes, además que la variable del área se encuentra por lo regular negativamente correlacionada con las variables dependientes.

Yu *et al.*, (2002) realizaron una búsqueda en la literatura sobre las variables que intervienen dentro del concepto “caudal mínimo” encontrando para ello diversos índices que ayudan a caracterizar el flujo, siendo los siguientes: (1) tiempo con caudal cero, (2) caudal base, (3) flujos mínimos anuales de 1 día, 10 días y 30 días, (4) la media de los déficits máximos anuales observados durante eventos consecutivos de caudal bajo por debajo de alguna descarga de referencia o umbral de referencia, (5) duración media anual máxima de los acontecimientos durante los cuales los caudales caen por debajo de un umbral de caudal y (6) cuantiles de flujo extraídos de una curva de duración de flujo (FDC) que se superan con cierta probabilidad.

En estudios relacionados con la evaluación de estructuras hidráulicas y falla de las mismas, las variables involucradas en el estudio son las siguientes de tipo geográfica (topografía), de tipo geomorfológica (área de la cuenca, perímetro de la cuenca, pendiente de la cuenca, pendiente del cauce principal, longitud del cauce principal, uso del suelo y tipo de vegetación), de tipo meteorológica (precipitación máxima en 24 horas), datos de la geometría de la estructura (tipo de cortina, altura de la cortina, diámetro de las obras de toma, elevación de la obra de toma, ancho del vertedor, altura del vertedor, batimetría de la presa, ancho de corona, ancho de base, talud aguas abajo y aguas arriba).

En el análisis de tratamientos de aguas, las variables físico-químicas mínimas a emplear son las siguientes, temperatura del agua, turbidez, grado de acidez y basicidad, conductividad eléctrica, concentración de los elementos, cantidad de cationes y de aniones.

En la modelación del tránsito de crecientes en cuencas urbanas las variables a emplear están relacionadas con la estructura hidráulica (tubería o alcantarilla) a analizar (coeficiente de rugosidad de Manning), propiedad hidrológica del suelo (conductividad hidráulica), y variables geomorfológicas (densidad de drenaje, radio de bifurcación, longitud de los tributarios, densidad de corrientes, relación de longitud y relación del área).

En el manejo de Redes Neuronales Artificiales enfocadas a la geometría de tormentas, las variables a emplear son del tipo geomorfológicas (longitud, latitud, elevación), de tiempo (fecha y horario, temperatura relativa y dirección del viento), climática o meteorológica (humedad relativa, punto de rocío, precipitación acumulada, precipitación, radiación solar y presión atmosférica), del tipo geomorfológica (topografía), en cuanto al análisis de las imágenes vía satélite (color de píxel, temperatura tope de la nube fría y geometría observada).

Para la delimitación de regiones homogéneas en estudios de planeación y manejo eco-hidrológico, y en la predicción de caudales eco-hidrológicos, así como la estratificación de régimen de caudal de sitios aforados y para la clasificación de las corrientes en regímenes no aforados, Sanborn y Bledsoe (2006) proponen un biomonitoreo de variables, además dedujeron que las variables relacionadas con la magnitud y duración de caudales en cuencas grandes se encuentran como características fisiográficas la pendiente y la red de drenaje, mientras que las variables relacionadas con la determinación del caudal mínimo son el área de almacenamiento y variables geológicas. En lo referente al caudal máximo y mínimo y en conjunto con la información sobre la magnitud del caudal y duraciones de caudal de los pulsos bajos proveen una aproximación de primer orden en cuanto a la perturbación relativa en las corrientes no aforadas, que es una influencia importante en el análisis de las comunidades de macroinvertebrados bentónicos en estudios de caudal ecológico.

Para los estudios regionales en climas húmedos en zonas montañosas y mesetas Singh *et al.* (2014) encontraron relevantes las características físicas de la cuenca, principalmente la elevación, seguido por el uso de suelo, y en menor importancia el estudio de variables climáticas. Los autores también encontraron que cuando la transferencia de parámetros se realizaba en regiones no contiguas las características variaban significativamente, aunque la mayoría estaba relacionada con clima, suelo y uso de suelo; mientras que la relación de escurrimiento resultó no ser una característica de transferencia de preponderancia.

En la delimitación de regiones homogéneas con el objetivo de la implementación de estaciones hidrométricas para el monitoreo de caudales medianos, además de las variables fisiográficas, Paris y Zuccarelli (2004) sugieren el empleo de variables hidráulicas como velocidad y profundidad y ancho de la sección transversal del cauce, además de la caracterización del tipo de lecho del río.

En el desarrollo del método de la avenida índice y métodos de regresión múltiple, la variable de mayor importancia es el área (Kult *et al.*, 2014; Nobert *et al.*, 2014). Para el método de la avenida índice Kult *et al.* (2014) proponen el empleo de variables como el perímetro, pendiente de la cuenca, longitud del canal principal y pendiente del canal principal. Mientras que otros autores como Nobert *et al.*, (2014) proponen la pendiente de la cuenca, la altitud media, y la lluvia media anual.

En el análisis de correlación canónica en estudios con poca cantidad de estaciones aforadas y en clima tropical, Ouarda *et al.* (2008) emplearon las siguientes variables: área de drenaje, altitud del centroide de la cuenca, latitud de la estación de aforo, longitud de la estación de aforo, longitud de la corriente principal, pendiente del canal principal, porcentaje del área de la cuenca cubierta por lagos y media anual del número de grados de días bajo 0°C.

En estudios relacionados con el cambio climático, y escenarios climáticos planteados a futuro, Peterson *et al.* (2011) recomiendan una primera evaluación para la comprensión de los patrones espacio-temporales de la variabilidad del caudal y su relación con las características hidrogeológicas (Johnston y Shmagin, 2008), y un posterior análisis para la evaluación del cambio de uso de suelo como respuesta ante las actividades antropogénicas (Peterson *et al.*, 2011).

En el análisis de las variables que influyen en la magnitud de las inundaciones, Ahn y Merwade (2016) clasificaron las inundaciones como moderadas, extremas y severas, llegando a la conclusión que la variable de la longitud de la cuenca interviene en avenidas extremas, mientras que para avenidas severas están influenciadas por la forma de la cuenca y el uso de suelo, adicional a esta investigación, los autores analizaron el efecto de la cubierta de suelo y el tamaño de la cuenca por medio de la prueba de Kruskal-Wallis, obteniendo que la cubierta de suelo afecta la ocurrencia de la inundación. El análisis se llevó a cabo en 94 estaciones en Indiana, EUA, por medio de un análisis de correlación simple, análisis de componentes principales (PCA), y regresión por pasos.

Adicionalmente se incluyen una serie de variables empleadas en el análisis regional hidrológico. En el ANEXO I se realizó con el objetivo de proporcionar al lector los tipos a variables básicas a emplear en algún estudio específico, además de las ecuaciones que describen la fisiografía de la cuenca.

7. ANÁLISIS PREVIO DE DATOS

Después de la selección de variables involucradas en el estudio es necesario aplicar pruebas para examinar la calidad de los datos, así como la identificación de datos atípicos. En la detección de datos atípicos, la discordancia estandarizada para los datos multivariados atípicos es posible medirla por medio de los cocientes de Momentos-L (Hosking y Wallis, 1997), es decir, variación, asimetría y curtosis, calculando la discordancia de los registros (Saf, 2009; Campos-Aranda, 2012). En la revisión de calidad de los datos, se analizan las series de datos en homogeneidad, estacionariedad, independencia y normalidad, una vez verificado lo anterior el investigador puede comenzar con la delimitación de regiones homogéneas.

Previo a un análisis regional, el investigador debe seleccionar cuidadosamente las variables que están relacionadas con el fenómeno a estudiar y de esta forma generar grupos homogéneos con significado físico (Ebisemiju, 1979; Nathan y McMahon, 1990; Caratti *et al.*, 2005). En la búsqueda de la literatura existen distintas pruebas que son aplicables al conjunto de datos, ya sean datos estadísticos, fisiográficos o climatológicos, como el análisis exploratorio de datos recomendado por Tukey (1977) que se ha implementado en diversos estudios desde su divulgación (Lobo *et al.*, 2006).

En todo proceso relacionado al estudio de series de tiempo, es necesario establecer una metodología para analizar los datos que intervienen en el problema o caso de estudio, la primera parte consiste en un análisis de datos inicial, es decir una descripción generalizada para proceder con un análisis estadístico formal, que bien podrían ser un análisis de regresión y modelación estadística (Ojeda-Ramírez *et al.*, 2011).

En el desarrollo de este capítulo se muestra la metodología del análisis inicial de los datos, continuando con la revisión de la consistencia de los datos y por último las pruebas de la calidad estadística que son aplicables en estudios hidrológicos.

Algunos autores (Ilorme y Griffis, 2013) han realizado un tratamiento previo a los datos, es decir, una prueba de normalidad por medio de la prueba de grafica de correlación y los tratamiento de datos por medio de la transformación de Box-Cox. En este tratamiento las variables se estandarizan, se escalan de 0 a 1 para evitar asignar mayor peso a cualquiera de ellas.

En otros estudios más recientes (Chebana *et al.*, 2017) se han realizado un análisis del cambio en las series de registros multivariados, cuando existe un cambio en la serie de datos, puede perjudicar la estacionariedad y la homogeneidad en los datos. Los autores realizaron una comparación de pruebas aplicables en la detección en el cambio de las series de datos, dentro de los resultados, encontraron que los factores que más influyen a las pruebas son el tamaño de la muestra, la amplitud del cambio, la magnitud de las series y la ubicación del cambio en la serie.

7.1. Estadística multivariada descriptiva

El análisis inicial de datos, también llamado análisis descriptivo o exploratorio, consiste en la obtención generalizada de los datos (Ojeda-Ramírez *et al.*, 2011).

En un Análisis Regional de Frecuencias (ARF) los datos no se obtienen de una muestra univariada en un punto específico, pudiendo ser éste una estación climatológica o hidrométrica, sino que, por el contrario, se analizan un conjunto de estaciones, por lo que el análisis que debe emplearse es un análisis descriptivo estadístico multivariado.

El análisis multivariado en el desarrollo de un análisis regional de crecientes juega un papel importante, debido a que no se analiza particularmente una muestra o conjunto de datos, sino que los m datos de las n estaciones seleccionadas deben ser analizados (Ojeda-Ramírez *et al.*, 2011).

A continuación, se muestran los estadísticos básicos para una muestra hidrológica, así como la estadística multivariada descriptiva que interviene en un estudio de ARF.

7.1.1. Estadísticos básicos de una muestra hidrológica

El objetivo de la estadística es conocer en primera instancia la descripción general del conjunto de datos con el cual se trabajará, por ello, es importante desarrollar la estadística descriptiva, con la cual es posible extraer la información esencial del conjunto de datos (Chow *et al.*, 1994; Ojeda-Ramírez *et al.*, 2011).

Para iniciar con la estadística descriptiva, primero se detalla a continuación dentro de la estadística univariada los tres principales parámetros estadísticos, denominados media, varianza y asimetría. Calculados de la siguiente forma (Chow *et al.*, 1994)

Parámetros característicos de la población (Chow *et al.*, 1994):

$$\mu = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad (7.1)$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 dx \quad (7.2)$$

$$\gamma = \frac{1}{\sigma^3} \int_{-\infty}^{\infty} (x - \mu)^3 f(x) dx \quad (7.3)$$

Estadística de la muestra (Chow *et al.*, 1994):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (7.4)$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (7.5)$$

$$C_s = \frac{n \sum_{i=1}^n (x_i - \bar{x})^3}{(n-1)(n-2)s^3} \quad (7.6)$$

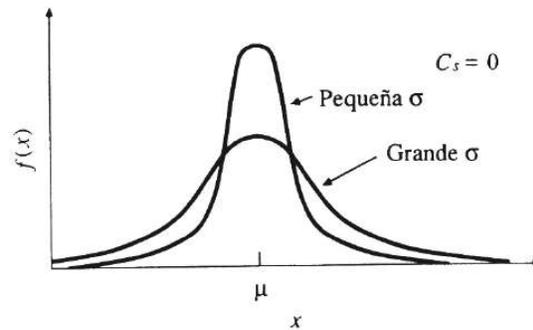


Figura 7-1. Desviación estándar en una muestra (Chow *et al.*, 1994)

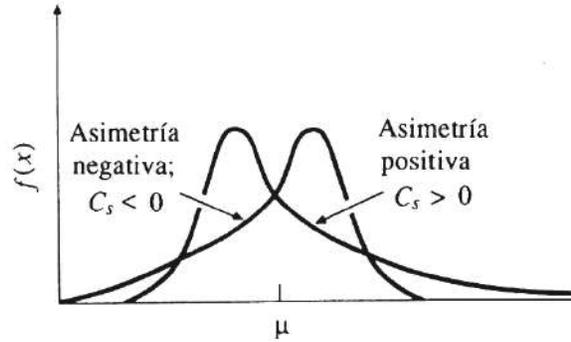


Figura 7-2. Coeficiente de asimetría en una muestra con datos hidrológicos (Chow *et al.*, 1994)

7.1.2. Estadística descriptiva multivariada

Como primera aproximación a los datos es necesario el análisis descriptivo, para ello se desarrollarán los estadísticos aplicados a los casos multivariados con el objetivo de la formación de grupos (Ojeda-Ramírez *et al.*, 2011).

A partir de una matriz de datos con n sitios (estaciones hidrométricas) y p variables (que bien pueden ser del tipo fisiográficas, climatológicas o hidrométricas) se forma la matriz X_{ij} .

Vector de medias (Ojeda-Ramírez *et al.*, 2011): Se obtienen las medias de las p variables que intervienen en el estudio

$$\bar{x}_p = \frac{\sum_{i=1}^n x_{ip}}{n} \quad (7.7)$$

El vector de medias se conforma por el vector columna de dimensiones $p \times 1$

$$\mathbf{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} \quad (7.8)$$

Matriz de varianzas y covarianzas:

$$s_{jk} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{n} \quad (7.9)$$

En el caso multivariado, con p variables, se obtiene p varianzas, una por cada variable, y $p(p-1)/2$ covarianzas, las cuales se ubican en una matriz S de varianzas-covarianzas de orden de $p \times p$.

$$\mathbf{S} = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{bmatrix} \quad (7.10)$$

La característica de esta matriz es que cuenta con la diagonal principal que contiene a las varianzas de cada una de las p variables y fuera de ésta se encuentran las covarianzas s_{jk} entre cada par de variables bajo estudio.

La matriz de correlación: medida descriptiva que cuantifica la correlación lineal de las variables X_j y X_k . El cual está comprendido entre -1 y 1. Cuando $r = 1$, se dice que la correlación es directa, mientras que, si el valor es cero, no existe correlación lineal, mientras que si la correlación es -1, se le denomina perfecta inversa o negativa.

$$r_{jk} = \frac{s_{jk}}{s_j s_k} \quad (7.11)$$

Matriz de correlaciones

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (7.12)$$

De esta forma la relación entre la matriz de correlaciones y la matriz de varianzas y covarianzas está dada por (Ojeda-Ramírez *et al.*, 2011):

$$\mathbf{R} = \mathbf{D}^{-1/2} \mathbf{S} \mathbf{D}^{-1/2} \quad (7.13)$$

$$\mathbf{S} = \mathbf{D}^{-1/2} \mathbf{R} \mathbf{D}^{-1/2} \quad (7.14)$$

donde \mathbf{D} es la diagonal $(s_{11}, s_{22}, \dots, s_{pp})$, es decir \mathbf{D} es una matriz cuyas entradas son las varianzas de las p variables de estudio, por lo que $\mathbf{D}^{-1/2} = (s_{11}^{-1/2}, s_{22}^{-1/2}, \dots, s_{pp}^{-1/2})$.

Medidas de variabilidad global: varianza total varianza media, varianza generalizada y la desviación típica generalizada.

Varianza total: Se estima como la traza de la matriz de varianzas y covarianzas:

$$VT = \sum_{j=1}^p s_j^2 \quad (7.15)$$

Varianza media: Varianza total entre el número de variables

$$VTm = \frac{1}{P} \sum_{j=1}^p s_j^2 \quad (7.16)$$

Varianza generalizada (VG) Es el determinante de la matriz de varianzas y covarianzas. $VG = |S|$

Desviación típica generalizada: Matriz cuadrada de la varianza generalizada.

7.1.3. Estadística descriptiva gráfica:

El correlograma es un gráfico es empleado para visualizar la dispersión en los datos, la característica de este gráfico es que es del tipo simétrico, en un estudio donde se cuente con p variables, se tendrán $p(p-1)/2$ gráficos de dispersión (Ojeda-Ramírez *et al.*, 2011).

7.2. Revisión de la consistencia de los datos

En un análisis regional basado en técnica de Momentos-L una etapa fundamental es la revisión y la inspección de los datos, para ello se debe garantizar la homogeneidad de las muestras demostrando la estacionariedad a través del tiempo y la independencia (Hosking y Wallis, 1997; Campos-Aranda, 2014a).

Para revisar la consistencia de los datos de las estaciones, se emplea la curva de masa doble Sine y Ayalew (Nobert *et al.*, 2014).

Para el análisis de frecuencia es necesario verificar que los datos de las muestras tengan la misma distribución de frecuencia.

Para la revisión de la estacionariedad estadística es necesario efectuar la prueba de Mann-Kedall (MK), que proporciona las bases para determinar si un conjunto de datos cambia con el tiempo.

7.3. Pruebas de calidad estadística a los datos

Posterior al análisis estadístico descriptivo, es necesario el procedimiento de pruebas de calidad estadística.

La importancia en el desarrollo de estas pruebas en series de tiempo hidrológicas permite verificar las características y supuestos requeridos para el procesamiento de los modelos estadísticos específicos (Adeloye y Montaseri, 2002; Machiwal y Jha, 2012).

Este análisis preliminar no se ha ejecutado con regularidad (Machiwal y Jha, 2012). Una prueba previa al análisis formal es un proceso que ha sido poco empleado y se ha descuidado (Lobo *et al.*, 2006; Machiwal y Jha, 2012). Machiwal y Jha (2006) encontraron que en diversos estudios solo se aplican las pruebas de tendencia lineal, siendo descartados los estudios de homogeneidad, estacionariedad, periodicidad y persistencia.

A continuación se muestra una síntesis de la información recopilada por Machiwal y Jha(2012) así como del documento técnico de la UNESCO realizado por Lobo *et al.*, (2006), se le propone al lector dirigirse a estos documentos en caso de ampliar la información al respecto.

Lobo *et al.*, (2006) resumen los siguientes siete puntos importantes a desarrollar en un Análisis Exploratorio de Datos:

1. Maximizar el entendimiento de los datos.
2. Describir la estructura subyacente de los datos (normalidad, multicolinealidad, asimetría, curtosis, linealidad, homogeneidad entre los grupos, homocedasticidad, entre otros).
3. Detectar variables importantes relacionadas con el fenómeno a estudiar.
4. Detectar valores atípicos.
5. Comprobar los supuestos implícitos en los análisis.
6. Determinar los ajustes óptimos de los factores.
7. Proporcionar al analista los elementos necesarios para concluir la significancia estadística de las variables.

7.3.1. Etapas de un análisis exploratorio de datos

Las seis etapas del análisis estadístico de datos consiste en lo siguiente (Lobo *et al.*, 2006):

1. Preparar los datos para hacerlos accesibles a cualquier técnica estadística.
2. Realizar un examen gráfico de la naturaleza de las variables individuales a analizar y un análisis descriptivo numérico que permita cuantificar algunos aspectos gráficos de los datos.
3. Realizar un examen gráfico de las relaciones entre las variables analizadas y un análisis descriptivo numérico que cuantifique el grado de interrelación existente entre ellas.
4. Evaluar, si fuera necesario, algunos supuestos básicos subyacentes a muchas técnicas estadísticas como, por ejemplo, la normalidad, linealidad y homocedasticidad.
5. Identificar los posibles valores atípicos y evaluar el impacto potencial que puedan tener en los análisis estadísticos posteriores.
6. Evaluar, si fuera necesario, el impacto potencial que pueden tener los datos faltantes sobre la representatividad de los datos analizados.

7.3.2. Valores atípicos, *outliers*

Los valores atípicos pueden definirse como “observaciones con características diferentes de las demás” (Lobo *et al.*, 2006). Hosking y Wallis (1997) identificaron dos tipos de error que se pueden presentar, aquellos relacionados durante la transcripción de datos y aquellos errores debido a la operación del instrumento de medición, consecuentemente esto se transcribe en un dato atípico dentro del conjunto de datos, adicional a los dos casos expuestos anteriormente. Lobo *et al.* (2006) clasificaron los datos atípicos de la siguiente forma: debidos a un error de procedimiento de transcripción de datos, los relacionados a un acontecimiento extraordinario, los relacionados a un rango de combinaciones de valores de la variable medida y aquellos valores relacionados a observaciones extraordinarias.

Lobo *et al.* (2006) proponen detectar todos estos valores en el filtrado de datos, así como analizar los resultados en presencia y ausencia de estos valores examinando la influencia de los mismos en los resultados del modelo.

La identificación de valores atípicos mediante la metodología propuesta por Tukey (1977) es la siguiente (Lobo *et al.*, 2006):

- a) Cálculo del rango intercuartil:

$$R1 = Q3 - Q1 \quad (7.17)$$

donde $R1$ es el rango intercuartil, $Q3$ es el tercer cuartil (75%) y $Q1$ es el primer cuartil (25%).

b) Cálculo del paso:

$$PASO = 1.5 * R1 \quad (7.18)$$

c) Cálculo de las cercas internas y externas, caso inferior y superior de la siguiente forma (Lobo *et al.*, 2006):

$$Cli = Q1 - PASO \quad (7.19)$$

$$Cls = Q3 + PASO \quad (7.20)$$

$$CEi = Q1 - 2 * PASO \quad (7.21)$$

$$CEs = Q3 + 2 * PASO \quad (7.22)$$

donde *Cli* es la cerca interna superior, *CEi* es la cerca externa superior.

Además, los valores ubicados entre la media y las cercas internas (inferior y superior) son denominados adyacentes, los valores entre las cercas internas y las cercas externas son considerados como alejados y los valores ubicados fuera de las cercas externas (inferior y superior) son considerados como muy alejados. Los valores ubicados fuera de las cercas externas son por lo general considerados como valores atípicos (Lobo *et al.*, 2006).

7.3.3. Pruebas de normalidad

Las pruebas de normalidad pueden ser tanto gráficas como modelos estadísticos. Dentro de los modelos gráficos que se han aplicado en la literatura son la gráfica de frecuencias, el histograma, gráfico de tallo y hoja, gráfico de cajas, gráfico de datos clasificados y gráfico de cuantiles. Los métodos estadísticos aplicados para verificar normalidad son Prueba de Chi-Cuadrado, Kolmogorov-Smirnov, Lilliefors, Anderson-Darling, Crámer-von-Mises, Shapiro-Wilk, prueba estadística de Filliben, prueba de rango estudiado y la prueba de Geary.

De todas las pruebas anteriores, la prueba de Shapiro-Wilk es la más popular y ha sido ampliamente usada en estudios hidrológicos (Machiwal y Jha, 2012).

7.3.4. Pruebas de homogeneidad

Al conjunto de pruebas para verificar la homogeneidad, también se les denomina pruebas de consistencia (Machiwal y Jha, 2012). En esta prueba el investigador debe verificar si los datos

observados pertenecen a una misma población, es decir, que los datos recolectados tengan una media invariable en el tiempo, por lo que todas las pruebas existentes se basan en la identificación de los cambios en el valor de la media.

Dentro de las pruebas más empleadas son las siguientes: Von Neumann, prueba de desviación acumulada, prueba bayesiana, prueba de Tukey, Link-Wallace, Dunnet y Barlett. De todas las pruebas mencionadas anteriormente, la prueba de Von Neumann es la más empleada en estudios relacionados a la rama de la hidrología (Machiwal y Jha, 2006), debido a que con las demás pruebas, es decir, el empleo de las pruebas como Tukey, Link-Wallace, Dunnet Barlett y la prueba de Hartley, requieren que “la población sea normalmente distribuida, lo que ocasiona que las pruebas sean paramétricas” (Machiwal y Jha, 2012).

El test de Von-Neumann es de la forma siguiente:

$$N = \frac{\sum_{t=1}^{n-1} (x_t - x_{t+1})^2}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (7.23)$$

Donde x_t es la variable hidrológica que constituye la secuencia en el tiempo, n es el número total de los registros hidrológicos y \bar{x} es el promedio de x_t . Los valores esperados para las series de tiempo homogéneas es de $N = 2$, con un valor de $N < 2$ se asume que es una serie de tiempo no homogénea.

7.3.5. Pruebas de revisión de estacionariedad

Cuando se verifica la estacionariedad en una serie de tiempo, se asume que las propiedades estadísticas son invariantes en el tiempo. La prueba general es determinar si las pruebas son paramétricas o no paramétricas (Machiwal y Jha, 2012).

Dentro de la revisión de la literatura realizada por Machiwal y Jha (2012), encontraron tres pruebas mayormente empleadas en la hidrología: la prueba T de Student, prueba T-Simple y la prueba de Mann-Whitney.

La prueba paramétrica de Mann-Whitney consiste en que la media de las sub-series es igual a la media de las segunda sub-serie. Se rechaza si el valor calculado obtenido es más grande que el valor crítico obtenido de tablas de distribución normal con el 5% de significancia. La ventaja de emplear esta prueba es que es una técnica más robusta debido a su origen no

paramétrico, por lo que es factible aplicarla tanto a series de tiempo normales como aquellas no normales (Machiwal y Jha, 2012).

Núñez *et al.* (2010) también proponen emplear cualquiera de las siguientes pruebas: Durbin-Watson o el análisis de tendencia pendiente de Sen o el uso del test de Mann-Kendall. Ouarda *et al.*, (2006) proponen emplear la prueba de Kendall para series de caudales diarios.

7.3.6. Métodos para detectar tendencia

Machiwal y Jha (2012) describen la tendencia como “una serie de valores sucesivos que a través del tiempo en el que hay un incremento o decremento” obteniendo de esta forma un valor lineal o no lineal.

En la revisión de la literatura se han registrado 15 pruebas distintas: Prueba de regresión, prueba de Spearman de orden de correlación clasificado, prueba de punto de inflexión, de fase de Kendall, prueba de Wald-Wolfowitz, suma de longitudes al cuadrado, prueba de adyacencia, prueba de diferencias de signo, prueba de Wilcoxon, pruebas de inversiones, prueba de correlación de Kendall, prueba de Mann-Kendall, prueba de estimación de pendiente Sen y pruebas de tendencia-homogeneidad.

Dentro de todas las pruebas descritas por Machiwal y Jha (2012), mencionan algunas pruebas que han dejado de aplicarse en las series hidrológicas. Para la prueba del punto de inflexión, los autores mencionan que ha dejado de ser práctico su uso, en la prueba de Wald-Wolfowitz “no es una técnica poderosa ni eficiente, no considera longitud de corridas grandes, por lo que hay información perdida durante el proceso” una alternativa más eficiente a la prueba de Wald-Wolfowitz es emplear la prueba de longitudes cuadradas.

Una ventaja al emplear las pruebas de Mann-Kendall, es la naturaleza de este test, al ser no paramétrico, analiza la tendencia de la serie ya sea normal o no.

Ouarda *et al.* (2006) proponen emplear el test de Wald-Wolfowitz para series de caudales diarios para verificar independencia.

7.4. Cantidad de variables a emplear en el análisis regional

En los estudios de análisis regional la cantidad de variables que se emplean no está especificada en la literatura, por lo que en ocasiones una alternativa es una aproximación de

prueba y error de las variables, la desventaja al emplear esta técnica es cuando la cantidad de variables es muy grande y la importancia de las variables descriptivas de la cuenca no se investiga de forma particular (Heřmanovský y Pech, 2013).

Para el empleo de variables óptimas existen diversas recomendaciones realizadas por diversos autores, Nathan y McMahon (1990) proponen para el método de agrupamiento ponderado, la identificación de variables más representativas por medio de un análisis de regresión escalonada. Otros autores como Malekinezhad *et al.* (2011) proponen un análisis factorial.

Heřmanovský y Pech (2013) realizaron una propuesta en la selección de variables descriptivas de la cuenca bajo el concepto de similitud física. Los autores descartaron trabajar con los enfoques de proximidad espacial y regresión, debido al argumento que recientemente que ha surgido sobre la poca efectividad al emplear la regresión lineal, debido a que las variables de importancia resultantes puedan explicar la relación no lineal entre los parámetros del modelo (Bao *et al.*, 2012). Mientras que el enfoque de proximidad espacial se ha criticado debido a que la cercanía no garantiza que el comportamiento hidrológico sea igual (Heřmanovský y Pech, 2013; Nobert *et al.*, 2014). Por lo que con la similitud física, ésta procesa la información de las variables características descriptivas de las cuencas que intervienen en el área de interés evaluando la cercanía espacial, por lo que este método es mejor (Heřmanovský y Pech, 2013; Nobert *et al.*, 2014).

El tratamiento previo a la delimitación de regiones homogéneas encontrado en la literatura son: análisis factorial (Malekinezhad *et al.*, 2011), combinación de análisis de correlación con ACP (Ahn y Palmer, 2016), y algoritmo basado en análisis de agrupamiento inverso (Heřmanovský y Pech, 2013).

El análisis factorial se ha empleado en métodos relacionados con el método de la Avenida índice (Malekinezhad *et al.*, 2011) para determinar la cantidad de variables que influyen en la magnitud del caudal, es decir de un grupo inicial de variables al aplicar el análisis factorial, se obtienen menor cantidad de variables, que se denominan factores, que explican la relación entre las variables independientes. Malekinezhad *et al.* (2011) aplicaron el análisis factorial en un estudio, para comparar el método de avenida índice con métodos de regresión múltiple, donde tenían 14 variables iniciales dependientes, mediante el análisis factorial,

obtuvieron cuatro factores: longitud del cauce precipitación media anual, coeficiente de Gravelius y temperatura media anual.

Agrupando el tipo de variables, éstas pueden ser de dos tipos, variables como descriptores hidrológicos y variables físicas como las variables fisiográficas y climáticas. En el enfoque por similitud física, se recomienda el empleo de los descriptores hidrológicos, es decir, de aquellos datos de registros que se relacionan con los registros de lluvia-escorrentamiento, aunque en la práctica es complicada la recabación de información, por lo que una solución es emplear las variables fisiográficas y las climáticas (Heřmanovský y Pech, 2013).

Heřmanovský y Pech (2013) desarrollaron un método para la selección óptima de las variables fisiográficas y climáticas, para ello desarrollaron una delimitación jerárquica de regiones homogéneas y un método de agrupamiento inverso. En las variables fisiográficas, emplearon cinco tipos de variables, i) descriptores del suelo, ii) descriptores geológicos, iii) cobertura del suelo, iv) descriptores morfológicos y, v) descriptores climáticos. En cuanto al método del agrupamiento inverso, desarrollaron una modificación para medir la distancia euclidiana, siendo de la siguiente forma:

$$S_A = \frac{X_A - \bar{X}}{s} \quad (7.24)$$

donde S_A es el valor estandarizado del descriptor de la cuenca A, X_A es el valor no estandarizado del descriptor de la cuenca A, mientras que \bar{X} y s son la media y desviación estándar para el descriptor de la cuenca A calculado para todas las cuencas del conjunto de datos.

$$d_{UG} = \left[\sum_{A=1}^N (S_A^U - S_A^G)^2 \right]^{\frac{1}{2}} \quad (7.25)$$

donde d_{UG} es la distancia euclidiana entre la cuenca no aforada U en el centroide de la región y la cuenca aforada G , mientras que S_A^U y S_A^G es el valor estandarizado CD para el descriptor A para la cuenca no aforada U (o para la cuenca aforada G).

Evaluaron la homogeneidad mediante las curvas de Andrews a partir del coeficiente de determinación, estimaron los parámetros del modelo de las cuencas no aforadas por medio de promedios ponderados.

Los autores identificaron dos ventajas, la primera consiste en que si hubo una reducción de la cantidad de descriptores con el algoritmo desarrollado en comparación con las pruebas de prueba -error convencionales. La segunda ventaja consiste en el conocimiento de la influencia, positiva o negativa de la adición de cada descriptor en cada paso iterativo. La primera desventaja se refiere a que la cantidad óptima de descriptores en comparación con el método de prueba error óptimo no es similar, la segunda desventaja consiste en que de forma arbitraia, se elige la jerarquía de importancia de los tipos de descriptores empleados, llegando a afectar el resultado final.

Otro método para la selección de características de las cuencas, es el propuesto por Ahn y Palmer (2016), metodología desarrollada para modelos de regresión. La metodología consiste en tres etapas: i) aplicar correlación de Pearson, con un criterio del coeficiente de determinación > 0.7 , ii) Análisis de Componentes Principales, para el decremento del total de las variables descriptivas, iii) Correlación de Pearson, con un criterio del coeficiente de determinación > 0.2 . Además en el caso de los modelos de regresión, es conveniente el empleo de las pruebas de normalidad de Shapiro-Wilk (Shapiro y Wilk, 1965; Ahn y Palmer, 2016) esta prueba es empleada con el 5% de significancia para determinar la normalidad en las características de la cuenca, como variables independientes. En esta prueba, es necesario realizar tres aproximaciones, exponencial, raíz cuadrada y logaritmos. Se adopta la transformación que más se aproxima a una distribución normal. Finalmente, todas las características de la cuenca transformada se normalizan, habiendo sustraído su valor medio y dividido por su desviación estándar.

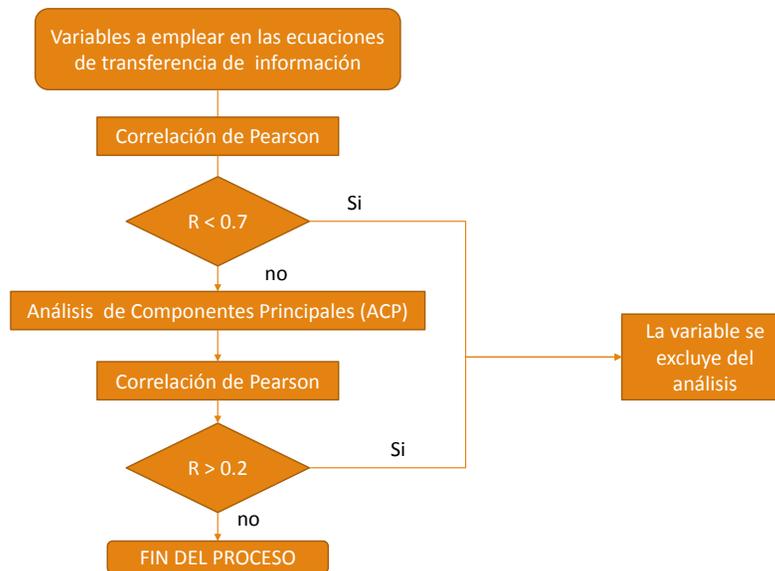


Figura 7-3 Esquema de tratamiento de datos en modelos de regresión (Ahn y Palmer, 2016).

7.5. Normalización de muestras

Si bien, es importante contar con las variables necesarias, es necesario un tratamiento previo a los datos, es decir, una normalización con el fin de remover la dependencia en la escala, en el caso de gastos mensuales es posible expresarlos como un porcentaje de gasto anual, mientras que para las mediciones de las descripciones del caudal los datos se pueden normalizar por medio del área de la cuenca (Haines *et al.*, 1988; Riggs y Harvey, 1990; Sanborn y Bledsoe, 2006).

7.6. Análisis estadístico multivariado

En la transferencia de información hidrológica, es común que el investigador trabaje con datos multivariados, es decir, un conjunto de datos que corresponden a las características tanto fisiográficas como climatológicas, por lo que el empleo de una técnica de Análisis de Componentes Principales es factible. Esta técnica es útil cuando se emplea un gran número de datos (Sanborn y Bledsoe, 2006).

La técnica ACP consiste en lo siguiente: a partir de un conjunto de datos, se crean combinaciones lineales de los mismos y se constituyen los vectores no correlacionados, es decir, los componentes principales cuya característica principal es que son ortogonales entre sí, además que en orden descendente conservan de la mayor a la menor varianza de los datos

originales, de manera que es posible identificar cuáles son las variables más representativas y descartar la multicolinealidad de las mismas (Sanborn y Bledsoe, 2006).

El análisis multivariado es una técnica que se ha empleado en la meteorología a finales de 1940 (Baeriswyl and Rebetez, 1997) y numerosas publicaciones se han hecho desde entonces, con estudios que han empleado datos de precipitación (Baeriswyl y Rebetez, 1997; Soltani y Reza, 2006; Vargas *et al.*, 2009; Satyanarayana y Srinivas, 2011). Las técnicas multivariadas se han identificado como herramientas adecuadas y poderosas al encontrar las regiones hidrológicamente homogéneas o en la clasificación de los datos meteorológicos (Soltani y Reza, 2006). Nathan y McMahon (1990) realizaron un análisis de agrupamiento jerárquico para la predicción de los gastos mínimos en el sudeste de Australia, encontrando que el método de Ward con la medida de similitud basada en la distancia euclidiana al cuadrado y lo proponen como método óptimo en la delimitación de regiones homogéneas (Soltani y Reza, 2006).

En caudales mínimos, para técnicas multivariadas, Laaha y Blöschl en el año 2006 emplearon 31 variables de características fisiográficas para las 325 sub-cuencas en Austria, con áreas de 7 a 963 km², las variables fueron: área de las subcuencas, elevación topográfica, pendiente topográfica, precipitación, geología, uso de suelo, y densidad de corrientes; donde cada una de éstas se relacionaron a detalle. La elevación topográfica es representada por la altitud de la estación de aforo, máxima altitud, rango de altitud, y altitud media. La pendiente topográfica, está representada por la pendiente media en porcentaje y por el porcentaje leve de pendiente, pendiente moderada y pendiente profunda. La precipitación representada por el promedio anual de la precipitación, la precipitación promedio en época de verano y en época de invierno. Mientras que para la geología se obtuvieron distintas variables dependiendo del tipo de formación geológica en la región, lo mismo para la variable del uso de suelo, donde la presencia de zonas urbanas, de agricultura, agricultura permanente, matorral, bosque, zonas rocosas, humedales, cuerpos de agua y zona glaciar se consideraron en el análisis.

Las técnicas multivariadas se han destacado por ser herramientas adecuadas y poderosas para encontrar regiones homogéneas y clasificar datos meteorológicos, como la lluvia (Soltani y Reza, 2006).

Nobert *et al.* (2014) emplearon el método de ACP para reducir la cantidad de variables, con ello obtuvieron la cantidad de variables importantes como perímetro, elevación media, área y lluvia media anual, como parámetros descriptivos en la cuenca del lago Victoria en Kenia,

En la investigación realizada por Baeriswyl y Rebetz (1997), obtuvieron 5 componentes principales cuyo eigenvalor es más grande de 1. Por lo que aplicaron el criterio de Kaiser que consiste en retener solo los factores con eigen valores superiores a uno.

El estudio de Matteau *et al.* (2009) surgió como continuación del interés que inició en el año de 1990 por la restauración de ríos a causa de las actividades antrópicas en Québec. Por lo que los autores proponen (1) La selección de variables hidrológicas que son modificadas por presas y (2) La identificación de los factores que influyen en la magnitud de los cambios hidrológicos. Los autores resolvieron esta problemática por medio de técnicas ACP así como Técnicas de Correlación Canónica. Obtuvieron que las variables que son principalmente afectadas por las presas son: por el primer componente principal con las corrientes de verano e invierno, tiempo de registro del caudal medio mensual máximo y la variabilidad estacional de las corrientes, mientras que para el segundo componente principal las variables obtenidas fueron la variabilidad anual de las corrientes mensuales y el promedio mínimo mensual de corrientes; y a través del análisis de correlación canónica la identificación de los efectos que son alterados principalmente por la acción de presas, debiéndose al régimen regulado y en un grado menor el tamaño de la cuenca.

7.6.1. Marco teórico de un ACP

El análisis de componentes principales (ACP) es un procedimiento estadístico esencialmente descriptivo, su objetivo principal es extraer (presentar) de una manera gráfica, el máximo de información contenida en una matriz de datos K , Pearson, en 1901 fue el primer investigador en desarrollar las nociones básicas de este método. Basado en dichos estudios, Hotelling, en 1933 fue el primero en publicar sobre el ACP. Sin embargo, la utilización del ACP es reciente, fue hasta 1980 con la ayuda de las micro-computadoras, cuando este procedimiento fue verdaderamente utilizado. La matriz de datos está formada por líneas, que corresponden a los sitios de medición (estaciones) y por columnas, que contienen las características o variables de medición (precipitación, uso del suelo, localización, cobertura vegetal, etc.). Estas variables deben representar de la mejor manera posible el fenómeno que se estudia.

Actualmente el ACP tiene muchas aplicaciones. Por ejemplo, se emplea para seleccionar las características (variables hidrológicas) que serán empleadas en un análisis regional. También para verificar, en forma cuantitativa, la correlación entre las diferentes características fisiográficas empleadas en el estudio. Incluso para formar grupos de variables que siguen un mismo patrón de comportamiento. En una forma más precisa, el ACP, permite conocer la estructura de las variables, cuáles de ellas están asociadas entre sí, cuales tienen la misma dirección o cuáles son los grupos que pueden formarse con ellas.

El principio de este procedimiento se basa en la idea de que las p variables cuantitativas de una matriz de datos; pueden ser representados en un espacio de p dimensiones. Como es lógico, no es posible crear ni visualizar dicho espacio. El ACP permite entonces, encontrar los espacios visibles, con las dimensiones más pequeñas; en donde las variables utilizadas pueden fácilmente ser visualizadas. Los más comunes, las rectas (una dimensión) o los planos de dos dimensiones. Dichos espacios visibles dependen también, del ángulo y la dirección desde las cuales son vistos.

Las rectas y los planos realizados en el ACP no se realizan con los valores iniciales de las variables; sino con los “índices sintéticos”. Estos índices se obtienen a partir de lo que se denomina en estadística, una combinación lineal de los valores iniciales. Empleando todos los índices posibles, el ACP busca aquellos que permitan visualizar de la mejor manera posible todos los sitios de medición; es decir, busca aquellos índices que maximicen la varianza en las series cronológicas de cada sitio. A este índice de maximización se le llama: “primer componente principal” el cual explica gran parte de la variación total, en los diferentes sitios de medición.

Enseguida se busca un segundo componente principal para complementar al primero. El proceso continua hasta encontrar los p componentes principales. Cabe señalar que el contenido de información en cada uno de los componentes, es cada vez menos preciso. Se puede decir que la fase esencial del ACP, consiste en transformar las p variables cuantitativas iniciales, todas más o menos correlacionadas entre ellas; en p nuevas variables cuantitativas no correlacionadas entre sí, llamadas componentes principales.

La transformación de las p variables cuantitativas iniciales que de alguna forma están correlacionadas entre sí; en las nuevas p variables cuantitativas, no correlacionadas, llamadas

componentes principales; comprende las siguientes dos etapas. Ya que los resultados del ACP son influenciados por el orden de magnitud de las variables. Las variables con mayor peso aparecerán de una manera más dispersa. Para remediar esta situación la primera etapa consiste en expresar las variables en las mismas unidades, si esto no es posible, se recomienda dar la misma importancia (peso) a las diferentes variables; es decir reducirlas dividiendo entre su desviación estándar. Este análisis puede emplear la matriz de varianza-covarianza o en su caso la matriz de correlación

$$\text{Estaciones} \begin{bmatrix} x_1, x_2, x_3, \dots, x_p \\ \text{Valores iniciales} \end{bmatrix} \Rightarrow \begin{bmatrix} z_1, z_2, z_3, \dots, z_p \\ \text{Componentes principales} \end{bmatrix}$$

La segunda etapa consiste en realizar una operación matemática que se le conoce con el nombre de “diagonalización”. Este procedimiento encuentra los coeficientes de las combinaciones lineares de las p variables iniciales; denominados “vectores propios”. Finalmente la varianza de cada una de las estaciones, sobre los ejes principales, estará dada por los valores de

$$\begin{bmatrix} x_1, x_2, x_3, \dots, x_p \\ \text{Matriz de correlación o} \\ \text{matriz de varianza-covarianza} \end{bmatrix}$$

[Diagonalización]

$$\begin{bmatrix} x_1, x_2, x_3, \dots, x_p \\ \text{Matriz de vectores propios} \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \lambda_p \end{bmatrix}$$

[Matriz diagonal de valores propios]

Para observar las estaciones (individuos), basta construir los planos a partir de los componentes principales que contengan la mayor cantidad de información. Lógicamente el primer plano a analizar será el plano de los ejes principales 1 y 2. Por definición y por construcción este plano es que contiene la mayor cantidad de información visible; es decir es el plano principal. Según la información obtenida a partir de este plano, será necesario o no,

examinar los siguientes planos complementarios. Sin embargo, no es suficiente examinar la distribución de las estaciones en estos planos; es importante también, identificar y verificar la repartición de las variables, es decir la correlación entre las variables de medición.

7.7. Comentarios y recomendaciones

El análisis previo de datos es un paso fundamental en la delimitación de regiones homogéneas. El analista debe recopilar información descriptiva de la zona de interés y estadística relacionada con la hidrometría. La primera se logra principalmente mediante datos de fisiografía y climatología, mientras que, para la segunda se toman como fuente de información los registros históricos hidrométricos de los aforos, este capítulo se enfocó en el tratamiento aplicado a estos registros. Para que una estación de aforo ingrese en el análisis regional, es necesario que cumpla con una longitud de registro aceptable, que puede ser como mínimo de 15 años.

Mediante el análisis previo de datos es posible conocer el comportamiento de las variables, así como la identificación de valores atípicos en las series de registros, ya sea mediante métodos gráficos o métodos estadísticos, para un posterior análisis de consistencia y de calidad de los mismos, que permitan la identificación de tendencias en las series.

Un primer análisis puede realizarse por el método gráfico de bloque de cajas que permite identificar las principales características de cada una de las muestras, al visualizar la media, valor máximo y mínimo y la dispersión de cada muestra respecto a las demás.

En la detección de valores atípicos de los registros se determinan a partir de rangos intercuartílicos, mientras que la prueba de Mann-Kendall es recomendable en el estudio de la consistencia de los datos. Para la calidad de los registros, se deben estudiar cuatro aspectos, los valores atípicos, las pruebas de normalidad, de homogeneidad, y de estacionariedad, por lo que se recomienda la aplicación de la prueba de Shapiro-Wilk, el test de Von-Neumann, la prueba de Mann-Whitney, y la prueba de Mann-Kendall, respectivamente.

Una vez realizado el análisis a las series históricas, posteriormente se trabajan los datos hidrométricos en conjunto con los datos fisiográficos, morfométricos, climatológicos, geológicos, entre otros, que estén relacionados con el objeto de estudio. Si bien, es importante contar con las variables necesarias, es necesario un tratamiento previo a los datos, es decir,

una normalización con el fin de remover la dependencia en la escala. La normalización de datos es un paso previo fundamental al ACP, para que las variables más importantes del estudio no se vean afectadas por la variabilidad en la escala de medición.

8. ANÁLISIS REGIONAL DE FRECUENCIAS ARF

Cuando la información histórica de la variable de interés está disponible en los registros históricos, es aplicable un Análisis de Frecuencias en el sitio (AF) que permita al hidrólogo conocer la magnitud del evento de diseño dado un periodo de retorno. Si por el contrario, se presenta ausencia en la cantidad y calidad de los registros de las variables, es necesario aplicar un Análisis Regional de Frecuencias (ARF), proceso que además requiere de la estimación de las características descriptivas de la zona de estudio (Mishra *et al.*, 2009) bajo un enfoque de Momentos L (LM), por lo que la metodología de ARF-LM ha cobrado importancia (Lilienthal *et al.*, 2018).

El Análisis Regional de Frecuencias (ARF) se ha empleado en la estimación de variables ya sea precipitación o caudal, sobre todo en zonas semiáridas y áridas (Hosking y Wallis, 1997; ICI-RAFT, 2013; Núñez *et al.*, 2016) y desde su divulgación (Hosking y Wallis, 1997) se ha combinado esta técnica bajo el enfoque de Momentos L (ARF-ML).

Las ventajas al emplear un análisis regional de frecuencias (ARF) contra un análisis de frecuencias (AF) ya se han demostrado anteriormente (Hosking y Wallis, 1997; Mishra *et al.*, 2007; Núñez *et al.*, 2010). Mishra *et al.* (2007) realizaron un estudio para representar la variabilidad de L-asimetría con el tamaño muestral, obteniendo una reducción en la estimación en la variabilidad al aumentar el número de observaciones (Núñez *et al.*, 2010). Núñez *et al.* (2010) realizaron una comparación gráfica de cuatro sitios con un ajuste de distribución Gumbel en el sitio, se presentaron cola inferior y superior totalmente insesgadas además de la presencia de valores atípicos; en un análisis posterior realizaron un ARF donde aplicaron un ajuste de distribución de probabilidades por región homogénea, demostrando un mejor ajuste de los datos a la distribución teórica, de esta forma Núñez *et al.* (2010) ejemplificaron la ventaja expuesta previamente por Hosking y Wallis (1997).

Esta técnica es útil en las diversas etapas del análisis regional, desde la identificación de valores atípicos o discordantes hasta la predicción de caudales por medio de una distribución de frecuencias regional. Este método se apoya de las técnicas ya desarrolladas en delimitación de regiones homogéneas, así como en métodos de transferencia de información hidrológica, como el método de la avenida índice. Por lo que se podría decir, que este método

es integral a los procesos aislados del análisis regional, ya que complementa y permite verificar cada proceso del análisis.

La superioridad al emplear el método de Momentos-L en vez de momentos regulares, radica en que Momentos-L es una técnica más robusta en cuanto a valores extremos, así como la facilidad en la caracterización de un gran número de distribuciones, además que cumplen con la característica de ser menos susceptibles al sesgo en su estimación (ICI-RAFT, 2013). El enfoque de Momentos-L con las distribuciones ajustadas a las series de datos, tienen un menor error que el método convencional de Momentos, lo anterior radica en la forma en la que se majean los datos, éstos deben ser ordenados linealmente y “ponderados de tal manera que imitan los momentos convencionales y sus relaciones” (ICI-RAFT, 2013). En el Tabla 8 1se muestra la simbología empleada en los dos tipos de metodologías.

Tabla 8-1. Tabla de simbología empleada en Momentos y Momentos-L (Hosking y Wallis, 1997).

DESCRIPTIVA	MOMENTO POBLACIÓN	MOMENTO MUESTRA	L-MOMENTO POBLACIÓN	L-MOMENTO MUESTRA
UBICACIÓN	μ	\bar{x}	λ_1	l_1
ESCALA	σ	s	λ_2	l_2
COEFICIENTE DE VARIACIÓN	C_v	C_v	τ	t
ASIMETRÍA	γ	g	τ_3	t_3
CURTOSIS	κ	k	τ_4	t_4

Si se empleara en una muestra pequeña, la metodología de Momentos resultaría una limitación algebraica de los estadígrafos como el sesgo y la curtosis, estos dos estadígrafos son empleados para establecer la proximidad de los valores observados en diferentes tipos de distribuciones, por lo que la ventaja de emplear Momentos-L por medio de la combinación lineal de los elementos de una muestra ordenada proporciona una distribución de frecuencias mejor ajustada a los datos observados (Naranjo-Toledo, 2011). Debido a que por medio de un método convencional de ajuste como el método de momentos o máxima verosimilitud son sensibles a datos extremos en la muestra, lo que ocasiona que con éstos se obtengan altos errores en la estimación de cuantiles (Núñez *et al.*, 2016).

8.1. Introducción a Momentos-L

La metodología de Momentos-L es un enfoque robusto en el análisis regional del análisis de frecuencias de variables climáticas e hidrológicas. Momentos-L son combinaciones lineales de estadísticos de orden, que se utilizan para resumir una distribución teórica de una muestra observada de una variable aleatoria (X). Hosking y Wallis (1997) definieron Momentos-L como funciones lineales de probabilidad de momentos ponderados (PWM's, por sus siglas en inglés), los cuales son robustos para los datos atípicos y poco sesgados en muestras pequeñas (Sarhadi y Modarres, 2011; Dodangeh *et al.*, 2014), por lo que se obtienen estimaciones más precisas, es decir, estimaciones más cercanas a los valores observados (Hosking y Wallis, 1997).

Las diferentes relaciones de Momentos-L se pueden definir como la medida de tendencia central, una medida de la escala y la dispersión, también denominado como coeficiente de variación, L-CV, la medida del sesgo, también denominado coeficiente L o L-CS, por último la curtosis, denominado L-coeficiente de kurtosis L-CK (Sarhadi y Modarres, 2011), por lo que una de las ventajas de este método es que proporciona la distribución de probabilidad regional por medio de L-asimetría y L-kurtosis (Dodangeh *et al.*, 2014).

Esta técnica ha cobrado popularidad desde la divulgación del método sobre todo en la identificación del tipo de distribución regional, así como en la prueba de las estaciones con datos atípicos y la homogeneidad regional por medio del diagrama de Momentos - L (Malekinezhad *et al.*, 2011).

Hosking y Wallis (1997) también obtuvieron tres estadísticos útiles en (i) la estimación del grado de homogeneidad de un grupo de sitios (medición de la heterogeneidad, H_i); (ii) la determinación de un sitio inusual en un grupo (medición de discordancia, D_i) y (iii) la identificación de la función de distribución regional (Z^{DIST}) (Sarhadi y Modarres, 2011).

El método estándar ARF-LM (Hosking y Wallis, 1997) sugiere la delimitación de regiones por métodos relacionados con las características descriptivas, como el método de clasificación jerárquica ascendente, y un posterior análisis de la verificación de regiones basado en las características estadísticas de las zonas delimitadas (Malekinezhad *et al.*, 2011).

La cantidad de elementos por región no debe exceder de 20 sitios, ya que el enfoque ARF-LM decrece en exactitud con más de 20 sitios en sus medidas estadísticas (ICI-RAFT, 2013).

Cuando se forman regiones homogéneas, se asume que dentro de la región las estaciones comparten la misma función de distribución de probabilidades, sin embargo no es factible delimitar regiones, como primera instancia, con base en su función de probabilidades ya que las características en el sitio con frecuencia no son confiables, por ello la recomendación consiste en emplear características físicas e hidrometeorológicas en la formación de regiones, postergando la estadística del sitio para verificar la homogeneidad en la región (ICI-RAFT, 2013).

La metodología permite emplear un total de 13 diferentes distribuciones (Hosking y Wallis, 1997) las cuales están constituidas por distribuciones de dos y tres parámetros. Las distribuciones de tres parámetros, toman L-asimetría como un parámetro, para la salida de valores L-kurtosis en función de L-asimetría. Esta función se puede representar como una curva en un gráfico de L-asimetría y L-kurtosis. La estadística Z de Hosking y Wallis, para la bondad de ajuste, se calcula sólo para distribuciones de tres parámetros. Se basa en la diferencia entre la L-kurtosis pronosticada de cada distribución y el valor en los datos; de forma que trazando promedios regionales como puntos en el mismo gráfico L-skewness / L-kurtosis da un análogo gráfico de esta estadística (ICI-RAFT, 2013).

La técnica de Momentos-L ha sido empleada en datos de intensidad de precipitación con base en curvas IDF (Trefry *et al.*, 2006), en datos de temperatura media del aire (Sarmadi y Azmi, 2016), caudal pico (Rao y Hamed, 1997), caudal mínimo (Dodanghted *et al.*, 2014), en datos de precipitación máxima anual (Rahman *et al.*, 2013), en análisis de sequías (Eslamian *et al.*, 2012; Feng *et al.*, 2014; Núñez *et al.*, 2016), en índices de sequía y evapotranspiración bajo enfoque de redes neuronales (Eslamian *et al.*, 2012) también esta técnica se ha empleado en datos de precipitación diaria, bajo enfoque multifractal (García-Marín *et al.*, 2015) y en el método de la avenida índice (Malekinezhad *et al.*, 2011; Wazneh *et al.*, 2013).

A continuación se describen algunos de los estudios más relevantes en los que se empleó la metodología regional de Momentos-L.

Para determinar las distribuciones regionales de probabilidad de los gastos máximos Saf (2009) empleó el método de avenida índice con Momentos-L. La diferencia de este método, es decir, la innovación de esta metodología a comparación de las demás es la prueba de tendencia que aplican en primera instancia a todo el conjunto de datos, con una prueba llamada Mann-Kendall y con ello se estudia la influencia de las presas en la tendencia de las series, además para la evaluación del modelo se aplicó una simulación de Montecarlo por medio de las bases de los datos del error medio cuadrático relativo y del error en la tendencia. El autor realizó un estudio para 45 estaciones aforadas en Turquía para dos cuencas fluviales, Río Kucuk y Río Buyuk, el período mínimo de análisis fue de 10 años. Para evaluar el desempeño del modelo se realizó una simulación de MonteCarlo donde se compararon el error medio cuadrático y la tendencia, además para la estimación de avenidas asociadas a distintos periodos de retorno, se realizó una relación entre el pico medio anual y el área de captación con el análisis regional de frecuencias para los sitios aforados.

En estudios de sequía Dodangeh *et al.* (2014) emplearon un análisis multivariado de agrupamientos y Momentos-L para datos de caudal mínimo por medio de un análisis de frecuencias, mientras que el enfoque bayesiano lo emplearon para estimar la incertidumbre de los parámetros del modelo y las estimaciones del caudal mínimo.

El objetivo principal de Feng *et al.* (2014) fue analizar la precipitación extrema después de eventos de sequía en magnitud y distribución espacial, empleando datos de precipitación diaria, encontrando que la precipitación es más intensa en el mes siguiente a una sequía, incluso con mayor intensidad que los valores medios a largo plazo.

8.2. Etapas del Análisis Regional de Frecuencias basado en Momentos-L

El procedimiento a seguir para la aplicación del análisis regional de frecuencias basado en los Momentos-L contempla generalmente cuatro pasos (Hosking y Wallis, 1997), los cuales son:

- 1) Análisis de datos (preparación y revisión de datos).
- 2) Identificación de regiones homogéneas.
- 3) Selección de la distribución de frecuencia regional.
- 4) Estimación de cuantiles (a nivel regional y/o de estación).

Una condición indispensable para desarrollar este análisis es que existan datos disponibles de un número significativo (al menos 5) de estaciones para la estimación de cuantiles a nivel regional y a nivel de las estaciones (Álvarez *et al.*, 1999; Naranjo-Toledo, 2011). Además, la cantidad de elementos por región homogénea no debe exceder de 20 sitios, ya que el enfoque ARF-LM decrece en exactitud con más de 20 sitios en sus medidas estadísticas (ICI-RAFT, 2013).

8.2.1. Revisión de datos

En un ARFC el análisis preliminar de datos consiste en verificar los supuestos de la metodología, debiendo analizar la homogeneidad de las muestras para asumir que la variable física de interés pertenece a un modelo probabilístico estacionario a través del tiempo, donde los datos no presentan correlación (Hosking y Wallis, 1997; Naranjo-Toledo, 2011)

Cuando el analista recolecta los datos las variables de interés, existen dos tipos de error que pueden presentarse, el primero está asociado a los valores de las muestras de datos que sean incorrectos, ya sea por el registro o el traspaso de los mismos. Mientras que el segundo error está relacionado con las condiciones en que los datos son obtenidos cambien a través del tiempo, como la ubicación de los instrumentos de medición, en los cuales puede presentarse alguna falla en el equipo (Hosking y Wallis, 1997; Naranjo-Toledo, 2011).

La verificación de datos consiste en el análisis individual de las estaciones de interés para la identificación de valores atípicos, también denominados *outliers*, así como la tendencia y cambio en los datos (Hosking y Wallis, 1997).

Además de la revisión común de los datos como la curva masa doble y la curva de cuantiles *QQ-plot*, Hosking y Wallis (1997) identificaron revisiones que se deben realizar a todo estudio que involucre el análisis estadístico: análisis individual de datos identificando valores repetidos o errores en la transcripción de datos, además de la revisión de la tendencia de las observaciones.

8.2.1.1. Obtener Momentos-L

La metodología de Momentos-L es un enfoque robusto en el análisis regional del análisis de frecuencias de variables climáticas e hidrológicas. Momentos-L son combinaciones lineales de estadísticas de orden, que se utilizan para resumir distribución teórica de una muestra

observada de una variable aleatoria (X). Los pesos ponderados se definen por medio de la siguiente ecuación (Greenwood *et al.*, 1979):

$$\beta_r = E\left\{X[F(X)]^r\right\} \quad (8.1)$$

Donde $F(X)$ es la función de distribución acumulada de X , y β_r es el r-ésimo orden de PWM calculado como:

$$\lambda_1 = \alpha_0 = \beta_0 \quad (8.2)$$

$$\lambda_2 = \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 \quad (8.3)$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (8.4)$$

$$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (8.5)$$

De forma general se tiene que:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^r P_{r,k}^* \alpha_k = \sum_{k=0}^r P_{r,k}^* \beta_k \quad (8.6)$$

Las diferentes relaciones de Momentos-L se pueden definir como λ_1 , como la medida de tendencia central, $\tau = \lambda_1 / \lambda_2$ es una medida de la escala y la dispersión, también denominado como coeficiente de variación, L-CV, la relación $\tau_3 = \lambda_3 / \lambda_4$ es una medida del sesgo, también denominado coeficiente L o L-CS, por último siendo la kurtosis $\tau_4 = \lambda_4 / \lambda_2$, denominado como L-coeficiente de kurtosis L-CK.

8.2.1.2. Medidas de discordancia

El concepto de medidas de discordancia fue concebido como aquellas estaciones que difieren al resto en un análisis regional, por lo que los cocientes de Momentos-L están alejados del conjunto de estaciones, esta definición se representa mediante una gráfica, donde una nube de puntos se concentra y está delimitada por elipses, en las cuales los datos atípicos se encuentran fuera de las mismas (Hosking y Wallis, 1997).

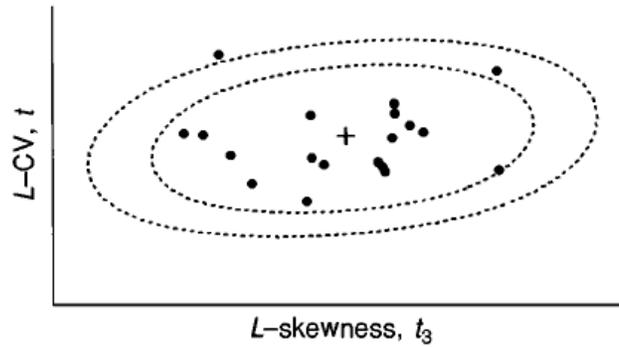


Figura 8-1. Esquema de medidas de discordancia (Hosking y Wallis, 1997)

Las medidas de discordancia que se emplean en el análisis regional se efectúan en dos etapas, la primera consiste en el análisis exploratorio preliminar de datos, para así identificar los datos atípicos, así mismo, después de la delimitación de regiones homogéneas un segundo análisis de discordancia es aplicado, donde si aún existen estaciones discordantes éstas deberán ser removidas del grupo al que fueron asignadas y agruparlas en uno diferente (Hosking y Wallis, 1997).

Las medidas de discordancia son empleadas con el fin de descartar datos atípicos considerando aquellas estaciones que son altamente discordantes de la totalidad del grupo de datos (Dodangeh *et al.*, 2014). Las medidas de discordancia son de la siguiente forma (Hosking y Wallis, 1997):

$$D_i = \frac{1}{3} (\mathbf{u}_i - \bar{\mathbf{u}})^T \mathbf{S}^{-1} (\mathbf{u}_i - \bar{\mathbf{u}}) \quad (8.7)$$

donde \mathbf{u}_i es el vector que se forma a partir de los coeficientes de variación, asimetría y curtosis, es decir, L-CV, L-asimetría y L-curtosis, respectivamente, para cada sitio i :

$$\mathbf{u}_i = [\tau^{(i)} \quad \tau_3^{(i)} \quad \tau_4^{(i)}]^T \quad (8.8)$$

mientras que el vector $\bar{\mathbf{u}}$ es el promedio del conjunto de coeficientes de la siguiente forma:

$$\bar{\mathbf{u}} = \frac{1}{N} \sum_{i=1}^N \mathbf{u}_i \quad (8.9)$$

donde \mathbf{S} se define como la matriz de sumas de cuadrados y productos cruzados

$$S = \sum_{i=1}^N (\mathbf{u}_i - \bar{\mathbf{u}})(\mathbf{u}_i - \bar{\mathbf{u}})^T \quad (8.10)$$

El primer enfoque para definir los límites de un sitio como discordante o no fue el propuesto por Hosking y Wallis (1993), posteriormente se identificaron los valores críticos (Hosking y Wallis, 1997) para regiones formadas a partir de 5 elementos y el valor crítico de 3 se asignó para toda región formada con más de 15 elementos (Tabla 8-2).

Tabla 8-2. Valores críticos para las medidas de discordancia (Hosking y Wallis, 1997)

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

8.3. Formación de regiones homogéneas

En la metodología del ARF-LM (Hosking y Wallis, 1997) se propone el empleo de métodos de delimitación basados en las características del sitio, y reservar los estadísticos para verificar la homogeneidad de las regiones formadas.

La agrupación de estaciones aforadas basadas en el estadístico de coeficiente de variación no es recomendable debido a tres principales problemas (Hosking y Wallis, 1997), donde la agrupación de regiones homogéneas se realizaría en función del mismo L-CV (coeficiente de variación):

- No habría un cambio importante en cuanto a la magnitud del coeficiente de variación regional en comparación con el L-CV por cada estación.
- Al agrupar los L-CV, se podrían llegar a agrupar valores atípicos
- Se deben formar los grupos homogéneos y verificar la homogeneidad con los mismos datos.

La formación de regiones homogéneas se realiza mediante un proceso subjetivo u objetivo, el primero se refiere a la partición de zonas por alguna razón ya sea administrativa o de

gestión de recursos (Rao y Hamed, 1997; Saf, 2009), mientras que el segundo tipo se refiere al empleo de los diversos métodos iterativos que existen en la literatura para delimitación de regiones basados en la proximidad espacial o en la similitud física (Hosking y Wallis, 1997).

Dentro de los métodos propuestos para la formación de regiones homogéneas en un ARF-LM (Hosking y Wallis, 1997) se encuentran el método de agrupamiento con ligadura promedio y el método de Ward (Hosking y Wallis, 1997) por mencioanr algunos métodos, pudiendo llegar a ser cualquier otro método de agrupación. Los autores proponen un máximo de 20 sitios por región homogénea, debido a que con el incremento de sitios decrece la precisión en la estimación de cuantiles.

Una vez delimitadas las regiones, es factible realizar modificaciones en la delimitación de regiones homogéneas mediante siete tipos de ajustes útiles (Hosking y Wallis, 1997), esto se aplica siempre y cuando existan elementos discordantes en la región o no se cumpla la homogeneidad en la agrupación de los elementos:

1. Mover uno o más sitios de una región a otra
2. Eliminar uno o más sitios del conjunto de datos
3. Subdividir la región
4. Separar una región, reasignando los sitios a las regiones restantes
5. Unir una región con otra
6. Unir dos o más regiones y redefinir los grupos
7. Obtener más datos y redefinir los grupos

El procedimiento recomendado inicialmente en el ARF-LM (Hosking y Wallis, 1997) es un proceso de delimitación por el método de Ward, sin embargo en la actualidad ya se han aplicado otros métodos para la delimitar a las regiones homogéneas, aunque numerosos estudios han seguido la metodología original (Malekinezhad *et al.*, 2011), mientras que otros estudios se han basado en propuestas de mejora o adaptaciones al método. Rao y Srinivas (2006) delimitaron por métodos híbridos con método de Ward y K-medias, mientras que Sarhadi y Modarres (2011) emplearon técnicas relacionadas con índice estacional. Sarmadi y Azmi (2016) emplearon dos pasos en la delimitación, primero bajo un enfoque de análisis factorial seguido de un análisis de agrupamiento.

Posterior a la delimitación de regiones homogéneas, éstas se verifican por los métodos relacionados al ARF (heterogeneidad y prueba de Langbein), por métodos estadísticos

(pruebas de calidad de cocientes estadísticos), así como métodos gráficos (curvas de Andrews).

8.3.1. Medidas de heterogeneidad propuestas por Hosking y Wallis

Varios autores han empleado las medidas de pruebas de homogeneidad basadas en la metodología de Momentos-L. Las medidas de discordancia y heterogeneidad se han aplicado a distintos métodos de delimitación en la evaluación de los grupos formados, algunos de éstos se han realizado por el método de K-medias en datos de sequías (Yoo *et al.*, 2012), en agrupación por método de Ward con índices de estacionalidad de precipitación (Malekinezhad *et al.*, 2011), con índices de estacionalidad de inundación (Sarhadi y Modarres, 2011) y con datos fisiográficos y climáticos de la cuenca (Burn *et al.*, 1997).

En la definición de la medida de heterogeneidad (Hosking y Wallis, 1997) se supone una región con N sitios, el sitio i tiene una longitud de registro de n_i observaciones, con los Momentos-L *ratios* o cocientes $t^{(i)}$, $t_3^{(i)}$ y $t_4^{(i)}$. Se obtienen los promedios regionales ponderados y asociados a una longitud de registro: t^R , t_3^R , y t_4^R , que corresponden a los cocientes de L-CV, L-Asimetría y L-Curtosis, respectivamente.

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i} \quad (8.11)$$

La desviación estándar de la muestra de L-CV se calcula de la siguiente forma:

$$V = \left\{ \frac{\sum_{i=1}^N n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^N n_i} \right\}^{1/2} \quad (8.12)$$

Con una distribución kappa de cuatro parámetros se ajustan los promedios regionales de los cocientes de Momentos-L.

Con el ajuste obtenido se simula para una muestra de N_{sim} para una región con N sitios. V se calcula para cada región homogénea y se determinan la media (μ_v) y la desviación estándar (σ_v). Con lo anterior se calcula la medida de heterogeneidad H

$$H = \frac{(V - \mu_v)}{\sigma_v} \quad (8.13)$$

La medida de homogeneidad entonces se obtiene al sustituir la medida V obtenida a partir de los Momentos-L regionales y la media y desviación estándar de la serie sintética. Por lo que el criterio de homogeneidad se basa en valores establecidos por Hosking y Wallis (1997), así como una posterior modificación de los rangos originales descritos por Wallis *et al.* (2007) (Tabla 8-3):

Tabla 8-3. Valores para medir el nivel de heterogeneidad (Hosking y Wallis, 1997)

Nivel de heterogeneidad	Valor establecido (Hosking y Wallis, 1997)	Valor establecido (Wallis <i>et al.</i>, 2007)
Homogénea	$H < 1$	$H < 2$
Posiblemente heterogénea	$1 < H < 2$	$2 < H < 3$
Heterogénea	$H > 2$	$H > 3$

8.3.2. Otras medidas de homogeneidad

8.3.2.1. Pruebas estadísticas

De este tema se habla en las técnicas de verificación de regiones homogéneas en el capítulo 9 DELIMITACIÓN DE REGIONES HOMOGÉNEAS en el subcapítulo 9.7 Verificación de regiones homogéneas sección 9.7.4 Pruebas estadísticas: cociente .

8.3.2.2. Test de Wiltshire

De este tema se habla en las técnicas de verificación de regiones homogéneas en el capítulo 9 DELIMITACIÓN DE REGIONES HOMOGÉNEAS en el subcapítulo 9.7 Verificación de regiones homogéneas en la sección 9.7.5 Test de Wiltshire.

8.4. Selección de una distribución regional de frecuencias

Este paso se realiza después de la delimitación de regiones homogéneas, donde es necesario encontrar una distribución que se ajuste a los datos y que proporcione cuantiles precisos, por lo que el primer paso es la identificación de las distribuciones que se puedan ajustar a los datos observados.

8.4.1. Pruebas de bondad de ajuste

Las pruebas de bondad de ajuste de una distribución están basadas en los gráficos de Q-Q, chi-cuadrada, Kolmogorov-Smirnov y Momentos. Sin embargo, con la aplicación del ARF-LM es factible la ejecución de pruebas de bondad de ajuste por medio de un estadígrafo de cocientes de Momentos-L lo que permite que la metodología sea más versátil (Hosking y Wallis, 1997).

La definición formal de la prueba de bondad de ajuste se describe de la siguiente forma (Hosking y Wallis, 1997): Se asume que la región tiene N sitios, por lo que el sitio i cuenta con una longitud de registro de n_i observaciones y los cocientes de Momentos - L $t^{(i)}$, $t_3^{(i)}$, $t_4^{(i)}$. Los promedios regionales por cada región homogénea serían los siguientes t^R , t_3^R , t_4^R , que corresponden L-CV, L-Asimetría y L-Curtosis, que se obtienen de la siguiente forma:

$$t^R = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i} \quad (8.14)$$

L-asimetría regional:

$$t_3^R = \frac{\sum_{i=1}^N n_i t_3^{(i)}}{\sum_{i=1}^N n_i} \quad (8.15)$$

L-curtosis regional

$$t_4^R = \frac{\sum_{i=1}^N n_i t_4^{(i)}}{\sum_{i=1}^N n_i} \quad (8.16)$$

Una vez obtenidos los coeficientes regionales de asimetría y curtosis (t_3^R y t_4^R) se grafican en el estadígrafo de cocientes de Momentos-L (Cong *et al.*, 1993) el cual está construido por medio de los puntos coordenados formados por L-asimetría y L-curtosis de la región, con la

relación de diferentes distribuciones de tres parámetros que pueden ser elegidas debido a la cercanía de los coeficientes regionales.

El procedimiento propuesto por Hosking y Wallis (1997) consiste en realizar un ajuste de distribuciones de tres parámetros, como la distribución Logística Generalizada (GLO), Generalizada de valores extremos (GVE), distribución de Pareto, lognormal y Pearson tipo III.

El procedimiento consiste en ajustar cada distribución a los promedios regionales de los Momentos-L, t_3^R , t_4^R , siendo L-CV, L-asimetría y L-curtosis, respectivamente. Se denota como t_4^{DIST} a L-Curtosis de la distribución ajustada, en este caso, *DIST* corresponde a la función de distribución con la cual se haya ajustado, pudiendo ser GLO, GEV, etc.

Se ajustan los cocientes regionales de Momentos-L (t_3^R , t_4^R) a una distribución kappa. Se simula un gran número de muestras, N_{sim} , de la región con N sitios, en cada sitio se desarrolla la distribución kappa y su distribución de frecuencias. De manera que las regiones simuladas con esta distribución, son homogéneas y no hay correlación cruzada o correlación serial, y tienen la característica de tener la misma longitud de registro que la muestra observada. Para la m -ésima región simulada, se calcula el promedio regional de L-asimetría $t_3^{[m]}$ y L-curtosis $t_4^{[m]}$ de la siguiente forma:

$$B_4 = N_{sim}^{-1} \sum_{m=1}^{N_{sim}} (t_4^{[m]} - t_4^R) \quad (8.17)$$

$$\sigma_4 = \left[(N_{sim} - 1)^{-1} \left\{ \sum_{m=1}^{N_{sim}} (t_4^{[m]} - t_4^R)^2 - N_{sim} B_4^2 \right\} \right]^{1/2} \quad (8.18)$$

La bondad de ajuste para cada distribución se calcula de la siguiente forma:

$$Z^{DIST} = \frac{(t_4^{DIST} - t_4^R + B_4)}{\sigma_4} \quad (8.19)$$

De forma que se considera como el mejor ajuste aquella Z^{DIST} cercana a cero, por lo que el valor crítico para descartar distribuciones de frecuencia no aptos son aquellos con la

condición $|Z^{DIST}| \leq 1.64$. Mientras que N_{sim} se simula mediante el método de Monte Carlo (She *et al.*, 2014).

8.5. Estimación de cuantiles

“La relación entre las distribuciones de frecuencia en diferentes sitios es la justificación del análisis regional de frecuencias. Lo anterior permite obtener estimaciones más precisas de los parámetros y cuantiles de la distribución mediante la combinación de datos de distintos sitios, que por el ajuste de distribución de los datos en cada sitio por separado” (Hosking y Wallis, 1997).

En la estimación de cuantiles se emplean métodos que ha resultado eficientes, como el método de la Avenida índice y el método de Máxima Verosimilitud, la ventaja al emplear el método de la avenida índice radica en la facilidad de su aplicación, además que al calcular las cocientes de Momentos-L se obtienen las características precisas del conjunto de datos regionales (Hosking y Wallis, 1997).

Los supuestos para aplicar el método de la avenida índice consisten en lo siguiente (Hosking y Wallis, 1997):

- No hay dependencia entre las muestras de los distintos sitios que conforman la región
- En los datos del sitio no hay observaciones con dependencia

La premisa de un análisis regional de frecuencias consiste en una vez determinada una región homogénea con su respectiva función de distribución de probabilidades de la variable de interés, todos los puntos dentro de la región homogénea se rigen por dicha curva en mayor o menor medida debido a un factor de escala (Hosking y Wallis, 1997):

$$\hat{Q}_i(F) = \mu_i \hat{q}(F) \quad (8.20)$$

donde $\hat{Q}_i(F)$ se estima por medio de la función de cuantiles regional adimensional (curva de crecimiento regional) estimada mediante el análisis regional de frecuencias y multiplicada por un factor de escala μ_i .

8.6. Comentarios y recomendaciones

El desarrollo de un Análisis regional de frecuencias de crecientes (ARFC) basado en técnicas de Momentos-L ofrece múltiples ventajas, debido a que permite una primera inspección de la calidad de los datos por medio de las medidas de discordancia para todo el conjunto de estaciones aforadas representativas, además que estas medidas pueden aplicarse después de la formación de regiones homogéneas para la verificación de resultados.

Una transferencia de información hidrológica exitosa se debe principalmente a la delimitación de regiones que garanticen la misma respuesta hidrológica, por lo que la agrupación de estaciones aforadas es primordial y con ello la verificación de homogeneidad del conjunto, para ello se describieron las pruebas de Momentos - L, pruebas estadísticas, avenida índice y el test de Wiltshire.

La distribución de frecuencias de la región también se puede seleccionar por medio de la técnica de Momentos-L, ya sea por un método gráfico o por la prueba estadística Z.

Desde la divulgación de la metodología de Momentos - L ésta ha cobrado importancia debido a la versatilidad del método desde la etapa de verificación de calidad inicial de datos, seguido por la verificación de homogeneidad de regiones, así como la selección de la distribución regional.

Por lo que se recomienda realizar en primera instancia un análisis de calidad de datos, posteriormente la verificación de las medidas de discordancia basada en la técnica de Momentos - L a todo el conjunto de estaciones aforadas, y una vez delimitadas las regiones homogéneas aplicar nuevamente esta técnica para verificar la heterogeneidad en cada zona.

9. DELIMITACIÓN DE REGIONES HOMOGÉNEAS

El ARFC implica dos aspectos principales, el primero consiste en la delimitación de regiones homogéneas y la determinación de un método de estimación regional para diferentes frecuencias de inundación, definiendo qué cuencas serán las donantes de información (Ouarda *et al.*, 2006; Malekinezhad *et al.*, 2011; Gado y Nguyen, 2016). En la segunda etapa implica la selección de un modelo probabilístico regional y transferencia de información, donde la distribución Generalizada de Valores Extremos (GVE) ha resultado ser la más apropiada en diversos estudios (Malekinezhad *et al.*, 2011; Nobert *et al.*, 2014). Sin embargo, se debe considerar una etapa inicial, que consiste en el paso fundamental del análisis de las variables para la transferencia de información (Nathan y McMahon, 1990; Caratti *et al.*, 2005) (ver capítulo 6).

Los métodos de delimitación requieren de una medida de similitud, que permita establecer el grado de semejanza entre las estaciones o cuencas (Wazneh *et al.*, 2015).

Se han desarrollado métodos de delimitación para la aplicación de un ARFC, y a partir de ellos se forman extensiones y modificaciones de los mismos (Burn *et al.*, 1997; Rao y Srinivas, 2006; Núñez *et al.*, 2016), los cuales son: i) método de residuales, ii) Análisis de Correlación Canónica ACC, iii) región de influencia iv) Análisis factorial v) Análisis de componentes principales vi) Análisis multivariado, como método de agregación jerárquico con enfoque de índice estacional, y los relacionados a los atributos geográficos y climáticos del área de estudio, vii) lógica difusa, viii) redes neuronales artificiales, como mapas auto-organizados ix) árboles de clasificación y regresión, y x) enfoques híbridos.

Los métodos de agrupación de zonas homogéneas que se han desarrollado recientemente son: análisis de agrupamiento (Burn, 1989; Nathan y McMahon 1990; Nathan y McMahon, 1992; GREHYS, 1996; Ramos, 2001; Yu *et al.*, 2002; Laaha y Blöschl, 2006; Sanborn y Bledsoe, 2006; Soltani y Reza, 2006; Kahya *et al.*, 2008; Lim *et al.*, 2009 ; Kar *et al.*, 2015; Dodangeh *et al.*, 2014), análisis de agrupamiento inverso (Heřmanovský y Pech, 2013); análisis de componentes principales, ACP, (Baeriswyl y Rebetez ,1987; Nathan y McMahon 1990; Nathan y McMahon,1992; Chokmani y Ouarda, 2004; Paris y Zucarelli, 2004; Sanborn y Bledsoe, 2006; Soltani y Reza, 2006; Kahya *et al.*, 2008; Matteau *et al.*, 2009; Singh *et al.*, 2014; Oueslati *et al.*, 2015), patrones de residuales de aproximación (Nathan y McMahon,

1990; Laaha y Blöschl, 2006), índice estacional (Laaha y Blöschl, 2006; Sarhadi y Modarres, 2011), regiones de influencia (GREHYS, 1996; Burn, 1990; Haddad *et al.*, 2012) árbol de clasificación y regresión (Laaha y Blöschl, 2006; Vezza *et al.*, 2010), análisis de correlación canónica, ACC, (GREHYS, 1996; Ouarda *et al.*, 2001; Caratti *et al.*, 2005; Matteau *et al.*, 2009; Tsakiris *et al.*, 2011) y Momentos-L (GREHYS, 1996; Shi *et al.*, 2010; Yang *et al.*, 2010; Dodangeh *et al.*, 2014; Núñez *et al.*, 2016).

De manera general todos los métodos antes mencionados, pueden ser clasificados en tres grupos principales (Schumann, 2011): el primer grupo corresponde a aquellos métodos en los que la agrupación se deba a alguna medida de similitud entre los objetos. El segundo tipo lo forman aquellos métodos en los que se establece una relación entre los estadísticos de la muestra y las características de la cuenca. Mientras que el tercer grupo lo forman aquellos métodos que involucran la geoestadística en lo que se “asume que los estadísticos de las crecientes están correlacionados espacialmente y emplean la correlación espacial para la interpolación de los estadísticos de las crecientes” (Schumann, 2011).

Los procesos de agrupación usados con mayor frecuencia en orden de importancia son: la agrupación con respecto a los valores medios (también denominado, *K-medias*) (Soltani y Reza, 2006 ; Vargas *et al.*, 2009; Ouyang *et al.*, 2010), seguido del vecino más próximo, aglomerado jerárquico y mapas auto-organizados (Ouyang *et al.*, 2010). Recientemente se ha registrado en la literatura algunos de los métodos más populares, seguidos de Ward y *K-medias*, siendo aquellos que consisten en la agrupación de cuencas basadas en la similitud de las características geográficas (Sarhadi *et al.*, 2011). Además, el método basado en la similitud de estacionalidad de crecientes ha ido ganando popularidad sobre todo en zonas áridas y semiáridas, así como en zonas montañosas y de planicie, donde la aplicación de técnicas de delimitación de regiones es compleja (Sarhadi *et al.*, 2011).

Uno de los métodos más populares y ampliamente usados es el método del análisis de aglomerados (Wazneh *et al.*, 2015). En el ARFC uno de los métodos comúnmente empleados es el análisis de clasificación jerárquica ascendente (Burn *et al.*, 1997) en el que se obtienen grupos no contiguos espacialmente. Otros métodos de grupos homogéneos formados que pueden llegar a ser no contiguos espacialmente, es el método de similitud de cuencas y el análisis de agrupamiento (Nathan y McMahon, 1990). El método

de Ward, ha demostrado ser una técnica que representa de forma óptima los patrones espaciales de lluvia (Soltani y Reza, 2006), caudales mínimos (Nathan y McMahon 1992), así como en estudios climáticos (Kahya *et al.*, 2008). El análisis de agrupamiento se ha enfocado en emplear diferentes esquemas para caudales máximos, mínimos y medios, analizando los efectos del agrupamiento en corrientes con regímenes similares en diferentes regiones geográficas (Paris y Zucarelli, 2004; Sanborn y Bledsoe, 2006; Peterson *et al.*, 2011; Dodangeh *et al.*, 2014).

Nathan y McMahon (1992), establecieron una metodología para el análisis de caudales mínimos, técnica que a través del tiempo fue tomada como base en el análisis regional y ha sido empleada en diversas partes del mundo y para distintos datos y objetos de estudio, como en caudales medianos (París y Zucarelli, 2004) y máximos (Álvarez-Olguín *et al.*, 2011) por citar algunos ejemplos. La metodología propuesta por Nathan y McMahon (1992) consiste en siete pasos: 1) recolectar variables empleadas en la predicción de caudales, 2) identificar las variables más relevantes en la predicción de caudales mínimos, por medio de un análisis de regresión paso a paso (*stepwise regression analyses*), 3) ponderar las variables de acuerdo a su importancia relativa, así como dependiendo de la magnitud del coeficiente beta derivado de las ecuaciones de la regresión paso a paso, 4) emplear distintos análisis de agrupamientos con distintos índices de proximidad y métodos de ligadura, 5) graficar los grupos de acuerdo a la metodología de graficación de multi-dimensional, 6) adecuar de grupos de acuerdo a la validación de los métodos gráficos, 7) refinar hasta llegar a la agrupación óptima.

Existen otros métodos menos populares que han ido adquiriendo difusión como métodos óptimos en la delimitación de regiones homogéneas. Como en el caso de redes neuronales (Chang *et al.*, 2010; Sahin y Cigizoglu, 2012; Ay y Kisi, 2014) y mapas auto-organizados (Lin *et al.*, 2005; Astel *et al.*, 2007; Lin y Wu, 2007; Fassnacht y Derry, 2010; Lin *et al.*, 2010; Farsadnia *et al.*, 2014; Nguyen *et al.*, 2015), para desarrollar procesos de delimitación por método de agrupamiento así como aplicaciones en hidrología estocástica (Maul y El-Shaarawi, 1993; Ouyang *et al.*, 2010; Santos *et al.*, 2010; Elsheikh *et al.*, 2013; Caraway *et al.*, 2014; Sarmadi y Azmi, 2016) y agrupamiento de entropía cruzada, CEC, por sus siglas en inglés (Tongal y Sivakumar, 2017).

En el desarrollo de redes neuronales artificiales bajo el enfoque de agrupación de variables se han desarrollado modelos híbridos para el pronóstico de inundaciones (Chang *et al.*, 2010). Sahin y Cigizoglu (2012) desarrollaron un análisis de régimen de precipitaciones y tipos de regiones de clima en Turquía, por medio del método de Ward y métodos neuro-difusos. Ay y Kisi (2014) en el área de planeación y tratamiento de aguas en Turquía aplicaron una combinación del método de K-medias con la metodología de redes neuronales para modelar la cantidad de concentración de demanda química de oxígeno.

El desarrollo de un ARFC con mapas auto-organizados (SOMs, por sus siglas en inglés) parte de un proceso de redes neuronales artificiales, ANN's (por sus siglas en inglés) ha sido práctico y se ha realizado bajo el enfoque de formación de grupos con C-medias difusas, K-medias y método de Ward (Farsadnia *et al.*, 2014). Un SOM se describe como “una técnica efectiva para extraer y visualizar características de la salida de un modelo y para resolver problemas de clasificación de variables” (Lin *et al.*, 2010). Los modelos SOM se han empleado principalmente en elaboración de hietogramas de diseño en sitios no aforados (Lin y Wu, 2007; Lin *et al.*, 2010), en definición de regiones homogéneas para datos de lluvia (Lin *et al.*, 2005) y en datos de telemetría de nieve y equivalencia de agua nieve (Fassnacht y Derry, 2010) en la evaluación de la calidad del agua superficial (Astel *et al.*, 2007) y en calidad de aguas subterráneas (Nguyen *et al.*, 2015).

En el área de la hidrología estocástica, también se han desarrollado estudios apoyados en bases de delimitación regional, Maul y El-Shaarawi (1993) elaboraron un estudio de calidad de agua mediante métodos estocásticos. Elsheikh *et al.*, (2013) desarrollaron el algoritmo K medias en la determinación de grupos en modelos de sub-superficie de flujo. En EUA Caraway *et al.* (2014) generaron muestras estocásticas de series de tiempo por medio del análisis de conglomerados y K-medias en datos de precipitación estacional, con datos en 66 estaciones con 29 años de registro. En estudios de sequía, se han empleado ACP e índices de sequía (Santos *et al.*, 2010; Sarmadi y Azmi, 2016). La investigación de Ouyang *et al.* (2010) fue realizado en China para un conjunto de datos de 1961 a 2000, con datos diarios de caudal, se emplearon cuatro estadísticos descriptivos: descarga media mensual, descarga máxima mensual, amplitud mensual, y desviación estándar mensual. Se empleó K-medias y como segundo enfoque se asumió un segundo tipo de medición entre clúster, el tiempo dinámico de deformación, DTW.

Para un total de 217 aforos en Australia, Tongal y Sivakumar (2017) emplearon una técnica de clasificación de cuencas llamada “Agrupamiento de entropía cruzada CEC” en caudales diarios, este método se basa en el análisis de series de tiempo en tres aspectos: dimensionalidad, complejidad y variabilidad cuyos parámetros de medición son dimensión en inmersión, entropía de la muestra y coeficiente de variación. Los autores realizaron combinaciones lineales y no lineales de los parámetros para explicar el fenómeno de variabilidad el caudal (Tongal y Sivakumar, 2017).

Un paso posterior a la formación de regiones homogéneas consiste en la verificación de la homogeneidad de los grupos formados.

Existen diversas técnicas de verificación de homogeneidad, por ejemplo por métodos gráficos multidimensionales, como en el caso de las curvas de Andrews (Nathan y McMahon, 1990; Paris y Zucarelli, 2004; Lahha y Blöschl, 2006; Heřmanovský y Pech, 2013) siendo un método visual en la verificación de regiones, otro método de revisión consiste en las técnicas de los rostros de Chernoff (Chernoff, 1973).

También existen los métodos estadísticos que permiten validar la homogeneidad regional son aquellos que emplean un cociente estadístico que relaciona la desviación estándar regional y el coeficiente de variación regional, el cual no debe exceder un valor umbral de 0.3 (Nobert *et al.*, 2014), otro método es el test de Langbein, cuando se aplica un método de avenida índice; en la bibliografía también se encuentra el test de Wiltshire (Campos-Aranda, 2010a); otro método empleado en el ARFC y que ha sido ampliamente usado es el método de Momentos L (Hosking y Wallis, 1993; Hosking y Wallis, 1997) éste por medio de dos pruebas, medidas de discordancia y grado de heterogeneidad.

Dentro del ARFC, el método de delimitación de regiones homogéneas por medio de Momentos-L ha sido ampliamente usado, desde su divulgación por Hosking y Wallis (1993) hasta la actualidad (Núñez *et al.*, 2016) ha demostrado ser un método con desempeño eficiente, que toma como bases el método de agrupación de clasificación jerárquica ascendente, es decir, el método de Ward. El método de Momentos-L ha sido ampliamente usado en el ARFC, para estudios relacionados con inundaciones, lluvias extremas, sequías, régimen natural de caudales, así como caudales mínimos (Dodangeh *et al.*, 2014). Básicamente esta técnica tiene las siguientes etapas: 1) Identificación de estaciones

discordantes, 2) Posterior a la aplicación de cualquier método de delimitación, se verifica la homogeneidad por medio de dos pruebas, medidas discordantes y grado de heterogeneidad; 3) Selección de una distribución de frecuencias regional, por medio de una técnica de bondad de ajuste de selección, y 4) Transferencia de información hidrológica por medio del método de avenida índice.

Después de la delimitación de regiones homogéneas, es importante la ubicación de la estación no aforada dentro de alguna de las regiones formadas, para lo cual se recurre a un análisis discriminante. El análisis discriminante se emplea cuando en un análisis regional, las características de las cuencas de cada región tienen una distribución multivariada normal, por lo que la cuenca no aforada se ubica dentro de alguno de los grupos homogéneos formados por medio de un análisis discriminante, empleando las características de la cuenca (Nathan y McMahon, 1990; Sanborn y Bledsoe, 2006). El análisis discriminante construye combinaciones lineales de variables denominadas funciones discriminantes y encuentra a los que se pueden dividir más efectivamente a los grupos predefinidos (Sanborn y Bledsoe, 2006).

A continuación se exponen algunos estudios realizados recientemente:

Métodos como el análisis de agrupamiento, análisis factorial, componentes principales y análisis de correspondencias, están dirigidos a la identificación de asociaciones o grupos de entidades similares, contemplando las particularidades de los datos observados (diferentes escalas de medida, unidades, órdenes de magnitud, etcétera). Además, las diferentes alternativas de representación de los resultados obtenidos (dendogramas, diagramas de cargas, distribución espacial de las marcas, curva de Andrews, etcétera), facilitan la interpretación de los mismos (Paris y Zucarelli, 2004).

Sanborn y Bledsoe (2006) realizaron un estudio en Colorado, Washington, y Oregon, con un total de 84 mediciones de caudal que se emplearon para calcular la gama de regímenes de flujo en la región. Se empleó un GIS para calcular los promedios mensuales de las variables climatológicas (Precipitación, evapotranspiración, nieve, temperatura, y radiación solar). Las variables fisiográficas incluyen la elevación de la cuenca, pendiente, relieve, y propiedades del suelo. Se utilizaron tres herramientas estadísticas, el ACP (para reducir la varianza en los datos), análisis de conglomerados (para identificar distintos tipos de régimen), y análisis

discriminante para predecir el régimen del flujo de un sitio sin aforo (Sanborn y Bledsoe, 2006). La metodología que presentaron los autores se basa en la estratificación de los regímenes de corrientes de las cuencas aforadas, para así clasificar los regímenes de las cuencas no aforadas, en este caso el enfoque de aplicación fue para la predicción del caudal ecológico. Debido al objeto de estudio que implica un bio-monitoreo del caudal, especies acuáticas y caudales ecológicos, los autores emplearon distintos esquemas de agrupamiento que fueron desarrollados empleando caudales máximos, medios y mínimos. Las herramientas estadísticas que emplearon se basan en las siguientes: ACP para el análisis de reducción de variables; Análisis clúster para identificar los distintos tipos de flujo de régimen y por último un análisis discriminante para la asignación del tipo de régimen.

Laaha y Blöschl (2006) realizaron una comparación de cuatro métodos de delimitación de regiones homogéneas, método de patrón de residuales, análisis de agrupamiento ponderado, árboles de regresión y clasificación y el método de similitud estacionalidad de crecientes, estos métodos se compararon en el desempeño de la predicción de caudales mínimos específicos al 95% , donde el método de similitud de estacionalidad de crecientes resultó con un mejor desempeño , ya que los demás métodos resultaron que subestimaron el caudal específico del 95% en las zonas húmedas donde se aplicaron. El método de estacionalidad de regiones obtuvo el 70% de la varianza en la validación cruzada. Los métodos fueron comparados mediante las siguientes pruebas: Análisis ANOVA, bondad de ajuste de los modelos de regresión, la evaluación del desempeño de las regresiones regionales para varios métodos de agrupamiento.

Mishra *et al.* (2009) realizaron un estudio relacionado con la integración de del número de curva (NC) de escurrimiento en la delimitación de regiones homogéneas. Dentro de un ARFC es requerida la estandarización de los atributos, por lo que los autores proponen una metodología sin estandarizar los atributos empleados en la delimitación, con lo anterior lograron al emplear las superficies delimitadas por el NC, en vez de emplear un método de agrupamiento, obteniendo menos datos discordantes y más homogéneos que por el método de agrupamiento.

En cuanto a la hidroclimatología, diversos estudios se han desarrollado para la clasificación de variables, como (Harris *et al.*, 2000) en Gran Bretaña emplearon técnicas multivariadas

para un conjunto de datos de gastos y temperatura medios mensuales, empleando registros de 20 años. Los datos fueron estandarizados, es decir, media cero y asimetría 1. Se empleó la técnica de ligadura promedio, la cual tiende a maximizar la varianza entre-*clusters*, mientras minimiza la varianza intra-*clusters*, el objetivo de los autores fue desarrollar un método para determinar la forma y magnitud de cada clase de régimen, así como de la temperatura del aire, por medio de análisis de aglomerados bajo la técnica de ligadura promedio.

Yokoo y Kazama (2012) identificaron la relación entre las características geográficas de la cuenca y el escurrimiento empleando una regionalización de parámetros, en zonas montañosas de Japón. Los autores analizaron las relaciones regionales entre las características entre los parámetros en un modelo físico, mediante un tanque de almacenamiento y las características geográficas, para después trasladarlas a relaciones regionales entre las características geográficas de una cuenca con el escurrimiento. Los autores también emplearon un modelo lluvia - escurrimiento físicamente distribuido para analizar la sensibilidad del efecto de las características geográficas y de las posibles relaciones entre variables y respuesta de la cuenca.

Isik *et al.* (2008) realizaron un estudio en Turquía para la transferencia de información de caudal en sitios no aforados. En su estudio analizaron 1410 estaciones y generaron 6 regiones homogéneas, por medio de método de Ward y método de K-medias, y curvas de duración de caudal, por lo que proponen para un conjunto de datos grandes, dividir los grandes grupos homogéneos en regiones sub-homogéneas y los flujos de flujo específicos diarios pueden usarse en el análisis FDC para obtener resultados más precisos.

9.1. Selección de variables

El primer aspecto importante en la metodología regional es el análisis de las variables disponibles, en el análisis regional se emplean dos tipos de datos, características de la zona de estudio, o los valores estadísticos de las series de tiempo de las variables hidrometeorológicas (D.F. Campos-Aranda, 2014a), aunque este último punto está bajo la observación al ser no recomendado en la metodología del Análisis Regional de Frecuencias por el hecho de emplear los estadísticos exclusivamente para la revisión y medición de la homogeneidad de las regiones formadas (Hosking y Wallis, 1997; Núñez *et al.*, 2016).

Autores como Nathan y McMahon (1990) y Caratti *et al.*, (2005) proponen considerar al inicio de la investigación qué variables están relacionadas con el fenómeno a estudiar, si lo anterior no se toma en cuenta es posible que se generen grupos homogéneos con poco significado físico, por lo que la importancia en el conocimiento de cuántos descriptores de la cuenca hay que emplear es una pregunta frecuente al iniciar un análisis de delimitación de regiones homogéneas.

En la transferencia de información hidrológica, es común que el investigador trabaje con datos multivariados, es decir, un conjunto de datos que corresponden a las características tanto fisiográficas como climatológicas, por lo que el empleo de una técnica de Análisis de Componentes Principales (ACP) es factible. Esta técnica es útil cuando se emplea un gran número de datos (Sanborn y Bledsoe, 2006). Las técnicas multivariadas se han destacado por ser herramientas adecuadas y poderosas para encontrar regiones homogéneas y clasificar datos meteorológicos (Soltani y Reza, 2006).

La representación gráfica de los resultados de un ACP se lleva a cabo con dos tipos de ilustraciones. La primera, es la representación de los elementos, sobre un plano de ejes coordenados, formados por cada uno de los componentes principales seleccionados a partir del criterio de Kaiser. La segunda, es lo que se conoce como círculo de correlaciones y se forma con los valores del coeficiente de correlación entre los componentes principales y las variables de medición, por lo que con esta inspección es factible la clasificación de las variables poco, medianamente y bien representadas, por lo que aquellas variables más próximas a la circunferencia se clasifican en las mejor representadas.

El gran conjunto de datos tanto fisiográficos, como climatológicos e hidrométricos, implica que la escala de cada variable sea distinta, por lo que el hidrólogo puede trabajar con datos normalizados, para que, en el proceso de regionalización este factor no altere los resultados de selección de variables, así como de formación de grupos homogéneos (Soltani y Reza, 2006).

Otro aspecto importante es la longitud de periodo de registro, el periodo de información mínima que se debe recabar para un análisis regional y propuesto por *The World Meteorological Organization* (WMO) es de 30 años, sin embargo, diversos estudios enfocados en la regionalización de la precipitación marcan la recolección de registros entre

un periodo de 10 a 20 años (GREHYS, 1996; Soltani y Reza, 2006; Ouarda *et al.*, 2006; Campos-Aranda, 2014b). Otra recomendación importante es la selección de la cantidad de registros dependiendo de los T años de periodo de registro a simular, la relación es 5T la cantidad de registros, es decir, cinco veces el periodo de retorno de interés (Reed *et al.*, 1999; Rao y Srinivas, 2006; Schumann, 2011).

Después de la selección de las estaciones aforadas es necesario realizar un análisis exploratorio de datos para identificar datos atípicos, además de verificar la normalidad, homogeneidad, estacionariedad y tendencia de las series de datos (ver el capítulo 7 ANÁLISIS PREVIO DE DATOS).

9.1.1. Datos faltantes

Jain y Dubes (1983) describen cuatro técnicas para el tratamiento de ausencia de datos, los cuales están organizados en vectores conformados por los elementos:

1. Borrar los vectores de datos o los elementos del vector de datos, siempre y cuando la cantidad de datos ausentes sea mínima, esta técnica debe emplearse con cuidado.
2. Suponer que el valor de la *j*-ésima característica del *i*-ésimo vector patrón no tiene valor. Encontrar el vecino más cercano K de x_i y reemplazarlo por el valor de x_{ij} por el valor promedio de la *j*-ésima característica de los vecinos más cercanos. El valor de K debe estar en función del tamaño de la matriz patrón.
3. La distancia entre dos vectores x_i y x_k , que contienen valores extraviados se calculan como sigue: primero se define la distancia d_j entre dos patrones a lo largo de la *j*-ésima característica:

$$d_j = \begin{cases} 0 & \text{Si } x_{ij} \text{ o } x_{kj} \text{ no tienen valor} \\ x_{ij} - x_{kj} & \text{de otra manera} \end{cases}$$

Por lo que la distancia entre x_i y x_k se escribe como:

$$d(i, k) = \frac{d}{d - d_0} \sum d_j^2 \quad (9.1)$$

donde d_0 es el número de características faltantes en x_i , x_k o en ambas. En dado caso que no hubiera datos faltantes la anterior expresión corresponde al cuadrado de la distancia euclidiana.

4. Se denota como \bar{d}_j al promedio de la distancia entre todos los pares de patrones alrededor del punto j y se define como sigue:

$$\bar{d}_j = \frac{2}{n(n-1)} \sum_{i=2}^n \sum_{k=1}^{i-1} |x_{ij} - x_{kj}| \quad (9.2)$$

Donde n el número de patrones. La distancia de los patrones a lo largo de la j -ésima característica se define como sigue:

$$d_j = \begin{cases} \bar{d}_j & \text{Si } x_{ij} \text{ o } x_{kj} \text{ son faltantes} \\ x_{ij} - x_{kj} & \text{de otra manera} \end{cases}$$

Finalmente, la distancia entre x_i y x_k se escribe como:

$$d(i, k) = \sum d_j^2 \quad (9.3)$$

9.2. Índice de proximidad – medidas de similitud

La presentación de datos multivariados puede ser de dos tipos, por matriz de patrones o por matriz de proximidad. En la primera clasificación existen tres tipos de patrones ya sean binarios, discretos y continuos para variables tanto cualitativas como cuantitativas. Mientras que, para la segunda clasificación, la matriz de proximidad puede contener medidas de similitud y de disimilitud para variables cualitativas como cuantitativas (Jain y Dubes, 1988).

Anderberg (1973) presentó una minuciosa revisión de las diferentes medidas de asociación y sus interrelaciones, sin embargo, no quedó del todo clara su aplicación en los campos de la ciencia; a no ser la de agrupar datos numéricos en grupos y subgrupos semejantes. A estas medidas se les conoce con el nombre de índices de proximidad y representan en forma matemática, la semejanza o desigualdad entre dos configuraciones, es decir, entre dos o más puntos. Su aplicación en hidrología consiste en obtener dichos índices a partir de configuraciones específicas, formadas por características hidrológicas medibles y continuas en una proporción de escala conocida (Jain y Dubes, 1988). Además, dado que, para la formación de grupos, ya sea por técnicas jerárquicas o divisivas, es necesario un índice de afinidad para clasificar la similitud o disimilitud entre pares de datos (Jain y Dubes, 1988).

El índice de proximidad entre dos puntos (configuración) “ i ” y “ k ” se denota como: $d(i, k)$ y debe de satisfacer las propiedades siguientes (Jain y Dubes, 1988):

- 1)
 - a) $d(i, i) = 0$ Para todos los i -ésimos puntos (no semejanza)
 - b) $d(i, i) \geq \max d(i, k)$ Para todos los i -ésimos puntos (semejanza)
- 2) $d(i, k) = d(k, i)$ Para todos los (i, k)
- 3) $d(i, k) \geq 0$ Para todos los (i, k)
- 4) $d(i, k) = 0$ Sólo si $x_i = x_k$
- 5) $d(i, k) \leq d(i, m) + d(m, k)$ Para todos los (i, k, m) , propiedad llamada “triángulo de la inequidad”

El índice de proximidad más empleado es el índice métrico de Minkowski, el cual mide la disimilitud. Los tres tipos más comunes del índice métrico de Minkowski son: distancia euclidiana, distancia Manhattan y sup-distancia o de Tchebychev.

9.2.1. Distancias

El índice métrico de Minkowsky r es de la siguiente forma:

$$d(i, k) = \left[\sum_{j=1}^d |X_{i,j} - X_{k,j}|^2 \right]^{1/r} \quad (9.4)$$

donde $X_{i,j}$ y $X_{k,j}$ es la j -ésima característica de la i -ésima y k -ésima configuración y $r \geq 1$

Hay tres principales índices que comúnmente son empleados y se mencionan a continuación:

9.2.1.1. Distancia euclidiana

La distancia euclidiana, con un índice de Minkowski de 2, es decir, $r = 2$ es de la siguiente forma:

$$d(i, k) = \left[\sum_{j=1}^d |X_{i,j} - X_{k,j}|^2 \right]^{1/2} \quad (9.5)$$

Esta distancia tiene también la variante de la distancia Euclidiana al cuadrado:

$$(d(i, k))^2 = \left[\sum_{j=1}^d |X_{i,j} - X_{k,j}|^2 \right] \quad (9.6)$$

Estas dos distancias son calculadas a partir de datos brutos y no de datos estandarizados. Esta distancia se ve afectada por la magnitud de las unidades de medición. Para corregir esto se puede intentar una estandarización antes de calcular las distancias (Jain y Dubes, 1988).

9.2.1.2. Distancia Manhattan

La distancia de Manhattan, también llamada *taxicab*, o distancia *city block* con un índice de Minkowski de 1, es decir, $r=1$ es de la siguiente forma:

$$d(i, k) = \sum_{j=1}^d |X_{i,j} - X_{k,j}| \quad (9.7)$$

Esta distancia es la diferencia media entre las magnitudes de las mediciones. En la mayoría de los casos esta distancia arroja resultados similares a la distancia Euclidiana. Sin embargo, es importante señalar que con esta distancia la diferencia entre mediciones es muy importante ya que la distancia no está elevada al cuadrado (Jain y Dubes, 1987).

9.2.1.3. Sup-distancia Tchebychev

La distancia de Tchebychev, también llamada sup-distancia, con un índice de Minkowski de r que tiende al ∞ es de la siguiente forma:

$$d(i, k) = \max |X_{i,j} - X_{k,j}| \quad (9.8)$$

$$1 \leq j \leq d$$

Esta distancia es empleada cuando se consideran dos objetos diferentes a partir de que alguna de sus magnitudes fuera medida en una escala diferente (Jain y Dubes, 1987).

9.3. Reglas de agregación

La primera etapa luego de que cada objeto forma una clase, es definir las distancias entre ellos. De esta forma se podrá tener una primera agrupación seleccionando los elementos más cercanos entre sí. Es necesario una regla de agregación para determinar en qué momento dos grupos son lo suficientemente similares para fusionarse en uno solo. Se pueden ligar dos

clases o elementos cuando sus distancias son próximas respecto a su algoritmo de agregación correspondiente (Jain y Dubes, 1987).

En un análisis de agrupamiento son necesarias las medidas de similitud como la distancia euclidiana, distancia euclidiana al cuadrado, coseno de vectores de variables, la suma de las diferencias absolutas en valores para cada variable (Nathan y McMahon, 1990). Los métodos de ligadura son: ligadura sencilla, ligadura completa, promedio de la ligadura entre grupos, promedio de ligadura dentro de los grupos, agrupamiento en el centroide y medianas, y el método de Ward (Nathan y McMahon, 1990). Existen dos tipos de técnicas de agrupamiento la técnica divisiva y la jerárquica (Soltani y Reza, 2006), la técnica jerárquica más popular es el agrupamiento por el método de Ward y algunos autores la señalan como el mejor método de delimitación de regiones homogéneas (Nathan y McMahon, 1990), mientras que la técnica divisiva más frecuentemente empleada es el agrupamiento por K-medias.

9.3.1. Método de agrupamiento por Ward

El método de Ward está enfocado al tratamiento de grandes muestras de datos, en la divulgación del método Ward (1963) recomendó su uso en el área de biología para la clasificación taxonómica, manejo de librerías, administración de documentos y todas aquellas actividades que emplean muestras de grandes datos (Ward, 1963).

El propósito del análisis de agrupamiento es colocar objetos en grupos de manera que los objetos en cada grupo tienen la mayor similitud entre sí, mientras que los objetos en diferentes grupos tienen la máxima disimilitud (Soltani y Reza, 2006), como resultado se obtiene un dendograma que es una representación gráfica donde es posible identificar “los grupos principales continuando con la discriminación de subgrupos, cuya vinculación es creciente en términos de similitud” (Paris y Zucarelli, 2004).

El uso del análisis de agrupamiento (*cluster analysis*) para la regionalización implica la agrupación de diversas observaciones y variables en grupos, de modo que cada grupo se compone de observaciones o variables con características similares tales como geográfico, físico, estadístico o comportamiento estocástico (Soltani y Reza, 2006). Los resultados en este método implican regiones homogéneas contiguas en el espacio, por lo que la ubicación de la cuenca no aforada se basa en las características de la cuenca y el paso final de este

método consiste en la estimación del caudal en la zona de interés mediante regresión múltiple entre los datos hidrométricos y las características de la cuenca (Nathan y McMahon, 1990).

Este método es uno de los más empleados en el análisis regional debido a las ventajas de su aplicación (Burn, 1989; Nathan y McMahon 1990; Nathan y McMahon, 1992; GREHYS, 1996; Baeriswyl y Rebetz, 1997; Ramos, 2001; Yu *et al.*, 2002; Laaha y Blöschl, 2006; Sanborn y Bledsoe, 2006; Soltani y Reza, 2006; Kahya *et al.*, 2008; Lim *et al.*, 2009 ; Kar *et al.*, 2015; Ilorme y Griffis, 2013; Dodangeh *et al.*, 2014).

En el método de agrupamiento, cada grupo contiene la menor varianza de variables (la disimilitud más pequeña). En el estudio realizado por los autores Sarhadi y Modarres (2011), la técnica de agrupamiento jerárquico, descrito por Kaufman y Rousseuw (1990), se aplica con el fin de clasificar las cuencas en regiones similares; se han propuesto varios métodos para el análisis de agrupamiento jerárquico, incluyendo la ligadura sencilla, media y completa, y el método de mínima varianza de Ward. Y la revisión de la literatura expresa que este método de clasificación es mejor que la mayoría de los métodos antes empleados (Sarhadi y Modarres, 2011).

En la regionalización de caudales, las técnicas del clúster análisis con técnica de Ward se han empleado en datos de caudales mínimos (Nathan y McMahon, 1990; Burn *et al.*, 1997; Yu *et al.*, 2002), caudales medianos (Paris y Zucarelli, 2004) y caudales máximos, también en precipitación (Baeriswyl y Rebetz, 1987; Nathan y McMahon, 1990; Arellano-Lara y Escalante-Sandoval, 2014), en variables climáticas como temperatura (Dodangeh *et al.*, 2014; Masud *et al.*, 2015), calidad del agua (O'Shea y Jankowski, 2006; Hynds *et al.*, 2014) y sequía (Harris *et al.*, 2000), entre otros.

En la regionalización de la precipitación Baeriswyl y Rebetz (1987) desarrollaron un estudio con el objetivo de analizar la distribución espacial de la lluvia para dos periodos de series históricas, donde el objetivo del estudio de los autores fue promover el estudio de la precipitación a través de tiempo como una medida de relación con el cambio climático.

Nathan y McMahon (1990) identificaron regiones homogéneas en 184 cuencas, con superficies de 1 a 250 km², en Australia, empleando para ello las características de caudales mínimos, con registros de duración de 17 años. Para la delimitación de regiones homogéneas

emplearon distintas técnicas: análisis de agrupamiento, análisis de regresión múltiple, análisis de componentes principales, y representación gráfica de datos multidimensionales, además desarrollaron una metodología que implica el análisis de regresión múltiple para la selección y ponderación de variables, el análisis clúster para la formación preliminar de los grupos, y la heterogeneidad de los grupos fue evaluada con métodos gráficos multidimensionales.

El método de Ward ha sido señalado por algunos autores como el método óptimo para la formación de regiones (Nathan y Macahon, 1990). Nathan y Macahon (1990) emplearon distintas técnicas: análisis de agrupamiento, análisis de regresión múltiple, análisis de componentes principales, y representación gráfica de datos multidimensionales, sin embargo, su evaluación general sugiere que el análisis de conglomerados ponderada (método de Ward basado en la distancia euclidiana al cuadrado) usando ponderaciones de acuerdo con los coeficientes del modelo de regresión inicial paso a paso presentando un mejor comportamiento

Burn *et al.* (1997) realizaron un análisis de regionalización de cuencas, por medio de clasificación jerárquica ascendente con la medición de la similitud de cuencas mediante las fechas de ocurrencia de las avenidas máximas anuales en la cuenca del río Saskatchewan-Nelson en la parte del oeste central de Canadá.

El análisis de aglomerados, también se ha empleado en casos de evaluación de sequías (Harris *et al.*, 2000). Harris *et al.* (2000) analizaron un conjunto de 500 corrientes de río en 11 países europeos, con el objetivo de analizar el coeficiente de sequía con el déficit de caudal, por medio del coeficiente de Índice de deficiencia regional (RDI) y el índice estandarizado de precipitación regional.

Yu *et al.* (2002) emplearon métodos multivariados (Método de Ward y Método de K medias) para regionalizar caudales mínimos a partir de variables hidrogeológicas, a partir de un análisis de componentes principales, donde además realizan un análisis discriminante y un análisis de agrupamiento, además de que emplearon tanto el enfoque de regresión tradicional así como regresión difusa para la regionalización de curvas de duración de caudal en sitios no aforados. Debido a que los autores recopilaron información sobre las variables que se ven

influenciadas por los caudales mínimos, identificando de esta manera que las variables hidrogeológicas si tienen cierta similitud con estas características.

Paris y Zucarelli (2004) aplicaron métodos estadísticos matemáticos multivariados para delimitar regiones hidrológicamente homogéneas en 11 cuencas en Argentina, por medio del uso de variables hidrométricas y características hidráulicas de los canales naturales, en donde el ancho de la sección contabilizó el mayor coeficiente de similitud, y establecieron las pautas para el diseño de una red de mediciones consideradas para la representatividad de las observaciones hidrométricas

Soltani y Modarres (2006) realizaron un análisis jerárquico y divisivo para categorizar los patrones de lluvia de Irán, así como la evaluación anual y temporal estacional, resultando ocho grupos homogéneos, estos grupos pueden ser clasificados en tres regímenes principales que fueron denominados como: invierno, verano-invierno y otoño, además el método de Ward representa de una mejor forma los patrones espaciales de la lluvia, en comparación con el promedio de los métodos de análisis jerárquico (Soltani y Reza, 2006). En ese mismo año, también se adaptó el método de Ward en el análisis de calidad del agua para la identificación de anomalías en su composición química (O'Shea y Jankowski, 2006).

En el análisis de caudal mínimo ecológico también ha sido aplicada esta técnica (Laaha y Blöschl, 2006), donde además, se realizó una revisión de homogeneidad por medio de métodos gráficos.

Isik *et al.* (2008) desarrollaron un aglomerado jerárquico con datos de caudal diario. Además esta técnica también se ha empleado en estudios de regionalización mensual de corrientes de ríos bajo técnicas estocásticas (Özçelik y Benzeden, 2010).

Gutiérrez-López *et al.* (2009) construyeron ecuaciones regionales de transferencia de información hidrológica, a partir de datos anuales de caudales medios, analizando la relación espacial que existe entre los parámetros de un modelo Ar (p) con las características orográficas de la región.

Se han realizado la formación de grupos homogéneos de acuerdo a las características del perfil fisicoquímico en muestras con contaminación de agua (O'Shea y Jankowski, 2006; Hynds *et al.*, 2014).

En la regionalización de caudales diarios, la curva de duración de caudal es la herramienta más popular empleada por los hidrólogos (Boscarello *et al.*, 2015; Chouaib *et al.*, 2018), mediante estas curvas es posible analizar la variabilidad en la corriente de un río ya que mediante estas curvas se puede presentar el caudal diario a lo largo del año, es muy útil cuando se requiere diseñar una obra y conocer el caudal mínimo. La regionalización de la duración de las curvas de caudal, FDC por sus siglas en inglés, se ha realizado bajo las técnicas de mapas auto-organizados, SOMs por sus siglas en inglés, bajo el enfoque de agrupamiento, donde las características físicas de la cuenca y los índices de caudal forman regiones homogéneas (Boscarello *et al.*, 2015) .

Masud *et al.* (2015) también lo emplearon en análisis de sequías en Canadá, con el objetivo de estudiar el impacto en la agricultura, energía y sector socio-económico. Emplearon los índices de SPEI (Índice de precipitación y evapotranspiración) y SPI (Índice de precipitación estandarizada) agrupando regiones por el método de agrupamiento de aglomerados.

Este método de agregación también ha sido exitoso en el análisis de variables climáticas, como en el caso de la variable de temperatura. Sarmadi y Azmi (2016) realizaron un estudio de regionalización de la temperatura media del aire por métodos multivariados y Momentos-L, para ello emplearon 44 estaciones sinópticas y climatológicas, los métodos multivariados empleados son análisis factorial (FA, por sus siglas en inglés) y análisis de agrupamiento (CA, por sus siglas en inglés) para después aplicar la técnica Momentos-L, con el objetivo de conocer la distribución espacial de la temperatura en Irán, considerando para ello el incremento en la mancha urbana dado un periodo de análisis.

El método de Ward bajo el enfoque de formación de regiones homogéneas propuesto por Nathan y McMahon (1990) consiste en los siguientes pasos (Laaha y Blöschl, 2006):

- Identificación de las variables o características descriptivas de las cuencas más relevantes, por medio de un análisis de regresión general escalonado (*stepwise regression*).
- Ponderar las características de las cuencas de acuerdo a la importancia relativa de cada una de ellas, determinadas por la magnitud del coeficiente β del modelo de regresión basado en las características estandarizadas de las cuencas.

- Empleando diferentes medidas de similitud y métodos de ligadura diseñar distintos *cluster* (agrupamientos).
- Realizar un análisis de prueba de homogeneidad gráfica, es decir, curvas de Andrews por cada clúster formado, identificando el grupo de curvas que contengan la menor variación.
- Remover los puntos dispersos del grupo óptimo seleccionado basados en las gráficas de las curvas de Andrews, deducir un conjunto de características de captación media para cada grupo homogéneo;
- Depurar hasta llegar al grupo óptimo formado derivado en el análisis de agrupamiento, comparando las características de cada cuenca con el valor medio del grupo, y reclasificar en caso de que sean muy distintos los valores.

9.3.1.1. Enfoque general del método de Ward

El método de Ward calcula la distancia entre dos grupos como la suma de los cuadrados entre dos grupos añadidos a lo largo de todas las variables. En cada generación clúster se minimiza la suma de cuadrados. Si C_K y C_L son dos grupos que se fusionaron para formar C_M clúster, la distancia entre el nuevo grupo y otro grupo C_J es (Soltani y Reza, 2006):

$$d_{J,M} = \frac{((n_J + n_K)d_{JK} + (n_J + n_L)d_{JL} - n_J d_{KL})}{n_J + n_M} \quad (9.9)$$

Donde n_L, n_K, n_J y n_M son el número de objetos en los grupos J, K, L y M , respectivamente y d_{JK} , d_{JL} y d_{KL} representan las distancias entre los cluster J y K y entre J y L , y K y L , respectivamente.

En el método de Ward dos clúster son combinados, como resultado tienen un mínimo incremento en el total de la suma de cuadrados dentro del clúster generado (Acreman y Sinclair, 1986). La unión de agrupamientos comienzan con m cluster, en cada “ciclo de fusión” los dos cluster más similares son fusionados, el procedimiento se detiene una vez que los grupos fusionados son menores a cierto valor específico (Acreman y Sinclair, 1986).

$$ss = \sum_{i=1}^m \sum_{j=1}^{n_j} \sum_{k=1}^p (x_{ijk} - z_i)^2 \quad (9.10)$$

Dados m clúster con n_j miembros descritos por p variables, donde z_i es el centroide i -ésimo cluster. El proceso es detenido cuando la probabilidad estadística alcanza el máximo.

9.3.1.2.Desarrollo del algoritmo de Ward

Ward (1963) desarrolló un método de agrupación por medio de la similitud entre los miembros que componen un grupo respecto a las características de los mismos, por medio de una agregación jerárquica de N miembros que se aglomeran debido al grado de similitud hasta formar J grupos homogéneos, el proceso concluye hasta llegar a un solo grupo. Para ello es necesario el cálculo de la distancia euclidiana, así como de la inercia entre centros de grupos.

Para la aplicación del método, se asume que se cuenta con P puntos en el espacio, los P puntos corresponden a las estaciones aforadas, de las cuales se calcula la matriz de distancia euclidiana al cuadrado, es decir con un índice de Minkowski de 2, de la siguiente forma:

$$(d(i, k))^2 = \left[\sum_{j=1}^d |X_{i,j} - X_{k,j}|^2 \right] \quad (9.11)$$

$$d_{i,i'}^2 = (x_{i'} - x_i)^2 + (y_{i'} - y_i)^2 \quad (9.12)$$

El método de Ward consiste en la asignación de peso a las variables que intervienen en el algoritmo de agrupación, por lo que se genera una primera matriz de inercia entre todos los P puntos a agrupar, mientras que en las etapas consecuentes del método se evalúan los pesos “ganados” en cada etapa de la clasificación jerárquica, logrando la unión de aquellos que generen una menor varianza. La matriz de inercias de los P puntos (estaciones) tiene la siguiente estructura:

$$\Delta I_{min} = (h, h') = \frac{m_h m_{h'}}{m_h + m_{h'}} d^2 (g_h, g_{h'}) \quad (9.13)$$

donde m son las masas de cada elemento h y h' , g son los centros de gravedad de cada elemento h y h' , d^2 es el cuadrado de la distancia euclidiana entre los centros de gravedad.

Este criterio representa el momento de inercia $Mg(g_h, g_{h'})$ entre los puntos (elementos) h y h' con sus respectivas masas m_h y $m_{h'}$, con relación a sus centros de gravedad g_h y $g_{h'}$:

$$g = \frac{(g_h m_h + g_{h'} m_{h'})}{m_h + m_{h'}} \quad (9.14)$$

A partir de la agregación jerárquica de elementos se deberá emplear la ecuación

$$\Delta I_{j,r} = \frac{1}{m_j + m_r} [(m_i + m_r) \Delta I_{ir} + (m_{i'} + m_r) \Delta I_{i'r} - m_r \Delta I_{ii'}] \quad (9.15)$$

donde $j = i + i'$, es la agregación del elemento, r es el elemento objetivo con respecto al cual se calcula la inercia, de la misma forma m_j será el total del peso de la variable debido a la suma de las masas $m_i + m_{i'}$ y m_r es el peso del elemento r .

En cada paso iterativo se elige como nueva agrupación a aquellos elementos que tengan una menor inercia entre sí, por lo que para el nuevo elemento creado se calcula la inercia que se genera a partir de ese grupo formado respecto a los elementos restantes del análisis.

5.1.1.1. Deducción del algoritmo de inercia

Un criterio de inercia representa la relación entre dos puntos y sus masas.

$$M_g^I = (g_h; g_{h'}) \quad (9.16)$$

Si se agregan sus masas m_h y $m_{h'}$, respectivamente y se ubican en un centro de gravedad g se tendrá:

$$g = \frac{(g_h m_h + g_{h'} m_{h'})}{(m_h + m_{h'})} \quad (9.17)$$

Efectivamente, si ahora tomamos en cuenta las distancias entre las masas como:

$$M_g^I = (g_h; g_{h'}) = m_h d^2(g, g_h) + \dots + m_{h'} d^2(g, g_{h'}) \quad (9.18)$$

Que corresponde a la distancia al centro de gravedad.

Si se utiliza un producto escalar entre dos puntos para determinar la distancia se tiene:
 $d^2(x, y) \leq x - y$, esto es: $x - y > d^2(g, g_h)$

$$d^2(g, g_h) \leq \frac{(m_h g_h + m_{h'} g_{h'})}{(m_h + m_{h'})} - g_h \quad (9.19)$$

$$\frac{(m_h g_h + m_{h'} g_{h'})}{(m_h + m_{h'})} - g_h \geq \frac{m_{h'}^2}{(m_h + m_{h'})^2} < g_{h'} - g_h \quad (9.20)$$

de donde:

$$d^2(g, g_h) = \frac{m_{h'}^2}{(m_h + m_{h'})^2} d^2(g_h, g_{h'}) \quad (9.21)$$

de esta forma se obtiene:

$$M_g^I(g_h, g_{h'}) = \frac{m_h \cdot m_{h'}^2}{(m_h + m_{h'})^2} d^2(g_h, g_{h'}) + \dots + \frac{m_{h'} \cdot m_{h'}^2}{(m_h + m_{h'})^2} d^2(g_h, g_{h'}) \quad (9.22)$$

Finalmente:

$$M_g^I(g_h, g_{h'}) = \frac{m_h \cdot m_{h'}}{(m_h + m_{h'})^2} d^2(g_h, g_{h'}) \quad (9.23)$$

9.3.2. K-Medias

El método de K-medias es una técnica de aglomerado no jerárquica y se obtienen resultados satisfactorios cuando se aplican en conjunto de datos grandes, este método es comúnmente empleado cuando se conoce el número de grupos a priori (Isik *et al.*, 2008). La principal diferencia con el método jerárquico, es que el método de K-medias no produce todos los aglomerados posibles a partir de n observaciones, sino que busca la mejor combinación de grupos a partir de la cantidad de grupos definidos (Afifi *et al.*, 2004; Isik *et al.*, 2008).

Las diferencias con el método jerárquico son las siguientes:

- El método de K-medias no crea una estructura de árbol, crea un nivel sencillo de cluster.

- El método cluster emplea los datos de las observaciones, no sólo la proximidad espacial, como el método jerárquico, que solo se basa en la medida de las distancias.

Algunos autores lo consideran como una mejora del método jerárquico (Isik *et al.*, 2008), una ventaja importante en el método K-medias es la factibilidad de aplicarlo en conjuntos grandes de datos (McNeil *et al.*, 2005).

El método de k-medias se ha empleado para la delimitación de regiones homogéneas en la aplicación de la técnica de L momentos, en datos de precipitación máxima anual para la elaboración de mapas iso-pluviales en zonas con poca red de monitoreo meteorológico (Rahman *et al.*, 2013) así como comparación entre K-medias y análisis de agrupamiento (Trefry *et al.*, 2006), en datos de caudales diarios (Isik *et al.*, 2008), en caudales mensuales (Parasuraman y Elshorbagy, 2007; Nourani *et al.*, 2013), así como para la predicción de datos mensuales del nivel del agua por medio del método de agrupamiento (Nourani *et al.*, 2013).

De Vos *et al.*, (2010) realizaron una propuesta para agrupar regiones homogéneas temporales identificando los periodos de similitud hidrológica y agrupación dinámica mediante K-medias.

En análisis de sequías en la formación de grupos homogéneos Yoo *et al.* (2012) emplearon el método de K-medias realizaron un análisis de frecuencias de sequías empleando función de distribución de probabilidades ponderada. La metodología de K-medias se ha empleado en datos de precipitación con relación a tifones para mejorar la predicción de avenidas a nivel meso-escala en Taiwán (Huang *et al.*, 2012).

Este método también ha sido empleado en la delimitación de regiones para clasificación de datos de caudal (Dikbas *et al.*, 2013). Dikbas *et al.*, (2013) realizaron la delimitación de región homogéneas para clasificar los datos de caudal máximo anual, estudio realizado en Turquía, donde las pruebas de homogeneidad se realizaron con la prueba de L momentos, por medio del método de K-medias y clasificación jerárquica ascendente, resultando que el método de K-medias ofrece buenos resultados. Para determinar el número óptimo de clúster formados se aplicó un criterio de la raíz media cuadrática del error de la desviación estándar y R-Cuadrada, posteriormente para conocer la homogeneidad en los grupos formados se

aplicó la prueba de heterogeneidad, donde se aplicó también una prueba llamada prueba de discordancia.

En análisis de suelos, se ha empleado para identificar la ubicación de muestras para el monitoreo de la humedad del suelo (Arkel y Kaleita, 2014).

García-Marín *et al.* (2015) desarrollaron una propuesta de agrupamiento bajo características multifractales empleando datos de precipitación diaria en 53 estaciones, por medio de un análisis regional de frecuencias bajo técnicas de Momentos-L y K-medias.

Esta metodología también se ha empleado en la delimitación de redes de drenaje y de valles a partir de un MDE, para ello Hooshyar *et al.* (2016) emplearon datos de topografía y manipularon el modelo digital de elevaciones para la formación de los grupos.

Técnicas recientes en la modelación de redes neurales basadas en algoritmos genéticos, también han empleado técnicas como la agrupación de K-medias, para el estudio del gasto pico de predicción en fallas y rompimiento de presas (Eghbali *et al.*, 2016).

9.3.2.1. Enfoque general del método K-Medias

El análisis de agrupamiento divisivo es un método común, basado en el algoritmo *K-medias*, que mide la proximidad entre los grupos utilizando la distancia euclidiana entre los centroides de los grupos. El algoritmo K-medias comienza a partir de una determinada partición especificada por el vector de asignación de $P(K, M)$ con las M observaciones asignados en K grupos (Soltani y Reza, 2006). El centroide de cada uno de los K grupos se calcula y la similitud entre las observaciones en los grupos se mide por una función de error e dado por:

$$e[P(M, K)] = \sum_K D[I, L(I)]^2 \quad (9.24)$$

donde $L(I)$ es el *cluster* contenido en el I -ésimo caso $D[I, L(I)]^2$ representa la suma de las desviaciones absolutas al cuadrado de los centroides de los grupos (media de las variables) más de todas las observaciones en el cluster y más de las variables. La asignación inicial de la observación en los K grupos se lleva a cabo normalmente por una partición al azar. El número de grupos que se mantiene es elegido por la función de error mínimo para un número

creciente de *clusters*. Otro método adecuado para seleccionar el número de grupos es calcular la estadística de pseudo F y t^2 .

$$pseudoF = \frac{\frac{T - P_G}{G - 1}}{\frac{P_G}{n - G}} \quad (9.25)$$

donde T es $\sum_{i=1}^n \|x_i - \bar{x}\|^2$, n es el número de observaciones, $P_G = \sum W_j$ que es la suma total respecto a los G agrupamientos en el G -ésimo nivel de jerarquía. La *pseudo t^2* para la unión de C_K y C_L es:

$$pseudot^2 = \frac{B_{KL}}{\frac{W_K + W_L}{N_K + N_L - 2}} \quad (9.26)$$

donde W_K es $\sum_{i \in C_K} \|x_i - \bar{x}_k\|^2$, W_L es $\sum_{i \in C_L} \|x_i - \bar{x}_l\|^2$, B_{KL} es $W_M - W_K - W_L$ si $C_M = C_K \cup C_L$

N_K y N_L son el número de observaciones en C_K y C_L , respectivamente.

Si tenemos dos grupos, C_K y C_L , para fusionarse y producir otro grupo, C_M , la distancia entre el C_M y otro grupo, C_J , se escribe como sigue:

$$d_{J,M} = \frac{n_K d_{JK} + n_L d_{JM}}{n_M} \quad (9.27)$$

donde n_L , n_K y n_M es el número de objetos previamente agrupados en L , K y M , respectivamente y $d_{J,M}$ son las distancias entre los cluster J y K y entre J y L , respectivamente.

9.3.2.2.Desarrollo del algoritmo

Este método de agrupación fue propuesto por Hartigan y Wong (1979) el cual produce K -grupos a partir de un conjunto de “ n ” elementos; cumpliendo con la condición de la función objetivo sea mínima:

$$E = \sum_{r=1}^K \sum_{x_i \in C_r} (x_i - g_r)^2 \quad (9.28)$$

donde C_r representa los grupos, x_i es un elemento dentro de un grupo y g_r es el “punto medio” del grupo C_r .

El procedimiento inicia con un número arbitrario de K-grupos. La distancia empleada en este método es la distancia euclidiana.

El procedimiento de cálculo es el siguiente:

1. Seleccionar al azar k-elementos, los cuales serán los centros de grupo iniciales,
2. Efectuar una partición en k-grupos a partir de la menor distancia de cada elemento y el centro de entre grupo elegido.
3. Calcular la calidad de la partición
4. Repetir los pasos (2) y (3) hasta obtener el menor valor de Qp .

9.4. Métodos gráficos de delimitación de regiones homogéneas

Los gráficos de probabilidad basados en el análisis experimental de datos fueron por primera vez usados por Hazen en el año 1914 para estudio de inundaciones (Everitt, 1978). El empleo de estos gráficos se usa comúnmente para realizar una primera aproximación informal de la muestra (Everitt, 1978). A continuación, se describen los métodos más empleados en el análisis regional.

9.4.1. Método geográfico

En este método se asume que a partir del punto de interés a un radio de 80 km las condiciones espaciales son homogéneas (Escalante-Sandoval y Reyes-Chávez, 2005).

9.4.2. Rostros de Chernoff

El método de rostros de Chernoff es un método gráfico empleado en la representación de datos multivariados en un espacio k -dimensional. En este método, un conjunto de datos multivariados se representa mediante una cantidad de rostros donde cada tipo de rostro agrupa las variables del conjunto de datos mediante la forma y distancia de las características

de un rostro, como lo son los ojos, cejas, nariz y boca (Chernoff, 1973), permitiendo de esta forma la inspección del gran conjunto de datos.

Chernoff (1973) empleó dos tipos de datos para la divulgación y aplicación de su método de agrupación por medio de inspección visual, en datos de fósiles y datos geológicos, trabajó con siete y doce variables representativas, respectivamente, agrupando de manera visual los rostros obtenidos.

Dentro de las ventajas al emplear un método gráfico Chernoff (1973) se enuncian cuatro principales, afirmando que por medio de esta técnica se cumplen las dos primeras ventajas:

- 1) Representa una ventaja en la comprensión del fenómeno a estudiar.
- 2) Establece una conexión entre el objeto de estudio mediante la inspección visual y las primeras conclusiones.
- 3) Facilita la comunicación de las primeras conclusiones
- 4) Permite el primer procesamiento informal en el cálculo de datos.

Por lo que la ventaja en la aplicación es claramente nemotécnica, el autor establece que “observando los rostros, los datos pueden ser fácilmente recordados en términos de emociones y apariencias”, en vez de recordar datos numéricos, además de procesar la relación entre los rostros y los datos numéricos (Chernoff, 1973).

El método propuesto originalmente por Chernoff (1973) podía tratar hasta 18 variables cuantitativas. Las recomendaciones del autor son las siguientes (Chernoff, 1973):

- 1) intentar diversas combinaciones de permutaciones de las variables
- 2) relacionar los procesos que se buscan describir y asociar las variables a los fenómenos.

Además, con la colección de los rostros resultantes de los datos multivariados es posible la identificación de formación de grupos y de valores atípicos (Everitt, 1978).

9.4.3. Curvas de Andrews

La técnica de curvas de Andrews se ha empleado en la regionalización como un método tanto para delimitación de regiones homogéneas, así como para verificación gráfica de la

homogeneidad regional (Nathan y McMahon, 1990; Paris y Zucarelli, 2004; Laaha y Blöschl, 2006; Heřmanovský y Pech, 2013) .

Una evaluación visual de los grupos formados se califica por medio de la solución donde se aprecian los elementos de cada grupo, estrechamente espaciados y lejos de otras agrupaciones (Sanborn y Bledsoe, 2006).

Este método fue propuesto por Andrews (1972) y consiste en el desarrollo de series Fourier para representar un espacio de multidimensional (Chernoff, 1973), por medio de curvas en dos dimensiones, con el empleo de las variables descriptivas de las zonas de estudio, graficadas en un rango de $-\pi$ a π , de la siguiente forma:

$$f(x) = \frac{x_1}{\sqrt{2}} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_4 \cos(2t) + \dots \quad (9.29)$$

donde x_1, x_2, \dots, x_n , representan cada una de las variables usadas para caracterizar la cuenca.

Andrews (1972) desarrolló la técnica empleando los componentes principales de la muestra. De forma que las variables x_1 a x_n se identifican como aquella variable más importante hasta la menos importante, por lo que aplicando componentes principales en conjunto con las curvas de Andrews, para x_1 llegaría a corresponder el valor del primer componente principal y así sucesivamente (Andrews, 1972; Chernoff, 1973; Everitt, 1978).

Las variables descriptivas, pudiendo ser meteorológicas o fisiográficas, se ingresan en orden de importancia estadística, comenzando con las de mayor importancia (Escalante-Sandoval y Reyes-Chávez, 2005).

9.4.4. Árbol de clasificación

La técnica de regresión de árboles, es una metodología aplicada en la minería de datos. El método de árboles de regresión y clasificación (CART, por sus siglas en inglés) fueron por primera vez empleados en regionalización hidrológica por Laaha (2002) aplicado a caudales mínimos (Laaha y Blöschl, 2006), de acuerdo con la información relacionada con Laaha y Blöschl (2006) la única aplicación previa en hidrología, es la aplicación de esta técnica en el procesamiento de imágenes satelitales para el análisis de la cobertura de nieve e interpolación

espacial de la misma en estudios realizados por Elder (1995) y por Rosenthal y Dozier (1996) (Laaha y Blöschl, 2006).

En este método se requieren establecer las variables dependientes e independientes del dominio de estudio, es decir, variables predictivas y variables objetivo, que pueden llegar a ser las variables descriptivas de la cuenca y las variables de predicción, respectivamente (Laaha y Blöschl, 2006). Este método depende totalmente de las variables descriptivas de la cuenca, por lo que se inicia con las variables más importantes hasta las menos importantes (Laaha y Blöschl, 2006). El proceso es una “partición binaria recursiva” (Laaha y Blöschl, 2006), donde los grupos se definen por condiciones binarias.

Las ventajas del método CART, son las siguientes (Laaha y Blöschl, 2006):

- Tienen una estructura no paramétrica
- La formación de árboles es visualmente fácil de interpretar.
- No hay sensibilidad global a valores atípicos, a diferencia de otros métodos.
- Hay un buen funcionamiento en el manejo de relaciones no lineales entre las variables.

Otra ventaja (Ragetti *et al.*, 2017) es que este método considera en su desarrollo tanto variables físicas así como proximidad espacial, por lo que es un método que puede llegar a ser más preciso al combinar ambos enfoques. Ragetti *et al.* (2017) recomiendan emplear la proximidad espacial como criterio de selección en la agrupación de regiones homogéneas, pero recomiendan omitir esta selección en caso de que no haya cuencas similares juntas. Por lo que la principal ventaja de este método de delimitación es la aplicación en zonas de estudio con topografía compleja, donde el enfoque de similitud espacial no es factible (Ragetti *et al.*, 2017).

Este método de delimitación es útil cuando se planea predecir el grupo apropiado de un evento basado en las características de eventos pasados (Mather y Johnson, 2015), el inconveniente del empleo de este método es cuando se generan árboles grandes debido a que es más complicada la interpretación, debido a ello es necesario “podar” el árbol generado para superar el problema del sobre ajuste de datos (Laaha y Blöschl, 2006).

Laaha y Blöschl (2006) realizaron una comparación de cuatro métodos de delimitación de regiones homogéneas, siendo los primeros en emplear un método de análisis de regresión de árbol el cual ofrece la ventaja de la ubicación automática de las cuencas no aforadas al modelo

en el agrupamiento (clúster) más apropiado, por medio de una clasificación y validación cruzada de 10 rutinas hasta llegar al tamaño y definición óptima de los grupos homogéneos.

El método de árbol de regresión para agrupación de regiones homogéneas consiste en los siguientes pasos (Laaha y Blöschl, 2006):

- Transformación de los datos a datos normalizados.
- Ajustar el árbol de regresión inicial a los datos.
- Determinar el tamaño óptimo del árbol por medio de 10 validaciones cruzadas.
- Podar el árbol inicial al tamaño del árbol de regresión derivado en el paso inmediato anterior.

Este método también se ha empleado en la identificación de la magnitud de las variables de respuesta en hidrogramas, al ser un método de fácil empleo (Ali *et al.*, 2010).

Schnier y Cai (2014) desarrollaron un estudio para la predicción de caudales basados en una frecuencia de excedencia dadas 25 características de 294 cuencas, con un rango de áreas de los 12 a los 88,000 km². Desarrollaron árboles de regresión, los cuales consideraron que tienen la misma respuesta hidrológica, donde las ecuaciones de regresión lineal múltiple se desarrollaron en cada región para obtener el modelo regional.

Mather y Johnson (2015) realizaron predicción de eventos de turbidez de corrientes empleando una combinación de aglomerados con árboles de regresión. La turbidez es una característica importante al determinar la calidad del agua, en análisis de sedimentos, en análisis de metales, es una característica importante al determinar, debido a que se influye o tiene relación con varios factores, como la forma del hidrograma y humedad antecedente.

Ragetti *et al.* (2017) realizaron un estudio para la predicción de eventos de avenidas repentinas en zonas montañosas no aforadas esto debido a los daños ocasionados por crecientes repentinas en zonas montañosas. El objetivo del estudio consistió en la búsqueda de parámetros descriptivos de transferencia para las zonas no aforadas mediante la metodología CART. Los autores evaluaron la efectividad del modelo CART con el método convencional de similitud física y climática.

9.5. Métodos estadísticos de delimitación de regiones homogéneas

9.5.1. Métodos de agrupación por fecha de ocurrencia: similitud estacional

Este método comenzó a aplicarse a inicios de los años 70 (Campos-Aranda, 2014 a, b) y está basado en el comportamiento estacional de alguna variable hidrológica. La estacionalidad, implica la delimitación de regiones homogéneas a partir de las fechas de ocurrencia. La efectividad en el empleo del método de similitud de crecientes a partir de un índice de estacionalidad puede garantizar la similitud en la respuesta hidrológica gen una cuenca, lo que implica un análisis de la distribución temporal y un análisis de semejanza en características fisiográficas y meteorológicas (Campos-Aranda, 2014a).

A inicios de la década de 1996, comenzó a emplearse con mayor frecuencia el análisis de estacionalidad de crecientes como una alternativa frente a los métodos tradicionales (Ouarda *et al.*, 2006).

Algunas ventajas de este método son la siguientes (Ouarda *et al.*, 2006):

- La principal ventaja de este método es que la similitud se describe con base en las fechas de ocurrencia, por lo que se podría asumir que la delimitación en función de los datos está libre de errores (Ouarda *et al.*, 2006).
- Otra ventaja del método es la factibilidad al separar distribuciones mezcladas de crecientes.

La similitud estacional ha sido empleada en el área de recursos hídricos, manejo de almacenamientos, pronóstico de crecientes y protección contra inundaciones (Ouarda *et al.*, 2006; Cundelrick *et al.*, 2004).

Existen tres enfoques de similitud que se han desarrollado, el enfoque gráfico basado en umbral sobre el pico, identificado como POT por sus siglas en inglés, método basado en estadísticos direccionales, denominados XY, y un método basado en frecuencias relativas de ocurrencia de avenidas, identificados como RF. También existe otro enfoque que es el de disimilitud estacional.

Básicamente este enfoque se lleva a cabo de la siguiente forma, se obtienen los índices de estacionalidad, se agrupan por método jerárquico y se revisa la homogeneidad mediante técnica de Momentos-L (Sarhadi y Modarres, 2011).

De acuerdo con Zrinji y Burn (1996) y Burn (1997), los índices de estacionalidad expuestos (DMC y r) pueden ser empleados para obtener la disimilitud entre cuencas (Campos-Aranda, 2014b). El índice de disimilitud de cuencas, IDE, combina al DMC y al r en una distancia Euclidiana ocurrencias a lo largo de todo el año.

Cunderlik *et al.* (2004) examinaron y compararon la incertidumbre en la estimación de la estacionalidad de crecientes, empleando datos de registros sintéticos con caudales máximos anuales, así como valores sobre cierto umbral de referencia. El estudio realizado consistió en determinar por medio de estadística direccional y por medio de la distribución mensual de la frecuencia los eventos de ocurrencia además, con las muestras sintéticas con 10 diferentes longitudes de registro se generaron distribuciones de estacionalidad uni, bi, y tri-modal.

Ouarda *et al.* (2006) realizaron una comparación de los tres métodos con el método tradicional de características fisiográficas con similitud geográfica en Canadá, con datos de caudal diario. Los resultados de los métodos en orden de desempeño en el siguiente orden: POT, RF y XY, en cuanto a la comparación de POT con el método tradicional, hubo una leve mejoría del método POT.

El objetivo del estudio de Laaha y Blöschl (2006) fue investigar el valor de los índices de estacionalidad en la regionalización de caudales mínimos. El proceso de regionalización empleado por los autores fue el modelo de regresión múltiple basado en las características de la cuenca y en los índices de estacionalidad, realizaron una evaluación mediante validación cruzada, con el propósito de evadir los problemas en la escala, se estandarizó el valor del caudal mínimo con el área de captación, por lo que los autores trabajaron con caudales específicos. Emplearon 31 características descriptivas de la cuenca a partir de siete clases: área de drenaje, elevación, pendiente topográfica, precipitación, geología, uso de suelo y densidad de drenaje.

En cuanto a la comparación de este tipo de medidas de similitud Sarhadi *et al.* (2011) realizaron una comparación en datos de caudal diario, encontrando que el método POT-DS conduce a óptimos resultados a diferencia de método RF y XY, pero en cuanto a la comparación de éste con el método con el método convencional de empleo de características descriptivas llevó a una mejoría leve en la estimación de cuantiles. El método óptimo resultó ser el método POT-CS, que llevó a una mejora del método POT-DS.

9.5.1.1. Medidas de estacionalidad

Este método surge debido a la importancia del análisis de la variabilidad anual del caudal, es por esta razón que el método se enfoca en dos estaciones verano e invierno, la primera etapa comprende el análisis del caudal del primero de abril al 30 de noviembre, mientras que la estación de invierno analiza los registros del primero de diciembre al 31 de marzo. Este método ha sido aplicado en caudales mínimos, por lo que se expresa de la siguiente forma (Laaha y Blöschl, 2006):

$$SR = \frac{q_{95s}}{q_{95w}} \quad (9.30)$$

donde los valores de $SR > 1$ indican la presencia de un régimen de caudal de invierno y un valor de $SR < 1$ indican la presencia de un régimen de caudal de verano.

Este método es útil en la inspección visual para la identificación de los patrones de caudal.

9.5.1.2. Similitud estacional, basada en frecuencias relativas de crecientes

Este método consiste en la identificación y agrupamiento de ocurrencias de fechas de crecientes que representan la información detallada de la temporalidad de crecientes. Este enfoque se utiliza generalmente para cada mes. En este método, las fechas de los sucesos de inundación se agrupan en meses y los FR de las ocurrencias de inundaciones se calculan por cada mes. Se realiza un ajuste al convertir el tiempo en ángulos. Mardia (1972) propone un método basado en la frecuencia en la que las frecuencias se pueden ajustar de modo que correspondan a 360 días con todos los meses que tiene la misma longitud y 1° corresponderá a 1 día. Es decir, en los meses con 31 días, la frecuencia se indica como 30/31 mientras que para indicar la frecuencia de febrero se indica como 30/28 o 30/29 para años bisiestos. Por lo que el año es reducido a 360 días pero la suma (S) de las frecuencias originales RF_i no es igual a la suma (S') de las frecuencias ajustadas RF'_i . Para preservar la suma S , al final se ajustan las frecuencias que son obtenidas RF'_i multiplicándolas por S/S' (Sarhadi y Modarres, 2011).

$$\sum_{i=1}^{12} RF_i = \sum_{i=1}^{12} RF'_i \frac{S}{S'} \quad (9.31)$$

El uso de este método puede representar los patrones de estacionalidad y similitud entre dos sitios basados en la frecuencia relativa FR. La similitud entre dos cuencas ($Sim_{i,j}$) se expresa:

$$Sim_{i,j} = \sqrt{\frac{\sum_{k=1}^{12} (RF_i^k - RF_j^k)^2}{12}} \quad (9.32)$$

donde RF_i^k y RF_j^k son las frecuencias relativas ajustadas RFs de la ocurrencia de avenidas en las cuencas i (j) para el mes k .

9.5.1.3. Similitud estacional: estadísticos direccionales

En este enfoque la estacionalidad de crecientes se logra por medio de la conversión de fechas de ocurrencia de crecientes a estadísticos direccionales, es decir, es la conversión de un día juliano de la fecha de ocurrencia a un valor angular. Cada una de las fechas individuales de las ocurrencias se define como una variable direccional al convertir un día juliano (JD_i) a un valor angular (θ_i) empleando la siguiente ecuación (Sarhadi y Modarres, 2011):

$$\theta_i = JD_i \frac{2\pi}{ND} \quad (9.33)$$

donde $0 \leq \theta_i \leq 2\pi$, además ND es el número de días en el año ($ND = 365$ o 366 en año bisiesto y el primer día de enero corresponde al día 1), θ_i es el valor angular que representa la fecha de la inundación (avenida) del evento i en radianes. Por lo tanto la fecha de ocurrencia representa un vector con magnitud unitaria y una dirección dada por θ_i . La dirección media, $\bar{\theta}$ (o el día medio de las ocurrencias de las avenidas) conocido como MDF , por sus siglas en inglés, se calcula de la siguiente forma (Sarhadi y Modarres, 2011):

$$\bar{\theta} = \tan^{-1} \left(\frac{\bar{y}}{\bar{x}} \right) \quad (9.34)$$

$$MDF = \bar{\theta} \left(\frac{ND}{2\pi} \right) \quad (9.35)$$

donde $\bar{x} \neq 0$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i) \quad (9.36)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n \sin(\theta_i) \quad (9.37)$$

y n es el número de muestras dado un sitio, la dirección media $\bar{\theta}$, representa la ubicación direccional de la magnitud de la muestra.

Una medida de dispersión adimensional de las fechas individuales de las fechas de ocurrencia alrededor del valor medio se puede definir como:

$$\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2} \quad (9.38)$$

Donde el valor de \bar{r} toma valores con $0 \leq \bar{r} \leq 1$, si el valor de \bar{r} está cercano a 1, indica que los eventos en la muestra se encuentran estrechamente agrupados en torno a la dirección media, mientras que el valor cercano a cero indica que la variabilidad en las fechas de ocurrencia es muy alta.

La dirección media y la medida de la varianza definidas en coordenadas polares por $\bar{\theta}$ y \bar{r} pueden ser representadas en un plano cartesiano, por el promedio de las coordenadas \bar{x} y \bar{y} . De esta manera las coordenadas pueden describir la similitud entre los sitios, a través de la distancia Euclidiana.

$$D_i S_{i,j} = \sqrt{(\bar{x}_i - \bar{x}_j)^2 + (\bar{y}_i - \bar{y}_j)^2} \quad (9.39)$$

Con $0 \leq \bar{r} \leq 1$, donde \bar{x}_i y \bar{y}_i se definen a partir de la ecuación (9.36) y (9.37).

Teniendo la medida de similitud $D_i S_{i,j}$ se delimitan las regiones homogéneas, por medio de un análisis de aglomerados por método de Ward, con los datos del vector medio r (Ouarda *et al.*, 2006; Sarhadi y Modarres, 2011), donde la validez de la formación de grupos se puede realizar por medio del coeficiente de correlación cofenético (Ouarda *et al.*, 2006).

9.5.1.4. Similitud estacional: *Post-DS Method*

La estacionalidad de crecientes se ha empleado para evaluar la similitud entre los grupos homogéneos formados, además con la determinación de la estacionalidad es útil en el pronóstico de caudales, en protección y manejo de cuencas, manejo en inundaciones y operación de vasos de almacenamiento (Cunderlik *et al.*, 2004).

El método POT, se refiere a la extracción de series de duración parcial por medio de un valor base o un valor umbral. Por esta razón ha sido un método que posee grandes ventajas respecto a aquellos métodos que impliquen trabajar con series de máximos anuales, esto se debe principalmente a una menor pérdida de información al fijar un valor base y con ello considerar un rango más amplio de datos, a diferencia de aquellos métodos que impliquen únicamente la toma de datos de máximos anuales (Sarhadi y Modarres, 2011).

En el estudio de análisis regional de frecuencias basado en el método POT-DS (*data series*), (Sarhadi y Modarres, 2011) emplearon un enfoque propuesto por Ouarda *et al.* (1993).

$$\xi_v = 0; \quad Q_v \leq Q_B \quad (9.40)$$

$$\xi_v = Q_v - Q_B; \quad Q_v \geq Q_B \quad (9.41)$$

donde Q_B es el nivel base, Q_v es el caudal en el río en el tiempo τ_v y ξ_v es la excedencia en el tiempo τ_v .

Para definido un nivel de truncamiento para la delimitación de la estacionalidad de inundación similar basado en el método POT, se selecciona la capacidad de carga del canal de río como el nivel de umbral. Sarhadi y Modarres (2011) proponen el cálculo de la capacidad del río a través del software HEC-GEO RAS mediante de la sección transversal, con una resolución del modelode alta calidad.

Una propuesta para este método sin tomar un valor de referencia de caudal de umbral, es la siguiente (Lang *et al.*, 1999; Ouarda *et al.*, 2001), donde (1) dos picos consecutivos de inundación tienen que ser separados por al menos tres veces el tiempo promedio para subir; y (2) la descarga entre dos picos consiguientes de inundación tiene que descender por debajo del valor de los dos tercios del menor de los dos picos.

En este método el año se divide en n estaciones, donde se asume que las n estaciones están distribuidas de forma idéntica, independientemente de su año de ocurrencia. “por lo tanto, con n estaciones, se obtienen n diferentes funciones de distribución para ajustarse a las superaciones registradas. El modelo estacional discreto hace dos suposiciones acerca de las características de las inundaciones. “La primera hipótesis es que diferentes tipos de tormentas producen diferentes características de inundación de una estación a otra, y la segunda es que, dentro de cada estación, la variación en la magnitud de la inundación es insignificante” (Sarhadi y Modarres, 2011).

En el modelo POT-DS el año se divide en n estaciones, donde las excedencias pertenecen a la K -ésima estación, es decir, $K= 1, \dots, n$ y se considera que son idénticamente distribuidos, independientemente de su año de ocurrencia. Por lo tanto, con n estaciones, se obtienen n diferentes funciones de distribución que se ajustan a los excedentes históricos. El modelo de estación discreto hace dos suposiciones acerca de las características de las inundaciones. La primera suposición es que los diferentes tipos de tormenta producen diferentes características de inundación de una estación a otra, y la segunda es que, dentro de cada temporada, la variación en la magnitud de las inundaciones es insignificante. La aplicación de este modelo consiste en el trazado del número medio de avenidas que exceden el umbral. Por lo tanto, este método gráfico puede ilustrar la variación estacional de las inundaciones y permite la agrupación de sitios en regiones gráficas que son homogéneas en la distribución de inundación estacional. Este método ayuda a la identificación de una estación específica que incluya eventos de inundación significativas. Los resultados obtenidos en este método son evaluados mediante las gráficas de excedencias trazadas en cada sitio de la región. Después de la extracción de la media del número de excedencias en cada estación, los sitios se agrupan de acuerdo a estas características (el número medio de eventos de inundación). La región de estudio se subdivide después en subregiones más pequeñas de acuerdo con la duración y ubicación de los eventos de inundación en una escala de tiempo (Sarhadi y Modarres, 2011).

Para buscar la homogeneidad en los resultados, es necesario realizar una evaluación de las gráficas de las excedencias de cada sitio en la región de estudio. El procedimiento consiste en la extracción en cada estación del número medio de excedencias (con respecto al valor base), los sitios se agrupan de acuerdo a sus características, es decir, al número medio de eventos, por lo que la región a su vez se subdivide de acuerdo a la duración y ubicación de

las estaciones en una escala de tiempo común, por medio de un método como el análisis de aglomerados (Sarhadi y Modarres, 2011).

9.5.1.5. Similitud estacional POT-CS METHOD

Existe otra variación del método POT, el cual está orientado en la identificación de la temporada crítica, es decir, se basa en las gráficas realizadas en un POT-DS para una temporada específica a través del número medio de excedencias sobre un umbral elegido, normalmente se elige aquella temporada que incluya el mayor número de excedencias, es decir, como juicio visual se elige la temporada con mayor pendiente (Sarhadi y Modarres, 2011).

El procedimiento (Sarhadi y Modarres, 2011) consiste en seleccionar la temporada crítica, excluir estos datos y por medio de la metodología de heterogeneidad de Momentos-L se revisa la homogeneidad del periodo crítico, mediante estas técnicas también se pueden obtener los diagramas de Momentos-L mediante el diagrama asimetría vs curtosis y asimetría contra variación para inspeccionar visualmente el grado de heterogeneidad.

9.5.2. Análisis de componentes principales (ACP)

El método ACP permite el conocimiento de las variables más importantes, por medio de la cantidad de varianza, además mediante el gráfico de los componentes principales que mejor representen al conjunto de datos originales es factible la delimitación de regiones homogéneas. Este método, aunque es meramente visual, ofrece al investigador una primera aproximación del comportamiento regional.

9.5.3. Análisis de Correspondencia (CA)

El Análisis de Correspondencia (AC) es una técnica particular del Análisis de Componentes Principales (ACP), normalmente éste es empleado para la representación de la dependencia entre variables (Bobée *et al.*, 1996). La diferencia entre AC y un ACP radica en el tratamiento a los datos, en un AC los datos son transformados en frecuencias que corresponden a las coordenadas espaciales de las observaciones, donde la normalización define la medición de Chi cuadrada como la principal característica de un AC, el principal objetivo en un AC es causar que las estructuras en las tablas de las variables descriptivas sean más claras (Bobée *et al.*, 1996).

Algunos autores apuntan que la técnica AC es una herramienta más efectiva y poderosa que la técnica empleando ACP, debido a Chi-cuadrada que es empleada para la estabilidad “ La distancia ponderada satisface el principio de distribución equivalente y asegura la invarianza de los resultados relativos para su codificación” (Bobée *et al.*, 1996).

9.5.4. Regiones de influencia

Este método surgió al hacer frente a la problemática de los métodos que involucraban regiones contiguas en el espacio, como en el caso del método de residuales, en estos métodos donde se obtienen regiones contiguas es factible la variabilidad a pequeña escala, por lo que en datos extremos la delimitación puede no ser precisa en estaciones que se encuentren a una gran distancia (Schumann, 2011).

En este método se asume que por medio de una medida de distancia existe similitud entre las estaciones que se encuentren dentro de la distancia acordada (Schumann, 2011). Se define de la siguiente forma (Burn, 1989):

$$D_{i_0} = \left[\sum_{l=1}^L W_l (Y_l(x_i) - Y_l(x_0))^2 \right]^{1/2} \quad (9.42)$$

donde D_{i_0} es la distancia entre la estación 0 y la estación i , W_l es el peso de la importancia relativa del atributo l del conjunto de L atributos y $Y_l(x_i)$ es el valor del atributo l de la estación i , tomando en cuenta el peso de la variable en cuestión así como el índice estacional.

En un análisis de regiones de influencia se asume que cada sitio genera una región homogénea, para el análisis de regiones de influencia es necesaria la medición de la distancia euclidiana. Una función ponderada es definida para reflejar la importancia relativa de cada sitio en la región de influencia para la estimación regional de los eventos de inundación extrema en el sitio, además la curva de crecimiento regional se pondera de acuerdo a la proximidad al sitio de interés (Rao y Srinivas, 2006; Malekinezhad *et al.*, 2011). Existen tres distintos enfoques dependiendo de la información con la cual se cuente, en la versión original (Burn, 1989) se sugiere un conjunto de atributos hidrológicos y fisiográficos de la estación aforada, como segundo enfoque para los sitios aforados, los atributos fisiográficos, hidrológicos y climatológicos, mientras que para los sitios no aforados existen otros enfoques basados en climatología y fisiografía.

En este método la distancia al sitio objetivo no debe exceder un valor de umbral de distancia (Rao y Srinivas, 2006; Malekinezhad *et al.*, 2011).

En este método se debe definir el conjunto de atributos descriptivos, así como la función de peso de cada estación dentro de una región (Escalante-Sandoval y Reyes-Chávez, 2005).

El método se expresa de la siguiente forma (Escalante-Sandoval y Reyes-Chávez, 2005):

$$D_{jb} = \left[\sum_{i=1}^p (C_j^i - C_b^i)^2 \right]^{1/2} \quad (9.43)$$

$$C_j^i = \frac{X_j^i}{S(X^i)} \quad (9.44)$$

$$C_b^i = \frac{X_b^i}{S(X^i)} \quad (9.45)$$

donde p es el número de atributos considerados, C_j^i y C_b^i son los valores estandarizados del atributo i para los sitios j y la estación base b , X_j^i es el valor del atributo i del sitio j , X_b^i son los valores del atributo i de la estación base b , $S(X^i)$ es la desviación estándar del atributo i considerando los sitios j y la estación base b .

El siguiente paso (Escalante-Sandoval y Reyes-Chávez, 2005) consiste en la identificación de la región de influencia, por medio de la asignación de un valor de tolerancia que acote la medida de la distancia. La propuesta de una función ponderada, reflejará la importancia de la estación respecto a la estación base en la estimación de la variable de interés.

$$w_{jb} = 1.0 - \left[\frac{\hat{D}_{jb}}{\wp} \right]^C \quad (9.46)$$

$$\sum w_{jb} = 1.0 \quad (9.47)$$

donde j es el número de estaciones en la región, w_{jb} es el factor de peso de la estación j en la región de influencia del sitio b , \wp y C son los parámetros a estimar, donde $C > 0$ y \wp

debe ser mayor o igual al valor de la tolerancia. Si ϕ es igual a la tolerancia, las estaciones en el umbral no tendrán contribución para la determinación de los gastos extremos en el sitio base. Los valores mayores incrementarán la ponderación de todas las estaciones incluidas en la región. La constante C determina la tasa de decremento de los pesos con la distancia (Escalante-Sandoval y Reyes-Chávez, 2005).

El paso siguiente consiste en una verificación de homogeneidad por medio de las medidas de discordancia (Escalante-Sandoval y Reyes-Chávez, 2005).

Campos-Aranda (2014a) aplicó un método que usa la estacionalidad de las crecientes, como medida de similitud de la respuesta hidrológica de la cuenca donde como resultado se obtuvieron dos grupos: de fecha tardía y promedio de ocurrencia de crecientes.

Una desventaja identificada por Bobeé y Rasmussen (1995) fue la no existencia de una solución matemática ante la selección y ponderación variables que intervienen en el algoritmo (Rao y Srinivas, 2006; Malekinezhad *et al.*, 2011).

9.5.5. Análisis de Correlación Canónica

En un ARF desarrollado con ACC “Las cuencas de drenaje se representan como puntos en los espacios de pares de variables canónicas no correlacionadas relacionadas con inundaciones y pares de variables canónicas no correlacionadas relacionadas con la cuenca para examinar la similitud en los patrones de puntos correspondientes en estos espacios” (Bobeé y Rasmussen, 1995; Rao y Srinivas, 2006).

El ACC es un análisis importante dentro del análisis estadístico multivariado, forma parte primordial dentro del análisis factorial discriminante, de la regresión multivariada, y el análisis de correspondencia, con el ACC es posible establecer las interrelaciones entre dos grupos de variables, por medio de este método se establecen combinaciones lineales de las variables que están más correlacionadas del primer grupo con las variables de un segundo grupo. En la literatura se tiene registrado que Decoursey en el año 1973 fue el primero en emplear las técnicas de ACC en hidrología (Ouarda *et al.*, 2001). El ACC se ha empleado para pronóstico de lluvias costeras mensuales, escurrimiento estacional del deshielo, así como para la predicción de precipitación (Ouarda *et al.*, 2001), también ha sido empleado para la identificación de variables relacionadas con el impacto causado por acciones

antrópicas en este caso por la implementación de presas (Matteau *et al.*, 2009), así como en la identificación de sitios con régimen similar de caudal para un sitio objetivo (Bobée *et al.*, 1996).

Una desventaja de este método es que los patrones de reconocimiento de grupos están basados en un juicio visual (Rao y Srinivas, 2006).

Una breve descripción del método se puede ejemplificar con el estudio desarrollado por GREHYS (1996) quienes clasificaron las variables en dos grupos, el primer grupo fue de las variables hidrológicas para resumir el comportamiento probabilístico de las avenidas y el segundo grupo con variables para caracterizar la geomorfología y meteorología de las cuencas. Donde las variables del grupo 1 se denomina vector Y, mientras que el segundo grupo, vector X. Por lo que el objetivo de esta metodología es que a partir del vector X, se realicen inferencias para el vector Y.

Otra aplicación de este método se realizó en México en la cuenca del río Balsas (Ouarda *et al.*, 2008), el objetivo del estudio fue desarrollar una comparación de métodos de delimitación validados por medio de la técnica de *Jack-knife*, error medio cuadrático y coeficiente de Nash. Los métodos de delimitación que compararon son los siguientes: método de clasificación jerárquica ascendente, análisis de correlación canónica, análisis de correlación canónica libre de parámetros y kriging canónico, los resultados apuntaron que el método de clasificación jerárquica ascendente conduce a cuantiles menos sesgados, sin embargo, los métodos relacionados con ACC conducen a un mejor desempeño del modelo al obtener menor valor del error cuadrático. Sin embargo, en todos los métodos se obtuvo un resultado común el cual consistió en que el desempeño de un modelo está en función de la cantidad y calidad de los datos.

9.5.6. Agrupación por residuales

En el método de delimitación de regiones por patrón de residuales las regiones se forman por medio de la extracción de los residuos del modelo de regresión empleando los valores de los signos positivos y negativos, por lo que las regiones formadas son espacialmente contiguas.

Este método parte de un modelo de regresión regional que relaciona el cuantil de inundación con las características descriptivas de la cuenca en cada uno de los puntos medidos y

observados, por lo que las regiones se forman por medio de la extracción de los residuos, los cuales tendrán un signo positivo o negativo (Thomas y Benson, 1970; Wandle, 1977; Glatfelter, 1984; Choquette, 1988, Rao y Srinivas, 2006).

Este método es empleado en el Servicio Geológico de los Estados Unidos (USGS, por sus siglas en inglés) (Rao y Srinivas, 2006). Rao y Srinivas (2006) señalaron que con el método de patrones residuales se obtienen resultados muy arbitrarios, donde en muchas de las ocasiones los resultados concuerdan con las fronteras espaciales de las características físicas descriptivas. Otra observación hecha por Laaha y Blöschl (2006) es la poca exactitud ya que el método se aplica a todo el dominio de interés, por lo que carece de sentido físico. Otro de los comentarios realizados de este método (Schaumann, 2011) consiste en la delimitación de regiones contiguas pueden representar una configuración compleja. La ventaja de este método es que mejora el modelo de regresión global (Laaha y Blöschl, 2006; Schnier y Cai, 2014).

En la búsqueda de la revisión de la literatura existen diversos métodos de delimitación de regiones siendo pocos los estudios relacionados con esta técnica, de forma general las etapas de este método son las siguientes (Laaha y Blöschl, 2006):

- Desarrollar una regresión escalonada (*stepwise regression*), obteniendo un modelo de regresión global.
- Graficar los residuales a partir del modelo de regresión global en un espacio geográfico.
- Si los patrones residuales son aparentes, delinear las regiones contiguas de los residuales de signo y magnitud similar.
- La zona no aforada se localiza en una de las regiones por su ubicación geográfica debido a que las regiones son espacialmente contiguas.
- La magnitud del caudal en la zona no aforada se estima por ecuaciones de regresión múltiple entre los datos observados de caudal y las características descriptivas de la cuenca en cada una de las regiones formadas independientes.

9.6. Cantidad óptimo de los grupos formados

Otro aspecto importante en la delimitación de cuencas es definir el número óptimo de grupos homogéneos y éste debe ser evaluado, en el análisis de agrupamiento se realizan tres pruebas estadísticas el criterio cúbico de agrupamiento (CCC por sus siglas en inglés, *Cubic Clustering Criterion*), la prueba pseudo-F- *statistic* y la desviación estándar media cuadrática y una regla de detención (Sanborn y Bledsoe, 2006). Otra prueba empleada es mediante el

criterio de la raíz media cuadrática del error de la desviación estándar y R-Cuadrada (Dikbas *et al.*, 2013).

En el análisis de aglomerados un tema sin resolver aún, es la cantidad óptima de grupos que deben ser formados, por lo que se proponen diversas pruebas estadísticas para detener la formación de grupos, estas pruebas estadísticas son las siguientes: desviación estándar del error medio cuadrático, R-cuadrada, y la semi-parcial R-cuadrada (Isik *et al.*, 2008).

Cuando se emplea el análisis de aglomerados por el método de Ward, la cantidad óptima de grupos formados se puede seleccionar por medio de los coeficientes de fusión. Este coeficiente se estima al considerar las distancias entre los grupos formados etapa por etapa, el valor del coeficiente de fusión se incrementa a medida que decrece el número de agrupamientos formados. Aunque para la elección del número óptimo de agrupaciones es arbitraria, es factible la construcción de la gráfica número de agrupaciones en el eje de las ordenadas, y en el eje de las abscisas el coeficiente de fusión, y mediante la comparación de las líneas de tendencia, una del lado derecho y otra del lado izquierdo es factible determinar la cantidad óptima de agrupaciones (Yoo *et al.*, 2011).

La validación de Silhouette o validación de silueta ha sido empleada en métodos de delimitación por K-medias. El método se define como sigue (Rousseeuw y Silhouette, 1987):

$$S_i = \frac{(b_i - a_i)}{\max\{a_i, b_i\}} \quad (9.48)$$

donde a_i es la distancia promedio del i-ésimo punto dentro del cluster, mientras que b_i es la distancia promedio del i-ésimo punto a los puntos de otro cluster. El valor de S_i estará dentro del rango de -1 a 1. En caso de obtener $S_i = 1$ significa que se obtuvo un cluster bien construido, si se obtiene $S_i = 0$ la muestra se puede asignar al clúster más cercano y si se llega a tener un valor de $S_i = -1$ significa que la clasificación se ha hecho incorrectamente. El número final agrupaciones se determina con el promedio general de todos los S_i calculados, el valor más grande de S_i determina el número representativo de agrupaciones.

9.6.1. Validez del agrupamiento

En la validez del agrupamiento jerárquico se recomienda emplear el coeficiente de correlación cofenético, el cual compara “la información de distancia generada por el algoritmo de enlace con la información de distancia calculada por la función de distancia en pares” (Ouarda *et al.*, 2006). Es un coeficiente que mide la distorsión en la clasificación “indica cómo los datos se ajustan a la estructura sugerida por la clasificación” (Ouarda *et al.*, 2006).

9.7. Verificación de regiones homogéneas

Para la revisión de la homogeneidad regional, en la revisión de la literatura se encuentran dos métodos (Campos-Aranda, 2012), la revisión de los estadísticos de asimetría y curtosis generados a través de la metodología de momentos L, propuestos por Hosking y Wallis (1997), así como las propiedades estadísticas adimensionales, con prueba de Langbein (Campos-Aranda, 2012), que se verá en capítulos siguientes.

9.7.1. Métodos gráficos: curvas de Andrews

Un método gráfico que ha sido empleado con frecuencia en la verificación de regiones homogéneas son las curvas de Andrews.

El método de las gráficas de Andrews, está justificado matemáticamente, los cuales poseen ciertas propiedades de invarianza (Khattree y Naik, 2002).

Khattree y Naik (2002) realizaron modificaciones a la metodología de las curvas de Andrews, donde los autores se basaron en la conexión que se puede presentar de las curvas de Andrews con los componentes principales, con la aplicación en el análisis de correspondencias.

Los autores (Garcia-Osorio y Fyfe, 2005) realizaron una herramienta computacional para analizar en un espacio tridimensional la formación de posibles grupos a través de las curvas de Andrews, bajo el concepto teórico de las curvas de Andrews, curvas de Wegman y una extensión de las curvas de Andrews, por lo que construyeron como base las curvas para un espacio bidimensional, construyendo por medio de éstas un método para tres dimensiones.

La metodología de las curvas de Andrews es una técnica ideal para visualizar la comparación de grupos homogéneos. Los grupos de cuencas de captación similares aparecerán como una banda de curvas estrechamente espaciados, perteneciendo a otro grupo las cuencas que no

estén cerca. De manera que puede emplearse esta metodología para evaluar otras técnicas de agrupamiento, donde la clasificación de grupos se identificará por la forma de las bandas de las curvas.

La función de Andrews también permite determinar la pertenencia a un grupo para las nuevas cuencas. La curva asociada con la nueva curva se compara con los diferentes grupos de curvas para elegir la más apropiada. Los autores Laaha y Blöschl (2006) apuntan que la verificación por medio de curvas de Andrews no es una opción viable cuando se trata de una validación cruzada en un estudio para un conjunto grande de datos, donde se requiere de un desarrollo automático en procesos iterativos. Por lo que sí podría emplearse cuando el conjunto de datos sea para muestras pequeñas.

9.7.2. Medidas de discordancia

La medición de la discordancia se aplica en un análisis preliminar de datos, es decir, en la etapa de análisis exploratorio de datos y en una segunda etapa, posterior a la delimitación de regiones homogéneas aplicando la metodología de medidas de discordancia a cada región creada.

Las medidas de discordancia son empleadas con el fin de descartar datos atípicos considerando aquellas estaciones que son altamente discordantes de la totalidad del grupo de datos (Dodangeh *et al.*, 2014).

Las ecuaciones y el procedimiento de cálculo se encuentran en el capítulo 8 ANÁLISIS REGIONAL DE FRECUENCIAS ARF, en la sección 8.2 Etapas del Análisis Regional de Frecuencias basado en Momentos-L número 8.2.1.2 Medidas de discordancia.

9.7.3. Medidas de heterogeneidad con la técnica de Momentos-L

Después de la formación de grupos otra prueba aplicable consiste en las medidas de heterogeneidad, propuestas por Hosking y Wallis (1997), la fundamentación teórica de esta metodología se encuentra en el capítulo 8 ANÁLISIS REGIONAL DE FRECUENCIAS ARF, en la sección 8.3.1 Medidas de heterogeneidad propuestas por Hosking y Wallis.

9.7.4. Pruebas estadísticas: cociente estadístico regional

Además de las pruebas gráficas y de homogeneidad por la técnica Momentos-L, existen pruebas de homogeneidad regional que se fundamentan en los estadísticos de la muestra, para

lo cual de cada región delimitada se estiman la media, desviación estándar y coeficiente de variación (Nobert *et al.*, 2014):

$$\bar{Q}_i = \frac{\sum_{j=1}^{n_j} Q_{ij}}{n_j} \quad (9.49)$$

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{n_j} (Q_{ij} - \bar{Q}_i)^2}{n_j - 1}} \quad (9.50)$$

$$CV_i = \frac{\sigma_i}{\bar{Q}_i} \quad (9.51)$$

donde \bar{Q}_i es el caudal medio en el sitio i [m^3/s], Q_{ij} es la tasa del caudal de la estación i en la región j en [m^3/s], σ_i es la desviación estándar y CV_i es el coeficiente de variación.

La media regional y la desviación estándar se calculan de la siguiente forma (Nobert *et al.*, 2014):

$$\overline{CV} = \sum_{i=1}^N \frac{CV_i}{N} \quad (9.52)$$

$$\sigma_{cv} = \sqrt{\frac{\sum_{i=1}^N (CV_i - \overline{CV})^2}{N}} \quad (9.53)$$

Donde N es el número de sitios dentro de la región homogénea.

Teniendo los anteriores parámetros se estima la siguiente expresión con la cual se puede declarar si una región es homogénea o no lo es (Sine y Ayalew, 2004; Nobert *et al.*, 2014):

$$CC = \frac{\sigma_{cv}}{\overline{CV}} \leq 0.3 \quad (9.54)$$

9.7.5. Test de Wiltshire

El test de Wiltshire está basado en el coeficiente de variación (CV), es decir, se obtiene mediante la relación de la desviación estándar del sitio j y su respectiva media. Para mayor información referirse al estudio realizado por Campos-Aranda (2010). El test se realiza para cada sitio de la región en estudio (Campos-Aranda, 2010a):

$$CV_j = \frac{S_j}{x_j} \quad (9.55)$$

donde n_j es la longitud de registro de datos del sitio j , se establece como U_j la varianza muestral del CV_j

$$U_j = \frac{V_R}{n_j} \quad (9.56)$$

donde V_R es la varianza regional del CV y a su vez, la variación total del número del CV dentro de la región de N sitios se designa por el estadístico S (Wiltshire, 1986a; Campos-Aranda, 2010b), que se caracteriza por una distribución χ^2 con $\nu = N - 1$ grados de libertad:

$$S = \sum_{j=1}^N \frac{(CV_j - CV_p)^2}{U_j} \quad (9.57)$$

Siendo CV_p , el valor ponderado del CV de la región, calculado de la siguiente forma:

$$CV_p = \frac{\sum_{j=1}^N \frac{CV_j}{U_j}}{\sum_{j=1}^N \frac{1}{U_j}} \quad (9.58)$$

Los valores críticos de χ^2 con un nivel de significancia del 5%, es decir, con una confiabilidad del 95% se encuentran en la siguiente Tabla 9-1 (Campos-Aranda, 2010a)

Tabla 9-1. Valores críticos de χ^2 (Campos-Aranda, 2010a)

ν	χ_c^2	ν	χ_c^2	ν	χ_c^2	ν	χ_c^2
6	12.60	10	18.30	14	23.70	18	28.90
7	14.10	11	19.70	15	25.10	19	30.10
8	15.50	12	21.00	16	26.30	20	31.40
9	16.90	13	22.40	17	27.60	21	32.70

La región es homogénea si $s < \chi_c^2$ y en caso contrario es heterogénea (Wiltshire, 1986a, 1986b; Rao y Hamed, 2000; Campos-Aranda, 2010b)

Continuando con el cálculo de la varianza regional V_R , se calculada con la siguiente ecuación:

$$V_R = \frac{1}{N} \sum_{j=1}^N n_j \cdot v_j \quad (9.59)$$

donde a su vez, v_j , es la estimación de la varianza truncada, calculada de la siguiente forma:

$$v_j = \frac{(n_j - 1)}{n_j} \sum_{i=1}^N \left[CV_{n-1}^i - \left(\frac{1}{n_j} \sum_{l=1}^N CV_{n-1}^l \right) \right]^2 \quad (9.60)$$

donde se asume que se tienen x_1, x_2, \dots, x_n que forman los n_j datos y que además cumplen con la característica de ser independientes e idénticamente distribuidos en cada sitio, por el hecho de pertenecer a una región homogénea. Por lo que CV_{n-1} será el coeficiente de variación calculado de una muestra de tamaño $n_j - 1$, consistente en los $x_1, x_2, x_{i-1}, x_{i+1}, \dots, x_{n_j}$ datos que corresponden a una serie de datos con la i -ésima observación suprimida.

9.7.6. Test de Langbein

Esta prueba es parte del procedimiento de cálculo del método regional de la Avenida índice, se encuentra la descripción en el capítulo 10 MÉTODOS REGIONALES en el subcapítulo 10.1 Métodos regionales sección 10.1.4 Avenida índice.

9.8. Comentarios y recomendaciones

En este documento se mostraron los distintos métodos para agrupar elementos que en primera instancia se consideran como heterogéneos y que por medio de los índices de proximidad y reglas de agregación permiten unir elementos en función de la similitud entre elementos.

Sin embargo, una delimitación funcional depende en gran medida de las variables que se emplean para la agrupación de elementos, por lo que el investigador deberá seleccionar todas aquellas variables relacionadas con el objeto de estudio, además del tratamiento de estandarización a las variables y su posterior análisis de componentes principales.

Existe una gran variedad de métodos aplicables para la formación de regiones homogéneas. Los métodos gráficos ofrecen una ventaja que consiste en la inspección visual de sitios discordantes a una agrupación de elementos, sin embargo se ha identificado que el método de rostros de Chernoff no ha sido muy eficaz en la formación de grupos homogéneos debido a que en primera instancia se necesitan cerca de 20 variables para realizar la caracterización, por lo que la agrupación puede recaer en variables no significativas, además que el juicio subjetivo del investigador al agrupar los elementos puede repercutir en la transferencia de información hidrológica hacia sitios no aforados. Por otro lado, el método de curvas de Andrews y el método ACP son métodos gráficos funcionales.

Una de las principales recomendaciones consiste en emplear los métodos gráficos como verificación después de la aplicación de algún método estadístico, ejemplos de éstos son el método de curvas de Andrews y el ACP, éste último mediante el gráfico de relación de componentes principales entre los ejes que muestren mayor varianza de las variables. Un ejemplo de lo anterior podría ser emplear métodos como el de Ward, K-medias, árbol de regresión o ACP y después realizar una inspección visual por el método de Andrews, con la desventaja que curvas de Andrews debe emplearse para un conjunto de datos pequeño.

En cuanto a métodos estadísticos, el método de agrupación por residuales es poco aplicada, ya que este método proporciona regiones delimitadas contiguas y para efectos de transferencia lo anterior no es recomendable, ya que la respuesta hidrológica no siempre es similar en relación con la continuidad espacial.

Por otro lado, los métodos ampliamente usados son el método de Ward y el de K-medias, el método de Ward se recomienda emplear como una primera aproximación en la búsqueda de la cantidad óptima de grupos, mientras que el método de K-medias se recomienda emplear como método de optimización en la búsqueda de los elementos intra-grupo, para ello se emplea la cantidad de número de grupo obtenido de la primera aproximación (Ward). Las medidas empleadas y los tipos de datos permiten agrupar por medio de similitud física y similitud estacional.

La similitud física también ha sido un enfoque ampliamente usado para la agrupación de elementos, y se han obtenido resultados satisfactorios, se recomienda este enfoque en

estudios relacionados con el análisis de cambio de uso de suelo y en análisis de eventos extremos.

Por otro lado, la agrupación por medio de similitud estacional ha demostrado ser muy efectiva, debido a que se trabaja con las fechas de ocurrencia de las crecientes máximas y fechas de crecientes a partir de cierto valor límite, por lo que la agrupación se basa en la estacionalidad de crecientes o precipitaciones, la aplicación de este enfoque depende del objeto de estudio, si el proyecto está encaminado a la búsqueda de caudal ecológico, análisis de eventos extremos y análisis del caudal base.

Métodos recientes como ACC y Árbol de regresión se han aplicado con éxito, por lo que el ACC se recomienda en estudios enfocados en el análisis de más de una variable dependiente, mientras que el método de Árbol de regresión se recomienda en el análisis de extremos máximos y mínimos bajo un enfoque de similitud física.

El último proceso en la formación de regiones homogéneas es la verificación de homogeneidad, por lo la prueba del cociente de variación es recomendable por ser práctica y de fácil empleo al conjunto de datos estadísticos. Por otro lado, si se emplea el método de la avenida índice, lo más recomendable es el empleo del método del test de Langbein de la versión corregida para verificación de homogeneidad.

10. MÉTODOS REGIONALES

La estimación precisa del caudal en un sitio no aforado ha representado un gran problema debido a la escasez de estaciones de aforo, por lo que la metodología regional para la estimación de crecientes implica la selección de variables descriptivas, delimitación de regiones, verificación de la homogeneidad regional continuando con la etapa de transferencia de información hidrológica, es decir, la asignación de información a un sitio no aforado. La búsqueda en la evaluación de metodologías que permitan la elección del método regional más eficiente ha sido una tarea común (Bobée et al., 1996), desde el año 1986 inició una etapa de desarrollo de procedimientos de regionalización, sin que a la fecha se haya establecido alguno como universalmente aceptable (Rao y Srinivas, 2006; Asadi *et al.*, 2018).

En un proceso de regionalización, primero se asume que se tiene una región heterogénea, por lo que el primer paso es la formación de grupos hidrológicamente homogéneos por medio de la adecuada selección de variables, el siguiente paso en la regionalización consiste en estimar la incertidumbre de los parámetros del modelo y la estimación de cuantiles, un enfoque tradicional es el trazo de los límites de confianza por medio de una aproximación asintótica (Dodangeh *et al.*, 2014).

Los primeros métodos de regionalización estaban basados el desarrollo de las ecuaciones empíricas, las cuales buscaban extrapolar el caudal de sitios aforados a no aforados por medio de parámetros morfométricos y climatológicos de las cuencas (Kite, 1977), otras metodologías estaban basadas en la estimación de los parámetros en cuencas no aforadas mediante técnicas de interpolación, lo anterior por medio de parámetros calibrados de cuencas aforadas más cercanas (Bao *et al.*, 2012).

Un ejemplo de ecuaciones empíricas es el método racional del proceso lluvia escurrimiento (Kite, 1977). Esta ecuación permite estimar el caudal en un sitio no aforado mediante las variables del área de la cuenca, intensidad de la lluvia y un coeficiente de escurrimiento que está en función del tipo y uso de suelo, así como de la cobertura vegetal en el área de la cuenca. Otra de las primeras propuestas (Kite, 1977) fue hecha por Coulson (1967) quien desarrolló una ecuación para estimar el evento de diseño para determinar el caudal anual en la parte sur de Ontario, en EUA, con un conjunto de 59 estaciones y en función de las isolíneas de escurrimiento medio anual, para lo cual desarrolló una ecuación en función del

caudal medio anual, un factor que depende de la distribución de frecuencias y del coeficiente de variación, como método de transferencia de información hidrométrica.

Sin embargo, el inicio del análisis regional comienza con el método de la avenida índice, este método fue desarrollado por Dalrymple (1960). Consta de dos etapas (Kite, 1977), la primera etapa consiste de un proceso regional de frecuencias, mientras que en la segunda etapa se transfiere la información por medio de las características fisiográficas descriptivas de la zona. Las variables propuestas en la metodología original desarrollada por Dalrymple incluyen dos tipos de variables, las fisiográficas y las hidrometeorológicas (Kite, 1979). El procedimiento de regionalización convencional en un análisis regional comprende la calibración de modelo y luego ajuste por regresión entre los parámetros del modelo y las características de las cuencas, utilizando algún tipo de función de transferencia (Singh *et al.*, 2012).

Después del método de la avenida índice, se desarrollaron métodos alternativos como el método de regresión múltiple, método estaciones-años, Box-Cox regional, método de la cuadrícula y métodos de distribución de frecuencias (Kite, 1977), aunque exista una gran variedad de enfoques distintos, desde su publicación, el método de la avenida índice ha sido el más popular (Mishra *et al.*, 2009) y se le han realizado adaptaciones a las pruebas de homogeneidad. Fill y Stedinger (1995) modificaron los valores de aceptación en la curva de verificación regional, otras modificaciones implican la consideración de funciones con dos poblaciones (Gutiérrez-López y Ramírez, 2005).

En el análisis regional existen tres enfoques que son aplicables en un análisis de transferencia de información hidrológica proximidad, similitud física y regresión (Merz y Blöschl, 2004; Oudin *et al.*, 2008; Zhang y Chiew, 2009; Bao *et al.*, 2012), aunque estos métodos implican un proceso de regionalización distinto tienen en común la selección de variables descriptivas de las cuencas aforadas.

En el concepto de similitud de cuencas se asume que la transferencia de información de un sitio aforado a uno no aforado es por medio de cuencas que producen procesos hidrológicos similares (Singh *et al.*, 2012), por lo que la transferencia de parámetros de cuencas aforadas hacia cuencas no aforadas se realiza mediante índices de similitud, un ejemplo de ello son los métodos de similitud de patrones de coseno o por medio de distancia euclidiana (Bao *et al.*, 2012), a su vez, en la similitud física existen dos enfoques: el enfoque de donante único,

donde se puede transferir un conjunto completo de parámetros de una cuenca de medición con los descriptores de captación más cercanos a la cuenca no homogénea (enfoque de donante único) o, alternativamente, el conjunto de parámetros para la cuenca no calibrada se puede estimar usando conjuntos de parámetros de algunas de las cuencas medias más parecidas (enfoque de donantes múltiples).

El enfoque de regresión es el método más popular en el análisis de regionalización hidrológica (*Bao et al.*, 2012), este enfoque consiste en estimar parámetros del modelo para cuencas no aforadas estimando ecuaciones de regresión múltiple que incluyen características descriptivas de las cuencas, éstas pueden ser variables del tipo geomorfológicas, propiedades hidrológicas, topográficas, geológicas, entre otras.

El enfoque de proximidad espacial parte de la hipótesis que las cuencas más próximas tienen un comportamiento similar (*Bao et al.*, 2012), este método ha sido criticado ya que la proximidad espacial no garantiza la similitud hidrológica (*Malekinezhad et al.*, 2011; *Boscarello et al.*, 2015).

Numerosos estudios se han realizado con el objetivo de definir cuál es el mejor enfoque de regionalización, sin embargo, se han obtenido resultados muy distintos (*Heřmanovský y Pech*, 2013). De acuerdo a los estudios realizados por *Bao et al.* (2012) identificaron que el orden de precisión en los métodos es el de proximidad espacial, seguido por similitud física y por último el enfoque por regresión, debido a que se le ha considerado al método de regresión que ha carecido de precisión en la estimación de caudales (*Heřmanovský y Pech*, 2013). Por otro lado, autores como *Heřmanovský y Pech* (2013) identificaron como método óptimo el enfoque de similitud física.

Algunos autores han desarrollado la combinación de enfoques de similitud con proximidad espacial, pero los resultados no han sido cuantitativamente mejores que la aplicación única de la proximidad espacial (*Merz y Blöschl*, 2004; *Zhang y Chiew*, 2009; *Heřmanovský y Pech*, 2013).

En modelos hidrológicos distribuidos se han empleado la comparación de estos tres enfoques, *Zhang y Chiew* (2009) realizaron una investigación de la comparación de enfoques de

regionalización en 210 cuencas en el sureste de Australia, siendo el enfoque de proximidad espacial más preciso que el método de similitud física.

La aplicación de diversas técnicas de regionalización ha llevado a obtener distintos resultados sobre cuál es el enfoque óptimo, sin embargo, Heřmanovský y Pech (2013) identificaron cuales son los principales factores que afectan la obtención de resultados distintos siendo las siguientes “La primera diferencia se debe a la cantidad y tipo de datos, la segunda diferencia en el modelo hidrológico empleado y la tercera causa está relacionada con el hecho de que cada enfoque de regionalización emplea un procedimiento diferente y que los estudios individuales difieren a este respecto” (Heřmanovský y Pech, 2013).

El método de la avenida índice es el enfoque comúnmente empleado en el análisis regional de crecientes (Saf, 2009; Malekinezhad *et al.*, 2011; Campos-Aranda, 2014a; Dodangeh *et al.*, 2014), el objetivo principal es la búsqueda de los valores de caudal de predicción, para ello debe estimarse la curva de crecimiento y los parámetros regionales, donde los parámetros regionales se calculan como los parámetros ponderados de los parámetros en el sitio (Wazneh *et al.*, 2013).

El término “avenida índice” fue descrito por Dalrymple en 1960, este procedimiento parte de la idea fundamental en el cual los sitios o puntos dentro de un región homogénea comparten la misma distribución de frecuencias y un factor de escala específico por sitio (Trefry *et al.*, 2006), cabe mencionar que la situación anterior se complica la estimación cuando el procedimiento se realiza en una zona árida y semiárida (Malekinezhad *et al.*, 2011).

El método de Avenida índice es un enfoque tradicional que implica el trazo de los límites de confianza por medio de una aproximación asintótica, para verificar la homogeneidad en las regiones (Dodangeh *et al.*, 2014). La avenida índice está relacionada con las características descriptivas de la cuenca (Mishra *et al.*, 2009). En diversos estudios alrededor del mundo, el área es la variable más importante, por lo que algunos autores (Nobert *et al.*, 2014) recomiendan incluir esta variable en el análisis de la avenida índice junto con las características de la cuenca.

Las 3 etapas básicas del método de la avenida índice son las siguientes (Mishra *et al.*, 2009): regionalización hidrológica homogénea, selección de una función regional de frecuencias, y

elección del factor de escala de avenida índice. Para determinar los cuantiles de inundación en cuencas no aforadas, tomando en cuenta el comportamiento de la escala una región homogénea en el Análisis Regional de Frecuencias de Crecientes, las propiedades hidrológicas se relacionaron empleando una función de escala espacial involucrando el área de drenaje (Skaugen y Væringstad, 2005).

En México el método de la avenida índice ha sido aplicado como respuesta a la escasez de estaciones de aforo (Gutiérrez-López y Ramírez, 2005; Campos-Aranda 2014 a, b). Algunos estudios (Gutiérrez-López y Ramírez, 2005; Campos-Aranda, 2014 a, b) han consistido en el análisis de estaciones hidrométricas bajo técnicas de L-Momentos y adecuación de la avenida índice para dos poblaciones

Algunos autores han optado por la aplicación de del método regional de la avenida índice a partir de la creación de zonas homogéneas, que si bien, pueden estar formadas por la estacionalidad de precipitaciones (Nobert *et al.*, 2014) y por métodos de clasificación jerárquica (Malekinezhad *et al.*, 2011), además de la verificación de homogeneidad regional por medio de la técnica de Momentos-L ha sido ampliamente usada (Lim *et al.*, 2009; Malekinezhad *et al.*, 2011; Dodangeh *et al.*, 2014), otros estudios (Nathan y McMahon, 1990g) verifican la homogeneidad mediante métodos gráficos, como la técnica de curvas de Andrews.

Laaha y Blöschl (2006) calcularon caudales mínimos, después de la delimitación de regiones homogéneas, transfirieron la información ubicando la zona no aforada dentro de las regiones homogéneas continuas y estimando el valor del caudal mínimo a partir de regresiones múltiples entre el caudal mínimo y las características de la cuenca. Diversos estudios comparan el método de la avenida índice con el método tradicional de regresión múltiple (Malekinezhad *et al.*, 2011; Nobert *et al.*, 2014)

Lim *et al.* (2009) también aplicaron la técnica de Momentos-L basada en la Avenida índice, aplicada al Río Rojo en la Cuenca Norte, localizada al Noroeste de Sur Dakota. En la delimitación de regiones homogéneas emplearon el método de Análisis Clúster, con las variables de la pendiente de la cuenca, área de contribución, latitud, longitud, y gasto específico pico (por área), donde obtuvieron tres regiones, el propósito de este estudio, fue la divulgación del método como una técnica factible en la delimitación de regiones.

Malekinezhad *et al.* (2011) realizaron una comparación del método de la avenida índice con el método de análisis de regresión múltiple mediante un análisis de L momentos, para analizar las variables más importantes, es decir, las principales características, realizaron un análisis factorial. Este método “busca descubrir si las variables observadas se pueden explicar en gran parte o en su totalidad en términos de un número mucho menor de variables denominadas factores” (Malekinezhad *et al.*, 2011). Los autores revisaron todas las estaciones que contenían más de 15 años de registro, primero realizaron un análisis factorial, obteniendo cuatro variables más importantes de 14 variables que consideraron como variables independientes, siendo la longitud de del cauce principal, precipitación media anual, coeficiente de compacidad (Gravelius), y temperatura media anual; después se realizó un agrupamiento por método Ward empleando la distancia euclidiana, se verificó la homogeneidad por medio de Momentos-L con la prueba estadística H. El siguiente paso fue la identificación de del mejor ajuste de distribución, que para este estudio fue la distribución GEV, distribución que fue elegida por medio de pruebas de bondad de ajuste (Z_{DIS}) y las relaciones de Momentos-L, posteriormente desarrollaron los modelos regionales del método de la avenida índice y el método de análisis de regresión múltiple donde lo validaron por medio de un análisis de RMSE (error medio cuadrático).

El análisis regional de frecuencias es una herramienta importante en el diseño de avenidas en cuencas no aforadas (García-Marín *et al.*, 2015). En las técnicas del Análisis Regional de Frecuencias de Crecientes, la distribución Generalizada de Valores Extremos ha resultado ser la más apropiada en diversos estudios (Malekinezhad *et al.*, 2011; Nobert *et al.*, 2014).

Nobert *et al.* (2014) desarrollaron el método de la avenida índice en la cuenca del Lago Victoria en Kenia, debido a la escasez de estaciones de aforo, para ello efectuaron un estudio multivariado donde descartaron las variables dependientes, donde el modelo de regresión se expresó con las variables de área, lluvia media anual y altitud, además emplearon la delimitación de regiones homogéneas basada en la identificación de las estaciones de la estacionalidad de precipitación, identificando los periodos de ocurrencia característicos de la zona.

Núñez *et al.*, (2016) aplicaron la metodología de L momentos para análisis de sequías con datos de precipitación. Aplicaron el estudio en tres países de América Latina, siendo Chile,

Colombia y México, los autores eligieron el país de México debido a sus características hiper-áridas a super-hídricas, ya que es el tercer país de América Latina que tiene mayor superficie árida con un 18% del total de la superficie del territorio mexicano. El conjunto de datos que emplearon fueron valores mensuales de precipitación, y la suma de estos les permitió conocer la precipitación anual, sin embargo, los autores no rellenaron datos de lluvia anual en los casos en los que no encontraron registros. El estudio se basó en cinco etapas: 1) el análisis de los datos, 2) identificación de regiones homogéneas, 3) análisis regional de una distribución de frecuencias, 4) estimación de parámetros de la función de cuantiles, y 5) mapeo espacial de las características de sequía.

También se han desarrollado otros enfoques como el método de estaciones – años, donde el gasto máximo anual de cada estación hidrométrica se divide entre su media aritmética (Garros–Berthet, 1994; Campos-Aranda, 2006).

En cuanto al método regional de momentos L ponderados también se ha aplicado por medio de la delimitación de regiones por índice de estacionalidad (Campos-Aranda, 2014a).

El enfoque de regresión regional es ampliamente usado, por ejemplo, en caudales mínimos, existen estudios (Laaha y Blöschl, 2006) en los cuales si el área de dominio es grande por lo que las condiciones heterogéneas del sitio son considerables, se recomienda dividir el área de dominio y aplicar modelos de regresión independientes, por lo que a esto se le denominaría enfoque de regresión regional. El análisis de regresión múltiple ha sido ampliamente usado (Yu *et al.*, 2002) con este método se establecen ecuaciones que relacionan las características físicas y climáticas de los sitios aforados. Por lo que este método se ha usado para predicción de caudales medios, varianza, autocorrelación, duración de caudal y volúmenes (Sanborn y Bledsoe, 2006).

10.1. Métodos regionales

10.1.1. Método estaciones-años

Método desarrollado por Cunnane (1988). El método de las estaciones-año consiste en un registro de eventos estandarizados con la media y la desviación estándar. La principal característica de este método consiste en que las muestras tengan una longitud de registro común (Cunnane, 1988) con al menos nueve años de registro (Escalante-Sandoval y Reyes-Chávez, 2002). Sin embargo, se recomienda que cada estación usada en el análisis cuente

con una longitud de registro de 10 años como mínimo (Escalante-Sandoval y Reyes-Chávez, 2002). Para un adecuado procedimiento y con propósitos de obtener una estimación confiable de un evento de diseño implica obtener una muestra que sea cinco veces el periodo de retorno de interés en años.

El procedimiento de cálculo es el siguiente (Escalante-Sandoval y Reyes-Chávez, 2002):

1. Para cada región delimitada como homogénea, se tienen registros de $i=1$ a n años en cada j estación. Por lo que el primer paso consiste en analizar la homogeneidad de las M estaciones que forman la región homogénea.
2. Para el conjunto de los n años y M estaciones, se obtienen los estadísticos de media, varianza, desviación estándar, asimetría, curtosis y coeficiente de variación.
3. Se revisa la independencia de eventos mediante la prueba de Anderson, para cada serie j .
4. Para cada serie j se genera una muestra modulada de la forma:

$$q_i^j = \frac{Q_i^j}{Q^j} \quad (10.1)$$

5. Se forma el registro llamado estaciones-años, de tamaño de muestra n_T con todos los eventos modulados q_i^j .

$$n_T = \sum_{j=1}^M n_j \quad (10.2)$$

6. El registro se ordena de mayor a menor y se le asigna un periodo de retorno y una probabilidad de no excedencia.

$$T = \frac{n_T - 1}{m} \quad (10.3)$$

$$P(X \leq x) = P = 1 - \frac{1}{T} \quad (10.4)$$

donde T es el periodo de retorno en años, n_T es el tamaño de la muestra del registro estaciones-año, m es el número de orden del registro, desde $m=1, \dots, n_T$, P es la probabilidad de no excedencia.

7. Al registro formado en el paso 5 se le ajustan las distintas distribuciones de probabilidad para el análisis de máximos y se selecciona aquel que proporcione el menor error estándar de ajuste.
8. Después de la selección de la distribución de mejor ajuste se estiman los eventos regionales para distintos periodos de retorno, se asume que éstos serán válidos para cualquier punto dentro de la región analizada.

$$(q)_i^j = \left(\frac{Q}{\bar{Q}} \right)_T^R, T = 2, 5, 10, 20, 50, 100, 500, 1000 \text{ y } 10000 \text{ años} \quad (10.5)$$

9. Para la estimación de un caudal \hat{Q}_T^j en un sitio j con escasa información, se debe multiplicar la relación regional por el correspondiente valor de \bar{Q}^j . En dado caso que se requiera estimar un caudal \hat{Q}_T^j en una zona no aforada (e) dentro de una región, se construye una relación $\bar{Q}^j - A_j$, donde $A_j, j=1, \dots, ms$ son las áreas drenadas, para que a partir de ella se pueda inferir el valor correspondiente de \bar{Q}^e para el área A_e del sitio (e), donde esta última relación puede ser del tipo lineal, potencial o polinomial.

Otra forma de emplear este método es el siguiente:

1. Se desarrollan los estadísticos descriptivos de cada región
2. Se estandarizan los datos históricos de cada estación por medio de los dos siguientes pasos :

$$z_i = \frac{x_i - \bar{x}}{\sigma} \quad (10.6)$$

$$z_i = \frac{x_i}{\bar{x}} \quad (10.7)$$

3. Se desarrolla una muestra total formada por los datos de todas las estaciones

$$n = n_1 + n_2 + \dots + n_n \quad (10.8)$$

4. A la muestra total se le aplica un análisis de frecuencias y se obtienen eventos de diseño para cada estación, por ejemplo, en caso de que sea una función de distribución Gumbel, quedaría de la siguiente forma:

$$F(x) = e^{-e^{\left(\frac{x-\mu}{\alpha}\right)}} \quad (10.9)$$

5. Despejando z_i para cada periodo de retorno Tr , des-estandarizando a z_i :

$$x_i = z_i \sigma + \bar{x} \quad (10.10)$$

10.1.2. Técnica de correlación y regresión múltiple

Este método surgió como una alternativa frente al método de la avenida índice (Kite, 1977). Este método permite asociar el periodo de retorno con las variables independientes, es decir, las características descriptivas de la zona de estudio.

La ecuación es de la siguiente forma (Kite, 1977):

$$Q_T = f(A^a, B^b, C^c, \dots, Z^z) \quad (10.11)$$

donde A, B, C, \dots, Z son variables independientes y a, b, c, \dots, z son constantes de derivación del análisis de regresión múltiple.

Una adaptación inicialmente propuesta en esta técnica (Kite, 1977), consiste en la selección una función de distribución de frecuencias que tuviera el mejor ajuste de los datos observados de las estaciones con registros, y una vez seleccionada se realizaba la regresión entre las variables descriptivas y los parámetros de la función seleccionada.

El procedimiento es el siguiente (Kite, 1977): Para cada una de las estaciones individuales de la región, se derivan los parámetros por medio de un análisis de frecuencias para cada sitio y después se desarrollan ecuaciones de regresión para estimar los parámetros de la distribución en un sitio requerido.

En este método regional se inspecciona la relación e independencia de las variables descriptivas mediante procesos de correlación y regresión discriminante para obtener todas aquellas variables independientes significativas estadísticamente y con ello obtener la

estimación precisa de la variable dependiente analizada (Escalante-Sandoval y Reyes-Chávez, 2005).

10.1.2.1. Selección de las variables independientes

La aplicación de un método regional consiste en la recopilación de las características descriptivas más representativas de la zona de estudio, sin embargo, esta selección debe evitar que las variables estén correlacionadas entre sí y que tengan multicolinealidad (Ahn y Palmer, 2011).

Un aspecto importante a desarrollar es conocer la cantidad de variables independientes que influyen en la estimación de la variable dependiente, es decir, el gasto de diseño asociado a un período de retorno. Para listar la cantidad de parámetros que intervienen en la ecuación de regresión existen diversos métodos: regresión lineal simple, regresión lineal múltiple, procedimiento de paso hacia adelante y hacia atrás (Kite, 1977; Sanborn y Bledsoe, 2006) otros autores han propuesto también un análisis de correlación canónica (Kite, 1977; Ouarda *et al.*, 2008).

Una cantidad límite de variables independientes en el modelo de regresión regional, no debe exceder un tercio el número de observaciones y con ello se preserva la confiabilidad en la estimación de la variable dependiente (Escalante-Sandoval y Reyes-Chávez, 2005).

Con el fin de evitar multi-colinealidad en los datos, se han desarrollado diversos métodos como el Factor de Inflación de varianza (VIF, por sus siglas en inglés) (Ahn y Palmer, 2016), el coeficiente C_p de Mallows (Laaha y Blöschl, 2006) y matrices de correlación (Nathan y McMahon, 1990).

Uno de los métodos para la selección de variables independientes es el método de regresión “*stepwise*”, método que emplea el coeficiente de Mallows C_p , como método de optimización (Laaha y Blöschl, 2006).

$$C_p = \frac{RSS_p}{\hat{\sigma}^2} + 2p - n \quad (10.12)$$

donde RSS_p es la suma residual de los cuadrados del modelo considerado con p coeficientes divididos por la varianza del error residual $\hat{\sigma}^2$ del modelo y corresponde a la

optimización relativa en términos del error del modelo. Además el término $2p - n$ implica la penalización de la complejidad del modelo. C_p es un criterio de selección penalizado que toma la ganancia de la varianza explicada, así como la parsimonia de modelos en cuenta y los modelos de campos que son óptimos en términos de errores de predicción. La variable de selección comienza con la selección arbitraria de una característica de la cuenca y subsecuentemente se añaden variables para minimizar C_p

Después de cada paso, se reemplazan una de las variables por cualquiera de las características, con el fin de que el valor de C_p muestre un decremento, el proceso continua hasta encontrar el mínimo valor de C_p , y de esta forma, las características añadidas para la obtención del C_p mínimo, se asumen como variables de control (Laaha y Blöschl, 2006).

Otro procedimiento convencional (Ahn y Palmer, 2016) consiste en realizar una regresión paso a paso para la selección de las variables, pudiendo llegar a desarrollarse por medio de mínimos cuadrados ordinarios (OLS, por sus siglas en inglés) o por medio de mínimos cuadrados generalizados (GLS, por sus siglas en inglés) con el objetivo de construir un modelo de regresión múltiple por lo que una variable a la vez es añadida. Las características que estas variables deben tener son: nivel de significancia $p\text{-value} \leq 0.05$.

Otro procedimiento para la cantidad óptima de variables a emplear en un enfoque de regresión, es el criterio de Akaike, AIC (Bao *et al.*, 2012).

Sarhadi y Modarres (2011) emplearon técnicas de análisis de aglomerados por método de Ward, con datos de índice estacional para agrupar regiones homogéneas, mientras que para la evaluación de la estimación de caudales en sitios no aforados emplearon la técnica Jack-knife, donde los cuantiles en el sitio se emplearon como variables dependientes y las características descriptivas de la cuenca, es decir, las variables fisiográficas y climatológicas, se emplearon como variables independientes. Una aportación distinta que realizaron (Sarhadi y Modarres, 2011) fue el hecho de seleccionar variables para los cuantiles bajos y otro conjunto de variables para cuantiles altos. Para cuantiles bajos se empleó la elevación máxima y la cubierta vegetal. Mientras que, para caudales de diseño altos, se consideró en

las ecuaciones de regresión el efecto del suelo desnudo y el área de captación, esta decisión fue principalmente por la región árida en la cual se desarrolló el estudio.

10.1.2.2. Procedimiento

El modelo es de la siguiente forma:

$$Y = a + b_1X_1 + \dots + b_sX_s \quad (10.13)$$

O también:

$$Y = aX_1^{b_1}, X_2^{b_2}, \dots, X_s^{b_s} \quad (10.14)$$

donde Y es la variable dependiente, X_k son las características fisiográficas y climatológicas ($k = 1, \dots, s$) y a, b_k es el conjunto de parámetros a estimarse de los datos.

El procedimiento de cálculo es el siguiente (Escalante-Sandoval y Reyes-Chávez, 2005):

1. Obtener la información de los eventos a ser analizados en cada una de las estaciones de la región, así como las características tanto fisiográficas como climatológicas de la zona de estudio.
2. Revisar la homogeneidad del conjunto de datos.
3. Obtener los estadísticos muestrales de cada serie Q_i^j : media, varianza, desviación estándar, asimetría, curtosis y coeficiente de variación.
4. Se revisa la independencia de eventos mediante la prueba de Anderson, para cada serie j .
5. Para cada serie Q_i^j se ajustan las diferentes funciones de probabilidad para el análisis de máximos (Normal, Log normal con 2 y 3 parámetros, Gamma con 2 y 3 parámetros, Log-Pearson tipo III, Gumbel, General de Valores Extremos, Valores Extremos de dos Componentes, Gumbel Mixta, GVE Mixta, y mezcla Gumbel-GVE), y se selecciona el método aquel que proporcione el mínimo error estándar de ajuste.
6. Una vez que se obtiene la distribución de mejor ajuste para cada uno de los sitios $j = 1, \dots, M$ es posible calcular los eventos \hat{Q}_i^j para los periodos de retorno $T = 2, 5, 10, 20, 50, 100, 500, 1000, 5000$ y 10000 .

7. Para cada periodo de retorno se plantea un sistema de ecuaciones, de la siguiente forma:

$$\hat{Q}_T = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 \quad (10.15)$$

Por ejemplo si se desea conocer el caudal para un periodo de retorno $T = 50$ años , para un conjunto de cinco sitios aforados y cuatro características fisiográficas, se tendría el siguiente conjunto de ecuaciones planteadas:

$$\hat{Q}_{50}^1 = a + b_1A^1 + b_2E_c^1 + b_3L_{cp}^1 + b_4D_d^1$$

$$\hat{Q}_{50}^2 = a + b_1A^2 + b_2E_c^2 + b_3L_{cp}^2 + b_4D_d^2$$

$$\hat{Q}_{50}^3 = a + b_1A^3 + b_2E_c^3 + b_3L_{cp}^3 + b_4D_d^3$$

$$\hat{Q}_{50}^4 = a + b_1A^4 + b_2E_c^4 + b_3L_{cp}^4 + b_4D_d^4$$

$$\hat{Q}_{50}^5 = a + b_1A^5 + b_2E_c^5 + b_3L_{cp}^5 + b_4D_d^5$$

8. El requisito para resolver las ecuaciones consiste en la normalidad de las series. Para probar la que las series involucradas son normales se deben obtener los estadísticos muestrales para cada una de las series del sistema. De lo contrario se debe proceder a transformarlas, por medio de la técnica Box-Cox.
9. Una vez que se plantea el sistema de ecuaciones normales, se debe realizar un análisis de varianza para determinar cuántas variables independientes deben incluirse finalmente en el modelo de regresión, como resultado es posible que la cantidad de variables propuesta inicial sea menor.
10. Si se requiere obtener un evento \hat{Q}_{50}^j en un sitio j que tiene escasa información, basta sustituir las características fisiográficas A^j , E_c^j , L_{cp}^j y D_d^j en la ecuación regional obtenida en el paso anterior. En caso de que se requiera calcular un evento \hat{Q}_{50}^j en un sitio no aforado (e) y dentro de la región homogénea, entonces, se sustituyen los valores fisiográficos de la estación no aforada A^e , E_c^e , L_{cp}^e y D_d^e en la misma ecuación. Si se requiere conocer otro valor estimado para un periodo de retorno en específico, se debe realizar nuevamente el procedimiento señalado previamente.

Procedimiento para simplificar el método:

Se consideran como variables dependientes los parámetros de la distribución característica de la región, por ejemplo para una distribución Gumbel, se tendrían dos parámetros, ubicación $\hat{\nu}$ y escala $\hat{\alpha}$, por lo que se tendrían que resolver los dos sistemas de ecuaciones normales:

$$\hat{\nu}^1 = a + b_1 A^1 + b_2 E_c^1 + b_3 L_{cp}^1 + b_4 D_d^1$$

$$\hat{\nu}^2 = a + b_1 A^2 + b_2 E_c^2 + b_3 L_{cp}^2 + b_4 D_d^2$$

$$\hat{\nu}^3 = a + b_1 A^3 + b_2 E_c^3 + b_3 L_{cp}^3 + b_4 D_d^3$$

$$\hat{\nu}^4 = a + b_1 A^4 + b_2 E_c^4 + b_3 L_{cp}^4 + b_4 D_d^4$$

$$\hat{\nu}^5 = a + b_1 A^5 + b_2 E_c^5 + b_3 L_{cp}^5 + b_4 D_d^5$$

$$\hat{\alpha}^1 = a + b_1 A^1 + b_2 E_c^1 + b_3 L_{cp}^1 + b_4 D_d^1$$

$$\hat{\alpha}^2 = a + b_1 A^2 + b_2 E_c^2 + b_3 L_{cp}^2 + b_4 D_d^2$$

$$\hat{\alpha}^3 = a + b_1 A^3 + b_2 E_c^3 + b_3 L_{cp}^3 + b_4 D_d^3$$

$$\hat{\alpha}^4 = a + b_1 A^4 + b_2 E_c^4 + b_3 L_{cp}^4 + b_4 D_d^4$$

$$\hat{\alpha}^5 = a + b_1 A^5 + b_2 E_c^5 + b_3 L_{cp}^5 + b_4 D_d^5$$

Al resolver el sistema y verificando que las cuatro características son significativas para la estimación de la variable dependiente (análisis de varianza), en este caso los parámetros $\hat{\nu}$ y $\hat{\alpha}$, se tendrían las siguientes ecuaciones regionales:

$$\hat{\nu}_R = \hat{a} + \hat{b}_1 A + \hat{b}_2 E_c + \hat{b}_3 L_{cp} + \hat{b}_4 D_d$$

$$\hat{\alpha}_R = \hat{a} + \hat{b}_1 A + \hat{b}_2 E_c + \hat{b}_3 L_{cp} + \hat{b}_4 D_d$$

Por lo que dado el caso de estimar un evento en un sitio con poca o nula información, será necesario sustituir las características fisiográficas y climatológicas en las ecuaciones regionales para obtener los parámetros de ubicación y escala de distribución Gumbel, como en este caso específico previo.

9.1.2.3.Revisión de la independencia del modelo de regresión múltiple

Al realizar un análisis de regresión, existe un supuesto importante que debe ser verificado, que es el de independencia en los residuales.

Existe una prueba gráfica para revisar la independencia, consiste en graficar los residuos contra el orden en el que se obtuvieron, los cuales no deben cumplir con un patrón específico (Gutiérrez-Pulido y De la Vara-Salazar, 2008). La prueba de Durbin-Watson “diagnostica la presencia de correlación entre los residuos consecutivos, que es una posible manifestación de la falta de independencia” (Gutiérrez-Pulido y De la Vara-Salazar, 2008).

La hipótesis de la prueba de Durbin-Watson (Gutiérrez-Pulido y de la Vara-Salazar, 2008) es:

Sea ρ el parámetro que representa la correlación entre residuos consecutivos ($Corr(e_t, e_{t-1}) = \rho; t = 2, 3, \dots, n$). La hipótesis en la prueba de Durbin-Watson es:

$$H_0 : \rho = 0$$

$$H_A : \rho > 0$$

donde la alternativa se toma en el sentido mayor ($>$) porque la autocorrelación positiva es la más frecuente en la práctica. Cuando los puntos se van alternando de negativos a positivos la autocorrelación es negativa. El estadístico de Durbin-Watson es:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n (e_i)^2} \quad (10.16)$$

donde los $e_i, i=1, 2, \dots, n$ son los residuos ordenados en el tiempo. La decisión sobre la hipótesis dada, consiste en la siguiente regla:

- | | |
|--------------------------|---------------------|
| Si $d < d_L$ | Se rechaza H_0 |
| Si $d < d_U$ | No se rechaza H_0 |
| Si $d_L \leq d \leq d_U$ | Sin decisión |

donde d_L y d_U son cotas que se leen en tablas teóricas, por lo que los datos necesarios a consultarlas serán el número de residuos n , nivel de significancia prefijado α y el número de variables explicativas del modelo, p .

Si la hipótesis de autocorrelación negativa $H_A: \rho < 0$ se utiliza el estadístico $d' = 4 - d$. En caso de querer desarrollar la hipótesis bilateral con alternativa $H_A: \rho \neq 0$, se combinan las dos pruebas unilaterales de tamaño α de manera que la prueba bilateral tenga el tamaño deseado 2α .

La desventaja de la prueba de Durbin-Watson consiste en la detección de la correlación solo en residuos consecutivos en el tiempo.

9.1.2.4. Calidad del modelo de regresión

Para determinar la calidad del modelo, es necesario analizar las gráficas de los residuales, el coeficiente de determinación y el coeficiente de determinación ajustado, estos dos últimos no deben ser menores a 0.70.

10.1.3. Método regional de la Transformación de Box-Cox

El método consiste en realizar un proceso de normalización de las series de registros observados, además la técnica también puede aplicarse a series homogéneas e independientes. De acuerdo con la siguiente expresión:

$$X_i = \frac{(Q_i)^\lambda - 1}{\lambda}, \lambda \neq 0, Q_i > 0 \quad (10.17)$$

$$X_i = \ln Q_i, \lambda \neq 0, Q_i > 0 \quad (10.18)$$

Los eventos \hat{Q}_T se obtienen por

$$\hat{X}_T = \hat{\mu}_x + \sigma_x U_T \quad (10.19)$$

$$\hat{Q}_T = (1 + \lambda \hat{X}_T)^{1/\lambda} \quad (10.20)$$

El procedimiento es el siguiente:

1. Para cada una de las series Q_i^j se obtiene el valor de la función normalizante $\hat{\lambda}_i$.
2. Debido a la variabilidad del coeficiente $\hat{\lambda}_i$ se obtiene un valor regional

$$\hat{\lambda}_r = \bar{\lambda} = \frac{1}{M} \sum_{i=1}^M \hat{\lambda}_i \quad (10.21)$$

3. Se generan las series $X_i^j, i=1,2,\dots,n_j; j=1,2,\dots,M$

$$X_i^j = \frac{(Q_i^j)^{\hat{\lambda}_r} - 1}{\hat{\lambda}_r}; \hat{\lambda}_r \neq 0 \quad (10.22)$$

$$X_i^j = \ln Q_i^j; \hat{\lambda}_r = 0 \quad (10.23)$$

4. Para cada una de las series X_i^j se obtienen $\hat{\mu}_x^j$ y $\hat{\sigma}_x^j \frac{1}{2}$
5. Si se requiere el evento \hat{Q}_T^j , entonces:

$$\hat{X}_T^j = \hat{\mu}_x^j + \sigma_x^j U_T, \hat{\lambda}_r \neq 0 \quad (10.24)$$

$$\hat{Q}_T^j = \exp(X_T^j), \hat{\lambda}_r = 0 \quad (10.25)$$

donde U_T es una variable normal estándar $N(0,1)$.

10.1.4. Avenida índice

Este método fue creado por Dalrymple (1960), el cual consiste en tres principales aspectos, dentro de cada región homogénea se tienen datos descriptivos, eliminación de estaciones, y un periodo común de registros, este último aspecto fue modificado, por lo que la restricción del periodo en común ya no fue aplicable.

En cuanto a la calidad de las estaciones, el autor propone emplear todas aquellas que tengan al menos un periodo de registro de 5 años y excluir del análisis aquellas estaciones que pertenezcan a caudales regulados, así como complementar datos faltantes por correlación entre estaciones, los cuales no son empleados en el análisis, sino únicamente para revisión y asignación de periodos de retorno.

El método de la Avenida Índice ha sido ampliamente usado en análisis regional, desde su divulgación hasta la actualidad y se le han hecho algunas adecuaciones al método de verificación de homogeneidad.

Este método se caracteriza por las dos etapas fundamentales en las que se desarrolla (Kite, 1977) para las cuales emplea dos tipos de variables hidrometeorológicas y fisiográficas. La primera etapa consiste en la construcción de curvas de frecuencia para la representación la relación del flujo para varias frecuencias con el caudal medio anual de cada estación aforada, y la segunda etapa: la relación desarrollada entre las características del área de drenaje y el caudal medio anual. Con lo anterior es factible la estimación del caudal en cualquier ubicación dentro de la región.

Una de las desventajas de la aplicación original de este método era la aplicación a un periodo de registro en común (Kite, 1977), sin embargo, Fill y Stedinger (1995) realizaron la adecuación al no requerir de un mismo periodo de registro.

La propuesta original (Kite, 1977) consistía en la identificación de estaciones con registros dentro de una región seleccionada con un periodo de registro en común, por lo que se eliminaban todas aquellas estaciones que tuvieran un registro mínimo de 5 años de recopilación de datos. La forma en la que se seleccionaba el periodo base de registro común era por medio de una gráfica de barras, el cual se planeaba en función del número máximo de estaciones-año, los datos que faltaban eran complementados por medio de la correlación entre estaciones (Kite, 1977).

10.1.4.1. Procedimiento del método original

El método de la avenida índice fue creado por Dalrymple en el año de 1960, este método parte del concepto en el que el caudal en distintos sitios dentro de una región se distribuye de forma idéntica, excepto por un factor de escala que está en función de las características fisiográficas de la cuenca.

El procedimiento de la avenida índice original (Kite, 1977) se explica a continuación:

Para cada una de las estaciones, se grafica periodo de retorno con la magnitud del evento de diseño en una sola gráfica, por lo que al final, se unen esos puntos por medio de una línea.

En previos estudios se ha probado que, bajo una distribución extrema tipo I, el periodo de retorno de 2.33 años es el evento medio anual para la estación en cuestión.

El periodo de retorno con $T = (n+1)/m$, posteriormente se grafica el periodo de retorno con la magnitud del evento, inicia con el orden del registro de las series de datos.

10.1.4.2. Pruebas de homogeneidad

La prueba de homogeneidad de Langbein se diseñó para emplear en el método de la avenida índice. A esta prueba se le han aplicado modificaciones. La versión original contempla la revisión para una población (Dalrymple, 1960), posteriormente se realizó una modificación (Fill y Stedinger, 1995) y se han desarrollado algunas adaptaciones de la versión original para dos poblaciones (Gutiérrez-López y Ramírez, 2005).

10.1.4.2.1. Prueba original de Langbein

Dalrymple en el año de 1960 (Kite, 1977) describió un test el cual podría ser usado en el procedimiento de la avenida índice para revisión de la homogeneidad hidrológica regional. La metodología está explicada por Kite (1977) de la siguiente forma:

El error estándar de la estimación de la variable reducida, y , en una distribución extrema tipo I está dada por:

$$\sigma_y = \frac{e^y}{\sqrt{n}} \sqrt{\frac{1}{T-1}} \quad (10.26)$$

Entonces se asume una distribución normal de las estimaciones, el 95% de las estimaciones recaerán en $\pm 2 \sigma_y$.

Los límites de confianza son:

$$2.25 \pm 6.33/\sqrt{n} \quad (10.27)$$

donde n es el número de años para el periodo de retorno.

Con la estimación de estos valores, se grafica la prueba de Langbein: Se grafican los límites de confianza inferior y superior, es decir, Tl y TU (Tabla 10-1), donde la prueba de Langbein trabaja con propiedades estadísticas adimensionales (Campos-Aranda, 2012)

Las estaciones que queden fuera de los límites de confianza serán descartadas, mientras que las estaciones que permanezcan dentro de los límites se asumen que cumplen con la homogeneidad.

Para cada estación se grafica en las ordenadas (TE) el periodo de retorno para 10 años, y en las abscisas como el número de registros anuales más la mitad del número de eventos calculados para cada estación. Las estaciones que queden fuera de los límites de confianza serán descartados, mientras que las estaciones que permanezcan dentro de la región homogénea: se deben de colocar los cocientes de los eventos de los distintos periodos de retorno con la media anual y se determinan las medianas de esos valores.

El siguiente paso consiste en graficar la mediana de los gastos contra el periodo de retorno para construir la curva de frecuencia regional.

El siguiente paso en el análisis de la avenida índice es graficar el área de drenaje contra el evento de la media anual.

Tabla 10-1. Límites de la prueba de homogeneidad propuesta por Dalrymple.

<i>Tamaño de la muestra</i> <i>n</i>	<i>Inferior</i> $y - 2\sigma_y$	<i>Límite inferior</i> T_u	<i>Superior</i> $y + 2\sigma_y$	<i>Límite superior</i> T_L
5	-0.59	1.20	5.09	160
10	0.25	1.80	4.25	70
20	0.83	2.80	3.67	40
50	1.35	4.40	3.15	24
100	1.62	5.60	2.88	18
200	1.80	6.50	2.70	15
500	1.97	7.70	2.53	13
1000	2.05	8.30	2.45	12

Kite (1977) identificó las deficiencias en el método de la avenida índice de la propuesta original de Dalrymple (1960), el primero de ellos es que este método no es recomendable para las estaciones que tienen un registro corto de tiempo, por lo que las curvas de frecuencia

no tienen un comportamiento ideal, sino que tienen valores altos y “el efecto del área de drenaje es relativamente alto para avenidas con altos intervalos de repetición” (Kite, 1977).

10.1.4.2.2. Modificación de la Avenida índice para dos poblaciones

Gutiérrez-López y Ramírez (2005) desarrollaron una adecuación de la avenida índice para dos poblaciones, debido al amplio uso de la distribución Gumbel en México y América Latina, así como la necesidad de adecuar la distribución a zonas donde la muestra en las cuencas costeras proviene de dos poblaciones.

El esquema propuesto parte del análisis de cada cuenca con su evento medio anual teórico, el cual estará representado por el evento medio anual del periodo de retorno de 2.33 años.

La modificación es aplicada en gastos máximos anuales, consiste en la modificación de los límites de la prueba de homogeneidad de Langbein, por medio de una función de probabilidad Doble Gumbel, a continuación se describe el procedimiento:

$$f(\bar{y}) = \lambda [\exp(-e^{-y_1})(e^{y_1})] + (1 - \lambda) [\exp(-e^{-y_2})(e^{y_2})] \quad (10.28)$$

donde el primer término representa la muestra de eventos para la temporada normal de lluvias y el segundo para eventos ciclónicos.

$$1 - \frac{1}{T} = \lambda e^{-e^{-y_1}} + (1 - \lambda) e^{-e^{-y_2}} \quad (10.29)$$

Cuando en la expresión λ es uno o cero, ésta se reduce a la expresión para una sola población.

Los autores tomaron como periodo de retorno de 10 años, verificaron que cada variable reducida para una y dos poblaciones de 2.2504 para cualquier parámetro de asociación λ . Identificaron que λ no influye en el cálculo de varianzas, por lo que los autores llegaron a establecer la hipótesis de que cada variable descrita por la distribución Gumbel univariada, contribuye en igual magnitud al cálculo de la desviación estándar (Tabla 10-2).

$$2\sigma y = 3.164/\sqrt{n} \quad (10.30)$$

Tabla 10-2. Límites de la prueba de homogeneidad para dos poblaciones.

<i>Años n</i>	$2\sigma y = 3.164/\sqrt{n}$	$y_{10} - 2\sigma$	<i>Ti (años)</i>	$y_{10} + 2\sigma$	<i>Ts (años)</i>
5	1.41	0.84	2.84	3.66	39.54
10	1.00	1.25	4.01	3.25	26.30
20	0.71	1.54	5.20	2.96	19.75
30	0.58	1.67	5.84	2.83	17.41
40	0.50	1.75	6.27	2.75	16.15
50	0.45	1.80	6.58	2.70	15.34
100	0.32	1.93	7.43	2.57	13.52
200	0.22	2.03	8.10	2.47	12.37
500	0.14	2.11	8.75	2.39	11.44
1000	0.10	2.15	9.10	2.35	10.99

10.1.4.2.3. Modificcaciones a la prueba de homogeneidad de Langbein original

En el método de la Avenida índice se considera que el área de la cuenca es el factor más importante para escalar las crecientes, sin embargo GREHYS (1996) apuntan que el método comenzó a perder popularidad debido a diversos estudios en los que se mostraba que el área de la cuenca no solo afectaba el valor de la escala, sino que también afecta a otras variables de interés como el coeficiente de variación y el coeficiente de sesgo, que de acuerdo a sus definiciones éstas deben ser constantes.

Wiltshire (1986) señaló que el método de Langbein no es eficiente ya que todas las regiones en el análisis resultaban homogéneas aunque tuvieran muchas diferencias entre ellas (Campos-Aranda, 2012).

La divulgación de esta metodología en México fue realizada por Campos-Aranda (2012), a continuación se extrae información de esa publicación:

Los aspectos encontrados por Fill y Stedinger (1995) respecto a la prueba desarrollada por Darlymple (1960) son los siguientes:

La prueba original no contempla la homogeneidad de la región como un todo, debido a que esta prueba se planteó para no permitir que existieran puntos fuera de las curvas de control, de ser así estos puntos debían eliminarse del análisis. Además, los autores dedujeron una expresión para el polinomio cuadrático del coeficiente de variación, por lo que la expresión para la varianza del estimador a comparación de la prueba original, son más estrechas.

Además de las correcciones debidas a las expresiones originales del test de Langbein, Fill y Stedinger (1995) modifican una desventaja del planteamiento original, que es la selección de un periodo común de datos, por lo que la estimación de datos faltantes ya no es útil.

A continuación se describe la versión de Test de Langbein modificada (Fill y Stedinger, 1995; Campos-Aranda, 2012):

Se calcula el coeficiente de variación de cada sitio i (Cv^i) y se obtiene su promedio ponderado por las longitudes de registro

$$Cv^i = \frac{S^i}{\bar{Q}^i} \quad (10.31)$$

donde:

$$\bar{Q}^i = \frac{\sum_{j=1}^{n_i} Q_j^i}{n_i} \quad (10.32)$$

$$S^i = \sqrt{\frac{\sum_{j=1}^{n_i} (Q_j^i - \bar{Q}^i)^2}{n_i - 1}} \quad (10.33)$$

$$Cv^R = \frac{\sum_{i=1}^m n_i \cdot Cv^i}{\sum_{i=1}^m n_i} \quad (10.34)$$

donde m es el número de estaciones hidrométricas de la región que es analizada, Cv^R es el coeficiente de variación regional.

El valor anterior se corrige por sesgo, sumándole el resultado de la ecuación siguiente:

sesgo

$$sesgo(Cv^R) = \frac{Cv^R \cdot \psi(Cv^R)}{\frac{1}{m} \sum_{i=1}^m n_i} \quad (10.35)$$

donde

$$\psi(Cv) = Cv^2 - 0.57Cv - 0.55 \quad (10.36)$$

Se determinan, con base en los Cv^i , los intervalos de confianza del periodo de retorno (T_s, T_i) necesarios para las longitudes de registro n_i , con base en las siguientes ecuaciones (Fill y Stedinger, 1995; Campos-Aranda, 2012):

$$T_{s/i} = \frac{1}{\left\{1 - \exp\left[-\exp\left(-\left(2.2504 + sesgoy_{10}^i \pm 1.96\sqrt{\text{var } y_{10}^i}\right)\right)\right]\right\}} \quad (10.37)$$

donde

$$sesgoy_{10}^i = 1.6732 \left[\frac{\varphi(Cv) - \psi(Cv)}{n_i} - \frac{\varphi(Cv) - m \cdot \psi(Cv)}{\sum_{i=1}^m n_i} \right] \quad (10.38)$$

y

$$\text{var } y_{10}^i = 2.80 \cdot \varphi(Cv) \cdot \left(\frac{1}{n_i} - \frac{1}{\sum_{i=1}^m n_i} \right) \quad (10.39)$$

$$\varphi(Cv) = Cv^2 - 1.14Cv + 1.10 \quad (10.40)$$

donde y_{10}^i es el estimador de la variable reducida de Gumbel de un periodo de retorno de 10 años en el sitio i .

Para cada sitio se calcula la estimación de la variable reducida:

$$y_{10}^i = 1.6732 \frac{Cv^R}{Cv^i} + 0.5772 \quad (10.41)$$

El periodo de retorno $T(Q_{10}^i)$ se calcula con la siguiente expresión:

$$T(Q_{i_0}^i) = 1 / \{1 - \exp[-\exp(-y_{i_0}^i)]\} \quad (10.42)$$

Se contabiliza el k número de veces que el valor anterior está fuera de los intervalos de confianza respectivos, calculados en el paso anterior (Campos-Aranda, 2012).

El número máximo de valores de k que son admisibles en una región homogénea se obtienen mediante la distribución de Bernoulli, considerando cada sitio como un experimento con probabilidad de éxito de 0.05 y de fracaso 0.95 (Fill y Stedinger, 1995; Campos-Aranda, 2012)

$$P(k \geq k_c) = \sum_{j=k_c}^m \frac{m!}{j!(m-j)!} (0.05)^j \cdot (0.95)^{m-j} = \alpha \quad (10.43)$$

donde α es la probabilidad de cometer error tipo I en el test, es decir, 5%. Las mejores aproximaciones son: $\alpha = 4.44\%$ y $\alpha = 5.03\%$, se obtuvieron con $m=7$ y $k_c=2$ y con $m=17$ y $k_c=3$, respectivamente. Para fines prácticos se puede aceptar $k_c=2$ para $5 \leq m \leq 10$ y $k_c=3$, cuando $11 \leq m \leq 19$. Entonces, cuando $k < k_c$ la región es homogénea, en caso contrario no es homogénea (Campos-Aranda, 2012).

En caso de aplicar una prueba con $\alpha = 1.0\%$, la aproximación que se obtiene es de $m=15$ y $k_c=2$ por lo que con una confiabilidad del 99% en regiones con más de 15 sitios es aceptable tener un valor de $k_c=2$, lo cual implica que con confiabilidad del 99% sólo en regiones con 15% o más sitios es aceptable tener un $k_c=2$. Por el contrario cuando en la prueba se requiera un $\alpha = 10.0\%$, sus tres mejores aproximaciones se obtienen con $m=6$, $k_c=2$, $m=11$, $k_c=3$ y $m=18$ y $k_c=4$, por lo cual, con confiabilidad de 90% se puede aceptar, de manera aproximada, que $k_c=2$ en regiones con menos de 7 sitios, $k_c=3$ cuando $8 \leq m \leq 16$ y $k_c=4$ regiones con más de 17 sitios (Campos-Aranda, 2012).

10.1.5. Método de la cuadrícula

Este método (Kite, 1977) se genera con series de máximos anuales para estimar la distribución del escurrimiento medio anual. Básicamente el método consta de la división de

la zona de estudio en una rejilla o malla con elementos individuales, en forma de cuadros, cada cuadro contiene datos en los vértices, así como datos en el centro de cada cuadrícula. Los datos que contiene cada elemento pueden ser del tipo fisiográfico o geomorfológico. La dimensión de la cuadrícula es variada de 1 a 10 km. Los datos de las celdas que contienen estaciones, ya sean climatológicas o hidrométricas se reservan para corregir la malla.

El procedimiento, descrito por Kite (1977), consiste en procesar los datos de temperatura media anual, después convertir los datos a evaporación por medio de la ecuación de Turc, que está en función de los datos de precipitación media anual, después calcular el escurrimiento y realizar una corrección con los datos observados provenientes de la estación hidrométrica. Cuando se tiene el parámetro óptimo de corrección se realizan ecuaciones de regresión en toda la malla, que contienen los datos de características fisiográficas.

El método de la cuadrícula o *square-grid method* involucra los siguientes pasos (Kite, 1977):

- a) Empleando datos de estaciones meteorológicas y una ecuación de regresión (que se establece entre la temperatura media anual en las estaciones con las características fisiográficas), además, un análisis similar se realiza para los datos de precipitación en las cuadrículas.
- b) Se estima la evaporación en la cuadrícula utilizando los datos de precipitación y temperatura:

$$E = \frac{P}{(0.9 + P^2/L^2)^{1/2}} \quad (10.44)$$

donde E es la evaporación [mm], P es la precipitación [mm], y T es la temperatura [°C].

$$L = 300 + 25T + 0.5T^2 \quad (10.45)$$

Empleando la ecuación (10.45), se asume que $E = P$ para la relación P^2/L^2 menor que 0.1 mm².

- c) Lo anterior es para estimar los valores preliminares del escurrimiento medio anual:

$$\text{Escurrimiento} = \text{precipitación} - \text{evaporación} \quad (10.46)$$

- d) Los resultados del inciso c) son empleados para calcular el promedio del escurrimiento promedio para cada cuenca, tendiendo registros de flujo y las razones (K) entre datos registrados (observados) y calculados.
- e) Se calcula una nueva precipitación por cuadro, empleando un factor de corrección K (todo el error se atribuye a la precipitación):

$$\text{precipitación}(\text{corregida}) = K * \text{escurrimiento} + \text{evaporación} \quad (10.47)$$

- f) Empleando los valores corregidos de la precipitación en cada cuadrado y los datos de precipitación de las estaciones, se establece una nueva ecuación de regresión entre: la precipitación y los factores fisiográficos, los datos de las estaciones de aforo de lluvia se ponderan 10 veces más que las estimaciones de precipitación en los cuadrados de “sin datos” (*ungaged*).
- g) El procedimiento se repite tantas veces como sea requerido para obtener una K aceptable lo más cercano a 1 (tanto como se considere razonable)
- h) Una vez calculados la precipitación, escurrimiento y evaporación en sitios aforados; la distribución del escurrimiento y precipitación en sitios no aforados se estima, empleando ecuaciones de regresión entre la precipitación finalmente el escurrimiento de las áreas aforadas y con datos fisiográficos.

El resultado final de esta técnica es un archivo de datos con Temperatura Media Anual (TMA), Precipitación y escurrimiento por cada cuadrícula.

10.1.6. Método derivado del uso de funciones de distribución de frecuencia estándar

Este método surgió como una alternativa al método de la Avenida índice en la elaboración de curvas de frecuencia, el método es empleado para las zonas de estudio donde la distribución extrema tipo I se representa bien con los datos observados (Kite, 1977).

Las características de este método son las siguientes (Kite, 1977):

- a) El método recomienda la selección de estaciones con al menos 15 años de registro, con caudal natural o con regulación mínima.
- b) Elaboración de “Curvas preliminares” para cada estación con la distribución extrema tipo I, para 100 años de período de retorno y se establecen los límites de confianza para cada curva preliminar usando la curva regional.
- c) Prueba de homogeneidad, por medio del test de Langbein con la metodología del método de la avenida índice.

- d) Para la creación de la curva por región, se asumen que las curvas preliminares calculadas en cada sitio forman una muestra regional. Por lo que las estimaciones son promediadas.
- e) Las curvas preliminares se adaptan como curvas adimensionales en determinados periodos, por medio de una proporción del caudal con el caudal medio anual.
- f) Para cada uno de los intervalos de ocurrencia, se calcula la media de los cocientes adimensionales de todas las estaciones. De esta forma se obtienen las relaciones de inundación para la curva regional. Los datos calculados se grafican en “papel Gumbel” (con la ordenada en escala aritmética) y la recta de mejor ajuste dibujada a través de ellos, la línea resultante se toma como la curva de frecuencia regional requerida.
- g) Cálculo de los límites de confianza para la curva regional. Un intervalo de recurrencia (por ejemplo 50 años) se selecciona arbitrariamente al igual que el ancho de la banda de confianza. El ancho es tomado de la distancia vertical entre la curva preliminar y el límite superior (o inferior). Los datos resultantes son divididos por la avenida media anual apropiada para producir cocientes, los cuales se definen como “errores” en curvas individuales. Los errores se combinan mediante el cálculo de la raíz cuadrada por el número de estaciones. El cociente resultante se toma como el “error” en la curva regional, es decir, es el ancho de los límites de confianza para el intervalo de recurrencia seleccionado. El mismo procedimiento se hace para otro intervalo de recurrencia. Los errores de los cálculos hechos con los intervalos de confianza se grafican en la curva regional colocándolas de acuerdo a los intervalos de recurrencia apropiadas en una dirección vertical a cada lado de la curva principal. El punto resultante está unido por líneas rectas para producir los intervalos de confianza requeridos. Note que los límites de confianza representan el mismo grado de confianza con el que fue empleado para calcular los límites de confianza de las curvas preliminares.
- h) Después de obtener la curva adimensional de la curva de frecuencia, con los límites de confianza, es necesario introducir una relación entre el caudal medio anual y las características de la cuenca:

$$\bar{Q} = f(A_u + \lambda^k A_c) \quad (10.48)$$

donde A_u y A_c son las áreas de drenaje no controladas y controladas por lagos y pantanos, respectivamente.

$$\lambda = \frac{A_c - A_u}{A_c} \quad (10.49)$$

donde A_L es la superficie total de los lagos y pantanos principales, k es una constante optimizada al minimizar el error estándar del caudal medio anual \bar{Q}

Posteriormente surgieron adaptaciones del método para otras funciones de distribución (Kite, 1977), con lognormal, Extrema tipo I y Pearson (tipo III). El procedimiento (Kite, 1977) es el siguiente:

De los registros históricos se calcula la media y la desviación estándar del dato del caudal pico.

Los parámetros estadísticos son ajustados en un periodo base mediante las correlaciones entre: la descarga pico en una estación de largo y corto plazo

$$\sigma_{1b} = \sigma_{1a} + (\sigma_{2b} - \sigma_{2a}) R^2 \sigma_{1a} / \sigma_{2a} \quad (10.50)$$

y

$$\mu_{1b} = \mu_{1a} + (\mu_{2b} - \mu_{2a}) R^2 \sigma_{1b} / \sigma_{2b} \quad (10.51)$$

donde μ y σ son la media y la desviación estándar, el subíndice 1 se refiere a la estación de corto tiempo, mientras que el subíndice 2 se refiere a la estación de largo plazo, a indica el periodo de corto plazo y b indica el periodo base, R es el coeficiente de correlación entre 1 y 2.

Si las ecuaciones son derivadas por estandarización las medias y las desviaciones estándar de los dos periodos de registro en cada estación, el R^2 no es estadísticamente correcto y debe ser omitido.

En el caso de la distribución Pearson tipo III, el dato del sesgo es necesario, las ecuaciones anteriores son empleadas para generar eventos a corto plazo para completar el registro para todos los años, empleando la siguiente relación:

$$x_1 = \mu_{1b} + R(x_2 - \mu_{2b}) \sigma_{2b} / \sigma_{1b} \quad (10.52)$$

donde x_1 es el caudal pico de descarga y debe ser estimado en la estación del periodo corto, x_2 es el caudal pico de descarga y debe ser medido en la estación del periodo largo.

Cuando el número completo de los eventos anuales está disponible para cada estación de los periodos cortos, los coeficientes de sesgo son calculados en cada estación.

Los parámetros de la distribución (media, desviación estándar, y para la distribución Pearson tipo 3, sesgo), se relacionan con las características de la cuenca por medio de un análisis de regresión múltiple. La transferencia de información para cualquier dado cierto periodo de retorno se estima por medio de la siguiente ecuación:

$$x_T = \mu + K\sigma \quad (10.53)$$

donde K es el factor de frecuencia.

10.1.7. Análisis regional de máximos

El procedimiento del método análisis regional de máximos (Kite, 1977), considera el dato máximo de cada estación.

Se considera que existe una serie de n valores independientes que contienen k valores extremos. El máximo evento de las n series se abstrae y se ordena del más grande al más pequeño en una nueva serie, llamada y .

Una analogía puede ser hecha con el caudal máximo anual registrado en estaciones aforadas dentro de una región hidrológicamente homogénea.

Por lo que la misma distribución de probabilidad se aplica a los registros de los eventos máximos en cada corriente.

Tal como en el método de la avenida índice los diferentes registros pueden ser reducidos en una idéntica distribución normalizando las avenidas medias anuales. El procedimiento de contorno toma n muestras independientes de k eventos y forma el tamaño de muestra nk , en los cuales las probabilidades pueden ser calculadas asociando periodos de retorno para nk años (en vez de considerar n años).

El problema es que los registros de caudal no son independientes, pero están fuertemente correlacionados. Esto reduce el período de máximo retorno disponible de nk para $f(R)nk$ donde $f(R)$ es una función del coeficiente de correlación, R , entre los registros de caudal. La expresión R , varía entre registros de caudal entre 1 (cuando $R=0$, registro independiente) a $1/n$ (cuando $R=1$ registros idénticos).

La probabilidad de otro evento aleatorio, y , excediendo el orden del registro máximo, y_1 , no puede ser determinado analíticamente cuando los registros no son independientes. Ordenando los registros máximos.

Por lo tanto, la ecuación empleada se desarrolla empleando la distribución normal de la siguiente forma (Kite, 1977):

$$x = B\varepsilon \quad (10.54)$$

donde x es una matriz de $n \times k$ de los eventos generados, B es una matriz de $n \times n$ que contienen los componentes principales y ε es una matriz de $n \times k$ de números aleatorios independientes normalmente distribuidos con media cero y varianza unitaria. El componente de la matriz B se deriva a partir de la matriz de correlación R de los n registros con k eventos hidrológicos, de la forma siguiente:

$$B = E\lambda \quad (10.55)$$

donde E es una matriz de $n \times n$ con valores propios obtenidos de R , y λ es una matriz de $n \times n$ con elementos diagonales siendo la raíz cuadrada de los eigenvalores de R y los elementos fuera de la diagonal, que son cero (Kite, 1977).

Después de la generación de x , los n máximos se seleccionan y se colocan en orden de magnitud y de probabilidades de excedencia empleando una aproximación a la distribución normal (Kite, 1977).

En resumen, el método extrae de una serie de registros de estaciones aforadas una matriz de correlación entre estaciones. Esta correlación es incorporada en un número de eventos generados por distribuciones extremas.

El método convierte la distribución-espacial en distribución-tiempo para eventos extremos.

10.2. Validación del método regional

En diversos estudios se ha aplicado el re-muestreo con el procedimiento denominado Jack-knife, básicamente este procedimiento implica la remoción de una estación aforada del conjunto de estaciones, asumiendo que ésta es una estación no aforada.

Gupta *et al.* (2008) proponen tres aspectos básicos para la evaluación del desempeño de modelos dinámicos (De Vos *et al.*, 2010): 1) Evaluación cuantitativa del modelo, es decir, la obtención de las mediciones estadísticas de las variables observadas y simuladas. 2) Evaluación cualitativa de la consistencia del modelo, es decir, realizar pruebas de análisis de sensibilidad e inspecciones visuales del comportamiento del modelo, y 3) evaluación cualitativa de la forma y función del modelo.

El análisis de incertidumbre se ha desarrollado el enfoque bayesiano el cual surge como un análisis de incertidumbre entre los grupos homogéneos, siendo pocos los estudios donde se ha desarrollado (GREHYS, 1996 y Dodangeh *et al.*, 2014). Dodangeh *et al.* (2014) apuntan que el principal mérito del enfoque bayesiano “es que la distribución posterior utilizada en la estimación puede reducir la incertidumbre de los parámetros del modelo y las estimaciones cuantitativas” además que “pocos son los estudios hidrológicos que emplean el enfoque bayesiano”.

Wazneh *et al.* (2013) proponen la validación del método por sesgo y error medio cuadrático. Donde la técnica de Jack-Knife ha sido empleada para evaluar el desempeño del modelo. El procedimiento es el siguiente (Ouarda *et al.*, 2006) en cada región, cada una de las estaciones se considera como no aforada, por lo que esta estación es removida de la base de datos. El siguiente paso consiste en evaluar los cuantiles de inundación con las cuencas o estaciones remanentes por medio de los métodos de delimitación que haya elegido el investigador, por citar un ejemplo, puede ser por medio de un ACC y por medio de regresión múltiple en el sitio no aforado, para la estimación del cuantil que corresponde a los distintos periodos de retorno de diseño, cabe mencionar que este procedimiento debe evaluarse para todo el conjunto de estaciones aforadas o de cuencas aforadas. El siguiente paso consiste en la comparación de los cuantiles mediante alguna función de probabilidad, GVE en el sitio, por

lo que las dos mediciones para evaluar el desempeño son el error medio cuadrático y la tendencia:

$$BIAS_T [\%] = \frac{1}{N} \sum_{i=1}^N \left(\frac{Q_T^i - Q_T^i}{Q_T^i} \right) \times 100 \quad (10.56)$$

$$RMSE_T [\%] = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{Q_T^i - Q_T^i}{Q_T^i} \right)^2} \times 100 \quad (10.57)$$

donde Q_T^i es el cuantil para el periodo de retorno T en el sitio i , Q_T^i es el valor observado del cuantil asociado al periodo de retorno T en el sitio i , N es el número de sitios en la región.

Otras medidas de validación implican el desarrollo de ecuaciones como el valor del criterio de *AKAIKE* (*AIC*) (Ahn y Palmer, 2016)

$$AIC_j = -2\Pi(y_j) - 2g \quad (10.58)$$

donde $\Pi(y_j)$ es el logaritmo de la verosimilitud de la función maximizada, g es el número de parámetros ajustados.

Otro ejemplo consiste en una vez removida una estación o cuenca a la vez, con las cuencas remanentes de la región, de cada región que se ha asumido como homogénea, se construyen ecuaciones de regresión múltiple con el fin de obtener los cuantiles de inundación para distintos periodos de retorno en el punto donde se asume no aforado. La estimación final de los cuantiles se obtiene mediante el método elegido como el óptimo junto con un procedimiento de estimación de regresión múltiple (Ouarda *et al.*, 2006).

El método de validación cruzada ha sido empleado para la medición del error en la predicción del caudal en zonas no aforadas, este parámetro se puede calcular por medio del error cuadrático promedio (Laaha y Blöschl, 2006). En el empleo de este método, algunos autores (Laaha y Blöschl, 2006) optan por dejar fuera una estación la que asumen como no aforada dentro de la región homogénea con el propósito de obtener una predicción más precisa, la ecuación es la siguiente:

$$V_{CV} = \frac{1}{n} \sum_{i=1}^n (\hat{Q}_i^{(-i)} - Q_i)^2 \quad (10.59)$$

donde n es el número total de cuencas, Q_i es el caudal observado para la cuenca i y $\hat{Q}_i^{(-i)}$ es el modelo de predicción sin emplear los gastos observados para la cuenca i .

La robustez del modelo es posible evaluarla por medio de la prueba estadística PRESS, *Prediction Residuals Sums of Squares*, es decir, de los datos se remueve una observación y se re-ajusta el modelo de regresión: Predicción residual = Valor predicho del modelo – valor en un cierto punto. Esta técnica fue aplicada por (Sanborn y Bledsoe, 2006) para validar el modelo resultante en caudales mínimos.

10.3. Comentarios y recomendaciones

En este documento se muestran los diversos métodos de transferencia de información hidrológica. Métodos como el de la avenida índice y regresión lineal múltiple son los más comúnmente empleados. En el método de regresión múltiple se recomienda realizar una selección de variables por medio de un Análisis de Componentes Principales, mientras que en el método de la avenida índice es necesario emplear la versión corregida del test de Langbein para verificación de la homogeneidad de los grupos formados y con ello se obtendría una clasificación óptima.

Otros métodos menos empleados, pero efectivos, son el método de estaciones-años que es recomendable emplear cuando el periodo de registro es común entre las muestras.

11. FUNDAMENTOS GENERALES EN EL DISEÑO DE UN SISTEMA DE MINIGENERACIÓN

11.1. Diseño de un sistema

El procedimiento para diseñar un sistema a nivel microcentral hidroeléctrico tiene cuatro etapas (Coz *et al.*, 1995):

1. Capacidad y estudio de la demanda: Consiste en cuantificar la energía necesaria en función de la demanda de la misma.
2. Estudio hidrológico e inspección del lugar: En esta etapa se establece el potencial hidroenergético del sitio elegido con base en la potencia disponible, así como la variación del caudal a lo largo del año y los usos del agua en la región.
3. Estudio de pre-factibilidad: En esta etapa se identifican de tres a cuatro opciones distintas, para satisfacer la demanda del consumidor, incluyendo sus ventajas y desventajas, estas opciones incluyen la propuesta de diseños distintos de microcentrales, la extensión de líneas de redes nacionales o el uso de un generador de diésel, así como sus costos comparativos. Además, en esta etapa se comparan los resultados de los estudios de demanda de energía con los estudios hidrológicos.
4. Estudio de factibilidad final: Esta etapa es de continuidad al seleccionar una de las opciones del estudio de pre-factibilidad, donde la regla de oro en el orden de importancia de los estudios a detalle es la siguiente: estudios de operación total y de mantenimiento, estudio económico y por último diseño de ingeniería.

En la etapa de factibilidad se establecen las tarifas del sistema, así como su implementación, es posible incluir medidas para el bienestar, medidas para la acumulación de fondos de desarrollo de usos finales de la hidroenergía, planes de contingencia, así como términos de referencia para comités de supervisión.

11.2. Microhidrogeneración

El agua almacenada en presas se emplea para distintos propósitos desde la generación eléctrica, abastecimiento de agua, navegación, producción pesquera, formación de centros de recreo, control de avenidas y proporcionar agua a las zonas agrícolas (Dolores-Juárez, 1992),

actualmente el interés por el desarrollo de energías limpias consiste en la adecuación de presas de usos distintos enfocadas a la mini y micro-generación (Palacios-Fonseca *et al.*, 2017)

Los esquemas de generación se clasifican en cuatro tipos los de gran generación (mayores a 30 MW de potencia), pequeña (menor a 30 MW), mini (de 1 a 5 MW) y micro (menor a 1 MW) (Tabla 11-1).

Tabla 11-1 Clasificación de las presas (CFE, 1980)

Categoría	Almacenamiento (10⁶m³)	Altura (m)	Pérdida de vidas	Daños
Grande	>60	>18	Considerables	Catastróficos
Intermedia	1.2 a 60	12 a 30	Pequeña	5 a 10 veces el costo de la presa
Pequeña	< 1.2	<15	Ninguna	Mismo orden de magnitud que el costo de la presa

De la misma tabla se obtuvieron los siguientes datos para una avenida de diseño de una obra con categoría de grande lo siguiente: Relación lluvia-escorrentamiento usando la precipitación máxima probable. Categoría intermedia: Relación lluvia-escorrentamiento basada en la tormenta más severa registrada en el pasado, maximizada por el punto de rocío, o periodo de retorno de 1000 años. Categoría pequeña: Periodo de retorno de 50 a 100 años.

Mientras que la clasificación de las plantas hidroeléctricas está en función de la posición de la casa de máquinas con respecto a la cortina, por lo que se pueden clasificar en tres tipos, las hidroeléctricas de río, las hidroeléctrica con casa de máquinas junto a la cortina y las hidroeléctricas con derivación (Dolores-Juárez, 1992).

La generación de la energía se procesa por la energía potencial almacenada debida a la diferencia de altura superficial y la base de la presa, donde se ubica la central. Por medio de los ductos se transforma de energía potencial en energía cinética donde la generación de electricidad se realiza por medio del movimiento de las turbinas. Por lo que la selección de la turbina estará en función del caudal, así como los ductos que transportarán el agua y que moverán las turbinas (González-García, 2014).

La importancia a nivel mundial sobre los proyectos en mini y micro centrales está relacionada con la solución de problemas de abastecimiento de energía en zonas alejadas de las grandes ciudades, así como el aprovechamiento del caudal en zonas estratégicas y donde su uso ha sido desaprovechado (González-García, 2014).

La ventaja de la generación hidroeléctrica en una microcentral está relacionada en el aspecto económico de la red, debido a que la instalación de esta central es independiente, por lo que el costo de las líneas de transmisión y el equipo no son tan elevados en comparación con los existentes de la red eléctrica (Coz *et al.*, 1985). Un esquema actual es la adecuación de presas para aprovechamiento de generación (Palacios-Fonseca *et al.*, 2017).

La propuesta de la generación a escala de micro generación ha ido cobrando relevancia en distintas partes del mundo China, Nepal, Himalaya, España, Bolivia, Perú y los Andes (González-García, 2014).

Desde el año de 1992 se realizó una modificación a Ley del Servicio Público de Energía Eléctrica donde se insta a retomar la estimación del potencial hidroeléctrico a pequeño, mini y micro escala, tanto para particulares, así como para CFE (González-García, 2014).

11.3. Los componentes de un sistema de hidrogenación

Las plantas hidroeléctricas cuentan con siete obras principales cuya función es la generación de energía eléctrica: retención, toma, derivación, obra puesta a presión, tubería a presión, casa de máquinas y desfogue, donde la cantidad de componentes dependen del tipo de central que se instale (Dolores-Juárez, 1992).

Un sistema típico de microhidrogenación (Figura 11-1) puede diferir de un lugar a otro, debido a que depende en gran medida de las condiciones del terreno. Por lo general el sistema contiene una barrera que desvía el caudal de agua a través de una apertura al costado del río (obra de toma) hacia un canal abierto. Por medio de un desarenador se separan las partículas de arena del agua. El canal sigue los contornos del cerro con el fin de mantener la elevación del agua derivada. A continuación, el agua ingresa a un depósito (cámara de carga) y luego pasa a una tubería cerrada (tubería forzada), la cual está conectada a una máquina hidráulica, denominada turbina, dentro de casa de máquinas, produciendo energía, donde la carga es el

conjunto de maquinaria y artefactos que serán abastecidos de energía por el sistema de hidro generación (Coz *et al.*, 1985).

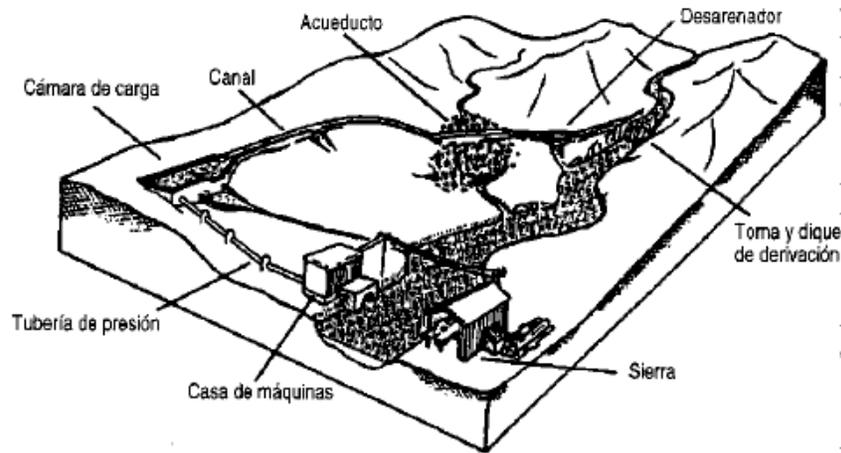


Figura 11-1. Componentes de un sistema de micro hidrogeneración (Coz *et al.*, 1985).

11.4. Energía a partir del agua y cálculo de potencial

En la etapa de diseño de una central hidroeléctrica es primordial el cálculo de la potencia instalable, para ello se revisa la potencia de un tramo de río, que comprende la obra de toma hasta la ubicación del desfogue.

$$\bar{P} = \gamma * \bar{Q} * h_b * \eta_s * g \quad (11.1)$$

donde \bar{P} es la potencia media en KW, \bar{Q} es el gasto medio aprovechable, γ es el peso específico del agua, h_b es la carga hidráulica disponible en metros que se calcula como la diferencia en metros del nivel del embalse y el nivel de la obra de toma, η_s es la eficiencia total de planta (del 75 al 90%) y g es la aceleración de la gravedad en m/s^2 .

Para que en un hidrosistema se genere una cantidad útil de energía en forma de electricidad o energía mecánica del eje, se requiere de un caudal de agua y una diferencia de altura, distancia vertical a lo largo de la cual cae el agua, denominada salto.

En el proceso de generación de energía, una parte de ésta se pierde en forma de fricción, calor, por lo que ningún sistema de conversión entrega la misma cantidad de energía útil, como la que absorbe.

La ecuación de conversión es la ecuación (11.2)

$$\text{Potencia de entrada} = \text{Potencia de salida} + \text{pérdidas} \quad (11.2)$$

$$\text{Potencia de entrada} = \text{Potencia de salida} * (\text{eficiencia de conversión}) \quad (11.3)$$

Considerando la eficiencia total del sistema

$$P_{\text{neta}} = P_{\text{disp}} * \eta_0 \quad (11.4)$$

donde P_{neta} es la potencia neta, potencia útil entregada, P_{disp} es la potencia disponible y η_0 es la eficiencia total del sistema.

La ecuación fundamental de la potencia hidráulica, donde la potencia disponible es igual al salto disponible (h_{disp}) multiplicado por el caudal Q , y multiplicado por el valor de la gravedad $9,81 \text{ m/s}^2$. La eficiencia total del sistema involucra todas las pérdidas del equipo y maquinaria del sistema de generación (Figura 11-2 y Ecuación (11.1)), si se resuelve la Ecuación (11.1) con los valores típicos del sistema, se obtiene un valor de $\eta_0 = 0.50$, la eficiencia varía de 0.40 a 0.60 .

$$P_{\text{neta}} = 9.81 * h_{\text{disp}} * Q * \eta_0 \quad (11.5)$$

donde P_{neta} es la potencia neta, potencia útil entregada (kW), h_{disp} es el salto disponible (m), Q es el caudal en m^3/s y η_0 es la eficiencia total del sistema en (%).

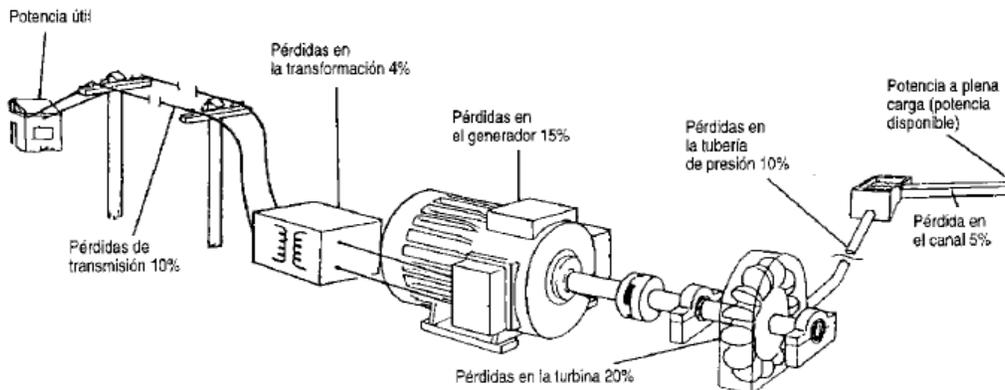


Figura 11-2. Pérdidas típicas del sistema para una microcentral operando a plena carga (Coz et al., 1995).

Para mayor información de la pérdida y eficiencia en los equipos consultar a Coz (1985)

Tabla 11-2. Pérdidas y eficiencias típicas a carga parcial

Equipo/maquinaria	Pérdidas	Eficiencia
Canal	5%	95%
Tubería de presión	10%	90%
Turbina	20%	80%
Generador	15%	85%
Transformación	4%	96%
Líneas de transmisión	10%	90%

$$\eta_0 = \eta_{obra\ civil} \times \eta_{tubería} \times \eta_{generador} \times \eta_{línea} \quad (11.6)$$

11.5. Generación media anual

Es la potencia por el número de horas de operación en un año, y está definida por el Factor de Planta, F.P., el cual cuando es cercano a uno, se considera que la central trabaja en base o firme, es decir, las 24 h del día los 365 días del año, también se puede estimar en función del valor regional de las centrales hidroeléctricas actuales de 40.1% (Palacios-Fonseca *et al.*, 2017).

La Generación Media Anual, G.M.A., se expresa en GWh/año

$$GMA = \bar{P} * 8640 * FP / 1000 \quad (11.7)$$

11.6. Factor de planta

El costo final de la energía producida por una microcentral, depende en gran medida del factor de planta. El factor de planta, también llamado factor de capacidad, que explica la relación existente entre la energía usada con la energía disponible en el sistema hidráulico (Coz *et al.*, 1985). El factor de planta es la relación entre la energía anual que produce la planta y la que podría generar si trabajara a plena carga todo el año (Dolores-Juárez, 1992).

Por lo que la recomendación esencial es diseñar para el factor de planta más alto posible, si se obtiene un bajo factor de capacidad significa energía costosa, indicando como alternativa otras fuentes de energía (Coz *et al.*, 1985). En el caso que se requiera que la planta trabaje únicamente cubriendo picos de demanda el factor de planta debería de ser bajo.

$$\text{Factor de planta} = \frac{\text{Potencia usada} \times \text{tiempo de potencia usada}}{\text{Potencia instalada} \times \text{periodo considerado}} \quad (11.8)$$

11.7. Factores de carga y de planta.

El término factor de carga se emplea cuando se planea un sistema de suministro de energía y sistema de tarifas. El factor de carga considera las características de la conducta de consumo y el consumo promedio de energía por consumidor (Coz *et al.*, 1985).

$$\text{Factor de carga} = \frac{\text{Energía total usada por los consumidores}}{\text{Capacidad total de energía conectada a los consumidores}} \quad (11.9)$$

11.8. Costo unitario de la energía

Cuando se planea un proyecto hidroeléctrico, una primera aproximación para conocer si la propuesta de un nuevo sistema hidroenergético es viable es si la electricidad que éste producirá será más barata o más cara que la electricidad producida por otros medios (Coz *et al.*, 1995). El costo unitario está expresado en \$/kWh

$$\text{Costo unitario de la energía} = \frac{C_{\text{anual}} + (O + M)}{P_{\text{instalada}} \times 8760 \times FP} \quad (11.10)$$

donde C_{anual} es el costo anual total (\$), $O + M$ es el costo de operación y mantenimiento (\$), $P_{\text{instalada}}$ es la capacidad de la potencia instalada en el sistema en (kWh), 8760 es la cantidad de horas al año, y FP es el factor de planta (adimensional).

11.9. Decisiones de costo-beneficio

Las categorías que integran los costos del sistema microhidroenergético son: los costos del capital y los costos corrientes (Coz *et al.*, 1985).

En cuanto a los costos de capital, se deben considerar los costos por: planteamiento y diseño, administración y finanzas, obras civiles, equipo electromecánico, distribución de electricidad, instrumentación y contingencias.

Para estimar los costos corrientes, se consideran los costos anuales de operación y mantenimiento; los costos variables, como pueden ser los costos de contratación de personal, capacitación, materiales, herramientas, equipos de repuestos, y por último las contingencias.

11.10. Capacidad y estudio de la demanda.

Este estudio establece cual es la demanda existente para una nueva fuente de energía y si existe un interés genuino y capacidad para pagar por el nuevo suministro propuesto. Además de los beneficios y desventajas que ocasionará a la población el nuevo sistema (Coz *et al.*, 1985).

Para realizar una proyección de demanda es necesario determinar las características de la zona, como el número de usuarios que se verían beneficiados con la implementación del sistema (INEA, 1997)

Por lo que una vez finalizado el estudio, se emite un informe que debe cubrir los siguientes tópicos (Coz *et al.*, 1985):

- a) Mapa de la región mostrando la ubicación de viviendas actuales, así como de las futuras actividades comerciantes o productivas.
- b) Informe de las encuestas a la población de cómo los sistemas propuestos pueden afectar a su seguridad económica y oportunidades en el futuro.
- c) Resumen de los organismos e instituciones que ayuden a la organización de la financiación, mantenimiento y operación de los nuevos sistemas.
- d) Descripción completa de los sistemas de irrigación actuales y su administración y planes futuros de irrigación, así como un informe completo de cómo esperan los pobladores que la energía hidráulica afecte sus planes de irrigación.
- e) Un diagnóstico de la capacidad de la organización local o individual para administrar el sistema complejo, incluyendo financiación, distribución de bienestar, operación y mantenimiento de maquinaria.
- f) Informe de la disposición y habilidad para pagar por el sistema propuesto, que tienen los habitantes, así como la proyección de la demanda futura.
- g) Por medio de diagramas de energía requerida diaria y anual, representar la cantidad de energía requerida, qué uso y cuándo se emplearía.
- h) Descripción de nuevos artefactos que puedan hacer uso de la fuente de energía y cómo pueden ser adquiridos y operados, así como las ganancias por su uso deben ser recaudadas y empleadas.
- i) Proponer algunas precondiciones organizacionales para elevar la capacidad a niveles requeridos.

- j) Un plan para un sistema de administración, explicando cómo las ganancias y tarifas serían recaudadas y acumuladas, cómo gastar, cómo se deberá organizar la operación y mantenimiento y planes de contingencias para acomodarse a las posibles dificultades futuras.

Informes de factibilidad

El informe de factibilidad puede llegar a tener al menos los siguientes temas principales (Coz *et al.*, 1985):

- a) Resumen
- b) Ilustraciones clave
- c) Demanda de la energía
- d) Opciones de suministro de energía
- e) Capacidad de administración
- f) Potencial hidráulico
- g) Diseño de la Microcentral hidroeléctrica
- h) Factor de planta: Comparación del suministro y la demanda.
- i) Operación y mantenimiento.
- j) Estructura de administración
- k) Provisiones de administración
- l) Costos de mantenimiento
- m) Plan de operaciones
- n) Análisis de costos
- o) Ingresos
- p) Bienestar
- q) Estructura tarifaria
- r) Análisis financiero
- s) Fuentes de financiamiento
- t) Viabilidad socioeconómica
- u) Monitoreo de planes de contingencia.

Estudios generales en presas pequeñas (anteproyecto)

Una vez que se defina al proyecto como factible, la etapa de estudios generales del sitio debe llevarse a cabo. Los estudios generales que deben contemplarse en la implementación de presas pequeñas (Marsal, 1974) son los siguientes:

Para estudios de la presa:

- Recopilación de datos pluviométricos, topográficos y geotécnicos (Camargo-Hernández y Franco, 1998)
- Trabajos complementarios de la información existente y su procesamiento.
- Anteproyectos del bordo, vertedor y obra de toma.

- Costos y programas de construcción y comparación de alternativas.

En estudios del aprovechamiento:

- Recopilación de datos del área a cultivar y calidad de la tierra, necesidades del agua potable, beneficios del control de avenidas, entre otros (Camargo-Hernández y Franco, 1998).
- Evaluación de datos
- Anteproyectos del sistema de riego de la planta de agua potable, según sea el caso de los alcances de la planta.
- Costos y programas de construcción
- Comparación de alternativas.

11.11. Evaluación del recurso hidroenergético

La capacidad de generación de energía mediante el empleo de agua está determinada por el salto o caída, que se pueda obtener y del caudal disponible. A su vez, el salto depende de la topografía del terreno y el caudal de las características del río o arroyo que se va a utilizar.

Análisis de la información hidrométrica

Para seleccionar algún método en específico para calcular la información hidrométrica, es necesario considerar los siguientes aspectos: a) el daño que produciría una falla y éste a su vez, está relacionado con el periodo de retorno, b) la cuantificación del volumen de la avenida y c) la disponibilidad de la información.

Para obras pequeñas, con orden menor de 100 años como periodo de retorno se propone el empleo de métodos estadísticos con el registro de gastos máximos (CFE, 1980).

5.1.1.1. Histograma de frecuencia de caudales

El proceso estadístico, a partir del dato de caudal ordenado de mayor a menor, es la estimación de la frecuencia absoluta de cada rango, la frecuencia relativa y la frecuencia relativa acumulada en porcentaje.

5.1.1.2. Curva de duración de caudales

La curva de frecuencias relativas permite visualizar a priori la mayor o mejor concentración de datos. Se calcula la frecuencia relativa de cada rango:

$$fr(\%) = \left(\frac{F}{N} \right) * 100 \quad (11.11)$$

El caudal medio se determina mediante la siguiente expresión

$$Q_m = \sum Q_m fr / 100 \quad (11.12)$$

La energía total generada al año para los diferentes caudales de diseño se obtiene al multiplicar la potencia por la duración:

$$Energía = \frac{Pot \times Duración \times 8760}{100} \quad (11.13)$$

donde *Energía* es la energía total generada en (kWh/año), *Pot* es la potencia en (kW), *Duración* está en % y 8760 es la cantidad de horas al año.

5.1.1.3. Curva duración de potencias

La cantidad de energía posible de generar se estima por medio del trazo de la curva de duración de potencias hasta convertir el eje de ordenadas de la curva de duración en eje de potencias multiplicando por $\gamma H \eta$.

5.1.1.4. Estimación de los caudales de avenidas

Los datos de avenidas servirán para adecuar y dimensionar las obras civiles, generalmente se trabaja con avenidas de 1 en 50 años o de 1 en 100 años (Coz *et al.*, 1985).

11.12. OBRAS CIVILES

Introducción

En la Figura 11-3 se muestran los distintos componentes de la ruta de abastecimiento de agua para una microcentral hidráulica, sin embargo, se deben de considerar los factores que intervienen en su diseño (Coz *et al.*, 1995):

- Barraje de derivación
- Toma de ingreso
- Compuertas de regulación
- Aliviaderos
- Canales de rebose
- Desarenador
- Canal

- Cámara de carga
- Tubería de presión
- Cruces de canal
- Apoyos de tubería
- Anclajes de tubería

Dependiendo del tipo de sistema las características de las obras de toma, conducción hidráulica golpe de ariete y pozos de oscilación tendrán alguna variación en sus componentes (Dolores-Juárez, 1992)

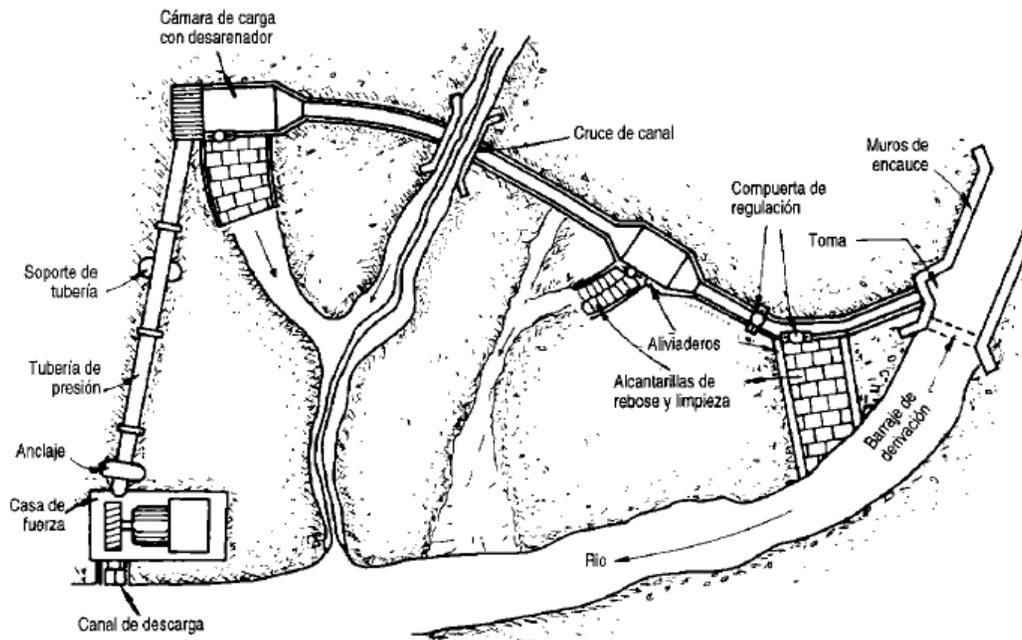


Figura 11-3. Componentes de una microcentral hidráulica (Coz et al., 1995).

5.1.1.5. Esquema general para una microcentral hidráulica

Tubería corta

Este tipo de tubería se elige cuando se tiene una pendiente escarpada, aunque el costo de la implementación de la misma sería más económico, las condiciones de operación y costo de mantenimiento se incrementarían si en esa zona llegara a presentarse la erosión y sedimentación.

Tubería larga

En estos casos la tubería se instala a lo largo del río, debido a que no existen las condiciones para la construcción de un canal.

Bocatomas

Las bocatomas son obras hidráulicas cuya función es regular y captar un determinado caudal de agua, para la producción de hidroenergía ya sea mecánica o eléctrica. Éstas permiten tomar el agua de los ríos y conducirla aprovechando la fuerza de gravedad (Figura 11-4).

Las bocatomas deben cumplir con las siguientes funciones (Coz *et al.*, 1995):

- Garantizar la captación de una cantidad constante de agua, especialmente en épocas de estío.
- Impedir, hasta donde sea posible, el ingreso de materiales sólidos y flotantes, haciendo que éstos sigan el curso del río o facilitando la limpieza.
- Proteger el resto del sistema de obras hidráulicas del ingreso de avenidas o embalses que pudieran producirse en las épocas lluviosas.

La ubicación más apropiada para una bocatoma es en los tramos rectos y estables del río, dependiendo de la topografía, la geología, el comportamiento de los suelos, así como de las variaciones hidrológicas del lugar que servirá de emplazamiento (Coz *et al.*, 1995).

La clasificación de las bocatomas es la siguiente (Coz *et al.*, 1995):

- Por el material del cual están hechas (concreto, roca, madera, etc.)
- Por su vida útil (permanentes y temporales)
- Por su forma y diseño (barraje total, sumergido, parcial).
- Por el método de construcción (gaviones, mampostería, emboquillado)

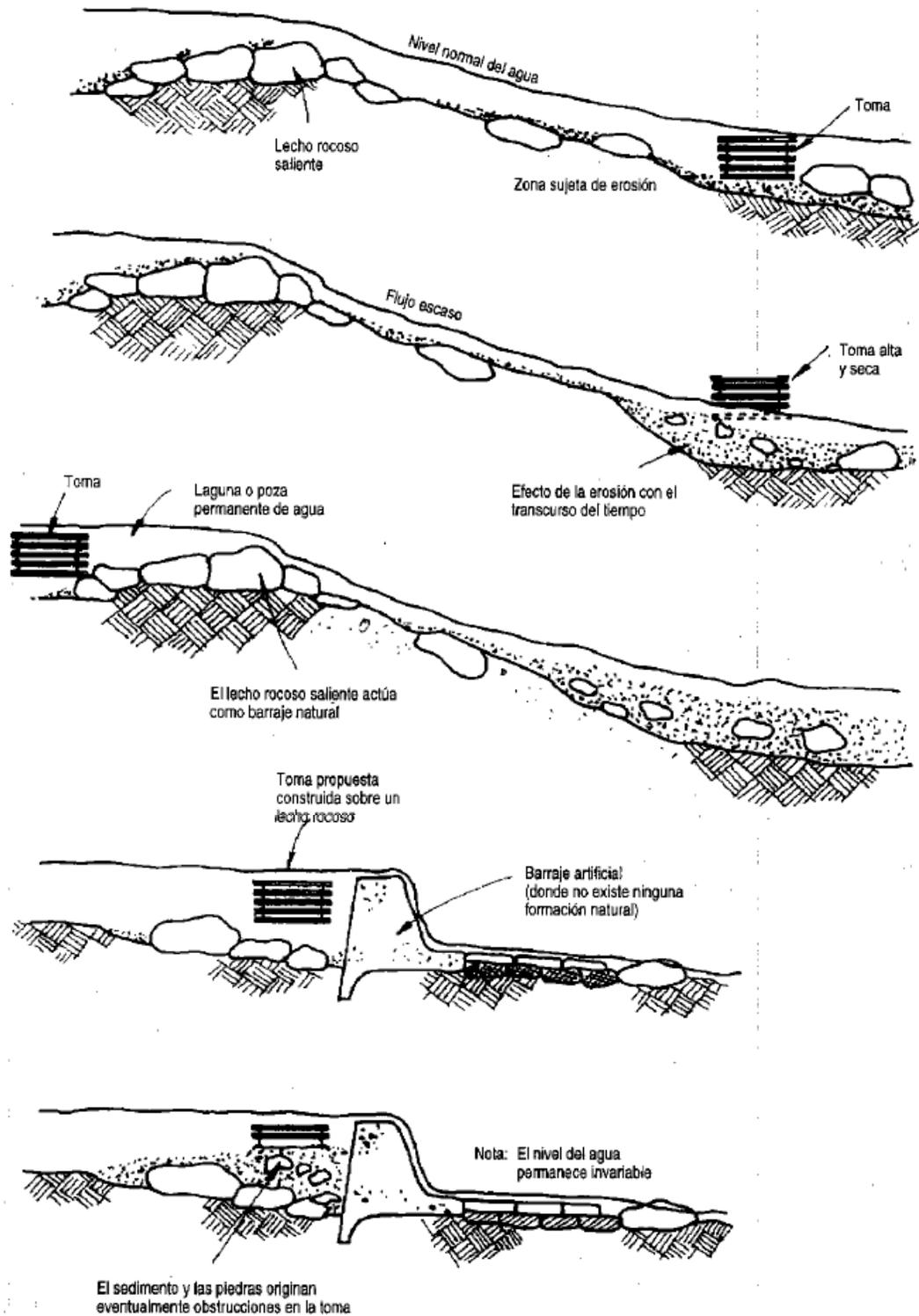


Figura 11-4. Ubicación de la bocatoma y empleo de un barraje de derivación
(Coz et al., 1995).

El diseño de una bocatoma depende de los siguientes parámetros:

- Curso del río (tramo curvo o tramo recto)
- Configuración del terreno (pendiente del cauce, ancho del valle)
- Caudal del río, ángulo de desvío
- Régimen del río
- Acarreo de materiales
- Geología del lugar
- Importancia de la obra, disponibilidad presupuestal.

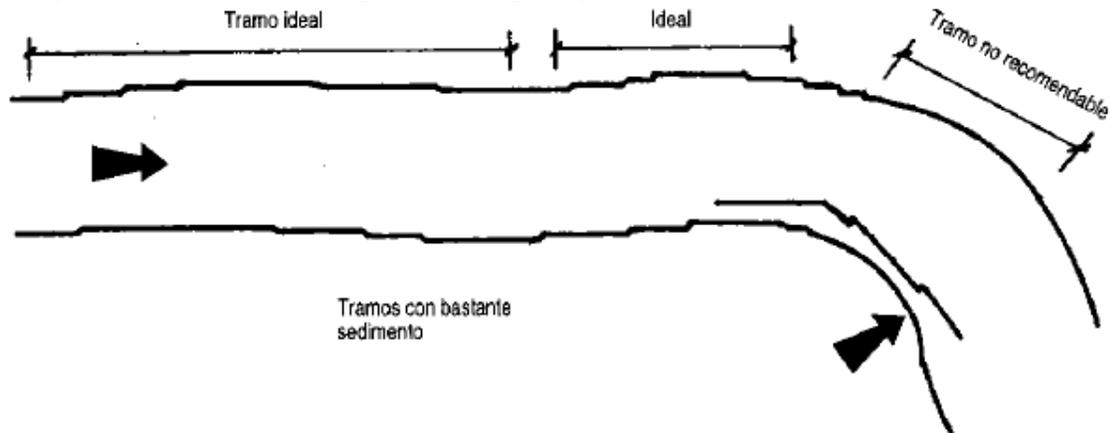


Figura 11-5. Ubicación de la bocatoma (Coz *et al.*, 1995)

Aliviaderos

El aliviadero es un tipo de vertedor, es un dispositivo hidráulico que consiste en una escotadura a través de la cual se hace circular el agua.

Desarenadores y cámara de carga

La principal función del desarenador es que las partículas abrasivas, como las arenas, que afectan a la turbina, se asienten en el fondo para posteriormente ser removidas.

Canales

Existen diferentes tipos de canales, dependiendo de las condiciones del proyecto:

- Canal de tierra sin revestimiento
- Canal de tierra con sello o revestimiento: aplicación de una capa delgada de material que reduce infiltración y fricción.
- Canal de mampostería o concreto
- Acueducto: hecho de planchas de acero galvanizado, madera o tubos cortados por la mitad.

Tuberías de presión

Son tuberías que transportan agua bajo presión hasta la turbina, y permiten aprovechar. Los principales componentes de la misma están en la Figura 11-6.

El costo de la tubería suele representar en gran parte el presupuesto total de la obra, por ello es necesario optimizar su diseño y para ello también existen pautas para seleccionar una tubería de presión.

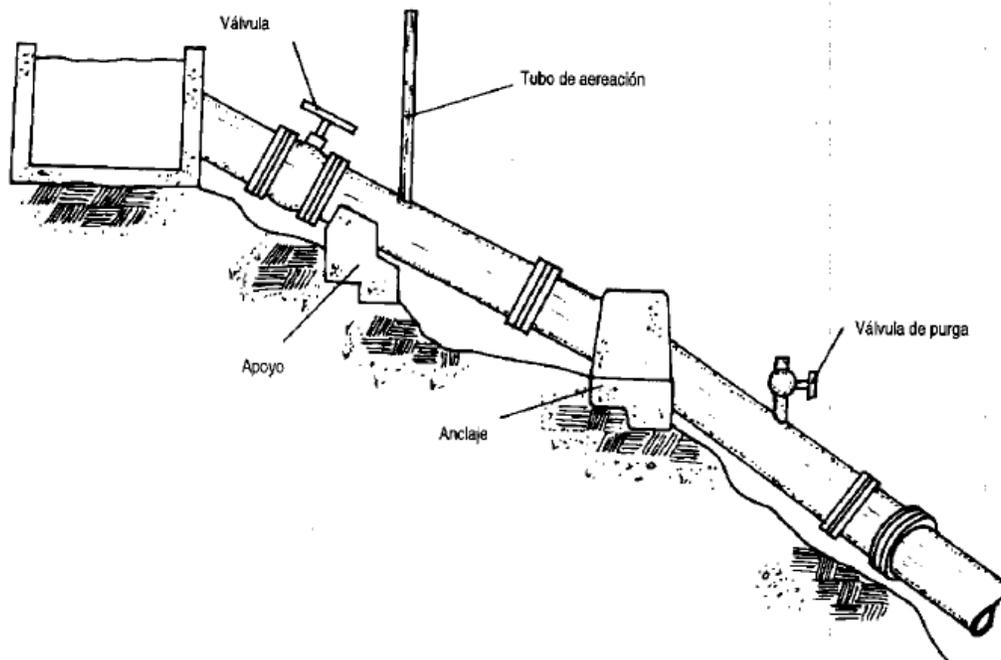


Figura 11-6. Componentes de una tubería de presión (Coz *et al.*, 1995).

Turbinas hidráulicas

5.1.1.6. Definición

Las turbinas hidráulicas son máquinas que transforman en movimiento giratorio la energía cinética o potencial del agua (Dolores-Juárez, 1992)

La clasificación de los diferentes tipos de máquinas que se emplean en el proceso de transformación de la energía son:

- a) Máquinas gravimétricas
- b) Maquinas hidrostáticas o de desplazamiento
- c) Máquinas-roto dinámicas o turbo-máquinas

5.1.1.7. Clasificación

Según la variación de la presión estática a través del rodete.

- a) Turbinas de acción o impulso, cuando la presión estática permanece constante entre la entrada y la salida del rodete (Coz *et al.*, 1985). Son las denominadas Pelton y la principal característica es que esta turbina aprovecha la energía cinética del agua, también se identifican dentro de este tipo a las turbinas Turgo y Michell-Banki (INEA, 1997).
- b) Turbinas de reacción, cuando la presión estática disminuye entre la entrada y la salida de rodete (Coz *et al.*, 1985). Las turbinas de este tipo emplean la energía potencial del agua, las de este tipo son las turbinas Kaplan, Hélice, Francis y los modelos derivados como las Kaplan doble álabe y las diagonales (Dolores-Juárez, 1992). También pertenecen a esta categoría la bomba roto-dinámica, turbina Deriaz, así como las turbinas axiales como la tubular, de bulbo y de generador periférico (INEA, 1997).

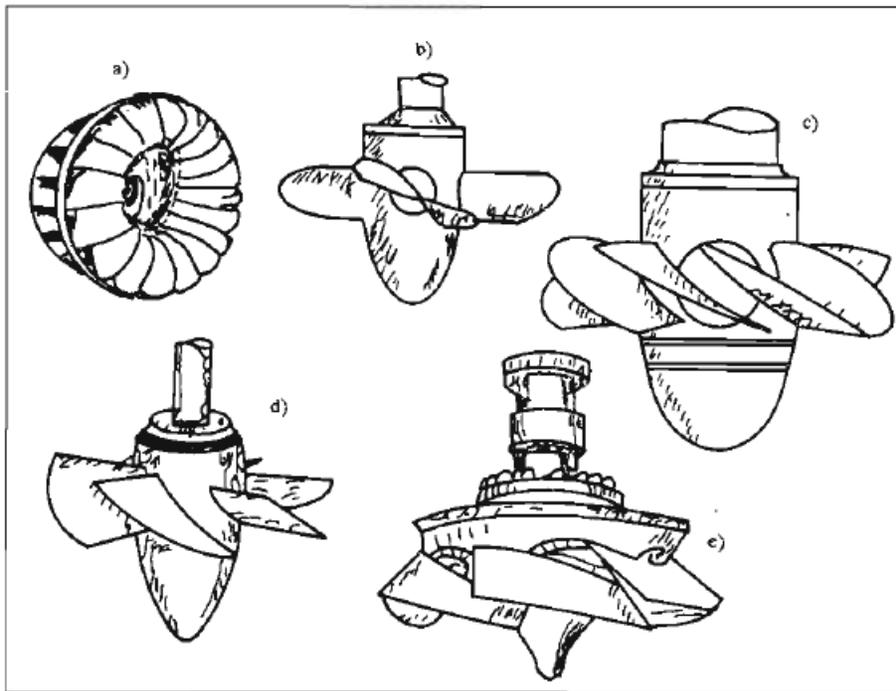


Figura 11-7. Turbinas hidráulicas a) Francis, b) Kaplan, c) Kaplan doble álabe, d) Hélice, e) Kaplan (Dolores-Juárez, 1992).

Tabla 11-3. Características de las turbinas (González-García, 2014)

Turbina	PELTON	TURGO	BANKI-MICHELL	FRANCIS	KAPLAN
Caudal (l/s)	0.5 - 250	90 - 450	2 - 1000	110-1100	100-3000
Caída (m)	30 - 500	30 - 100	1 - 200	10-100	3-15

Según la dirección del flujo a través del rodete

- a) Turbinas de flujo tangencial
- b) Turbinas de flujo radial
- c) Flujo semi-axial
- d) Flujo axial

Según el grado de admisión del rodete

- a) Turbinas de admisión parcial
- b) Turbinas de admisión total.

5.1.1.8. Partes de una turbina hidráulica

Los elementos fundamentales de una turbina hidráulica son los siguientes:

- 1. El distribuidor
- 2. El rodete
- 3. Otros elementos: Tubo de aspiración y carcasa

11.13. Tipos de turbinas hidráulicas

En esta sección se describen las características de las turbinas hidráulicas, que se clasifican en dos grandes grupos, en turbinas de acción y de reacción. Cualquiera de estos dos grupos corresponde a turbinas modernas que se emplean en centrales hidráulicas.

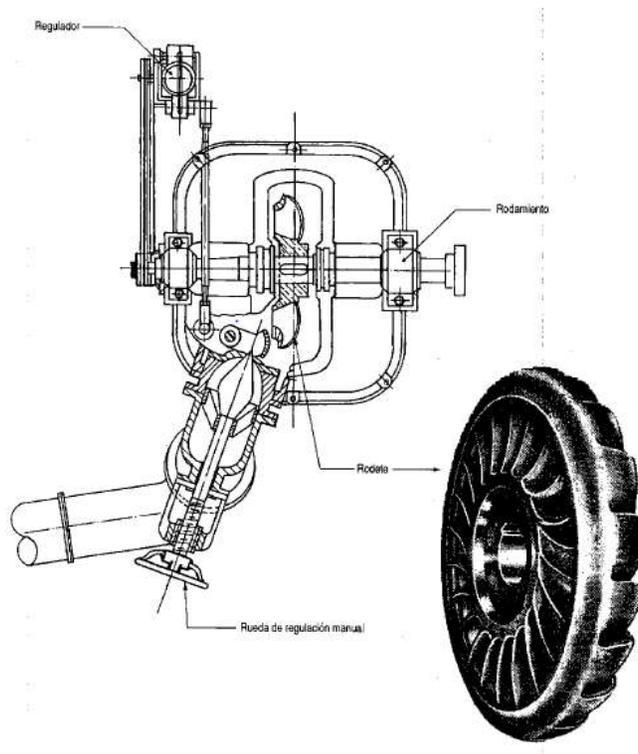


Figura 11-8. Turbina Turgo

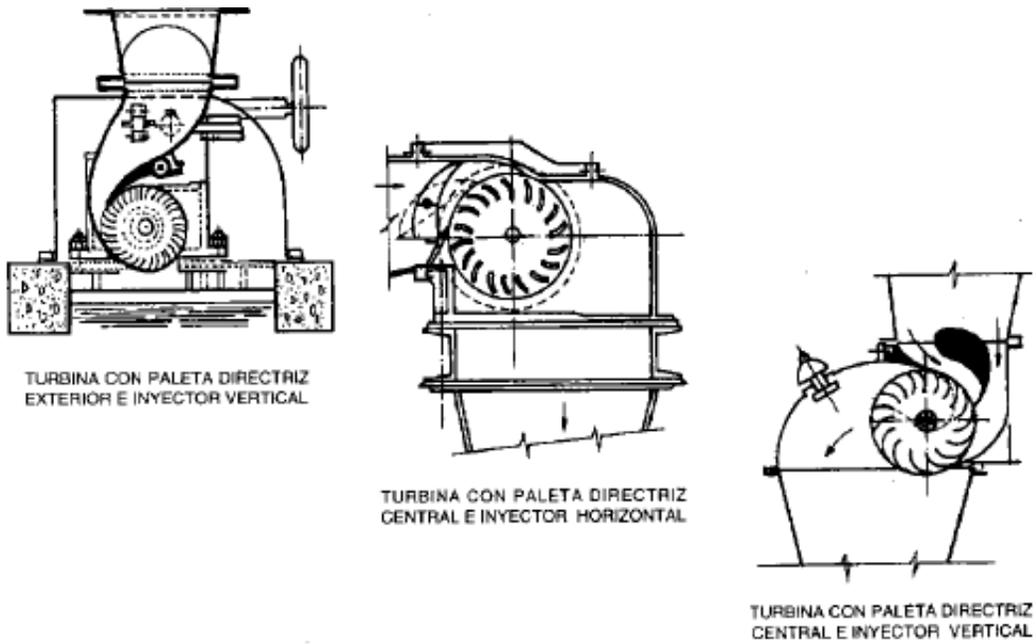


Figura 11-9. Turbina Michell-Banki.

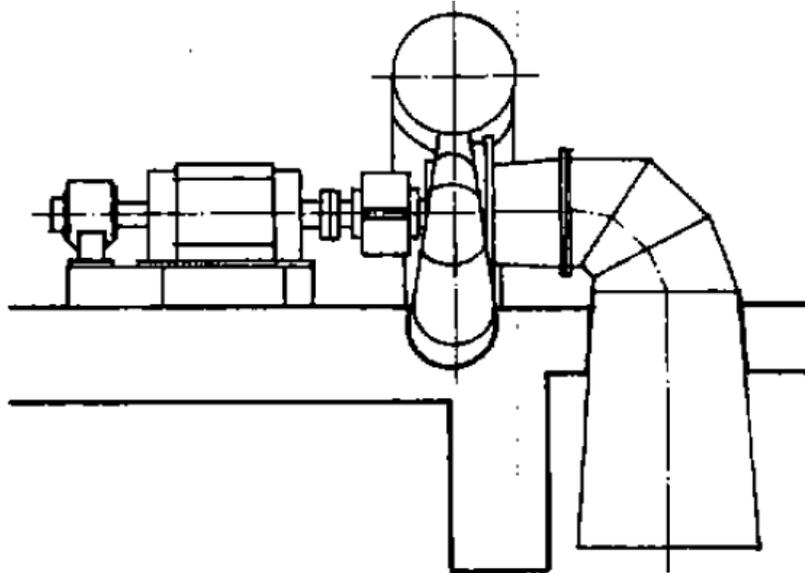


Figura 11-10. Turbina Francis.

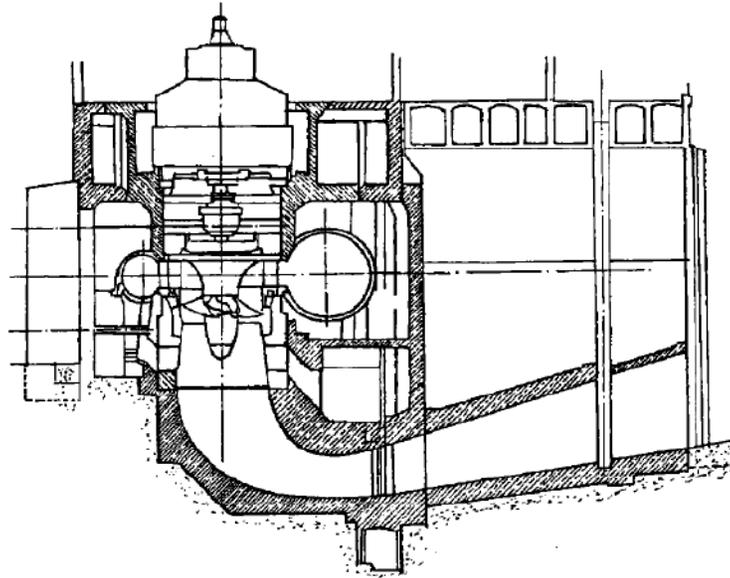


Figura 11-11. Turbina Kaplan

Selección de turbina

En la selección del tipo de turbina existen distintos factores a considerar, uno de ellos es la potencia de la turbina, para lo cual los fabricantes de las mismas, elaboran una tabla de relación de eficiencia de la turbina con la potencia de la misma (Coz *et al.*, 1985).

Otros factores son el costo de la turbina, el aprovechamiento del salto y caudal disponible, la colocación del eje de forma vertical u horizontal, así como la velocidad angular que debe ser lo más alta posible (INEA, 1997).

Para la selección de una turbina se emplea un diagrama de selección (Figura 11-12) en función del salto neto, potencia, caudal, potencia y una eficiencia promedio (Tabla 11-4).

Tabla 11-4. Eficiencia del grupo de generación en porcentaje (Coz *et al.*, 1985)

Potencia (kW)	TIPO DE TURBINA			
	PELTON	MICHELL-BANKI	FRANCIS	AXIAL
<50	58-65 (%)	54-62 (%)	59-65 (%)	58-66 (%)
51-500	65-69 (%)	62-65 (%)	66-70 (%)	66-70 (%)
501-5000	69-73 (%)	65 (%)	70-74 (%)	70-74 (%)

En un sistema de microcentral, la potencia generada se obtiene de la siguiente forma (Coz *et al.*, 1985):

$$H = \frac{P}{\rho g \eta} \times \frac{1}{Q} \quad (11.1)$$

$$\log H = \log \left(\frac{P}{\rho g \eta} \right) - \log(Q) \quad (11.2)$$

Para la selección de la turbina se asume una eficiencia promedio de una relación lineal entre la carga y el caudal, para una potencia constante en coordenadas logarítmicas (Figura 11-12).

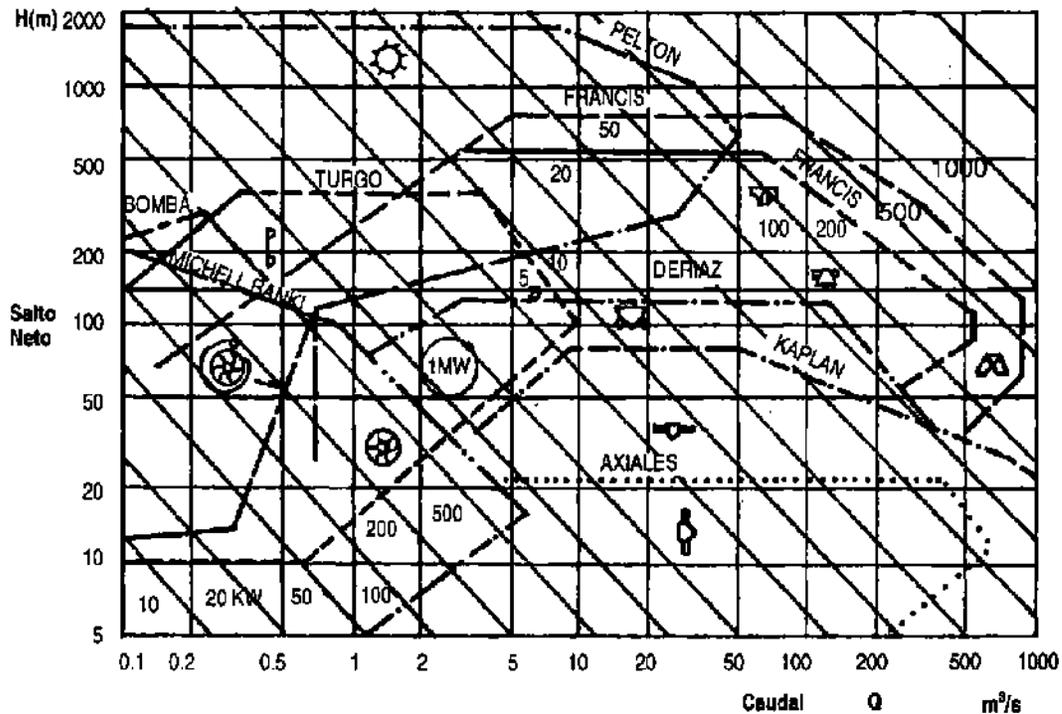


Figura 11-12. Diagrama de selección de turbinas hidráulicas (Coz et al., 1985).

Otra forma de selección es una vez calculada la potencia y la el salto, ubicar en las abscisas la potencia en MW y en las ordenadas el salto en metros (Figura 11-3).

Instalación de turbinas

La instalación de turbinas reactivas requiere de los siguientes elementos básicos (Dolores-Juárez, 1992):

- Cámara espiral o caracol: Sirve para proporcionar la alimentación homogénea a la turbina.
- Estator de la turbina

- Aparato de regulación de la turbina. Regula el gasto de la turbina de modo que se pueda modificar la potencia de la misma según se requiera.
- Rueda y cámara de la turbina.
- Tubo de succión o de aspiración. Este tubo permite aprovechar gran parte de la energía cinética del agua que ya pasó por la turbina y también situarla a un nivel superior del desfogue sin tener pérdidas de carga estática.

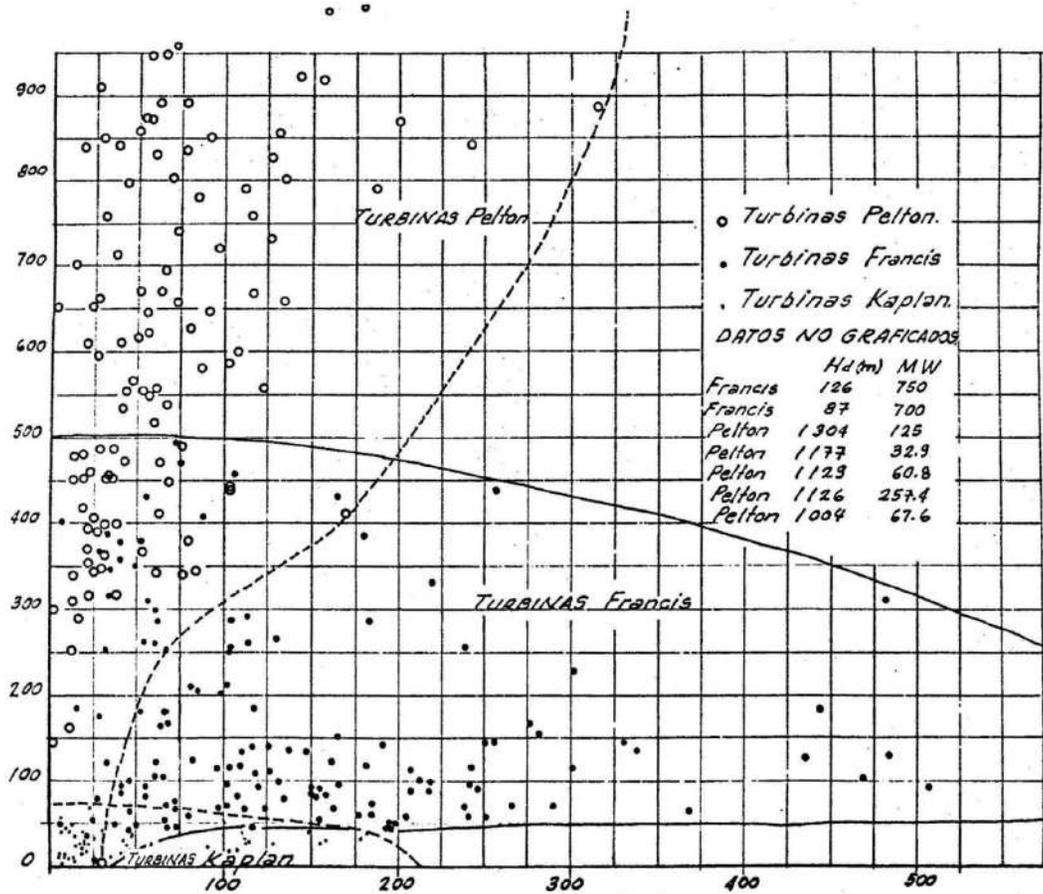


Figura 11-13 Diagrama de selección de turbinas (Nava-Mastache y García-Gutiérrez, 2013)

Curvas características

Las curvas características de una turbina hidráulica, representan el comportamiento de la maquina frente a las condiciones variables del servicio, donde el ensayo constituye la base para las pruebas de recepción de una central (Coz *et al.*, 1985).

La operación de las turbinas se describe mediante las curvas características de eficiencia-potencia (Figura 11-14), universales (Figura 11-15) y de operación (Figura 11-16) (Dolores-Juárez, 1992).

Las curvas características de eficiencia-potencia, representan la eficiencia de las turbinas en función de la potencia, considerando la carga estática y a la velocidad como constantes.

La característica universal, muestra la eficiencia del modelo en las coordenadas de caudal y velocidad referidos, así como las líneas con igual coeficiente de cavitación y apertura de paletas del aparato de regulación.

La característica de operación se utiliza en las plantas hidroeléctricas que están produciendo energía eléctrica, por lo que las curvas de eficiencia se dan en función de la altura y de la potencia, considerando la velocidad constante.

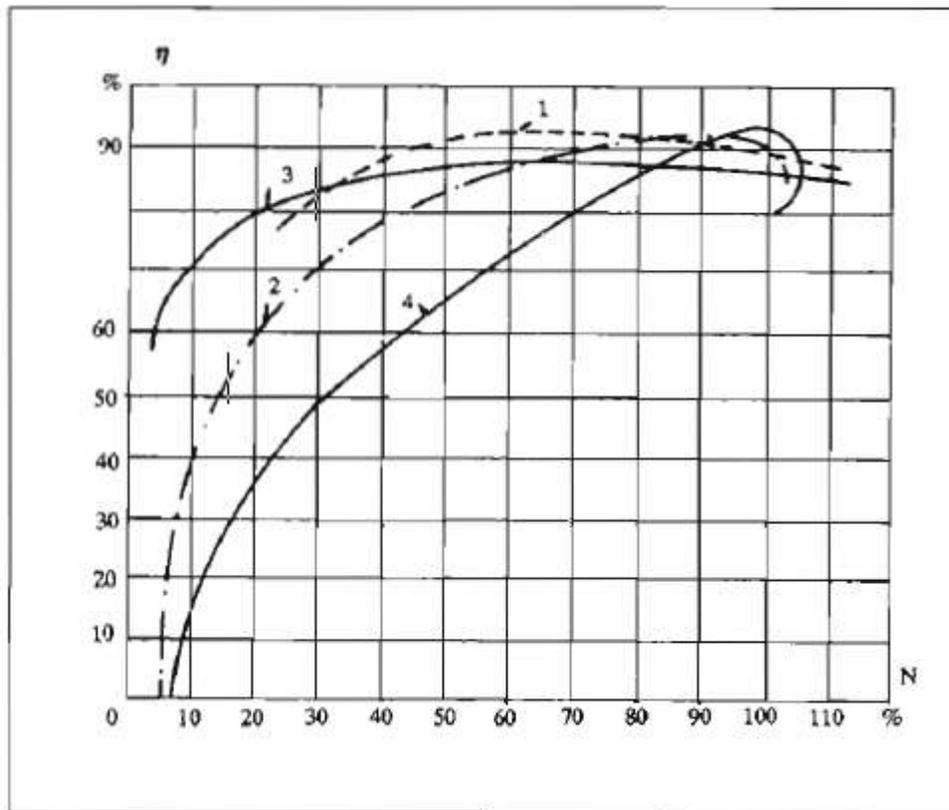


Figura 11-14. Curvas de eficiencia potencia 1. Kaplan, 2. Francis, 3. Pelton, 4. Hélice (Dolores-Juárez, 1992).

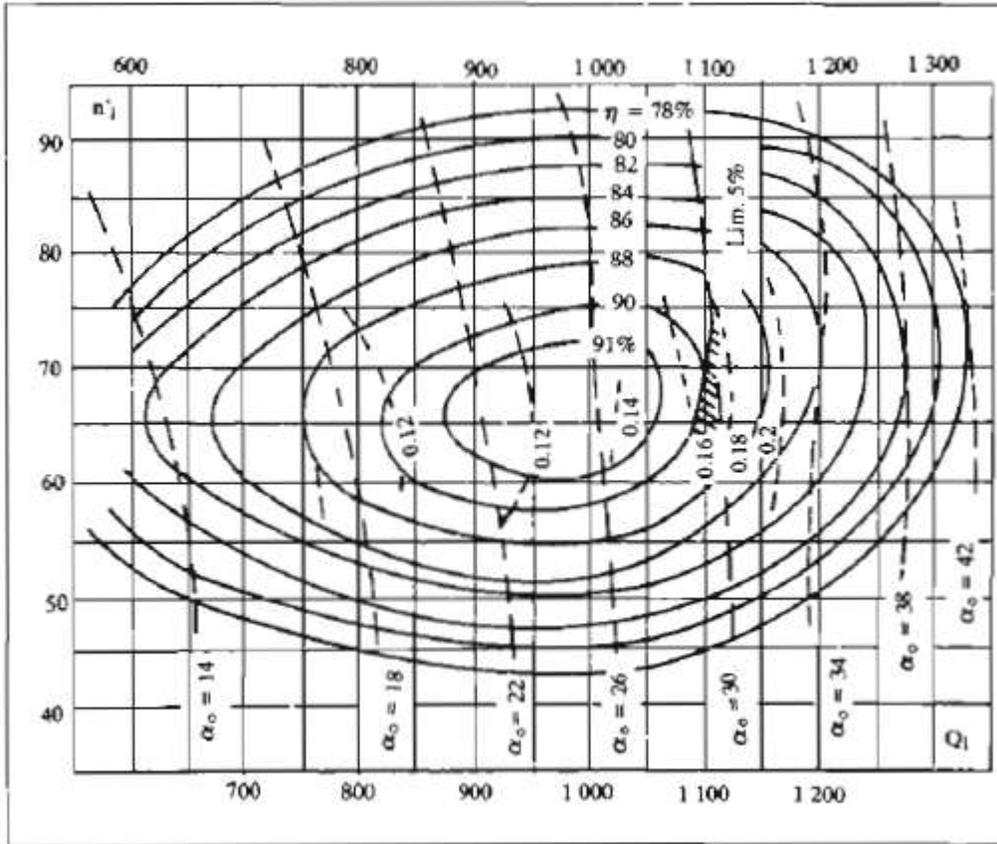


Figura 11-15. Característica universal de turbina Kaplan (Dolores-Juárez, 1992).

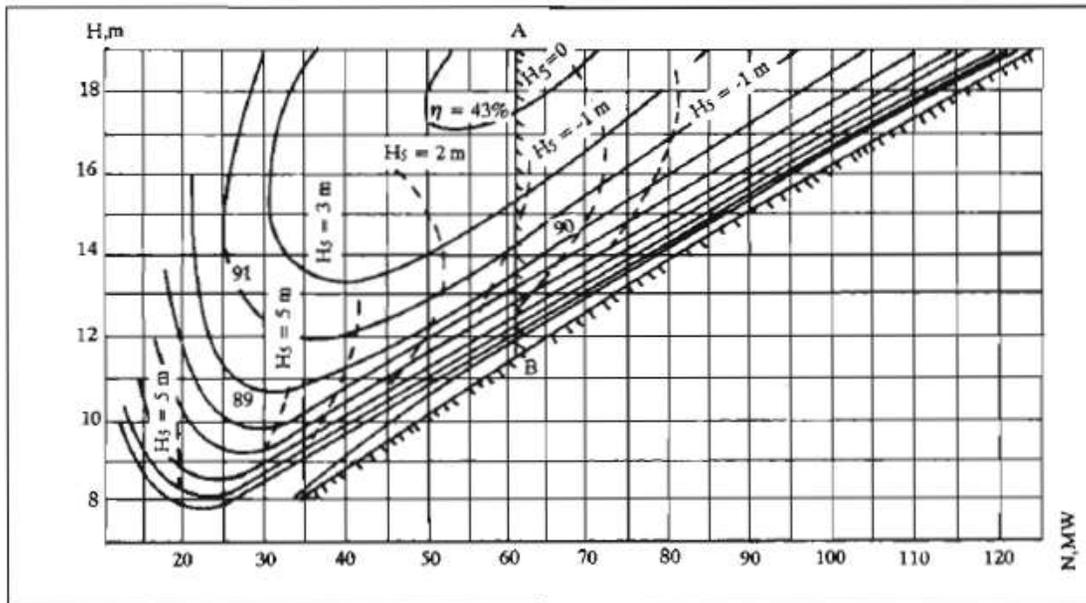


Figura 11-16. Características de operación de una turbina Kaplan de diámetro de 880 cm y velocidad de 62.5 r.p.m. (Dolores-Juárez, 1992).

11.14. Comentarios

En este documento se muestra la metodología básica para la selección de turbina en una central micro-eléctrica, además de las recomendaciones para el análisis de datos de caudal, en caso que se requiera profundizar más sobre el tema, se le invita al lector consultar la diferente bibliografía.

12. METODOLOGÍA

12.1. Revisión de la literatura

Esta etapa consiste en la revisión de la literatura de los principales temas que involucran el desarrollo de un análisis regional. Estos temas involucran las variables descriptivas empleadas en el análisis regional, el análisis previo de calidad de datos, el análisis regional de frecuencias de crecientes, así como los principales métodos de delimitación y técnicas de regionalización aplicadas en el análisis regional de frecuencias de crecientes y el diseño de mini-centrales hidroeléctricas.

12.2. Selección de estaciones hidrométricas

12.2.1. Selección de estaciones de aforo

La selección de estaciones se realiza mediante la extracción de datos de hidrometría mensual y anual, así como de precipitación anual por medio del Banco Nacional de Datos, y del extractor rápido de información climatológica, ERIC, creados por la Comisión Nacional del Agua (IMTA, 1988; Ouarda *et al.*, 2008). Se consideran todas las estaciones hidrométricas que cumplen con un periodo de registro mínimo de 20 años (Campos-Aranda, 2014a) y que registran un caudal no regulado.

12.2.2. Selección de estaciones discordantes

Por medio de la metodología del Análisis Regional de Crecientes y Momentos L (ARFC-ML) se analiza la medida de discordancia del conjunto de estaciones hidrométricas que cumplen con el periodo de registro y con caudal no regulado.

La fundamentación teórica se aplica de acuerdo al capítulo 8 ANÁLISIS REGIONAL DE FRECUENCIAS ARF, 8.1 Introducción a Momentos-L y 8.2 Etapas del Análisis Regional de Frecuencias basado en Momentos-L.

12.2.3. Calidad de los registros

Se aplica la prueba de tendencia de Mann-Kendall para el conjunto de estaciones hidrométricas. La prueba de Mann-Kendall es una prueba paramétrica que analiza la tendencia de una serie de datos, en este caso, de una serie de registros históricos, mediante el parámetro del valor-p, es decir, del nivel de significancia del 5% nominal, donde la hipótesis nula (H_0) establece que no se presenta tendencia lineal en la serie de datos, por medio de un

valor-p menor a 5%, si el valor $p < 5\%$ se rechaza la hipótesis nula y se infiere que la serie presenta tendencia lineal aceptando la hipótesis alternativa.

12.3. Delimitación de cuencas y selección de características de la cuenca

12.3.1. Delimitación de cuencas

En la caracterización fisiográfica de la zona de estudio se extraen datos del Instituto Nacional de Estadística y Geografía, INEGI, como el modelo digital de elevaciones que se obtiene del Continuo de Elevaciones Mexicano (CEM) con una resolución de 30 metros por píxel, así como los archivos en formato *shape* de uso y cobertura de suelo y cuerpos de agua.

Los puntos de salida de cada cuenca se generan a partir de la ubicación de las estaciones de aforo hidrométricas, por medio de una revisión de la ubicación geográfica del dato en formato *shape* con la ubicación geográfica publicada en el *catálogo de* estaciones hidrométricas, obtenidos de la página web de la CONAGUA: <http://www.conagua.gob.mx/CONAGUA07/Contenido/Documentos/Portada%20BANDAS.htm>

12.3.2. Selección de variables descriptivas más representativas de la cuenca

Por cada cuenca generada se obtienen las características descriptivas más importantes, la selección de éstas se realiza en función de la revisión de la literatura de las variables que intervienen en estudios similares de análisis de caudales máximos (descritas en el ANEXO I, “Características de la cuenca”).

En la obtención de los datos de precipitación por cuenca, se realiza la estimación por medio de la generación de polígonos de Thiessen.

Ahn y Palmer (2011) desarrollaron una metodología de selección de variables descriptivas más preponderantes para la aplicación de métodos de análisis de regresión múltiple, donde el criterio de selección en la primera etapa consiste en un análisis de la correlación simple de Pearson, donde toda variable que sea igual o supere el valor de 0.7 pasa a ser una variable representativa. La segunda etapa consiste en la aplicación de técnica de ACP y la tercera y última etapa consiste en una verificación donde nuevamente se aplica correlación simple de Pearson.

Para este estudio se tomó como base la metodología de Ahn y Palmer (2011) en la primera etapa con la consideración adicional formulada para este trabajo de tesis basada en la revisión del valor de significancia, valor p , al 5% nominal, con el cual se busca obtener una optimización en la selección de las variables. Posteriormente aquellas variables significativas se ingresan para un ACP, donde todas aquellas variables definidas dentro del círculo de correlación al valor como límite inferior de 0.8 se verifican por medio de correlación simple de Pearson de manera que toda variable que cumpla con el valor p mayor de 5%, es significativa.

12.4. Delimitación de regiones homogéneas.

La delimitación de regiones se realiza por medio de dos métodos, el primer método que se aplica es del de Ward y se optimiza mediante el método de K-medias en una segunda etapa (como se muestra en el capítulo 9, en la sección 9.3.1 Método de agrupamiento por Ward y 9.3.2 K-Medias). Posteriormente, se verifican la homogeneidad y la discordancia de las regiones generadas.

Los insumos para la agrupación de las estaciones homogéneas se realizan bajo un enfoque de similitud estacional y físico, debido a que se pretende agrupar estaciones aforadas que compartan la misma respuesta hidrológica para la implementación de minicentrales, por lo que se trabaja con las características más significativas de la zona de estudio y el gasto medio mensual histórico.

12.5. Análisis Regional de Frecuencias de Crecientes (ARFC)

12.5.1. Selección de una distribución de frecuencias regional

Mediante la metodología del ARFC y Momentos-L, se define la distribución regional aplicable en cada región homogénea, lo anterior mediante el diagrama de Momentos - L (Hosking y Wallis, 1997).

El siguiente paso consiste en realizar el análisis de frecuencias de todas las estaciones hidrométricas, así como el resultado del evento de diseño de la extrapolación de datos asociada a periodos de retorno de 5 a 500 años con la distribución regional obtenida.

12.5.2. Generación de ecuaciones de regresión

En cada región homogénea se seleccionan como variables independientes de predicción aquellas obtenidas en el paso “Selección de variables descriptivas más representativas de la cuenca” (sección 12.3.2), es decir, aquellas determinadas como variables de predicción bien representadas en el análisis multivariado y además con un valor de correlación de Pearson mayor o igual a 0.7. La variable dependiente de cada región es el caudal medio anual observado. La ecuación de regresión se ajusta por mínimos cuadrados, los valores del error relativo serán las medidas para estimar la precisión de cada ecuación generada con los valores observados.

12.5.3. Derivación de curvas de crecimiento regional

En cada región homogénea se genera una curva regional promedio, ésta se deriva de las curvas adimensionales de crecimiento de todas las estaciones, es decir, el caudal asociado a un periodo de retorno por medio de la distribución regional seleccionada (sección 12.5.1).

Por lo tanto en un sitio no aforado, el caudal asociado a cierto periodo de retorno se determina por medio de la siguiente ecuación:

$$X_T = \mu_i Z_T \quad (12.1)$$

donde μ_i es el caudal medio del sitio no aforado y Z_T es el cuantil asociado a la curva de crecimiento regional adimensional.

12.6. Generación de curvas de duración de caudal (CDC)

En la estimación del potencial hidroeléctrico de un sitio se emplea el cálculo de las Curvas de Duración de Caudal (CDC), estas curvas permiten visualizar frecuencia de ocurrencia de un gasto además “representan una relación entre el caudal y el porcentaje de tiempo en que un valor dado del caudal se iguala o se excede durante un período histórico” (Isik *et al.*, 2008), por lo que el un valor de un caudal Q está asociado a cierta probabilidad de excedencia p :

$$p_i = 1 - P\{Q_i \leq q\} \text{ o } p_i = 1 - P\{Q_i > q\} \quad (12.2)$$

donde $p_i = i / (i + 1)$ es la probabilidad de excedencia, Q_i es el valor del caudal equivalente de las observaciones q_i .

Por cada región homogénea se deduce una curva regional de duración de caudal específico resultante de un ajuste de todas las curvas adimensionales que comprenden la región homogénea, esto mediante una ecuación polinómica. Lo anterior permite generar la CDC en un sitio no aforado mediante la siguiente ecuación (Isik *et al.*, 2008):

$$Q_{up} = q_p A_u \quad (12.3)$$

donde Q_{up} es el caudal asociado a una probabilidad de excedencia o igualdad a cierto valor de percentil p , q_p es el caudal específico en la región homogénea asociado a un valor de percentil p o probabilidad de excedencia, y A_u es el área de la cuenca no aforada en km^2 .

12.7. Aplicación de la metodología propuesta a un sitio de interés

Mediante la aplicación de metodología expuesta se selecciona un sitio probable para la construcción de una minicentral hidroeléctrica.

Cálculo de la capacidad a instalar:

$$P = \gamma * g * h_{disp} * Q * \eta_s \quad (12.4)$$

donde P es la potencia media en kW, Q es el caudal medio aprovechado en m^3/s , γ es el peso específico del agua, h_{disp} es la carga disponible en m, g es la aceleración de la gravedad de 9.81 m/s^2 , η_s es la eficiencia total de la planta, del 75 al 90%.

La Generación Media Anual (GMA) se calcula mediante la siguiente ecuación:

$$GMA = \bar{P} * 8640 * FP / 1000 \quad (12.5)$$

donde GMA se expresa en GWh/año, y el factor de planta en porcentaje.

13. ZONA DE ESTUDIO

13.1. Hidrología

La República Mexicana está dividida en 13 Regiones Hidrológico-Administrativas y en 37 regiones hidrológicas. La zona de estudio seleccionada corresponde a la Región Hidrológica número 10 (RH10) y pertenece a la Región Hidrológico-Administrativa III (RHA III), denominada Sinaloa y Pacífico Norte, respectivamente.

La RH10 (Figura 13-1) se encuentra localizada al noroeste de México, tiene una superficie de 103055.56 km², pertenecen a esta zona la totalidad del estado de Sinaloa y parte de los estados de Durango, Chihuahua y Sonora.

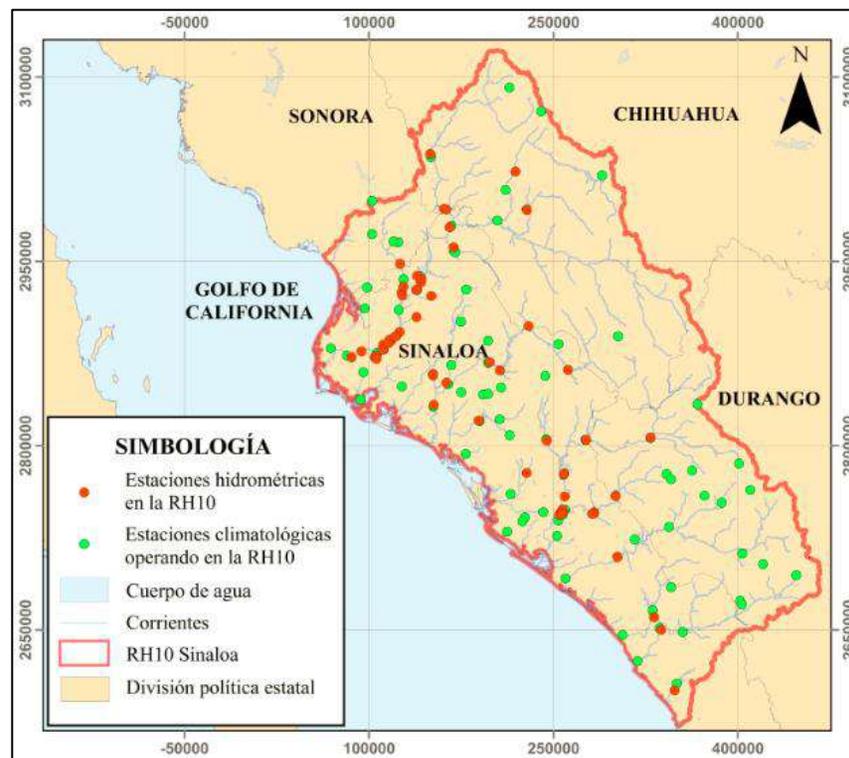


Figura 13-1 Mapa de ubicación de la Región Hidrológica 10, Sinaloa, México.

Las principales corrientes superficiales de la RH10 son el río Fuerte, Sinaloa, Mocorito, Culiacán, San Lorenzo y Piaxtla. Las principales presas son Luis Donaldo Colosio (Huites), Miguel Hidalgo y Costilla (El Mahone), Josefa Ortíz de Domínguez (El Sabino), Guillermo Blake Aguilar (El Sabinal), Gustavo Díaz Ordaz (Bacurato), Adolfo López Mateos (El

Humaya), Sanalona (Sanalona), Ing. Aurelio Benassini Viscaíno (El Salto) y José López Portillo (El Comedero).

Los consejos de cuenca existentes en la RH10 son el Consejo de cuenca de los Ríos Fuerte y Sinaloa y el Consejo de Cuenca de los ríos Mocorito al Quelite. El déficit en acuíferos en hectómetros cúbicos anuales (hm^3) debido a la carencia de escurrimientos superficiales se presenta en mayor cantidad en el río Culiacán, seguido del río Mocorito con 24.6 y 0.7 hm^3 , respectivamente y la sobre-explotación con un volumen considerable con 139.6 hm^3 en río Culiacán y 72.9 hm^3 en el río Mocorito. Además, las demandas agrícolas de la región se cubren con los escurrimientos de las cuenas de los ríos Fuerte, Culiacán, Sinaloa y San Lorenzo (CONAGUA, 2012).

Existe un monitoreo de variables hidrométricas por medio de 67 estaciones administradas por el Instituto Mexicano de Tecnología del Agua (IMTA), cuya base de datos es proporcionada por el Banco Nacional de Datos de Aguas Superficiales (BANDAS) por medio de la CONAGUA, además de un monitoreo climatológico por medio del Servicio Meteorológico Nacional (SMN) a través de la CONAGUA, con 87 estaciones activas actualmente en la RH10 (Figura 13-1).

13.2. Clima

Los climas preponderantes en la región hidrológica son templado sub-humedo y cálido subhúmedo (Figura 13-2). La temperatura media anual en la región presenta fuertes variaciones, entre los 10 y los 26°C. En la zona norte de la región se presentan temperaturas más cálidas y secas, aunque en las partes serranas pueden presentarse temperaturas bajo cero en ciertas épocas del año (CONAGUA, 2012).

13.3. Precipitación

En la Región Pacífico Norte la precipitación media anual (PMA) es de 747 mm, con un 70% de estacionalidad entre los meses de julio a septiembre (CONAGUA, 2012). Mientras que en la RH10 Sinaloa la PMA los valores de la precipitación van de los 125 a los 2000 mm (Figura 13-3).

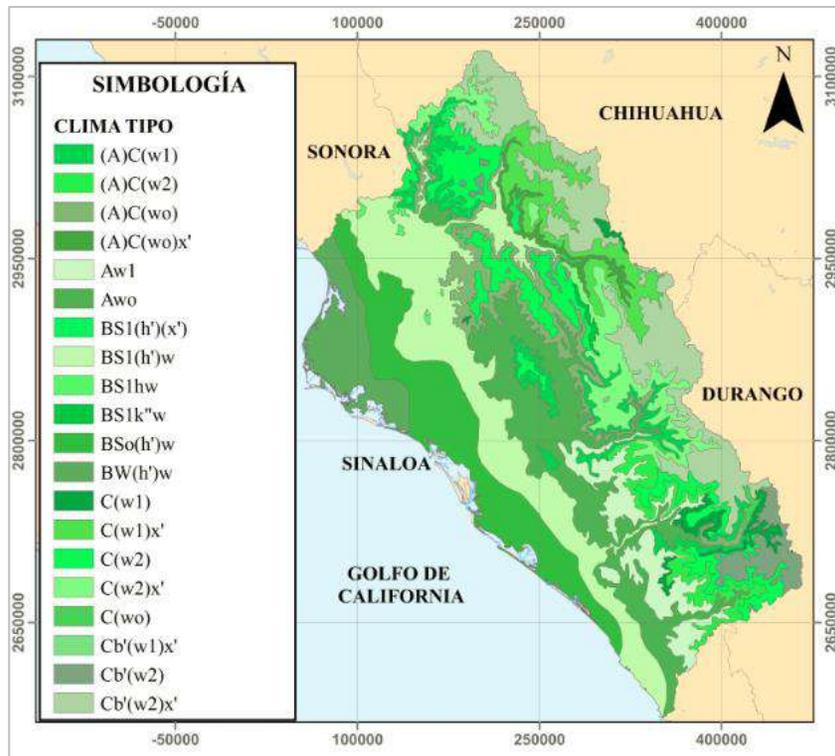


Figura 13-2 Mapa que muestra el clima en la Región Hidrológica 10, Sinaloa, México.

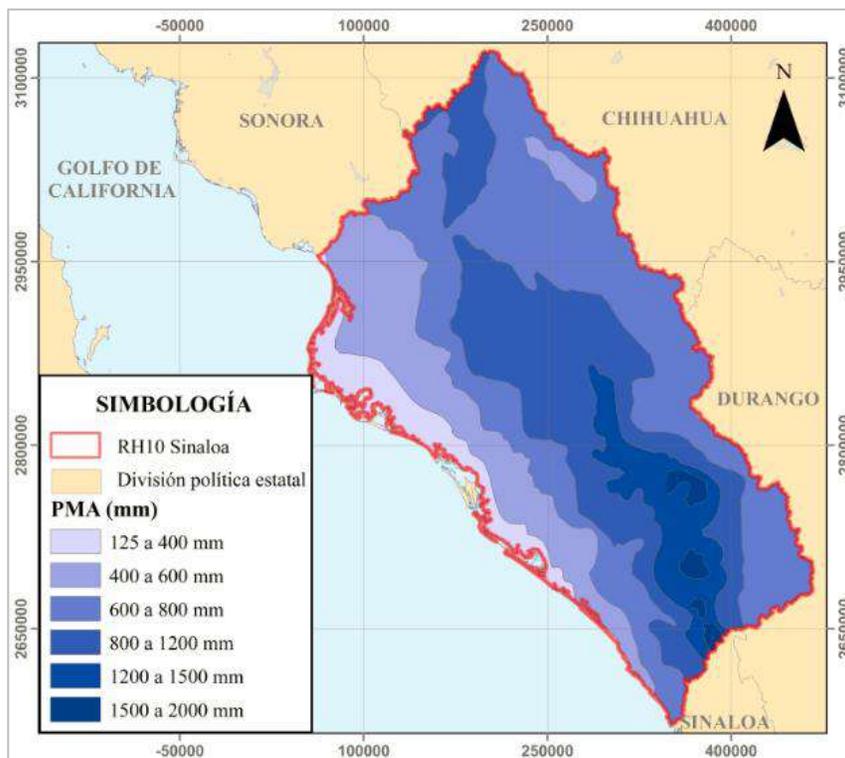


Figura 13-3 Precipitación Media Anual (PMA) Región Hidrológica 10, Sinaloa, México

13.4. Centrales hidroeléctricas existentes en la RH10

De acuerdo con la información publicada por la CONAGUA (2012) en la RHA III se tienen 164 presas registradas en el Sistema Informático de Seguridad de Presas (SISP), la capacidad total es de 23600 hm³, 18 embalses concentran el 99% de la capacidad de almacenamiento total. Mientras que en la RH10 se cuenta con cuatro embalses que cuentan con un almacenamiento mayor a 1000 hm³ (Tabla 13-1).

Tabla 13-1 Presas con almacenamiento mayor a 1000 hm³ (CONAGUA, 2012)

Nombre oficial	Nombre común	Río	Capacidad (hm ³)
Luis Adolfo López Mateos	El Humaya	Culiacán	2,952
Luís Donaldo Colosio	Huites	Fuerte	2,408
Miguel Hidalgo	El Mahone	Fuerte	2,571
José López Portillo	El Comedero	San Lorenzo	2,100
Gustavo Díaz Ordaz	Bacurato	Sinaloa	1,670

13.5. Estudios recientes

El Centro Mario Molina (CMM) realizó en el año 2011 un estudio del potencial de generación minihidroeléctrica en México como una aproximación operativa, como parte de las un proyecto denominado: Estrategias regionales y sectoriales para lograr un desarrollo sustentable y de baja intensidad de carbono en México. Actualmente existe una plataforma web de consulta de estos sitios, generada por el Inventario Nacional de Energías Renovables (INERE) <https://dgel.energia.gob.mx/inere/>.

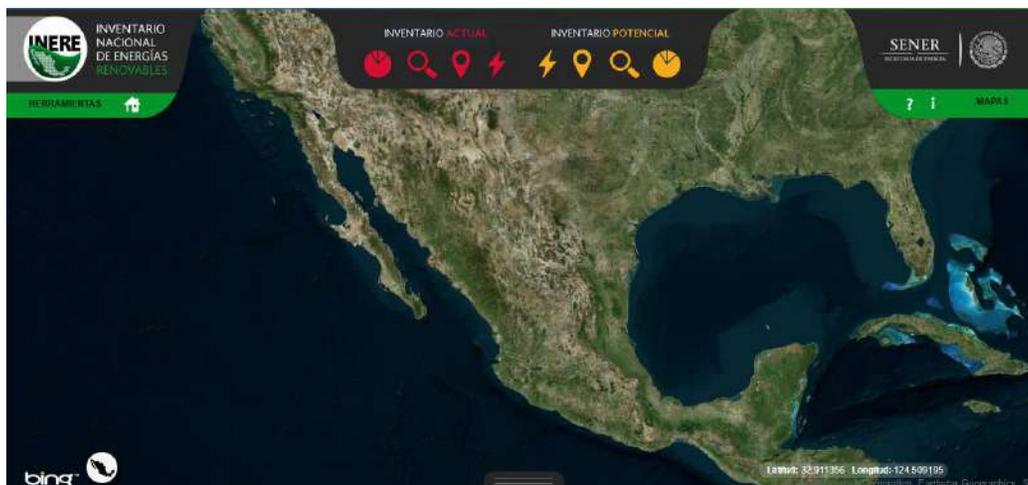


Figura 13-4 Página de consulta INERE

El estudio consistió en la estimación del potencial de generación en 26 cuencas para la implementación de centrales de paso, incluidos los principales ríos del Pacífico, se identificaron un total de 489 sitios, la caracterización de estos sitios estuvo basada en una etapa previa desarrollada por la Comisión Federal de Electricidad (CFE), basada en criterios técnicos y de sustentabilidad.

En criterios de sustentabilidad se buscó asegurar la permanencia del gasto ecológico en épocas de estiaje, por lo que el aprovechamiento del caudal fue del 60% del caudal medio.

En cuanto al aspecto técnico, se buscó por medio del análisis del cauce aquellos sitios con la mejor caída posible, incluyeron también un análisis de pequeños túneles con longitud máxima de 4 km que permitieran la ganancia de desnivel en sitios aprovechables. La carga mínima aprovechable fue de 20 metros, con obras de contención de altura máxima de 20 metros, en algunos sitios hubo excepciones debidas a la topografía, por lo que la obra de contención y la carga aprovechable se estimó con 10 metros.

El factor de planta fue del 25%, dado que se diseñó para centrales de paso, por lo que con el objetivo de mantener el escurrimiento en todo el tramo del río y en épocas de estiaje. Además, se consideró como equipamiento 2 unidades de generación para la operación del 50% del caudal aprovechable.

Un estudio actualizado en la RHA III Pacífico Norte, es el Programa Hídrico Regional Visión 2030 que fue realizado por la CONAGUA (2012). En este estudio se muestra la situación actual de la Región en el aspecto ambiental e hídrico, así como la problemática de la insuficiencia de infraestructura urbana, la sobre explotación de acuíferos y la etapa de sequías en la zona. Uno de los estudios más recientes aplicados en la RH10 es el realizado por Campos-Aranda (2014 a, b) donde agrupa estaciones homogéneas en función de métodos por regiones de influencia. En cuanto a la precipitación, estudios previos en la región (Arellano-Lara y Soto-Escalante, 2014; Reséndiz-Mendoza y Gutiérrez-López, 2017) permiten la caracterización de la zona y como parte del proceso lluvia escurrimiento la agrupación de estaciones climatológicas por métodos multivariados.

14. RESULTADOS Y DISCUSIÓN

14.1. Revisión de la literatura

El resultado de esta etapa se muestra en los capítulos del 6 al 11.

14.2. Selección de estaciones hidrométricas aplicables en el Análisis Regional

En total se seleccionaron 26 estaciones hidrométricas que cuentan con un periodo de registro histórico no aforado en las corrientes de la RH10. Se aplicaron tres filtros por medio de las medidas de discordancia de la metodología de Momentos-L, debido a la cantidad de datos (estaciones), el valor de discordancia D_i menor a 3, por lo cual se eliminaron del análisis a las estaciones Jesús Cruz y Urique, quedando un total de 24 estaciones (Tabla 14-1).

La prueba de Mann-Kendall se aplicó a las 24 estaciones hidrométricas resultantes del análisis de discordancia. Se presentó tendencia en las estaciones hidrométricas Guamúchil, Jaina, Santa Cruz, Acatitlán y Guasave, pero para efectos de predicción se tomaron como menos precisos, a diferencia de aquellos en lo que se aceptó la hipótesis nula. Este resultado es similar al de estudios previos realizados en la misma región hidrológica (Campos-Aranda, 2014b).

Tabla 14-1 Resultados de la prueba de Mann-Kendall para 24 estaciones.

n	ID	ESTACIÓN	N	Di			valor-p
				Filtro 1	Filtro 2	Filtro 3	
1	10027	El Bledal	57	0.0247	0.0593	0.1317	0.874
2	10029	Naranjo	47	0.5648	0.7449	0.8713	0.856
3	10031	Guamúchil	33	0.1631	0.1567	0.1572	< 0.0001
4	10034	Zopilote	64	0.9105	0.8946	0.8522	0.174
5	10036	Jaina	63	0.2090	0.1884	0.1709	< 0.0001
6	10037	Huites	51	1.0037	0.9997	1.0658	0.17
7	10040	Santa Cruz	59	0.5891	0.5589	0.5361	< 0.0001
8	10057	Bamícori	33	0.9044	1.0145	1.2162	0.278
9	10064	Chinipas	33	0.6528	0.7912	0.8935	0.2
10	10065	Ixpalino	49	0.5475	0.7066	0.8933	0.7
11	10066	Choix	41	0.2600	0.3260	0.4075	0.259
12	10070	Acatitlán	50	1.0806	1.1943	1.1919	0.002
13	10077	Palo Dulce	26	1.2867	1.2427	1.1890	0.533
14	10078	La Tina	24	2.3162	2.2317	2.1648	0.677
15	10079	Badiraguato	42	1.6158	1.8210	1.8837	0.223

16	10083	El Quelite	36	0.2367	0.2392	0.2359	0.693
17	10086	Pericos	31	0.3096	0.5635	0.9713	0.08
18	10087	Tamazula	36	2.0169	1.9703	1.9536	0.573
19	10090	Chico Ruíz	26	0.9190	0.8740	0.8578	0.908
-	10100	Urique II	29	1.6020	3.0660	-	-
20	10110	Tohayanna	27	1.5117	1.4509	1.3809	0.255
21	10112	Guatenipa II	40	0.7407	0.7168	0.7044	0.012
22	10113	La Huerta	28	1.6617	1.9777	2.5983	0.114
-	10119	Jesús Cruz	15	3.8922	-	-	-
23	10120	Guasave Puente Carretera	22	0.2791	0.5301	0.9729	< 0.0001
24	10137	Tecusiapa	48	0.7016	0.6813	0.7000	0.242

n: Número de estación; ID: Clave de la estación; N: longitud de periodo de registro; Di: Medida de discordancia en cada estación.

14.3. Delimitación de cuencas y selección de características de la cuenca

En la delimitación de cuencas se consideró el punto de salida de las mismas como la ubicación de la estación de aforo, por lo que las variables descriptivas se obtuvieron por medio de la información obtenida de INEGI y con el modelo digital de elevaciones de tamaño de pixel de 30 x 30 (Figura 14-2).

Mediante la revisión de la literatura se emplearon las características de la cuenca aplicables en la descripción de morfometría, topografía, uso de suelo, red de drenaje, pendiente y climatología, siendo un total de 32 variables (Tabla 14-2), mientras que para el dato de precipitación correspondiente a climatología se desarrolló mediante el análisis de estaciones climatológicas y se procesó el dato mediante polígonos de Thiessen empleando las 77 estaciones climatológicas que cumplen con la calidad del registro de datos (Figura 14-2).

Tabla 14-2 Características 24 estaciones.

N	Tipo de variable y descripción	Unidades	Nomenclatura	Valor mínimo	Valor máximo	Valor promedio
	<u>Morfometría</u>					
1	Perímetro	km	P	109.45	1711.53	419.6
2	Área	km ²	A	209.64	17826.63	2900.83
3	Índice de compacidad	%	IC	0.010	0.023	0.015
4	Parámetro de forma	km ⁻¹	PF	0.066	0.576	0.239
	<u>Topografía</u>					
5	Elevación punto de salida	msnm	EPS	22	620	205.68
6	Longitud	m	X	138459	347107	228701.44
7	Latitud	m	Y	2601030	3037620	2834997.6

8	Elevación de la estación de aforo	msnm	EEA	9	620	205.16
9	Elevación máxima en la cuenca	msnm	T_EMX	683	3286	2224.12
10	Elevación mínima en la cuenca	msnm	T_Emn	9	565	171.36
11	Elevación media en la cuenca	msnm	T_Em	150.98	2274.26	971.39
12	Rango de altitud	m	RALT	634.00	3179.00	2049.00
13	Relieve relativo de la cuenca	-	RRM	0.881	0.661	0.338
<u>Uso de suelo</u>						
14	Zona agrícola	%	UZA	0.42	76.31	19.94
15	Bosque	%	UB	0.45	91	50.54
16	Pastizal	%	UP	0.11	5.76	1.97
17	Selva	%	US	0.5	91.63	37.17
18	Zona urbana	%	UZU	0	4.89	0.78
19	Cuerpo de agua	%	UCA	0	1.65	0.47
20	Sin vegetación	%	USV	0.01	0.01	0.01
21	Selva- área agrícola	%	USAA	2.44	20.54	11.49
22	Matorral	%	UM	9.11	9.11	9.11
<u>Red de drenaje</u>						
23	Máximo orden	-	ORD	90.89	100	99.64
24	Longitud	km	LCR	1	4	2.2
25	Longitud del cauce principal	km	LMR	25.32	1389.22	221.73
26	Pendiente del cauce principal	adim	PMR	31.08	400.76	126.56
27	Elevación en el punto más alto del cauce	msnm	EUP	0.01	0.03	0.02
28	Elevación en el punto más bajo del cauce	msnm	EDS	338	3156	1865.16
29	Densidad de corrientes	km/km2	DC	9	565	176.52
<u>Pendiente</u>						
30	Pendiente media	grados (°)	Pmed	0.05	0.19	0.09
31	Pendiente máxima	grados (°)	pmax	4.45	26.43	17.16
<u>Climatología</u>						
32	Precipitación Media Anual	mm	PMA	460.4	2094.5	954.88

matorral, máximo orden de corrientes, longitud de corrientes, longitud del cauce principal y elevación en el punto más alto de la corriente.

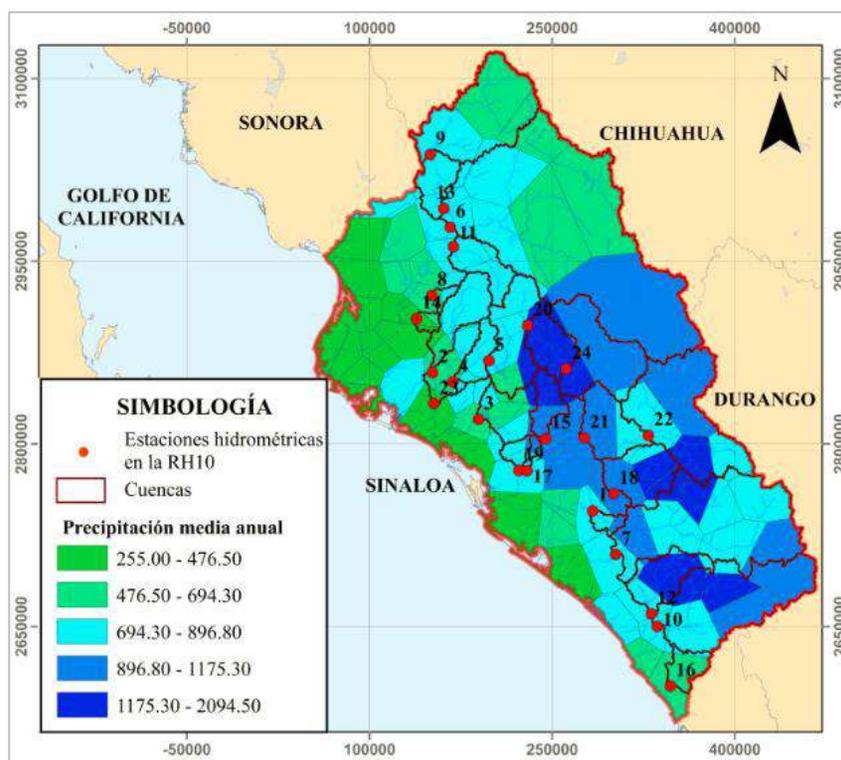


Figura 14-2 Polígonos de Thiessen para el cálculo de la precipitación media anual.

Tabla 14-3 Primer filtro de variables.

N	Variable	Descripción	R ²	valor-p
1	P	Perímetro de la cuenca	0.698	0
2	A	Área de la cuenca	0.737	0
3	IC	Índice de compacidad	-0.277	0.19
4	PF	Parámetro de forma	-0.585	0.003
5	EPS	Elevación del punto de salida	0.179	0.402
6	X	Longitud del punto de salida	0.076	0.724
7	Y	Latitud del punto de salida	0.092	0.67
8	EEA	Elevación de la estación de aforo	0.18	0.401
9	T_EMX	Elevación máxima de la cuenca	0.6	0.002
10	T_Emn	Elevación mínima de la cuenca	0.008	0.97
11	T_Em	Elevación media de la cuenca	0.504	0.012
12	RALT	Rango de altitud	0.647	0.001
13	RRM	Relieve relativo de la cuenca	0.472	0.02
14	UZA	Porcentaje de zona agrícola	-0.408	0.048
15	UB	Porcentaje de bosque	0.571	0.004

16	UP	Porcentaje de pastizal	0.168	0.432
17	US	Porcentaje de selva	-0.416	0.043
18	UZU	Porcentaje de zona urbana	0.214	0.315
19	UCA	Porcentaje de cuerpo de agua	0.175	0.414
20	USV	Porcentaje de suelo sin vegetación	-0.067	0.755
21	USAA	Porcentaje de selva-área agrícola	0.246	0.246
22	UM	Porcentaje de matorral	0.683	0
23	ORD	Máximo orden de corrientes	0.638	0.001
24	LCR	Longitud de corrientes	0.743	0
25	LMR	Longitud del cauce principal	0.703	0
26	PMR	Pendiente del cauce principal	-0.31	0.141
27	EUP	Elevación en el punto más alto de la corriente	0.49	0.015
28	EDS	Elevación en el punto más bajo de la corriente	0.051	0.813
29	DC	Densidad de corrientes	-0.273	0.197
30	Pmed	Pendiente media de la cuenca	0.479	0.018
31	pmax	Pendiente máxima de la cuenca	0.52	0.009
32	PMA	Precipitación media de la cuenca	0.313	0.136

Se procedió a elaborar un segundo filtro con las 16 variables más representativas por medio de un ACP. Como resultado se obtuvo que mediante dos valores propios se pudo representar en un 84.40 % la varianza de los datos respecto a la matriz de datos originales. La selección de estos dos valores propios estuvo en función del criterio de Kaiser, donde el valor propio que contribuye a una mayor varianza de los datos son todos aquellos que tengan un valor igual o superior a 1 (Tabla 14-4).

Tabla 14-4 Valores propios para 16 variables.

λ_i	Valor propio	% Total de varianza	Valor propio acumulado	% Varianza acumulada
1	11.25369	70.33559	11.25369	70.3356
2	2.25072	14.06703	13.50442	84.4026
3	0.94069	5.87932	14.44511	90.2819
4	0.49814	3.11337	14.94325	93.3953
5	0.34876	2.17978	15.29201	95.5751
6	0.24238	1.51486	15.53439	97.0899
7	0.18633	1.16455	15.72072	98.2545
8	0.10789	0.67431	15.82861	98.9288
9	0.06142	0.38389	15.89003	99.3127
10	0.05059	0.31621	15.94063	99.6289
11	0.03845	0.24034	15.97908	99.8693
12	0.01542	0.09638	15.99450	99.9656
13	0.00396	0.02474	15.99846	99.9904

14	0.00072	0.00450	15.99918	99.9949
15	0.00045	0.00279	15.99963	99.9977
16	0.00037	0.00233	16.00000	100.0000

Mediante los dos valores propios, se graficó el círculo de correlaciones con los dos componentes principales (Figura 14-3).

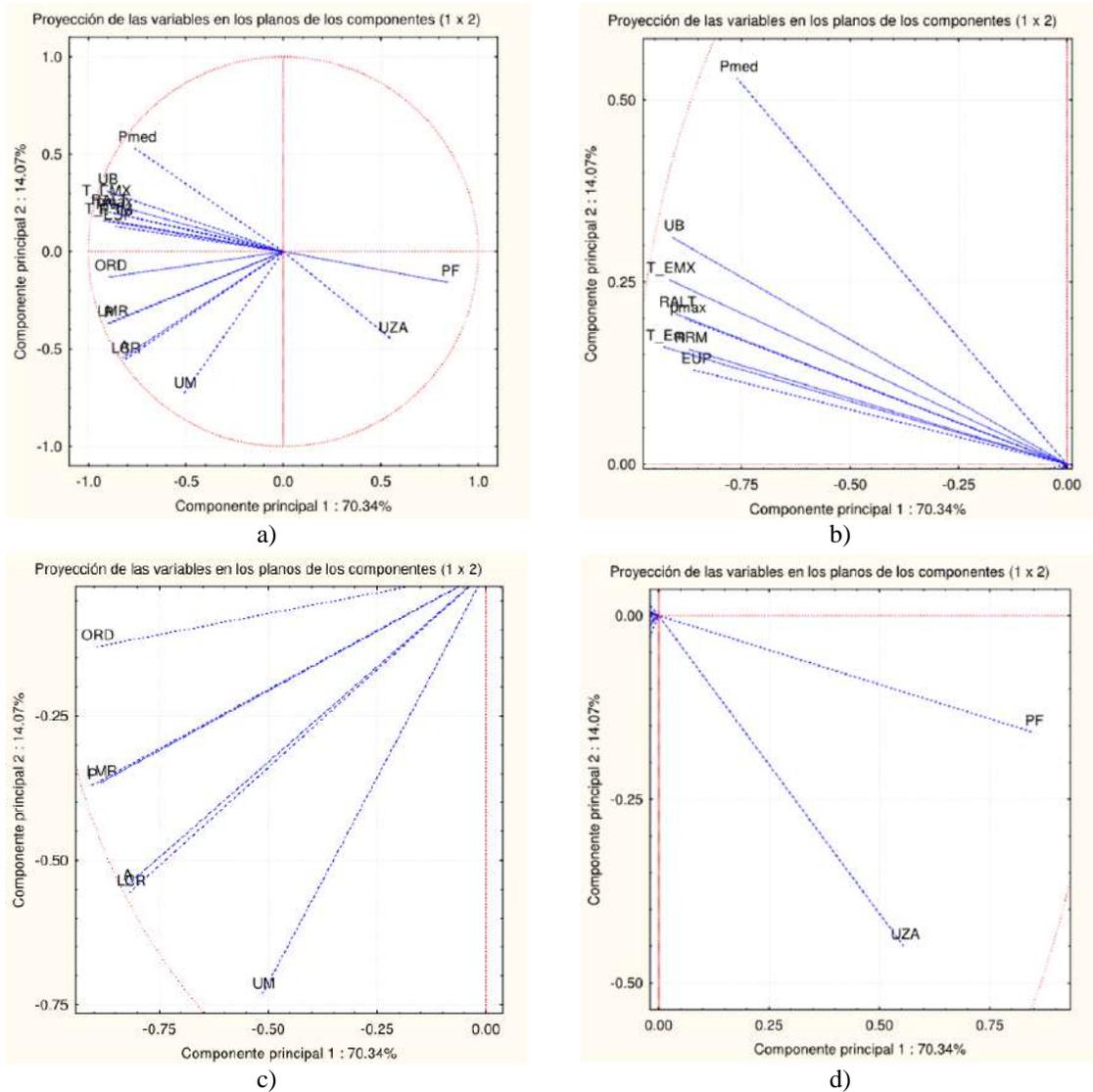


Figura 14-3 Círculo de correlación entre los componentes 1 y 2. a) cuatro cuadrantes, b) cuadrante II, c) cuadrante III, d) cuadrante IV.

Dado que la varianza se explicó mejor en los dos primeros valores propios, se obtuvieron las variables que están mejor representadas bajo esta proyección, es decir, todas aquellas variables con un coeficiente de correlación mayor a 0.8, siendo en total 13 variables, que se

verificaron mediante la correlación simple de Pearson, resultando que todas estas son variables significativas por medio del valor de significancia p (Tabla 14-5), mostradas en negritas en la tabla.

Tabla 14-5 Correlación de los dos primeros componentes principales

Variable		Factor 1	Factor 2	R ²	p-valor
Perímetro de la cuenca	P	-0.908027	-0.371238	0.698	0
Área de la cuenca	A	-0.827377	-0.545501	0.737	0
Parámetro de forma	PF	0.846815	-0.158442	-0.585	0.003
Elevación máxima de la cuenca	T_EMX	-0.917153	0.253392	0.600	0.002
Elevación media de la cuenca	T_Em	-0.929382	0.160815	0.504	0.012
Rango de altitud	RALT	-0.901923	0.206403	0.647	0.001
Relieve relativo de la cuenca	RRM	-0.871707	0.157862	0.472	0.020
Porcentaje de zona agrícola	UZA	0.554037	-0.449049	-	-
Porcentaje de bosque	UB	-0.910661	0.311540	0.571	0.004
Porcentaje de matorral	UM	-0.514420	-0.732981	-	-
Máximo orden de corrientes	ORD	-0.896579	-0.130341	0.638	0.001
Longitud de corrientes	LCR	-0.818639	-0.555481	0.743	0
Longitud del cauce principal	LMR	-0.886275	-0.366041	0.703	0
Elevación en el punto más alto de la corriente	EUP	-0.859171	0.128908	0.490	0.015
Pendiente media de la cuenca	Pmed	-0.761398	0.529920	-	-
Pendiente máxima de la cuenca	pmax	-0.878051	0.199274	0.520	0.009

La metodología tradicional, consiste en el tratamiento de análisis multivariado a todo el conjunto de variables, que en este caso fue de 32 variables, por lo que una varianza aceptable, igual o mayor al 80%, se obtiene con el valor propio acumulado de 6 valores propios, por lo que la gráfica bidimensional de la proyección sobre los ejes considerados con mayor varianza, es decir, del valor propio 1 y 2, consistiría en la representación del 58% de varianza (Figura 14-4). A diferencia del valor obtenido con la metodología propuesta, se obtuvo el valor de 84% (Figura 14-5), por lo que el método propuesto es viable.

Tabla 14-6 Valores propios para 32 variables.

λ_i	Valor propio	% Total de varianza	Valor propio acumulado	% Varianza acumulada
1	14.13090	44.15908	14.13090	44.1591
2	4.43831	13.86973	18.56922	58.0288
3	3.11343	9.72948	21.68265	67.7583
4	2.01110	6.28469	23.69375	74.0430

5	1.82535	5.70423	25.51911	79.7472
6	1.42301	4.44690	26.94212	84.1941
7	1.08452	3.38913	28.02664	87.5832
8	0.85602	2.67506	28.88266	90.2583
9	0.81427	2.54458	29.69692	92.8029
10	0.65936	2.06049	30.35628	94.8634
11	0.48053	1.50166	30.83681	96.3650
12	0.35574	1.11168	31.19255	97.4767
13	0.24793	0.77479	31.44048	98.2515
14	0.17255	0.53923	31.61303	98.7907
15	0.14658	0.45806	31.75961	99.2488
16	0.09041	0.28253	31.85002	99.5313
17	0.06343	0.19821	31.91345	99.7295
18	0.04308	0.13463	31.95653	99.8642
19	0.02928	0.09149	31.98581	99.9557
20	0.01063	0.03321	31.99644	99.9889
21	0.00222	0.00694	31.99866	99.9958
22	0.00118	0.00370	31.99984	99.9995
23	0.00016	0.00050	32.00000	100.0000

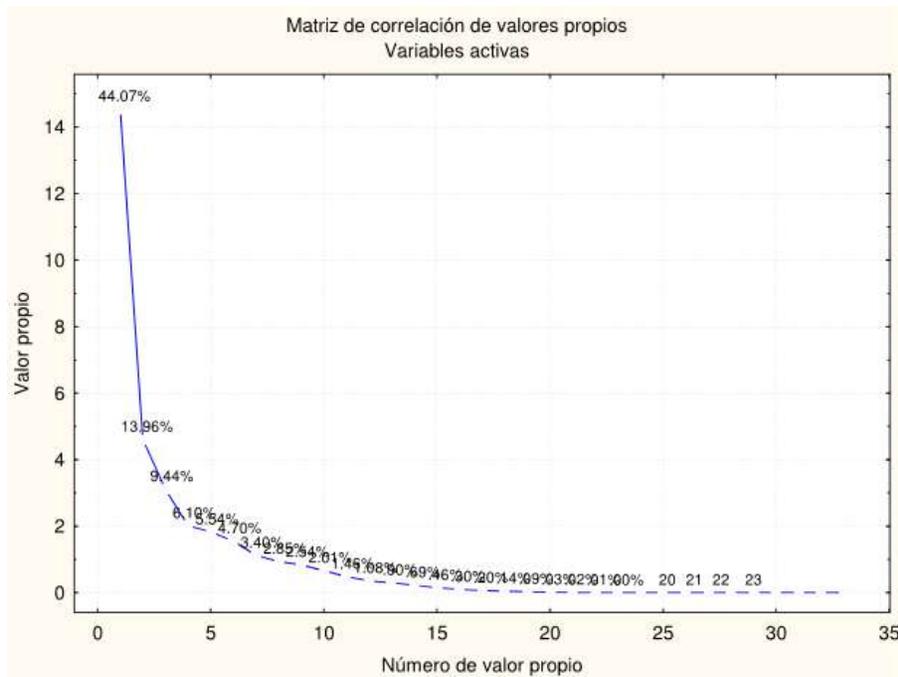


Figura 14-4 Valores propios de la matriz de correlación de 32 variables.

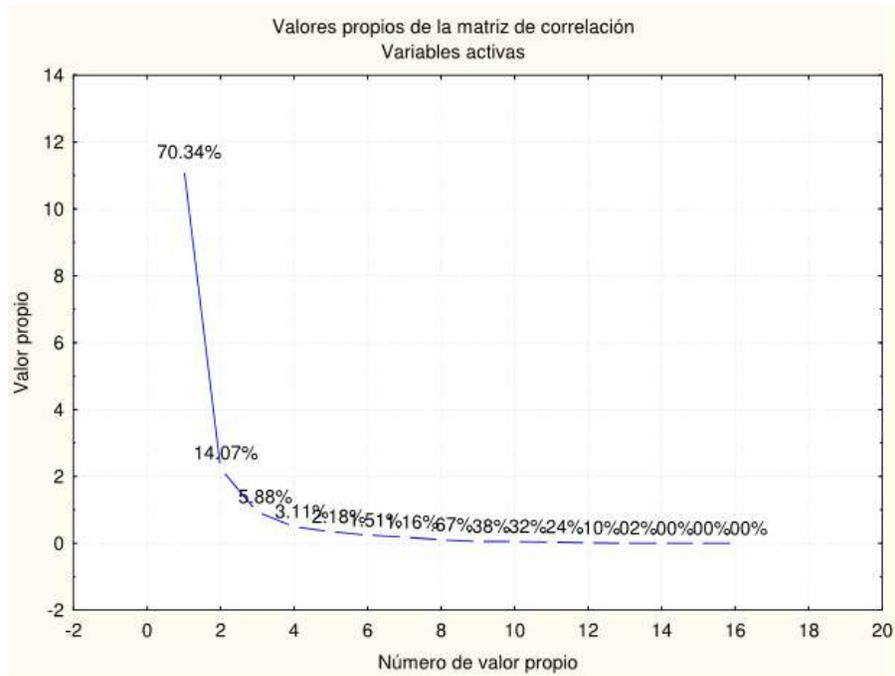


Figura 14-5 Valores propios de la matriz de correlación de 16 variables

14.4. Delimitación de regiones homogéneas

En la delimitación de regiones homogéneas un aspecto importante es la selección de los datos que serán empleados en la agrupación de elementos, dado que en este trabajo se tienen los objetivos de la formación de grupos homogéneos con el propósito de transferencia de información hidrológica así como del análisis del potencial hidroeléctrico para la implementación de mini centrales en sitios no aforados, la delimitación estuvo basada en dos tipos de datos, el caudal mensual medio y las características descriptivas más representativas para el conjunto de las 24 estaciones que cumplieron con la calidad de los datos.

En cumplimiento con lo anterior, la matriz de datos para la delimitación estuvo formada por 12 variables hidrométricas, que son los caudales mensuales medios específicos de cada estación y las 13 variables representativas para el conjunto de estaciones hidrométricas (sección 14.3), es decir, resultó una matriz de 24 renglones por 25 columnas.

La delimitación de regiones homogéneas se desarrolló en dos etapas, el primer método que se aplicó fue el método de delimitación jerárquica ascendente y el método de K – medias, éste último conocido por ser el método de optimización en la agrupación de elementos.

Mediante la metodología de Ward, se obtuvieron 4 grupos de acuerdo al criterio de la gráfica de distancia de ligadura a través de los pasos (Figura 14-6), por medio de esta gráfica fue posible la selección de la distancia euclidiana al cuadrado de agrupación de elementos, la cual no excedió de 70 km acumulada.

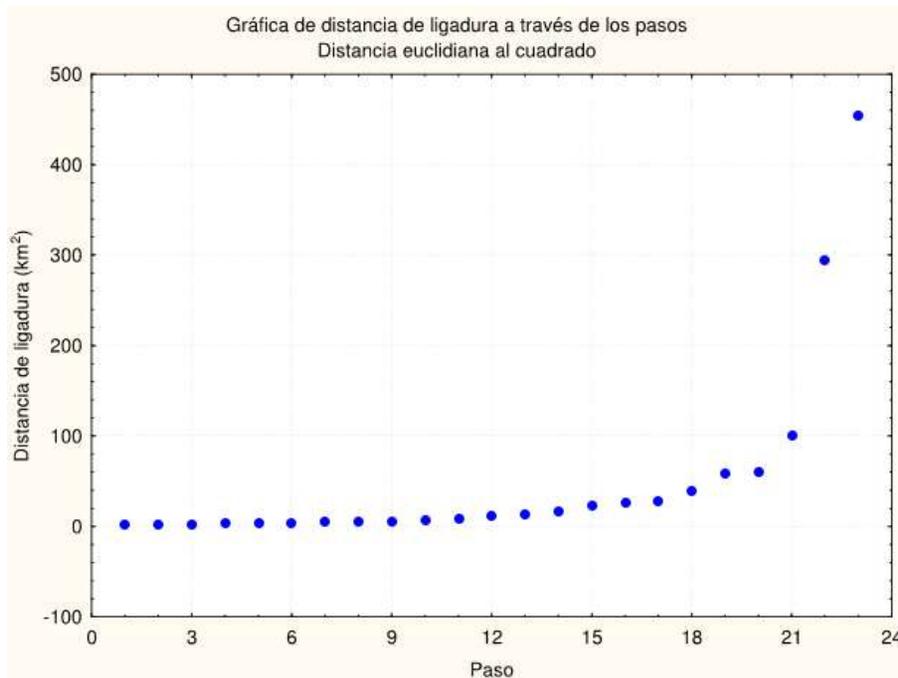


Figura 14-6 Gráfica de distancia de ligadura a través de los pasos de agrupamiento de elementos.

Debido a lo anterior, se obtuvo como primera aproximación la formación de cuatro grupos homogéneos con distancia de ligadura euclidiana al cuadrado menor a 70 km (Figura 14-7). Mediante la metodología de optimización de formación de grupos homogéneos, K-medias (Figura 14-8), se obtuvo la agrupación de elementos recomendable dentro de la RH10.

Se verificó la medida de discordancia entre cada agrupación generada, en la Tabla 14-7 se muestra el valor del elemento del grupo con mayor discordancia y el valor de discordancia crítico (Hosking y Wallis, 1997), así como el valor de verificación de homogeneidad (CC), las cuales no excedieron el valor crítico de 0.3.

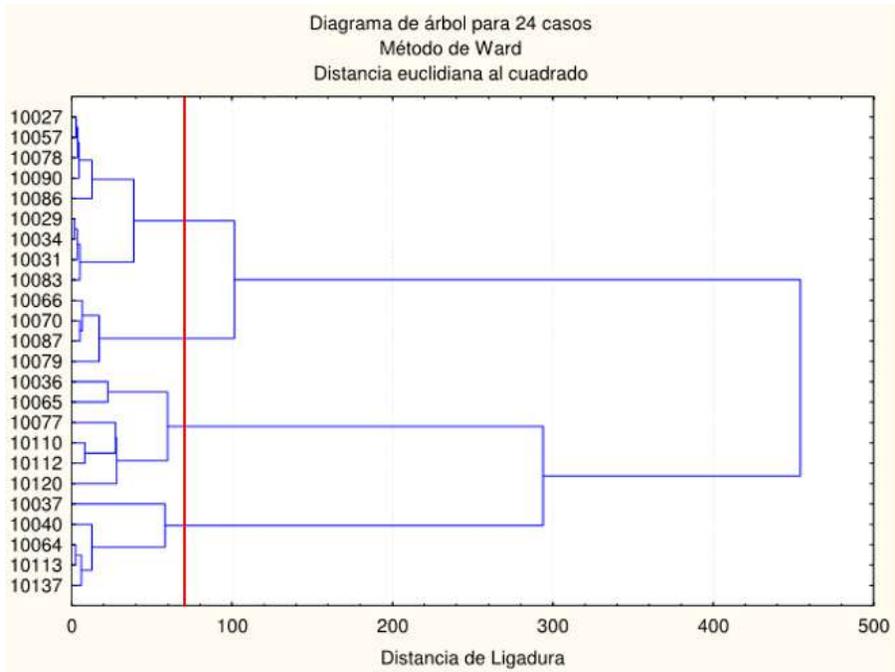


Figura 14-7 Dendrograma de Ward para 25 variables.

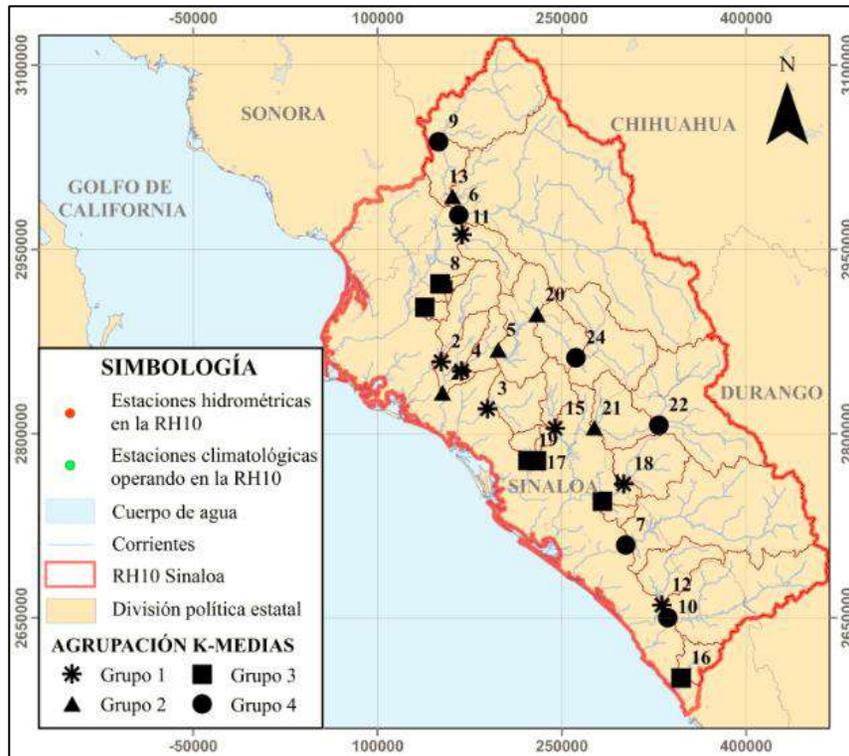


Figura 14-8 Agrupación por el método de K-medias

Tabla 14-7 Resultados obtenidos por región homogénea

RH	e	Mayor D_i en la RH	D_i crítico	CV_R	σ_R	CC	τ_3^R	τ_4^R	DR
1	7	1.399	1.917	1.1839	0.2803	0.2367	0.4279	0.2907	GVE
2	5	1.309	1.333	0.8973	0.2138	0.2383	0.3854	0.2492	LN3
3	6	1.611	1.648	0.9416	0.2557	0.2715	0.0563	0.0347	GLO
4	6	1.463	1.648	0.8593	0.2124	0.2471	0.3735	0.2455	LN3

RH: Región homogénea, e: Elementos por grupo, D_i : Medida de discordancia en la región i, DR: Distribución regional

14.5. Análisis Regional de Frecuencias de Crecientes (ARFC)

14.5.1. Selección de una distribución de frecuencia regional

Mediante el diagrama de Momentos-L (Hosking y Wallis, 1997), se obtuvo la distribución regional de frecuencias para cada agrupación homogénea, mediante los datos de los cocientes de asimetría-L y curtosis-L, es decir, τ_3^R y τ_4^R (Tabla 14-7).

En la RH1 la distribución regional aplicable es la General de Valores Extremos (GVE) (Figura 14-9), en la RH3 la distribución regional aplicable es la Logística Generalizada (GLO) (Figura 14-11), mientras que para las RH2 y RH4 la distribución Log Normal de tres parámetros es la distribución aplicable (LN3) (Figura 14-10 y Figura 14-12).

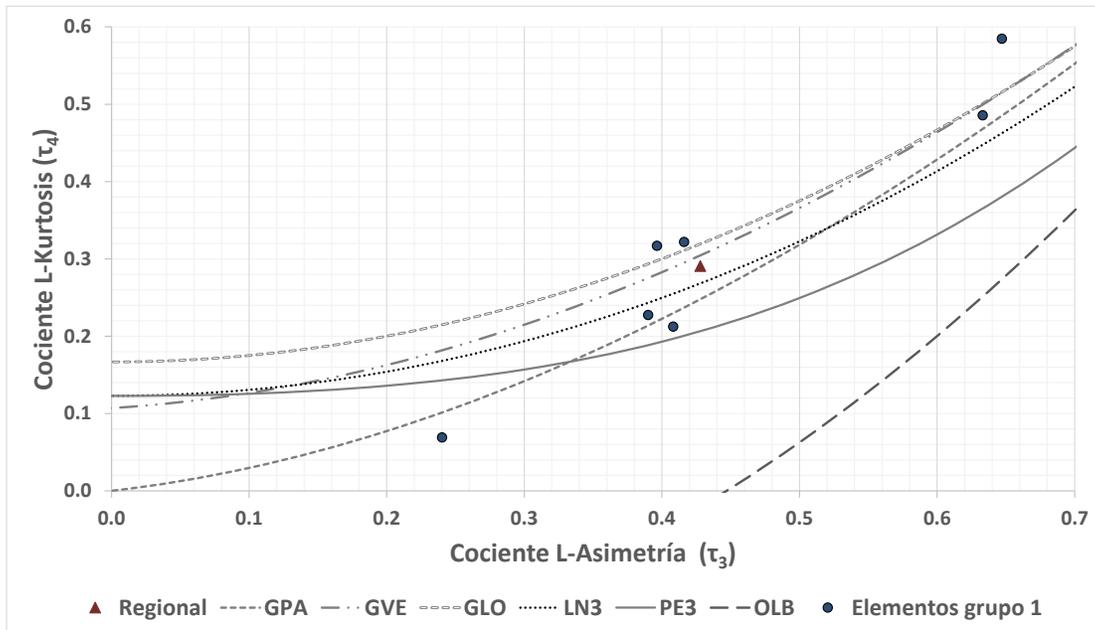


Figura 14-9 Diagrama de cocientes de Momentos-L para la RH1.

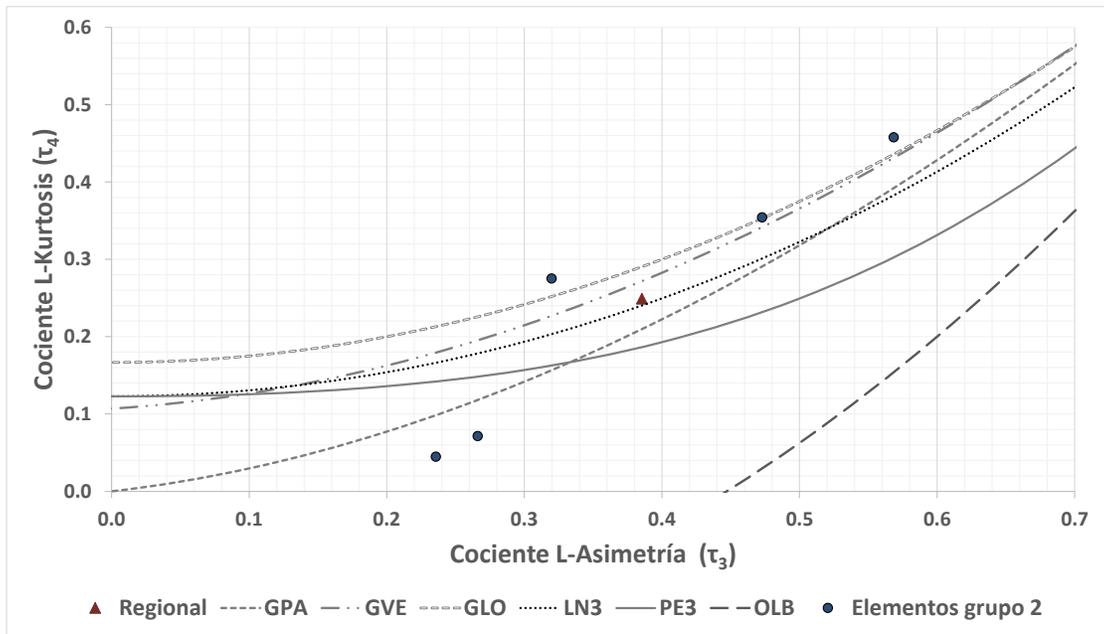


Figura 14-10 Diagrama de cocientes de Momentos-L para la RH2.

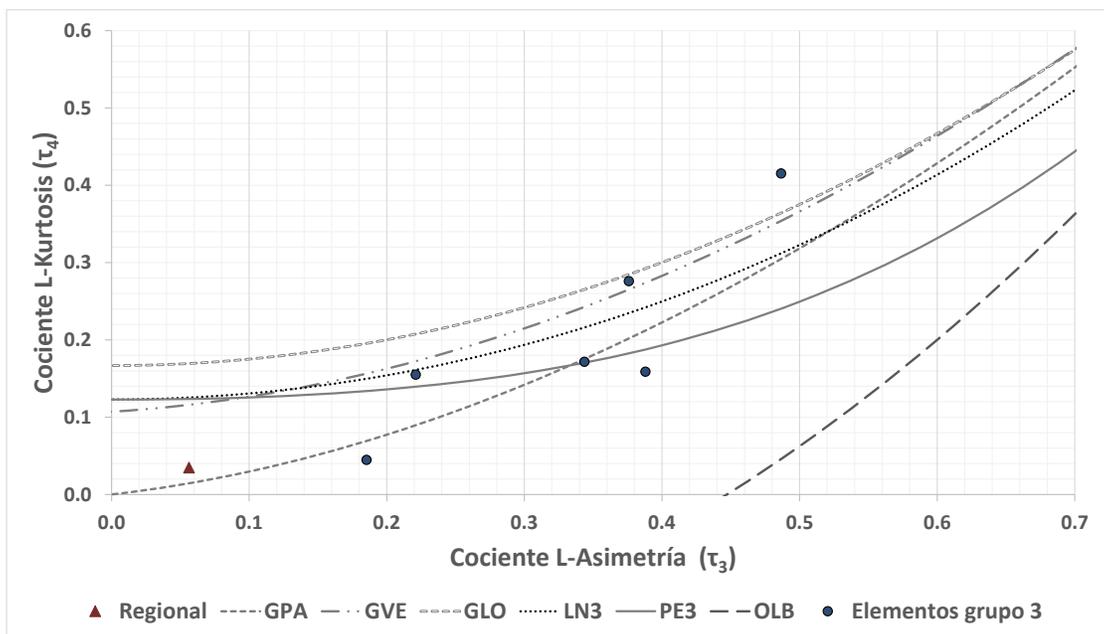


Figura 14-11 Diagrama de cocientes de Momentos-L para la RH3.

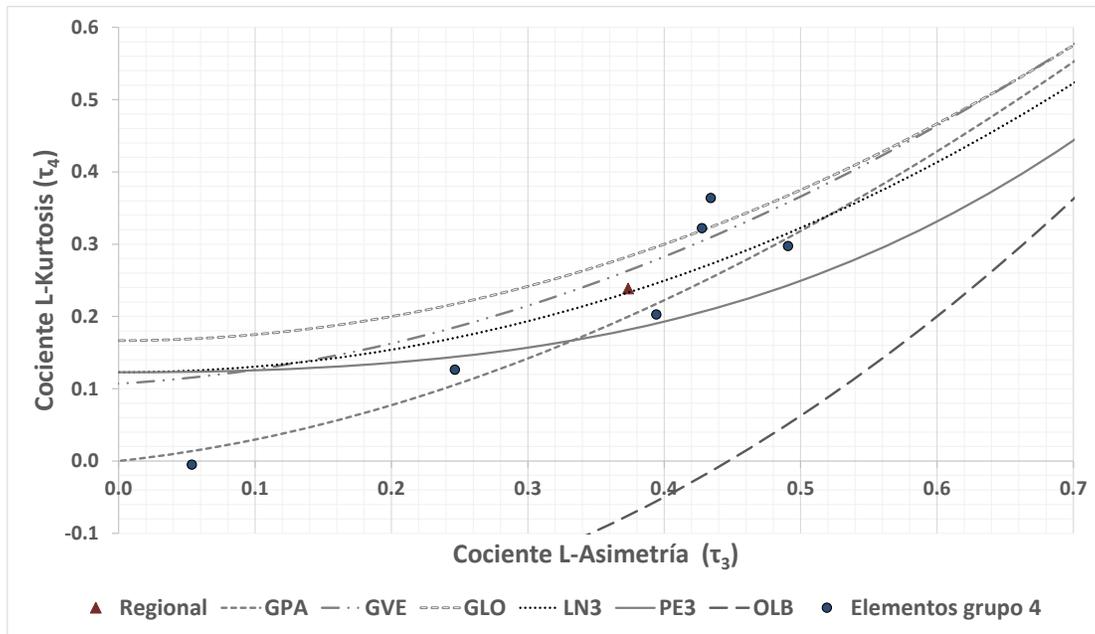


Figura 14-12 Diagrama de cocientes de Momentos-L para la RH4.

14.5.2. Generación de ecuaciones de regresión

Con el objetivo de analizar la correspondencia entre la avenida índice y las características de la cuenca se analizó mediante mínimos cuadrados a las ecuaciones que describen cada sitio aforado dentro de cada región homogénea.

Una suposición dentro del análisis regional de la avenida índice consiste en que todas las series de máximos anuales de todos los distintos sitios dentro de la región homogénea están idénticamente distribuidos excepto por un factor de escala, que se denomina avenida índice (Nobert *et al.*, 2014), por ello, fue necesario establecer la relación entre las características de la cuenca más representativas y la avenida observada que fue tomada del registro histórico.

Las características más representativas de cada estación se tomaron del análisis del último filtro de la selección de características (ver sección 14.3), siendo todas aquellas que tuvieron mayor correlación: precipitación, área, longitud de corrientes y longitud del cauce principal (Tabla 14-8). La avenida índice fue definida como el caudal medio anual.

Mediante un ajuste por mínimos cuadrados se obtuvieron las ecuaciones regionales (Tabla 14-9).

Tabla 14-8 Variables independientes y dependientes de cada estación.

Estación	P (km)	A (km ²)	LCR (km)	LMR (km)	Caudal medio observado (m ³ /s)
Región Homogénea 1					
10029	367	1833	145.52	144	606.73
10031	322	1652	92.04	83	585.56
10034	247	927	76.03	90	321.16
10066	327	1455	91.60	106	334.94
10070	361	1908	154.19	143	781.98
10079	265	1017	101.40	89	984.88
10087	332	2263	187.25	104	758.53
Región Homogénea 2					
10036	466.7	2946.5	229.20	118.7	946.73
10077	275.3	1221.1	82.60	86.0	1083.23
10110	273.8	1512.9	74.20	98.2	1078.52
10112	336.4	2126.7	174.90	95.3	1441.30
10120	313.5	1309.6	110.30	104.6	946.73
Región Homogénea 3					
10027	157.28	373.34	25.32	50.83	288.88
10057	126.29	219.08	37.87	37.87	189.13
10078	109.45	277.94	31.08	31.08	104.87
10083	242.07	847.79	54.34	83.75	451.63
10086	120.20	209.64	40.05	40.05	242.64
10090	146.41	398.13	43.21	44.57	198.29
Región Homogénea 4					
10037	1308.95	19824.15	1519.87	400.76	3328.20
10040	944.94	8992.74	677.08	293.54	934.61
10064	629.30	5132.80	373.59	223.41	980.79
10065	730.99	6166.82	474.91	255.41	1185.65
10113	722.99	6109.00	466.57	153.50	944.79
10137	559.14	3796.70	280.17	158.14	987.85

Tabla 14-9 Ecuaciones de regresión para cada región homogénea.

Región homogénea	Ecuación de regresión
1	$\bar{Q}_{R1} = 16.3398P - 2.1221A + 18.4107LCR - 19.3661LMR - 1766.8598$
2	$\bar{Q}_{R2} = -7.7054P + 0.2549A + 6.7552LCR - 2.8179LMR + 2578.0781$
3	$\bar{Q}_{R3} = -2.1551P - 0.3219A - 0.5932LCR + 15.7414LMR - 38.4833$
4	$\bar{Q}_{R4} = -5.7537P - 0.6577A + 13.5673LCR + 1.8062LMR + 2550.7777$

Se realizó la comparación del valor observado con el valor estimado mediante las ecuaciones obtenidas (Tabla 14-9), donde el error relativo es considerablemente pequeño (Tabla 14-10).

Mientras que en la Figura 14-13 se muestran las gráficas de dispersión del caudal observado y estimado.

Tabla 14-10 Comparación de caudal observado y estimado en cada región homogénea.

Estación	Nombre	Caudal medio observado Qm (m³/s)	Caudal medio estimado (m³/s)	Error relativo
Región Homogénea 1				
10029	Naranja	606.73	589.40	-0.029
10031	Guamúchil	585.56	639.52	0.092
10034	Zopilote	321.16	313.09	-0.025
10066	Choix	334.94	361.05	0.078
10070	Acatitlán	781.98	709.29	-0.093
10079	Badiraguato	984.88	959.03	-0.026
10087	Tamazula	758.53	817.74	0.078
Región Homogénea 2				
10036	Jaina	946.73	946.73	3.60252E-16
10077	Palo Dulce	1083.23	1083.23	2.09903E-16
10110	Tohayanna	1078.52	1078.52	2.1082E-16
10112	Guatenipa II	1441.30	1441.30	4.73268E-16
10120	Guasave Puente Carretera	946.73	946.73	3.60252E-16
Región Homogénea 3				
10027	El Bledal	288.88	287.50	-0.005
10057	Bamícori	189.13	192.46	0.018
10078	La Tina	104.87	106.97	0.020
10083	El Quelite	451.63	453.03	0.003
10086	Pericos	242.64	241.65	-0.004
10090	Chico Ruíz	198.29	193.81	-0.023
Región Homogénea 4				
10037	Huites	3328.20	3324.52	-0.001
10040	Santa Cruz	934.61	915.25	-0.021
10064	Chinipas	980.79	1025.91	0.046
10065	Ixpalino	1185.65	1193.16	0.006
10113	La Huerta	944.79	979.98	0.037
10137	Tecusiapa	987.85	923.06	-0.066

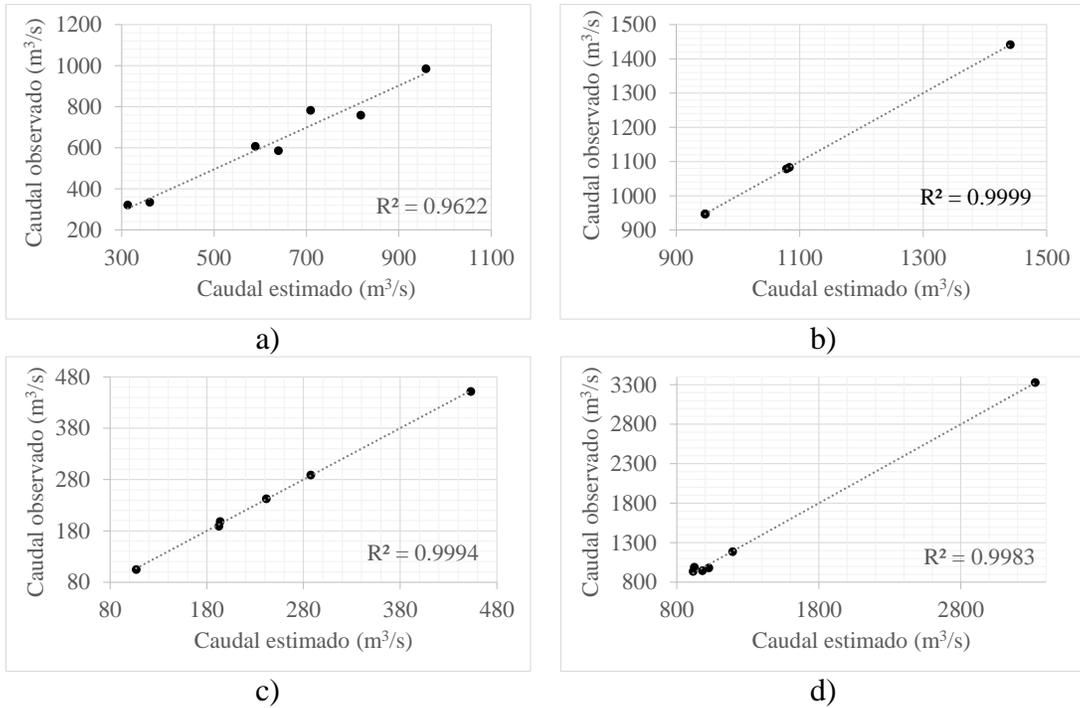


Figura 14-13 Gráficas de dispersión del caudal observado y estimado en las cuatro regiones (a - d)

14.6. Generación de curvas de duración de caudal (CDC)

Las gráficas de curvas de duración de caudal obtenidas en cada región homogénea se crearon a partir del caudal mensual específico, la curva media regional de cada región fue ajustada a una ecuación polinómica de grado 10 (Figura 14-14 a Figura 14-17), los coeficientes de la ecuación se encuentran en la Tabla 14-11.

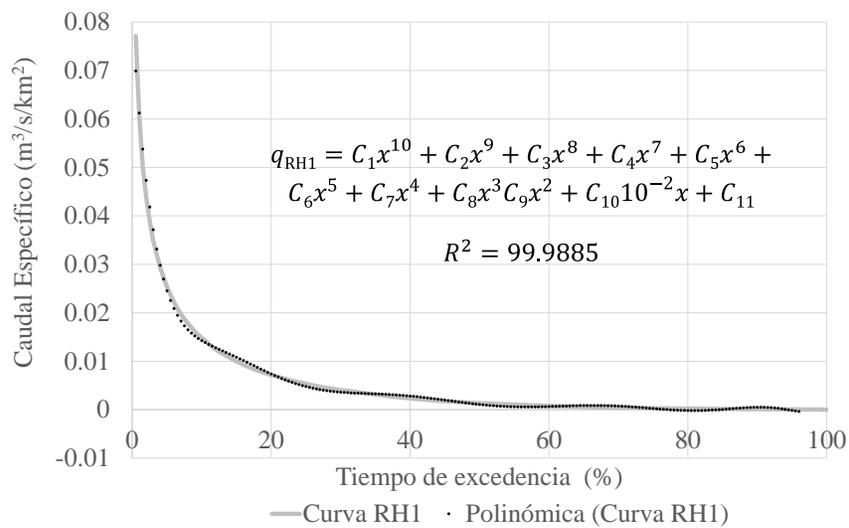


Figura 14-14 Curvas de duración de caudal para la región homogénea 1

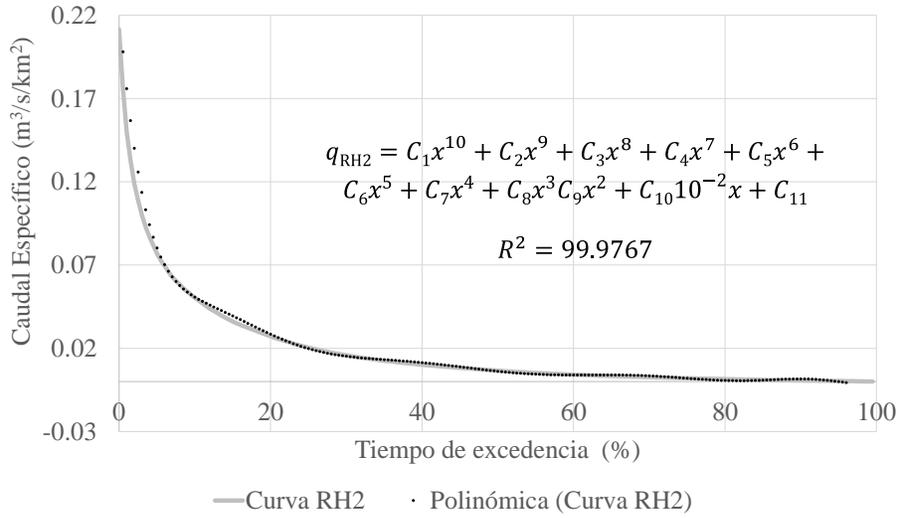


Figura 14-15 Curvas de duración de caudal para la región homogénea 2

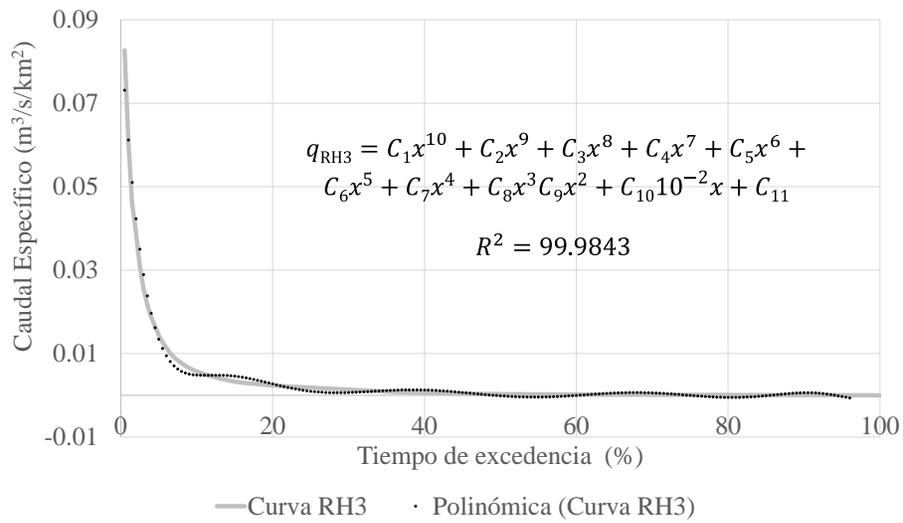


Figura 14-16 Curva de duración de caudal para la región homogénea 3.

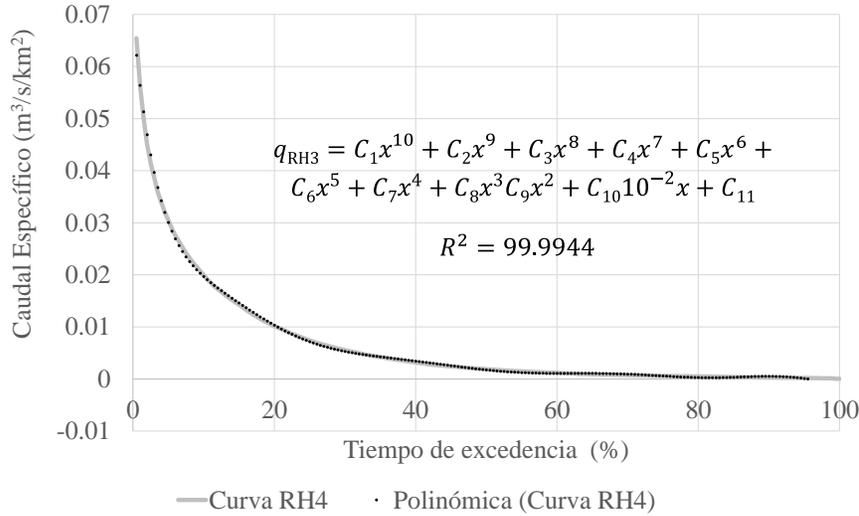


Figura 14-17 Curva de duración de caudal para la región homogénea 4

Tabla 14-11 Valores de los coeficientes de cada ecuación de CDC regional

Coefficientes	1	2	3	4
C1	0.0021515	0.0050347	0.0031529	0.0012150
C2	-0.0023182	-0.0059217	-0.0034636	-0.0013919
C3	-0.0126720	-0.0289990	-0.0184040	-0.0070568
C4	0.0118350	0.0302840	0.0175680	0.0070972
C5	0.0264100	0.0585780	0.0377870	0.0145010
C6	-0.0205850	-0.0526340	-0.0296700	-0.0125190
C7	-0.0216480	-0.0450010	-0.0313580	-0.0110570
C8	0.0119530	0.0291300	0.0177730	0.0062010
C9	0.0079258	0.0176440	0.0098194	0.0055269
C10	-0.0042437	-0.0133560	-0.0035339	-0.0039236
C11	0.0010545	0.0061047	-0.0001059	0.0017129

Nota: en las ecuaciones $x = (t-50.25)/28.94$, donde $t=0, 1, \dots, 100$, que indican el valor del percentil de 0 a 100%.

14.7. Selección de un sitio para la aplicación de la metodología regional

Mediante la metodología expuesta se procedió a la elección de un sitio probable para la implementación de una minicentral hidroeléctrica, debido a que existe un déficit y sobre – explotación considerable de acuíferos en los ríos Mocerito y Culiacán, no se eligió al sitio de implementación en estas cuencas.

Por lo que la región homogénea seleccionada para implementar la metodología fue aquella región homogénea que estuviera dentro del Consejo de Cuenca de los Ríos Fuerte y Sinaloa, esta selección estuvo basada en la situación reciente en la cuenca, donde la degradación de

la calidad del agua se encuentra en un nivel no crítico, tanto en la parte alta y media de la misma, además existe riesgo de inundación en la parte baja de la cuenca y en áreas cercanas al margen del río (CONAGUA, 2012).

En la revisión de estudios previos, el CMM (2011) realizó una investigación apoyada de sitios propuestos por la CFE de aquellos puntos de interés para la implementación de centrales de pequeña escala. Por lo que para este trabajo de investigación se consultaron los sitios de interés dentro de la RH10 y del consejo de cuenca de los ríos Fuerte y Sinaloa, seleccionando la región homogénea 4 por las características geográficas que comparte con la Cuenca de los ríos Fuerte y Sinaloa. Mediante la plataforma web de la SENER (<https://dgel.energia.gob.mx/inere/>) que muestra la investigación del CMM (2011) se seleccionó el proyecto clasificado como probable denominado “Río Verde”.

En la plataforma web se publican tanto la capacidad instalable y el potencial de generación de la central probable a implementar.

En el resumen general de consulta del INERE se mencionan algunos criterios tanto técnicos como de sustentabilidad (ver la sección 13.5 Estudios) que fueron empleados en el estudio, por lo que una actividad a desarrollar en este trabajo y debido a la escasez de acceso a mayor información del mencionado proyecto, fue obtener la capacidad instalable publicada en la plataforma web (1.55 MW) empleando los criterios técnicos que se mencionan en la “síntesis del estudio potencial hidroeléctrico sustentable con base en pequeñas centrales” cuya dirección de acceso es la siguiente: https://dgel.energia.gob.mx/documentos/POTENCIAL/MINIHIDRAULICA/Pot_Minihidro_CMM.pdf, obteniendo la capacidad instalable por medio de los valores adquiridos del caudal de diseño con la metodología propuesta de este trabajo.

14.7.1. Gasto medio anual

Se trazó la cuenca (Figura 14-18) a partir del sitio de implementación “Río Verde” y se obtuvieron los parámetros descriptivos: perímetro de la cuenca de 575.44 km, área de la cuenca 4420.13 km², longitud del cauce principal de 146.51 km y longitud de todas las corrientes de 294.60 km. Y utilizando la ecuación obtenida en la Tabla 14-9, se obtuvo el caudal medio anual de 594.01 m³/s.

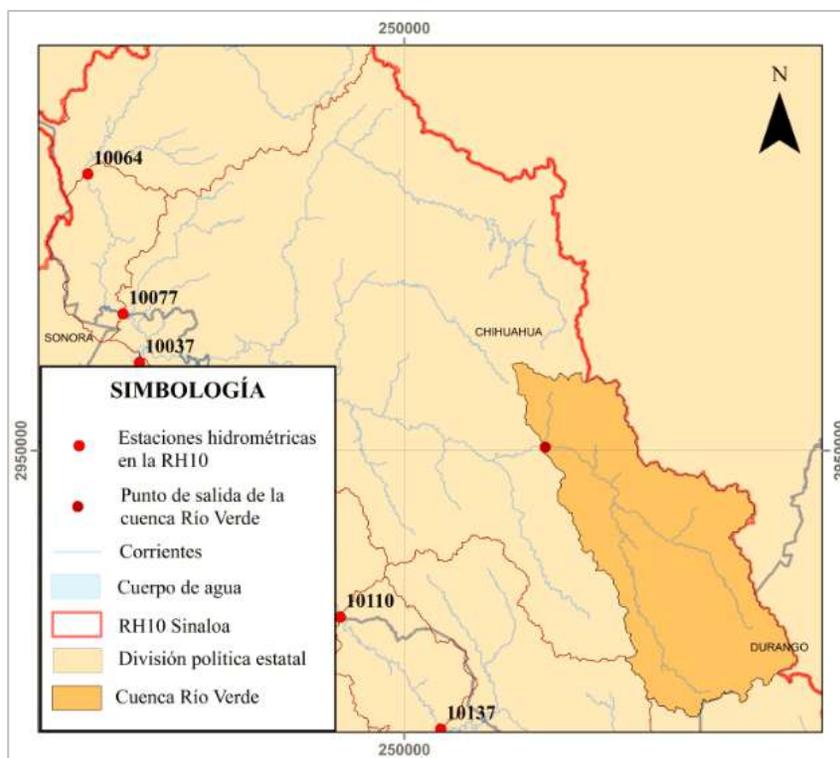


Figura 14-18 Ubicación del proyecto Río Verde

14.7.2. Curva regional

En el desarrollo de la metodología y la aplicación en un sitio elegido como no aforado es importante la comparación de los valores obtenidos con aquellos obtenidos en estudios previos (Campos-Aranda 2014 a,b), que aunque distan del periodo de registros obtenidos, si aportan al investigador un panorama de revisión a nivel general. Debido a lo anterior se realizó una comparación del estudio actual con un estudio previo que cuenta con registros de datos al año 2002 (D.F. Campos-Aranda, 2014a) (Tabla 14-12) resultando satisfactorio el resultado.

Tabla 14-12 Comparación de predicciones históricas de crecientes

ID	MA	EEA (m ³ /s)	Periodos de retorno (m ³ /s)				
			5	10	50	100	500
10037	*LN3 (mv)	965.44	5870.42	8848.98	18522.55	24124.73	41333.63
	**LP3(dl)	793.90	4368.00	6679.00	15613.00	21776.00	45130.00
10040	*LN3 ECS	391.25	1472.10	2190.84	4207.92	5252.81	8160.89
	**LP3(dl)	399.30	1525.00	2274.00	4447.00	5583.00	8721.00
10064	*LN3 (2/3) ECS	119.83	1424.35	1770.34	2513.48	2824.37	3546.26
	**GVE(mL)	139.00	1231.00	1626.00	2714.00	3283.00	4919.00

10065	*LN3 (2/3) ECS	354.35	1702.00	2367.23	4201.71	5139.16	7717.53
	**LP3(dl)	343.40	1635.00	2311.00	4431.00	5651.00	9450.00
10113	*LN3 (2/3) ECS	108.36	1410.13	1672.57	2151.06	2325.52	2687.96
	**GVE(ml)	96.30	1431.00	1718.00	2223.00	2393.00	2709.00
10137	**LN3 (2/3) ECS	209.31	1544.13	2078.99	3356.64	3939.40	5393.75

Donde: ** estudio actual, * estudio previo, ID: clave de la estación hidrométrica, MA: Método de ajuste (dl: momentos dominio logarítmico, ml: momentos-L, mv: máxima verosimilitud), EEA: Error Estándar de Ajuste

La curva regional para el grupo homogéneo número 4 (Figura 14-19) se realizó con la información dada de la distribución regional seleccionada, LN3, del cuantil regional medio (de todas las estaciones involucradas en la región) asociado a un evento de diseño (Tabla 14-13).

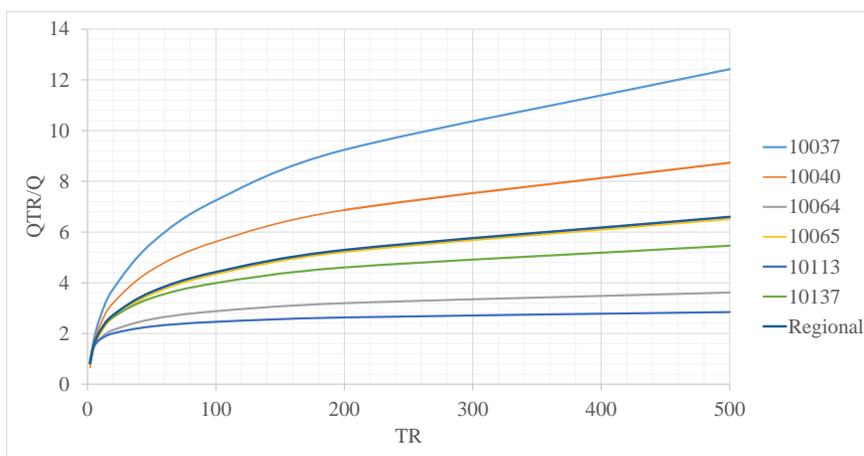


Figura 14-19 Curvas de crecimiento regional trazadas con la curva de crecimiento regional en la RH4.

El caudal esperado en la zona de proyecto (Tabla 14-13) se obtuvo por medio de la ecuación (12.3), la curva regional esperada en la zona de proyecto se muestra en la Figura 14-20.

Tabla 14-13 Estimación de cuantiles regionales

Periodo de retorno T_r	Cuantil regional Z_T	Caudal esperado en el sitio Q_{TR} (m^3/s)
2	0.8237	489
5	1.5471	919
10	2.1132	1255
20	2.7282	1621
50	3.6415	2163
100	4.4221	2627
200	5.2914	3143
500	6.5968	3919

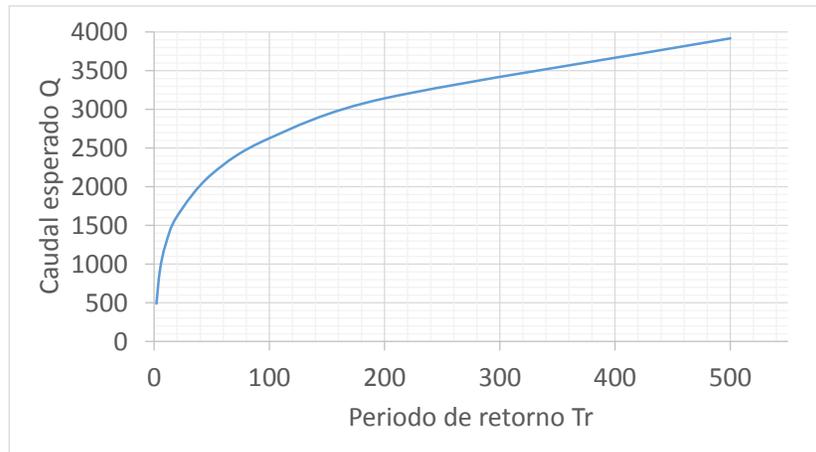


Figura 14-20 Curva del caudal esperado en la RH4

El comportamiento del caudal específico medio mensual histórico en la RH4 se incrementa en los meses de junio a noviembre, hay menor escurrimiento en los meses diciembre a abril, y el mes de mayo es el de menor escurrimiento (Figura 14-21).

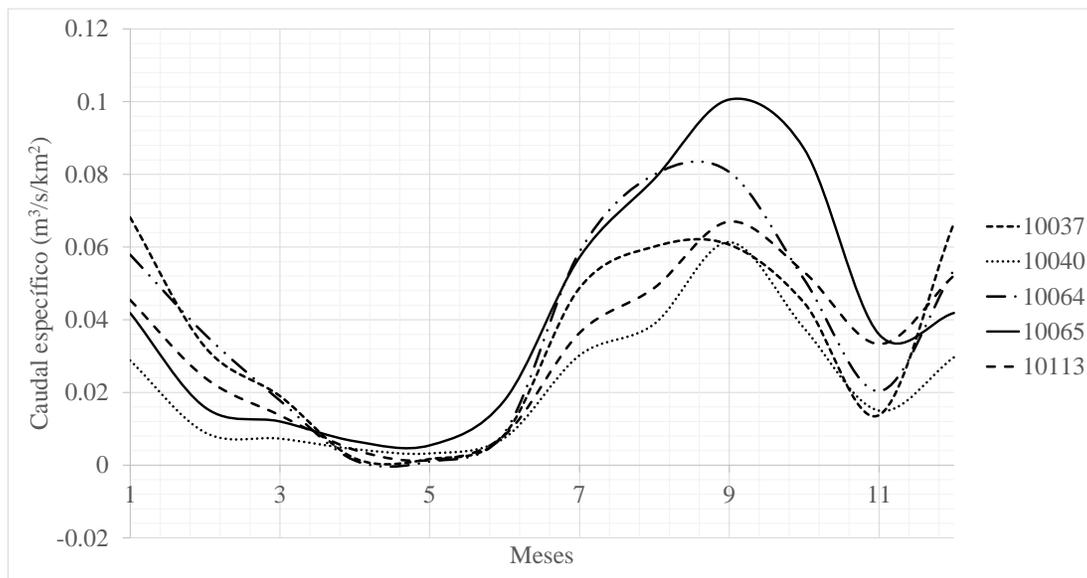


Figura 14-21 Caudal mensual específico en la región homogénea 4

14.7.3. Curva de Duración de Caudal en la zona de interés

La CDC de la RH4 se definió a partir de los coeficientes de la Tabla 14-11. Con el objetivo de conocer la potencia y generación media anual se asignaron valores tanto propuestos como valores encontrados en la revisión de la literatura aplicados en la zona de estudio. Estos

valores se refieren al estudio del Centro Mario Molina (CMM) donde proporcionaron sitios puntuales en todo el país con potencial hidráulico para la implementación de minicentrales de paso, algunos datos publicados que se emplearon en este proyecto son el factor de planta (FP) del 25%, y carga aprovechable de 10 a 20 metros como valores mínimos, mientras que el valor propuesto para este estudio es la eficiencia del sistema (η) que se adoptó como 85%, se emplearon las ecuaciones (12.4) y (12.5).

Debido a lo anterior se propusieron dos casos, el caso 1 (C1) con una carga de 25 m, FP de 25% y $\eta=85\%$; mientras que para el caso 2 (C2) los datos procesados son carga de 14 m, FP de 25% y $\eta=85\%$ (consultar el siguiente *link*)

(https://dgel.energia.gob.mx/documentos/POTENCIAL/MINIHIDRAULICA/Pot_Minihidro_CMM.pdf)

Los resultados se muestran en la Tabla 14-14 y Figura 14-22.

La selección del tipo de turbina empleando el diagrama Figura 11-13 con una carga de 25 metros y para la potencia calculada en la Tabla 14-14 sería la tipo Kaplan.

Tabla 14-14 Cálculo de potencia instalable y generación media anual en la RH4.

P (%)	x	CAP* (m³/s/km²)	Q_{PR} (m³/s)	P_{C1} (MW)	GMA_{C1} (GWh/año)	P_{C2} (MW)	GMA_{C2} (GWh/año)
5	-1.5636	0.0301	132.9122	27.7072	59.8476	15.5160	33.5146
10	-1.3908	0.0197	86.9713	18.1302	39.1613	10.1529	21.9303
15	-1.2180	0.0146	64.5098	13.4479	29.0474	7.5308	16.2665
20	-1.0453	0.0103	45.6997	9.5267	20.5776	5.3349	11.5235
25	-0.8725	0.0072	31.6531	6.5985	14.2527	3.6951	7.9815
30	-0.6997	0.0053	23.3958	4.8772	10.5347	2.7312	5.8994
35	-0.5270	0.0043	18.8241	3.9241	8.4761	2.1975	4.7466
40	-0.3542	0.0034	15.1974	3.1681	6.8431	1.7741	3.8321
45	-0.1814	0.0026	11.3177	2.3593	5.0961	1.3212	2.8538
50	-0.0086	0.0017	7.7229	1.6099	3.4774	0.9016	1.9474
55	0.1641	0.0012	5.4633	1.1389	2.4600	0.6378	1.3776
60	0.3369	0.0011	4.7833	0.9971	2.1538	0.5584	1.2061
65	0.5097	0.0011	4.7598	0.9922	2.1432	0.5556	1.2002
70	0.6824	0.0009	4.1233	0.8595	1.8566	0.4813	1.0397
75	0.8552	0.0006	2.5589	0.5334	1.1522	0.2987	0.6452
80	1.0280	0.0003	1.2205	0.2544	0.5496	0.1425	0.3078
85	1.2008	0.0003	1.4541	0.3031	0.6547	0.1697	0.3667

90	1.3735	0.0005	2.2748	0.4742	1.0243	0.2656	0.5736
95	1.5463	0.0001	0.4391	0.0915	0.1977	0.0513	0.1107
100	1.7191	0.0006	2.4629	0.5134	1.1090	0.2875	0.6210

Nota: CAP: Curva de ajuste polinómica de la RH4, QPR: Caudal asociado al periodo de excedencia-percentil en m³/s, P_{C1}: Capacidad a instalar condición 1, GMA_{C1}: Generación Media Anual caso 1, PC2: Capacidad a instalar caso 2, GMA_{C2}: Generación Media Anual caso 2.

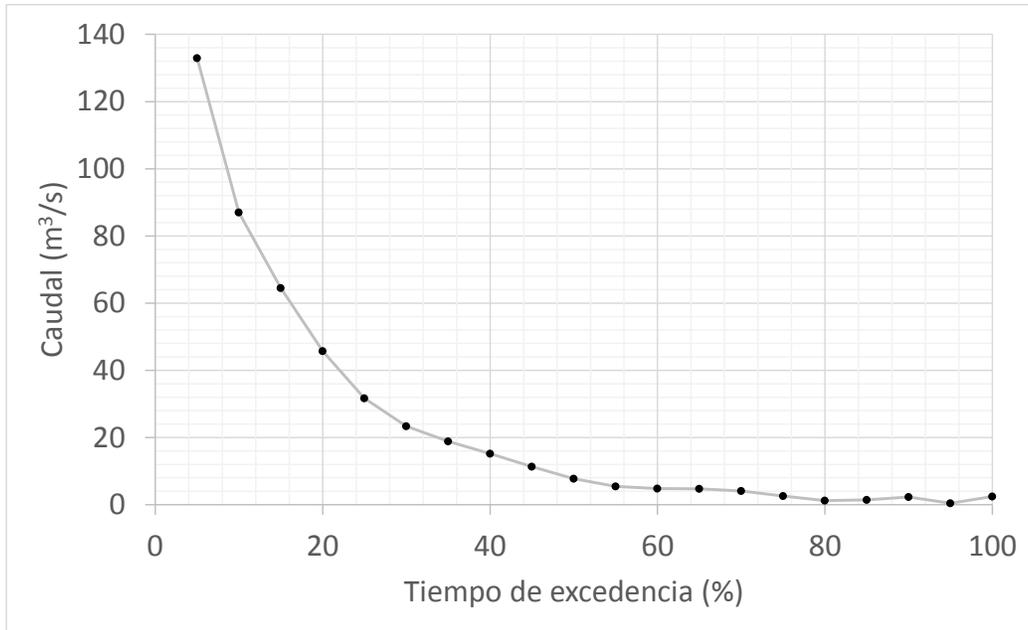


Figura 14-22 CDC en la RH4.

15. CONCLUSIONES

La estimación del caudal en sitios no aforados mediante métodos regionales permite la estimación más precisa de un evento de diseño asociado a cierto periodo de retorno, por lo que se mitiga el sub-dimensionar y sobre-dimensionar alguna obra hidráulica a implementar en cierta zona.

Este trabajo de investigación se desarrolló bajo dos enfoques: teórico y práctico.

En cuanto al aspecto teórico, la elaboración del compilado de la revisión de la literatura y del estado del arte que proporcionan al lector un panorama de los distintos métodos y técnicas de aplicación en el análisis regional así como la obtención de las características de la cuenca, análisis previo de datos, ARFC, métodos de delimitación, métodos regionales y aspectos generales en el diseño de una minicentral. En cuanto al aspecto práctico, se aplicó la metodología propuesta a un sitio.

La escritura de capítulos proporciona una formación básica tanto teórica como práctica en el desarrollo de un análisis regional, debido a la fundamentación teórica, así como a los ejemplos de aplicación realizados, que son explicados paso a paso. Además el orden de los seis capítulos representa el desarrollo general de un análisis regional.

La necesidad de la verificación de la cantidad de estaciones a emplear en un análisis regional hidrológico es importante debido a que en caso de omitir este procedimiento, ocasionaría la formación de grupos no del todo homogéneos por lo que la agrupación de elementos llevaría un re-proceso de información y no garantizaría la homogeneidad en cada grupo. Derivado de esta investigación el orden propuesto de pruebas a aplicar es el siguiente: cantidad de registro de datos, medidas de discordancia y verificación de tendencia en las series de registros.

La selección de variables descriptivas más importantes es un paso fundamental en el análisis regional hidrológico, por lo que la propuesta inicial es la selección de todas aquellas variables relacionadas con el objeto del estudio, esto permite la formación de agrupaciones homogéneas con significado físico. La misma recomendación ocurre para las variables hidrométricas. En este trabajo se realizó la agrupación por medio del caudal medio mensual junto con las características descriptivas de cada cuenca, debido a que el objetivo fue

establecer la similitud entre regiones por medio de la permanencia del caudal a través del tiempo, para de esta manera obtener las CDC características.

El análisis multivariado se propuso como una herramienta para la selección de variables más representativas por medio de un ACP. En un ACP se analizan todas las variables, en este caso las M variables descriptivas de la zona de estudio, donde se busca presentar de forma gráfica, el máximo de la información contenida en la matriz de información original M, esto es posible medirlo mediante el criterio de Kaiser y la varianza acumulada debida a cada valor propio, es decir mayor o igual al 80% de la varianza total.

En este trabajo se propuso una metodología para el filtro de variables, comparándola con la metodología tradicional para 32 variables. En los resultados obtenidos la propuesta resultó satisfactoria al mejorar en un 26.40% la varianza acumulada, con la metodología propuesta la mayor varianza y valores propios mayores a 1 se proyectan en dos ejes con un 84.40%, mientras que con la metodología tradicional la varianza acumulada aceptable, es decir, mayor del 80%, se explicó en 7 valores propios, lo que implicó la representación de únicamente el 58% de la varianza acumulada en dos ejes respecto a la matriz de datos originales. Por lo que se adoptó la metodología propuesta resultando 13 variables más representativas del total de 32 variables de la matriz original.

Lo anterior permitió la agrupación de elementos homogéneos, es decir, de estaciones hidrométricas, por medio de una matriz de 25 variables, 12 hidrométricas, que corresponden al caudal medio mensual y las 13 variables más representativas. Se recomienda el empleo del método de agrupación jerárquica, en dos fases, la primera de forma jerárquica, es decir, mediante el método de Ward, y el último por medio de un método de optimización, método de K-medias.

En cuanto al Análisis Regional de Frecuencias de Crecientes (ARFC) aplicado con la teoría de Momentos-L (ML) permitió la selección de una distribución regional de cada agrupación homogénea. Por lo que la teoría de Momentos-L es una técnica eficiente, desde la medición de la discordancia en la etapa inicial de selección de estaciones hidrométricas así como en la verificación de la homogeneidad de la agrupación resultante, además que proporciona la selección de una distribución regional aplicable mediante el diagrama de Hosking y Wallis (1997), lo anterior se utiliza para la estimación del cuantil regional asociado a un periodo de

retorno que es empleado en la curva regional para la determinación del caudal esperado en la zona no aforada.

Las ecuaciones de regresión regionales para la avenida índice esperada en los sitios no aforados, se derivaron de las variables independientes de predicción y la variable dependiente, que para este caso fue el valor del caudal observado, por lo que la selección de las variables predictivas estuvo con base a las variables descriptivas más representativas y de correlación mayor o igual a 0.7, obteniendo un error relativo menor al 10% y en la región número 4, la región seleccionada para la aplicación de la metodología propuesta, se obtuvo en promedio un error relativo menor al 7%. En el análisis de dispersión del caudal estimado y observado, se obtuvieron valores de ajuste por región de 0.96 a 0.99. De esta forma en la selección de un sitio no aforado, basta con la identificación de las características de la zona no aforada (área, perímetro, longitud de corrientes y longitud del cauce principal), de manera que éstas multiplicadas por un coeficiente proporcionarán el valor medio esperado en el sitio sin aforos.

Con el objetivo de caracterizar dentro de la RH10 el potencial hidroeléctrico, se desarrollaron las CDC regionales con gastos medios mensuales, esta creación de curvas se realizó por medio de un ajuste del tipo polinómico y de grado 10 con coeficientes de correlación de ajuste confiables desde 99.97 a 99.99, lo que permitió construir la CDC para el sitio no aforado seleccionado, donde se obtuvieron la capacidad a instalar y la generación media anual, donde la capacidad a instalar corresponde en la permanencia del caudal del 50%.

Finalmente, se puede concluir que la metodología propuesta resultó satisfactoria por medio de la evaluación en cada una de sus etapas, proporcionando fundamentación teórica y práctica para futuros estudios en análisis regional hidrológico.

16. REFERENCIAS

- Acreman, M. C., y A. Werritty. 1987. Flood frequency estimation in Scotland using index floods and regional growth curves. *Trans. R. Soc. Edinburgh Earth Sci.* 78:305–313.
- Acreman, M. C., y C. D. Sinclair. 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. *J. Hydrol. (Amsterdam, Neth.)* 84:365–380.
- Adeloye, A.J. y M. Montaseri. 2002. Preliminary streamflow data analyses prior to water resources planning study. *Hydrol. Sci. J.* 47: 679-692.
- Afifi, A., V. A. Clark, y S. May. 2004. *Computer-aided multivariate analysis*. 4th ed. Chapman & Hall/CRC, Boca Raton, Fla.
- Ahn, K. H., y R. Palmer. 2016. Regional flood frequency analysis using spatial proximity and basin characteristics: Quantile regression vs. parameter regression technique. *J. Hydrol.* 540:515–526. doi: 10.1016/j.jhydrol.2016.06.047
- Ahn, K. H., y V. Merwade. 2016. Role of Watershed Geomorphic Characteristics on Flooding in Indiana, United States. *J. Hydrol. Eng.* 21:5015021-1-5015021–12. doi: 10.1061/(ASCE)HE.1943-5584.0001289
- Ali, G. A., A. G. Roy, M. C. Turmel, y F. Courchesne. 2010. Multivariate analysis as a tool to infer hydrologic response types and controlling variables in a humid temperate catchment. *Hydrol. Process.* 24:2912–2923.
- Álvarez, M., J. Puertas, y B. Soto. 1999. Análisis regional de las precipitaciones máximas en Galicia mediante el método del índice de avenida. *Ingeniería del Agua.* 6:283-290.
- Álvarez-Olguín, G., F. Sustaita-Rivera, y N. Hotait- Salas. 2011. Identificación de regiones hidrológicas homogéneas mediante análisis multivariado. *Ing. Investig. y Tecnol.* 12:277–284.
- Anderberg, M.R. 1973. *Cluster analysis for applications*. Academic Press. Inc. New York.
- Andrews, D.F. 1972. Plots of high dimensional data. *Biometrics.* 28:125-136

Arellano-Lara, F., y C. A. Escalante-Sandoval. 2014. Multivariate delineation of rainfall homogeneous regions for estimating quantiles of maximum daily rainfall: A case study of northwestern Mexico. *Atmósfera*. 27:47–60.

Arkel, V. Z., y A. L. Kaleita. 2014. Identifying sampling locations for field-scale soil moisture estimation using K-means clustering, *Water Resour. Res.* 50:7050–7057. doi:10.1002/2013WR015015.

Asadi, P., Engelke, S., y Davison, A. C. 2018. Optimal regionalization of extreme value distributions for flood estimation. *J. Hydrol.* 556:182–193. doi:10.1016/j.jhydrol.2017.10.051

Astel, A., S. Tsakovski, P. Barbieri, y V. Simeonov. 2007. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Resour. Res.* 41:4566–4578. doi:10.1016/j.watres.2007.06.030

Ay, M., y O. Kisi. 2014. Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques. *J. Hydrol.* 511:279–289. doi:10.1016/j.jhydrol.2014.01.054.

Baeriswyl, P. A., y M. Rebetez. 1997. Regionalization of precipitation in Switzerland by means of principal component analysis. *Theor. Appl. Climatol.* 58:31–41.

Bao, Z., J. Zhang, J. Liu, G. Fu, G. Wang, R. He, X. Yan, J. Jin, y H. Liu. 2012. Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions. *J. Hydrol.* 466–467:37–46.

Beran, M.A., M. Brilly, A. Becker, y O. Bonacci. 1990. Regionalization in hydrology. International Association of Hydrological Sciences. IAHS Publication no. 191, p. 260.

Bobée, B., y P. Rasmussen. 1995. Recent advances in flood frequency analysis. US National Report to International Union of Geodesy and Geophysics (1991–1994). *Reviews in Geophysics*. doi:10.1029/95RG00287.

Boscarello, L., G. Ravazzani, A. Cislighi, y M. Mancini. 2015. Regionalization of Flow-Duration Curves through Catchment Classification with Streamflow Signatures and

Physiographic–Climate Indices. *J. Hydrol. Eng.* 19:1–8. doi:10.1061/(ASCE)HE .1943-5584.0001307.

Brath, A., A. Castellarin, M. Franchini, y G. Galeati. 2001. Estimating the index flood using indirect methods. *Hydrol. Sci. J.* 46:399–418.

Burn, D. H. 1989. Cluster analysis as applied to regional flood frequency. *Journal of Water Resources Planning and Management* 115: 567–582.

Burn, D. H. 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. *J. Hydrol.* 202:212-230

Burn, D. H., Z. Zolt, y M. Kowalchuk. 1997. Regionalization of catchments for regional flood frequency analysis. *J. Hydrol. Eng.* 2:76–82.

Camargo-Hernández J.E. y V. Franco. 1998. Adquisición de datos. Capítulo 2 del manual de ingeniería de ríos. Series del Instituto de Ingeniería número 607. ISSN 0185-2345.

Campos-Aranda, D.F. 2006. Contraste de métodos regionales de estimación de crecientes en la cuenca del río Guayalejoen Tamaulipas. *Tlálloc.* 37:14-24

Campos-Aranda, D. F. 2010a. Discusión al artículo “Una modificación simple y útil del método de la avenida índice.” *Tecnol. y Ciencias del Agua* 1:119–121.

Campos-Aranda, D. F. 2010b. Verificación de la homogeneidad regional mediante tres pruebas estadísticas. *Tecnol. y Ciencias del Agua* I:157–165.

Campos-Aranda, D. F. 2012. Descripción y aplicación de la versión corregida del Test de Langbein para verificar homogeneidad regional. *Ing. Investig. y Tecnol.* 13:411–216.

Campos-Aranda, D. F. 2014a. Análisis Regional de Frecuencia de Crecientes en la región hidrológica no. 10 (Sinaloa), México. 1: índices de estacionalidad y regiones de influencia. *Agrociencia* 48:147–158.

Campos-Aranda, D. F. 2014b. Análisis Regional de Frecuencia de Crecientes en la región hidrológica no. 10 (Sinaloa), México. 2. Contraste de predicciones locales y regionales. *Agrociencia* 48:255-270.

Caratti, J. F., J. A. Nesser, y C. L. Maynard. 2005. Watershed Classification Using Canonical Correspondence Analysis and Clustering Techniques : a Cautionary Note. *J. Am. Water Resour. Assoc. (JAWRA)* 40(5):1257-1268.

Caraway, N. M., J. L. McCreight, y B. Rajagopalan. 2014. Multisite stochastic weather generation using cluster analysis and k-nearest neighbor time series resampling. *J. Hydrol.* 508:197–213. doi:10.1016/j.jhydrol.2013.10.054.

Cevallos-Flores J. E. 2015. Modelación de la Propagación de Hidrogramas Parabólicos en la Cuenca Urbana del Dren Peñuelas. Disertación de maestría (Posgrado-UAQ).

CFE. 1980. Manual de diseño de obras civiles hidrotecnia A.1.10. Avenida de diseño. Comisión Federal de Electricidad. Instituto de Investigaciones Eléctricas. (Ed.) Linatti. México. (NOTAS: elaboración por Domínguez-Mora, R. revisión: Maza-Álvarez, J.A. y por Frías-Alcaraz, M).

CFE. 1981. Manual de diseño de obras civiles hidrotecnia A.1.3. Escurrimiento. Comisión Federal de Electricidad. Instituto de Investigaciones Eléctricas. (Ed.) Linatti. México. (NOTAS: elaboración por Domínguez-Mora, R. revisión: Maza-Álvarez, J.A. y por Frías-Alcaraz, M).

Chang, L. C., H. Y. Shen, Y. F. Wang, J. Y. Huang, y Y. T. Lin. 2010. Clustering-based hybrid inundation model for forecasting flood inundation depths. *J. Hydrol.* 385:257–268. doi:10.1016/j.jhydrol.2010.02.028.

Chebana, F., M. A. B. Aissia, y T. B. M. J. Ouarda. 2017. Multivariate shift testing for hydrological variables, review, comparison and application. *J. Hydrol.* 548:88–103. doi:10.1016/j.jhydrol.2017.02.033.

Chernoff, H. 1973. The use of faces to represent points in k-dimensional space graphically. *J. Am. Stat. Assoc.* 68:361–368.

Chokmani, K., y T. B. M. J. Ouarda. 2004. Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Water Resources Research* 40:W12514, doi:10.1029/2003WR002983.

- Choquette, A.F.1988. Regionalization of peak discharges for streams in Kentucky. Water-Resour. Invest. Rep. (U. S. Geol. Surv.) 87- 4209.
- Chow, V. Te, D. R. Maidment, y L. Mays. 1994. Hidrologia Aplicada. Editorial McGraw Hill, México.
- Cong, S., Y. Li , J.L. Vogel, y J.C. Schaake. 1993. Identification of the underlying distribution form of precipitation by using regional data. Water Resour. Res 29:1103-1111.
- Coulson, A. 1967. Flood Frequencies of Nova Scotia Streams Technical Bulletin No. 4, Water Resources Branch, Ottawa.
- Coz, F., T. Sánchez, B. Viani, J. Segura, L. Rodríguez, H. Miranda, E. Castromonte, J.Guerra, L. Quiroz, J. Gaitán, L. Moreno e I. Muñoz. 1995. Manual de mini y microcentrales hidráulicas, una guía para el desarrollo de proyectos. ITDG-Perú. 274. Perú
- Cunderlik, J. M., T. B. M. J. Ouarda, y B. Bobée. 2004. On the objective identification of flood seasons. Water Resour. Res. 40(1) W01520. doi:10.1029/2003WR002295.
- Cunnane, C. 1988. Methods and merits of regional flood frequency analysis. Journal of Hydrology 100(1-4): 269-290.
- Dalrymple, T. 1960. Flood frequency analysis. Water Supply Paper 1543-A, V.S. Geological Survey, Washington, D.C.
- De Vos, N. J., T. H. M. Rientjes, y H. V. Gupta. 2010. Diagnostic evaluation of conceptual rainfall-runoff models using temporal clustering. Hydrol. Process. 24:2840–2850.
- DeCoursey, D. G. 1973. Objective regionalization of peak flow rates. Floods and Droughts. Proceedings of the Second International Symposium in Hydrology. Fort-Collins, CO, Sept.11-13. Water Resour. Publ., Fort-Collins, CO.
- Dikbas, F., M. Firat, A. C. Koc, y M. Gungor. 2013. Defining Homogeneous Regions for Streamflow Processes in Turkey Using a K-Means Clustering Method. Arab. J. Sci. Eng. 38:1313–1319.

Dodangeh, E., S. Soltani, A. Sarhadi, y J.-T. Shiau. 2014. Application of L-moments and Bayesian inference for low-flow regionalization in Sefidroud basin, Iran. *Hydrol. Process.* 28:1663–1676. doi:10.1002/hyp.9711.

Ebisemiju, F. S. 1979. An Objective Criterion for the Selection of Representative Basins. *Water Resour. Res.* 15:148–158.

Eghbali, A. H., K. Behzadian, F. Hooshyaripor, R. Farmani, y A. P. Duncan. 2016. Improving Prediction of Dam Failure Peak Outflow Using Neuroevolution Combined with K -Means Clustering. *J. Hydrol. Eng.* 5:1–14.

Elder, K. 1995. Snow distribution in alpine watersheds. PhD dissertation. p 309. University of California, Santa Barbara.

Elsheikh, A. H., M. F. Wheeler, y I. Hoteit. 2013. Clustered iterative stochastic ensemble method for multi-modal calibration of subsurface flow models. *J. Hydrol.* 491:40–55. doi:10.1016/j.jhydrol.2013.03.037.

Erwin, D., D. Juanda, S. Notosiswoyo, y P. Soemintadiredja. 2009. Hydrogeochemistry of volcanic hydrogeology based on cluster analysis of Mount Ciremai , West Java , Indonesia. *J. Hydrol.* 376:221–234. doi:10.1016/j.jhydrol.2009.07.033.

Escalante-Sandoval C.A y L. Reyes-Chávez. 2005. Técnicas estadísticas en hidrología. (2nd Ed.) p 298. UNAM, Facultad de Ingeniería, México.

Eslamian, S., H. Hassanzadeh, J. Abedi-Koupai, y M. Gheysari. 2012. Application of L-Moments for Regional Frequency Analysis of Monthly Drought Indexes. *J. Hydrol. Eng.* 17:32–42.

Everitt, B. S. 1978. Graphical techniques for multivariate data. North-Holland. New York, USA. ISBN 0-444-19461-4.

Farsadnia, F., M. Rostami Kamrood, A. Moghaddam Nia, R. Modarres, M. T. Bray, D. Han, y J. Sadatinejad. 2014. Identification of homogeneous regions for regionalization of watersheds by two-level self-organizing feature maps. *J. Hydrol.* 509:387–397. doi:10.1016/j.jhydrol.2013.11.050.

Fassnacht, S. R., y J. E. Derry. 2010. Defining similar regions of snow in the Colorado River Basin using self-organizing maps. *Water Resour. Res.* 46:W04507. doi:10.1029/2009WR007835.

Feng, J., D. Yan, C. Li, Y. Gao, y J. Liu. 2014. Regional Frequency Analysis of Extreme Precipitation after Drought Events in the Heihe River Basin , Northwest China. *J. Hydrol. Eng.* 19:1101–1112. doi:10.1061/(ASCE)HE.1943-5584.0000903.

Fill, H.D., y J.R. Stedinger. 1995. Homogeneity Test Based Upon Gumbel Distribution and a Critical Appraisal of Dalrymple's Test. *J. Hydrol.* 166: 81-105.

Gado, T. A., y V. Nguyen. 2016. Comparison of Homogenous Region Delineation Approaches for Regional Flood Frequency Analysis at Ungauged Sites. *J. Hydrol. Eng.* 21:1–10.

Ganora, D., P. Claps, F. Laio, y A. Viglione. 2009. An approach to estimate nonparametric flow duration curves in ungauged basins. *Water Resour. Res.* 45:1–10.

García-Marín, A. P., J. Estévez, C. Sangüesa-Pool, R. Pizarro-Tapia, J. L. Ayuso-Muñoz, y F. J. Jimenez-Hornero. 2015. The use of the exponent $K(q)$ function to delimit homogeneous regions in regional frequency analysis of extreme annual daily rainfall. *Hydrol. Process.* 29:139–151.

Garcia-Osorio, C., y C. Fyfe. 2005. Visualization of High-Dimensional Data via Orthogonal Curves. *J. Univers. Comput. Sci.* 11:1806–1819.

Glatfelter, D.R. 1984. Techniques for estimating magnitude and frequency of floods on streams in Indiana. US Geological Survey, Water-Resour. Invest. Rep. 84-4134.

Greenwood, J.A., J.M. Landwehr, N.C. Matalas, y J.R. Wallis. 1979. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research* 15: 1049–1054.

GREHYS. 1996. Presentation and review of some methods for regional flood frequency analysis. *J. Hydrol.* 186:63–84.

- Gubareva, T. S. 2012. Classification of river basins and hydrological regionalization (as exemplified by Japan). *Geogr. Nat. Resour.* 33:74–82. doi:10.1134/S187537281201012X.
- Güler, C., M. A. Kurt, M. Alpaslan, y C. Akbulut. 2012. Assessment of the impact of anthropogenic activities on the groundwater hydrology and chemistry in Tarsus coastal plain (Mersin, SE Turkey) using fuzzy clustering, multivariate statistics and GIS techniques. *J. Hydrol.* 414–415:435–451. doi:10.1016/j.jhydrol.2011.11.021.
- Gupta, H. V., T. Wagener, y L. Yuqiong. 2008. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrol. Process.* 22:3802–3813. doi:10.1002/hyp.6989.
- Gutiérrez-López A., R. Becerra, J. Esparza, J. López, y P. Gallardo. 2009. Estimación de caudales en cuencas no aforadas, empleando modelos estocásticos y multivariados. *Revista Aqua-LAC.* 1: 72-81.
- Gutiérrez-López, A., y A. I. Ramírez. 2005. Predicción hidrológica mediante el Método de la Avenida Índice para dos poblaciones. *Ing. Hidráulica en México* 20:37–47.
- Gutiérrez-Pulido, H., y R. De la Vara-Salazar. 2008. Análisis y diseño de experimentos. 2nd ed. México.
- Haddad, K., A. Rahman, y J. R. Stedinger. 2012. Regional flood frequency analysis using Bayesian generalized least squares: A comparison between quantile and parameter regression techniques. *Hydrol. Process.* 26:1008–1021. doi:10.1002/hyp.8189.
- Haines, A.T., B.L. Finlayson, y T.A. McMahon. 1988. A global classification of river regimes. *Appl. Geogr.* 8, 255–272.
- Harris, N. M., A. M. Gurnell, D. M. Hannah, y G. E. Petts. 2000. Classification of river regimes: A context for hydroecology. *Hydrol. Process.* 14:2831–2848.
- Hartigan, J.A., y M.A. Wong. 1979. Algorithm AS 136.A K-means clustering algorithm. *Applied Statistics.* 28(1): 100–108.
- Heřmanovský, M., y P. Pech. 2013. Selection of catchment descriptors for the physical similarity approach. Part II: Application. *Soil Water Res.* 8:186–194.

- Hooshyar, M., D. Wang, S. Kim, S. C. Medeiros, and S. C. Hagen. 2016. topographic curvature and k -means clustering of contours. *Water Resour. Res.* 52:1–22. doi:10.1002/2015WR018479.
- Hosking, J. R. M., y J. R. Wallis. 1993. Some statistics useful in regional frequency analysis. *Water Resources Research.* 29(1): 271–281.
- Hosking, J. R. M., y J. R. Wallis. 1997. *Regional Frequency Analysis: An approach based on L-moments.* 2nd. ed. Cambridge University Press. Cambridge, New York.
- Huang, J. C., C. K. Yu, J. Y. Lee, L. W. Cheng, T. Y. Lee, y S. J. Kao. 2012. Linking typhoon tracks and spatial rainfall patterns for improving flood lead time predictions over a mesoscale mountainous watershed. *Water Resour. Res.* 48:W09540. doi:10.1029/2011WR011508.
- Hynds, P., B. D. Misstear, L. W. Gill, y H. M. Murphy. 2014. Groundwater source contamination mechanisms: Physicochemical profile clustering, risk factor analysis and multivariate modelling. *J. Contam. Hydrol.* 159:47–56. doi:10.1016/j.jconhyd.2014.02.001.
- ICI-RAFT. 2013. ICI-RAFT Regional Analysis of Frequency Tool Quick Start Guide. Version. International Center for Integrated Water Resources Management (ICIWaRM) Institute for Water Resources (IWR) United States Army Corps of Engineers (USACE). <https://www.iwr.usace.army.mil> (Accessed 20 Apr 2017).
- Ilorme, F., y V. W. Griffis. 2013. A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation. *J. Hydrol.* 492:151–162. doi:10.1061/j.hydrol.2013.03.045.
- INEA. 1997. *Guía de diseño de pequeñas centrales hidroeléctricas.* Instituto de Ciencias Nucleares y Energías Alternativas. República de Colombia. Ministerio de Minas y Energía. Academic Press, Santafé de Bogotá, Colombia.
- Isik, S., V. P. Singh, y F. Asce. 2008. Hydrologic Regionalization of Watersheds in Turkey. *J. Hydrol. Eng.* 13:824–834. doi: 10.1061/(ASCE)1084-0699(2008)13:9(824).

J. Lilienthal, R. Fried, y A. Schumann. 2018. Homogeneity testing for skewed and cross-correlated data in regional flood frequency analysis. *Journal of Hydrology*. 556:557-571. doi:10.1016/j.jhydrol.2017.10.056.

Jain, A.K., y R. C. Dubes. 1988. *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliff, NJ 07632.USA. Michigan State University. ISBN: 0-13-022278-X.

Johnston, C. A., y B. A. Shmagin. 2008. Regionalization, seasonality, and trends of streamflow in the US Great Lakes Basin. *J. Hydrol.* 362:69–88. doi:10.1016/j.jhydrol.2008.08.010.

Kahya, E., S. Kalaycı, y T. C. Piechota. 2008. Streamflow Regionalization : Case Study of Turkey. *J. Hydrol. Eng.* 13:205–214. doi:10.1061/(Asce)1084-0699(2008)13:4(205).

Kar, A. K., A. K. Lohani, N. K. Goel, y G. P. Roy. 2015. Rain gauge network design for flood forecasting using multi-criteria decision analysis and clustering techniques in lower Mahanadi river basin, India. *J. Hydrol. Reg. Stud.* 4:313–332. doi: 10.1016/j.ejrh.2015.07.003.

Kaufman, L., y P. J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.

Khattree, R., y D. N. Naik. 2002. Andrews plots for multivariate data: Some new suggestions and applications. *J. Stat. Plan. Inference* 100:411–425.

Kirk, R., y R. E. Williams. 1985. Discrimination of Groundwater Sources Using Cluster Analysis , MANOVA , Canonical Analysis and Discriminant Analysis. 21:1149–1156.

Kite, G.W. 1977. *Frequency and Risk Analyses in Hydrology*. Water Resources Publications, Fort Collins. Colorado, USA. ISBN: 0-918334-64-0.

Koutroulis, A. G., I. K. Tsanis, y I. N. Daliakopoulos. 2010. Seasonality of floods and their hydrometeorologic characteristics in the island of Crete. *J. Hydrol.* 394:90–100. doi:10.1016/j.jhydrol.2010.04.025.

Kult, J. M., L. M. Fry, A. D. Gronewold, y W. Choi. 2014. Regionalization of hydrologic response in the Great Lakes basin: Considerations of temporal scales of analysis. *J. Hydrol.* 519:2224–2237. doi:10.1016/j.jhydrol.2014.09.083.

Laaha, G., y G. Blöschl. 2006. A comparison of low flow regionalisation methods–catchment grouping. *J. Hydrol.* 323:193–214.

Lachance, M., y B. Bernard. 1979. Characterization of the Water Quality in the Saint Lawrence River: Determination of Homogeneous Zones by Correspondence Analysis. *Water Resour. Res.* 15:1451–1462.

Lang, M., T. B. M. J. Ouarda, y B. Bobée. 1999. Towards operational guidelines for over-threshold modelling. *J. Hydrol.* 225:103–117.

Lee, H., N. McIntyre, H. Wheeler, y A. Young. 2005. Selection of conceptual models for regionalisation of the rainfall-runoff relationship. *J. Hydrol.* 312:125–147. doi: 10.1016/j.jhydrol.2005.02.016

Lim, Y. H., M. Asce, y D. L. Voeller. 2009. Regional Flood Estimations in Red River Using L-Moment-Based Index-Flood and Bulletin 17B Procedures. *J. Hydrol. Eng.* 14:1002–1016. doi: 10.1061/(ASCE)HE.1943-5584.0000102.

Lin, G. F., L. H. Chen, y S. C. Kao. 2005. Development of regional design hyetographs. *Hydrol. Process.* 19:937–946. doi:10.1002/hyp.5550

Lin, G. F., y M. C. Wu. 2007. A SOM-based approach to estimating design hyetographs of ungauged sites. *J. Hydrol.* 339:216–226. doi: 10.1016/j.jhydrol.2007.03.016.

Lin, G.-F., M. C. Wu, G. R. Chen, y S. J. Liu. 2010. Construction of design hyetographs for locations without observed data. *Hydrol. Process.* 24:481–491. doi:10.1002/hyp.7500.

Lobo, D., D. Gabriels, F. Ovalle, F. Santibañez, M. Moyano, R. Aguilera, R. Pizarro, C. Sanguesa, y N. Urrea. 2006. Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América Latina y el Caribe. ISBN: 9290890886.

Machiwal, D., y M. K. Jha. 2012. *Hydrologic Time Series Analysis: Theory and Practice*. Springer, New York, USA.

- Malekinezhad, H., H. P. Nachtnebel, y A. Klik. 2011. Comparing the index-flood and multiple-regression methods using L-moments. *Phys. Chem. Earth* 36:54–60. doi:10.1016/j.pce.2010.07.013
- Mamun, A. A., A. Hashim, y J. I. Daoud. 2010. Regionalisation of low flow frequency curves for the Peninsular Malaysia. *J. Hydrol.* 381:174–180. doi:10.1016/j.jhydrol.2009.11.039.
- Mardia, KV. 1972. *Statistics of Directional Data*. Academic Press, San Diego, Calif.
- Marsal, J.R. 1974. Presas pequeñas notas sobre diseño y construcción. Boletín 326. Instituto de Ingeniería. Universidad Nacional Autónoma de México. Ciudad universitaria, México 20, D.F.
- Masud, M. B., M. N. Khaliq, y H. S. Wheater. 2015. Analysis of meteorological droughts for the Saskatchewan River Basin using univariate and bivariate approaches. *J. Hydrol.* 522:452–466. doi:10.1016/j.jhydrol.2014.12.058.
- Mather, A. L., y R. L. Johnson. 2015. Event-based prediction of stream turbidity using a combined cluster analysis and classification tree approach. *J. Hydrol.* 530:751–761. doi:10.1016/j.jhydrol.2015.10.032.
- Matteau, M., A. A. Assani, y M. Mesfioui. 2009. Application of multivariate statistical analysis methods to the dam hydrologic impact studies. *J. Hydrol.* 371:120–128. doi:10.1016/j.jhydrol.2009.03.022.
- Maul, A., y El-Shaarawi. 1993. Stochastic Models Applied to Cluster Analysis of Censored Water Quality Data. *Water Resour. Res.* 29:2705–2711.
- McNeil, V. H., M. E. Cox, y M. Preda. 2005. Assessment of chemical water types and their spatial variation using multi-stage cluster analysis, Queensland, Australia. *J. Hydrol.* 310:181–200. doi:10.1016/j.jhydrol.2004.12.014.
- Merz, R., y G. Blöschl. 2004. Regionalisation of catchment model parameters. *J. Hydrol.* 287:95–123. doi:10.1016/j.jhydrol.2003.09.028.

Mishra, B. K., K. Takara, y Y. Tachikawa. 2009. Integrating the NRCS Runoff Curve Number in Delineation of Hydrologic Homogeneous Regions. *J. Hydrol. Eng.* 14:1091–1097. doi: 10.1061/(ASCE)HE.1943-5584.0000101.

Mishra, B., Y. Tachikawa, y K. Takara. 2007. Suitability of sample size for identifying distribution function in regional frequency analysis. *Annuals of Disas. Prev. Res. Inst., Kyoto Univ.* N° 50 B, 2007. doi:10.1061/(ASCE)HE.1943-5584.0000101.

Naranjo-Toledo, R. C. A. 2011. Análisis regional de frecuencia mediante el método de los L-momentos en las regiones de Valparaíso y Metropolitana para la generación de cartografía probabilística de sequía meteorológica. *Prensa Académica Universidad de Chile.*

Nathan, R. J., y T. A. McMahon. 1990. Identification of homogeneous regions for the purposes of regionalisation. *J. Hydrol.* 121:217–238.

Nathan, R. J., y T. A. McMahon. 1992. Estimating Low Flow Characteristics in Ungauged Catchments. *Water Resour. Manag.* 6:85–100.

Nava-Mastache, A., y H. García-Gutiérrez. 2013. Selección y dimensionamiento de turbinas hidráulicas para centrales hidroeléctricas. Disponible en: http://www.ingenieria.unam.mx/~deptohidraulica/publicaciones/pdf_publicaciones/SELEC_yDIMENSIONAMIENTOdeTURBINAS.pdf (Accessed Jul 30, 2018)

Neykov, N. M., P. N. Neytchev, P. H. A. J. M Van Gelder, y V. K. Todorov. 2007. Robust detection of discordant sites in regional frequency analysis. *Water Resour. Res.* 43(6): W06417. doi:10.1029/2006WR005322.

Nguyen, T. T., A. Kawamura, T. N. Tong, N. Nakagawa, H. Amaguchi, y R. Gilbuena. 2015. Clustering spatio-seasonal hydrogeochemical data using self-organizing maps for groundwater quality assessment in the Red River Delta, Vietnam. *J. Hydrol.* 522:661–673. doi:10.1016/j.jhydrol.2015.01.023.

Nobert, J., M. Mugo, y H. Gadain. 2014. Estimation of design floods in ungauged catchments using a regional index flood method. A case study of Lake Victoria Basin in Kenya. *Phys. Chem. Earth* 67–69:4–11. doi:10.1016/j.pce.2014.02.001.

Nourani, V., M. T. Alami, y F. D. Vousoughi. 2013. Hybrid of SOM-Clustering Method and Wavelet-ANFIS Approach to Model and Infill Missing Groundwater Level Data. *J. Hydrol. Eng.*:1–19. doi:10.1061/(ASCE)HE.1943-5584.0001398.

Núñez, J., K. Verbist, J. Ramírez H., y M. Hallack-Alegría. 2010. Guía metodológica para la aplicación del Análisis Regional de Frecuencia de Sequías basado en L-momentos y resultados de aplicación en América Latina. Available from: http://www.cazalac.org/2015/fileadmin/templates/documentos/Guia_Metodologia_Atlas_de_Sequia.pdf (Accessed May 20, 2017).

Núñez, J., M. Hallack-Alegría, y M. Cadena. 2016. Resolving regional frequency analysis of precipitation at large and complex scales using a bottom-up approach: The Latin America and the Caribbean Drought Atlas. *J. Hydrol.* 538:515–538. doi:10.1016/j.jhydrol.2016.04.025.

O'Shea, B., y J. Jankowski. 2006. Detecting subtle hydrochemical anomalies with multivariate statistics: An example from “homogeneous” groundwaters in the Great Artesian Basin, Australia. *Hydrol. Process.* 20:4317–4333. doi:10.1002/hyp.6177.

Ojeda-Ramírez, M. M., F. Velasco-Luna, C. Cruz-López, y P. Tapia-Blázquez. 2011. Metodología estadística aplicada a las finanzas públicas. Veracruz, México. ISBN: 978-607-00-5322-1.

Okur, A. y U. Sorman. 2000. Application of at site regional regional frequency analyses by using the L-moments technique. *Teknik Dergi.* 11: 2199-2218.

Ouarda, T. B. M. J., C. Girard, G. S. Cavadias, y B. Bobée. 2001. Regional flood frequency estimation with canonical correlation analysis. *J. Hydrol.* 254:157–173.

Ouarda, T. B. M. J., F. Ashkar, y N. El-Jabi. 1993. Peaks over threshold model for seasonal flood variations. In: *Engineering Hydrology* (Ed.) C. Y. Kuo, p 341–346. ASCE Publications, New York, USA.

Ouarda, T. B. M. J., J. M. Cunderlik, A. St-Hilaire, M. Barbet, P. Bruneau, y B. Bobeé. 2006. Data-based comparison of seasonality-based regional flood frequency methods. *J. Hydrol.* 330:329–339. doi:10.1016/j.jhydrol.2006.03.023.

Ouarda, T. B. M. J., K. M. Bâ, C. Diaz-Delgado, A. Cârsteanu, K. Chokmani, H. Gingras, E. Quentin, E. Trujillo, y B. Bobée. 2008. Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study. *J. Hydrol.* 348:40–58.

Oudin, L., V. C. Andréassian, C. Perrin, C. Michel, y N. Le Moine. 2008. Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.* 44:W03413. doi:10.1029/2007WR006240.

Oueslati, O., A. M. De Girolamo, A. Abouabdillah, T. R. Kjeldsen, y A. Lo Porto. 2015. Classifying the flow regimes of Mediterranean streams using multivariate analysis. *Hydrol. Process.* 29:4666–4682. doi: 10.1002/hyp.10530.

Ouyang, R., L. Ren, W. Cheng, y C. Zhou. 2010. Similarity search and pattern discovery in hydrological time series data mining. *Hydrol. Process.* 24:1198–1210. doi:10.1002/hyp.7583.

Özçelik, C., y E. Benzedden. 2010. Regionalization approaches for the periodic parameters of monthly flows: A case study of Ceyhan and Seyhan River basins. *Hydrol. Process.* 24:3251–3269. doi: 10.1002/hyp.7752

Palacios-Fonseca, A. A., N. Peña-García, E. A. Cervantes-Carretero, A. Güitrón de los Reyes, y M. López-Pérez. 2017. Bases para un Centro Mexicano en Innovación de Energía Hidroeléctrica. Primera ed. México.

Panda, U. C., S. K. Sundaray, P. Rath, B. B. Nayak, y D. Bhatta. 2006. Application of factor and cluster analysis for characterization of river and estuarine water systems - A case study: Mahanadi River (India). *J. Hydrol.* 331:434–445. doi:10.1016/j.jhydrol.2006.05.029.

Parajka, J., G. Blöschl, y R. Merz. 2007. Regional calibration of catchment models: Potential for ungauged catchments, *Water Resour. Res.* 43:W06406. doi:10.1029/2006WR005271.

Parasuraman, K., y A. Elshorbagy. 2007. Cluster-Based Hydrologic Prediction Using Genetic Algorithm-Trained Neural Networks. *J. Hydrol. Eng.* 12:52–62. doi:10.1061/(ASCE)1084-0699(2007)12:1(52).

Paris, M., y G. Zucarelli. 2004. Regionalización de caudales. Propuesta metodológica para la identificación de regiones homogéneas. *Ing. Hidráulica en México* 19:5–19.

Peterson, H. M., J. L. Nieber, y R. Kanivetsky. 2011. Hydrologic regionalization to assess anthropogenic changes. *J. Hydrol.* 408:212–225. doi: 10.1016/j.jhydrol.2011.07.042.

Ragettli, S., J. Zhou, H. Wang, C. Liu y L. Guo. 2017. Modeling flash floods in ungauged mountain catchments of China: A decision tree learning approach for parameter regionalization. *Journal of hydrology.* Vol 555, pp. 330-346. doi:10.1016/j.jhydrol.2017.10.031.

Rahman, M. M., S. Sarkar, M. R. Najafi, y R. K. Rai. 2013. Regional extreme rainfall mapping for Bangladesh using L-moment technique. *J. Hydrol. Eng.* 18:604–615. doi:10.1061/(ASCE)HE.1943-5584.0000663.

Ramos, M.C. 2001. Divisive and hierarchical clustering techniques to analyse variability of rainfall distribution patterns in a Mediterranean region. *Atmospheric Research.* 57(2): 123–138.

Rao, A. R., y K. H. Hamed. 1997. Regional Frequency Analysis of Wabash River Flood Data by L-Moments. *J. Hydrol. Eng.* 2:169–179.

Rao, A.R., y K.H. Hamed. 2000. Regional homogeneity and regionalization. In: *Flood Frequency Analysis.* (Ed.) CRC. pp. 47- 52. Press LLC, USA.

Rao, A. R., y V. V. Srinivas. 2006. Regionalization of watersheds by hybrid-cluster analysis. *J. Hydrol.* 318:37–56. doi:10.1016/j.jhydrol.2005.06.003.

Reed, D. W., D. Jakob, y A. J. Robson. 1999. Selecting a pooling group. In: Robson, A.J. y Reed, D.W. Ed. *Statistical procedures for Flood Frequency Estimation.* Flood Estimation Handbook. Institute of Hydrology, Wallingford, UK.

Requena, A. I., F. Chebana, y L. Mediero. 2016. A complete procedure for multivariate index-flood model application. *J. Hydrol.* 535:559–580. doi:10.1016/j.jhydrol.2016.02.004.

Riggs, H.C., y K.D. Harvey. 1990. Temporal and spatial variability of streamflow. In: Wolman, M.G., Riggs, H.C. Eds. *Surface Water Hydrology*. Geological Society of America, Boulder, CO. pp. 81–96.

Rosenthal, W., y J. Dozier. 1996. Automated mapping of montane snow cover at subpixel resolution from the Landsat Thematic Mapper. *Water Resour. Res.* 32:115–130.

Rousseeuw, P.J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics.* 20:53–65.

Saf, B. 2009. Regional Flood Frequency Analysis Using L Moments for the Buyuk and Kucuk Menderes River Basins of Turkey. *J. Hydrol. Eng.* 14:783–794. doi: 10.1061/(ASCE)1084-0699(2006)11:5(392).

Saharia, M., P. E. Kirstetter, H. Vergara, J. J. Gourley, y Y. Hong. 2017. Characterization of floods in the United States. *Journal of Hydrology.* 548:524–535. doi:10.1016/j.jhydrol.2017.03.010.

Sahin, S., y K. H. Cigizoglu. 2012. The sub-climate regions and the sub-precipitation regime regions in Turkey. *J. Hydrol.* 450–451:180–189. doi:10.1016/j.jhydrol.2012.04.062.

Sanborn, S. C., y B. P. Bledsoe. 2006. Predicting streamflow regime metrics for ungauged streams in Colorado, Washington, and Oregon. *J. Hydrol.* 325:241–261.

Sánchez-Bribiesca, J.L. y O. A. Fuentes-Mariles. 1997. Dimensionamiento de una presa derivadora. Capítulo 4 del Manual de Ingeniería de Ríos. Series del Instituto de Ingeniería número 595. Instituto de Ingeniería UNAM. Instituto de ingeniería. Universidad Autónoma de México. Ciudad Universitaria. Coyoacán 04510. México D.F.

Santos, J. F., I. Pulido-Calvo, y M. M. Portela. 2010. Spatial and temporal variability of droughts in Portugal. *Water Resour. Res.* 46:1–13. doi:10.1029/2009WR008071.

Santos, J. F., M. M. Portela, y I. Pulido-Calvo. 2013. Dimensionality reduction in drought modelling. *Hydrol. Process.* 27:1399–1410. doi:10.1002/hyp.9300.

Sarhadi, A., y R. Modarres. 2011. Flood seasonality-based regionalization methods: A data-based comparison. *Hydrol. Process.* 25:3613–3624. doi: 10.1002/hyp.8088.

Sarmadi, F., y M. Azmi. 2016. Regionalizing Mean Air Temperature in Iran by Multivariate Analysis and L-Moment Methods. *J. Hydrol. Eng.* 21:1–8. doi: 10.1061/(ASCE)HE.1943-5584.0001280.

Satyanarayana, P., y V. V. Srinivas. 2011. Regionalization of precipitation in data sparse areas using large scale atmospheric variables - A fuzzy clustering approach. *J. Hydrol.* 405:462–473. doi:10.1016/j.jhydrol.2011.05.044.

Schrnie, S., y X. Cai. 2014. Prediction of regional streamflow frequency using model tree ensembles. *J. Hydrol.* 517:298–309. doi:10.1016/j.jhydrol.2014.05.029.

Schumann, H. A. 2011. *Flood Risk Assessment and Management: How to Specify Hydrological Loads, Their Consequences and Uncertainties.* Springer Dordrecht Heidelberg London New York. doi:10.1007/978-9-481-9917-4.

She, D. X., J. Xia, D. Zhang, A. Z. Ye, y A. Sood. 2014. Regional extreme-dry-spell frequency analysis using the L-moments method in the middle reaches of the Yellow River Basin, China. *Hydrol. Process.* 28:4694–4707. doi:10.1002/hyp.9930.

Shi, P., X. Chen, S. M. Qu, Z. C. Zhang, y J. I. Ma. 2010. Regional frequency analysis of low flow based on L moments: Case study in karst area, Southwest China. *Journal of Hydrologic Engineering.* 15: 370–377.

Sine, A., Ayalew, S., 2004. Identification and Delineation of Hydrological Homogeneous Regions-The case study of Blue Nile River Basin, In: *Proceedings, Lake Abaya Research Symposium.*

Singh, R., S. A. Archfield, y T. Wagener. 2014. Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments - A comparative hydrology approach. *J. Hydrol.* 517:985–996. doi:10.1016/j.jhydrol.2014.06.030.

Singh, S. K., A. Bárdossy, J. Götzinger, y K. P. Sudheer. 2012. Effect of spatial resolution on regionalization of hydrological model parameters. *Hydrol. Process.* 26:3499–3509. doi:10.1016/j.jhydrol.2014.06.030.

- Singh, S. K., H. McMillan, A. Bárdossy, y F. Chebana. 2016. Non-parametric catchment clustering using the data depth function. *Hydrol. Sci. J.* 61:2649-2667. doi: 10.1080/02626667.2016.1168927
- Skaugen, T., y T. Væringstad. 2005. A methodology for regional flood frequency estimation based on scaling properties. *Hydrol. Process.* 19:1481–1495. doi: 10.1002/hyp.5588.
- Sohoulande Djebou, D. C., V. P. Singh, y O. W. Frauenfeld. 2014. Analysis of watershed topography effects on summer precipitation variability in the southwestern United States. *J. Hydrol.* 511:838–849. doi:10.1016/j.jhydrol.2014.02.045.
- Soltani, S. M., y Reza. 2006. *Journal of Spatial Hydrology* *Journal of Spatial Hydrology*. *J. Spat. Hydrol.* 6:1–14. doi:10.1017/CBO9780511806049.
- Thomas, D.M., y M. A. Benson. 1970. Generalization of streamflow characteristics from drainage-basin characteristics. US Geological Survey Water-Supply Paper 1975. US Government Printing Office, Washington, DC.
- Tongal, H., y B. Sivakumar. 2017. Cross-entropy clustering framework for catchment classification. *J. Hydrol.* 552:433–446. doi:10.1016/j.jhydrol.2017.07.005.
- Trefry, C. M., D. W. W. Jr, y D. Johnson. 2006. for the State of Michigan. *J. Hydrol. Eng.* 10:437–449. Doi: 10.1061/(ASCE)1084-0699(2005)10:6(437)
- Tsakiris, G., I. Nalvantis, y G. Cavadias. 2011. Regionalization of low flows based on canonical correlation analysis. *Advances in WaResources* 34(7): 865-872.
- Tukey, J.W. 1977. *Exploratory Data Analysis*. Addison-Wesley.
- Vargas, B. J., C. L. De la Fuente, A. Gutiérrez-López, y J.L. Arumi. 2009. Análisis geostadístico para describir la distribución espacial de la precipitación de una cuenca Patagónica. *Revista Aqua-LAC*. 1: 72-81.
- Veza, P., C. Comoglio, M. Rosso, y A. Viglione. 2010. Low flows regionalization in North-Western Italy. *Water Resources Management* 24:4049-4074.

Viglione, A., Laio, F., y P. Claps. 2007. A comparison of homogeneity tests for regional frequency analysis. *Water Resour. Res.* 43:1-10. doi:10.1029/2006WR005095.

Wafa Chouaiba, Peter V.Caldwell y Younes Alila. 2018. Regional variation of flow duration curves in the eastern United States: Process-based analyses of the interaction between climate and landscape properties. *Journal of Hydrology*. doi:10.1016/j.jhydrol.2018.01.037.

Wallis, J., M. Schaefer, B. Barker, y G. Taylor. 2007. Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington States. *Hydrol. Earth Syst. Sci.* 11:415- 442.

Wandle, Jr., S.W. 1977. Estimating the magnitude and frequency of floods on natural-flow streams in Massachusetts. US Geological Survey Water Resources Investigations Report 77-39.

Ward, J. H. 1963. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* 58:236.

Wazneh, H., F. Chebana, y T. B. M. J. Ouarda. 2013. Depth-based regional index-flood model. *Water Resour. Res.* 49:7957-7972. doi:10.1002/2013WR013523.

Wazneh, H., F. Chebana, y T. B. M. J. Ouarda. 2015. Delineation of homogeneous regions for regional frequency analysis using statistical depth function. *J. Hydrol.* 521:232-244. doi:10.1016/j.jhydrol.2014.11.068.

Wiltshire, S.E. 1986a. Regional flood frequency analysis I: Homogeneity statistics. *Hydrol. Sci. J.* 31:321-333.

Wiltshire, S.E. 1986b. Identification of homogeneous regions for flood frequency analysis. *J. Hydrol.* 84:287-302.

Wong, H., y B. Q. Hu. 2013. Application of interval clustering approach to water quality evaluation. *J. Hydrol.* 491:1-12. doi:10.1016/j.jhydrol.2013.03.009.

Yang, T, Q. Shao, Z-C Hao, X. Chen , Z. Zhang, C-Y. Xu, y L. Sun. 2010. Regional frequency analysis and spatio-temporal pattern characterization of rainfall extremes in the Pearl River Basin. *Journal of Hydrology* 380:386-405.

Yokoo, Y., y S. Kazama. 2012. Numerical investigations on the relationships between watershed characteristics and water balance model parameters: Searching for universal relationships among regional relationships. *Hydrol. Process.* 26:843–854. doi:10.1002/hyp.8299.

Yoo, C., H. Ku, y K. Kim. 2011. Use of a Distance Measure for the Comparison of Unit Hydrographs: Application to the Stream Gauge Network Optimization. *J. Hydrol. Eng.* 16:880–890. doi:10.1061/(ASCE)HE.1943-5584.0000393.

Yoo, J., H. H. Kwon, T. W. Kim, y J. H. Ahn. 2012. Drought frequency analysis using cluster analysis and bivariate probability distribution. *J. Hydrol.* 420–421:102–111. doi:10.1016/j.jhydrol.2011.11.046

Yu, P., y T. Yang. 1996. Synthetic regional flow duration curve for Southern Taiwan. 10:373–391.

Yu, P.-S., T.-C. Yang, y C.-W. Liu. 2002. A regional model of low flow for southern Taiwan. *Hydrol. Process.* 16:2017–2034. doi: 10.1002/hyp.399.

Zhang, Y., y F. H. S. Chiew. 2009. Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resour. Res.* 45:1-13. doi:10.1029/2008WR007504.

Zrinji, Z., y D. H. Burn. 1996. Regional flood frequency with hierarchical region of influence. *J. W. Res. Plann. Manag.* 122: 245-252.

ANEXO I

Características de la cuenca

Geomorfología de la cuenca

N	VARIABLE	UNIDAD	AUTOR(ES)
MORFOMETRÍA DE LA CUENCA			
1	Área (A) <i>Area (A)</i>	km ²	Acreman y Sinclair (1986) Acreman y Werritty (1987) Ahn y Palmer (2016) Álvarez–Olguín <i>et al.</i> (2011) Brath <i>et al.</i> (2009) Dodangeh <i>et al.</i> (2014) Gubareva (2012) Heřmanovský y Pech (2013) Kult <i>et al.</i> (2014) Laaha y Blöschl (2006) Malekinezhad <i>et al.</i> (2011) Mamun <i>et al.</i> (2010) Nathan y McMahon (1990) Nobert <i>et al.</i> (2014) Ouarda <i>et al.</i> (2001) Ouarda <i>et al.</i> (2006) Ouarda <i>et al.</i> (2008) Paris y Zucarelli (2004) Requena <i>et al.</i> (2016) Sanborn y Bledsoe (2006) Sarhadi y Modarres (2011) Schnier y Cai (2014) Singh <i>et al.</i> (2014) Singh <i>et al.</i> (2016) Skaugen y Væringstad (2005)
2	Perímetro de la cuenca (P) <i>Basin perimeter (P)</i>	m	Álvarez–Olguín <i>et al.</i> (2011) Gubareva (2012) Nobert <i>et al.</i> (2014) Requena <i>et al.</i> (2016) Yu <i>et al.</i> (2002)
3	Relación de compacidad (C) <i>Basin compactness (C)</i> $C = \frac{A / P^2}{100}$	-	Ahn y Palmer (2016) Kult <i>et al.</i> (2014) Malekinezhad <i>et al.</i> (2011) Rahman <i>et al.</i> (2013)
4	Elongación de la cuenca (E) Diámetro de un círculo que contenga la misma magnitud que el área de la cuenca $\text{Elongación} = 2\sqrt{A / \pi} / L$	m ⁻¹	Brath <i>et al.</i> (2009) Rahman <i>et al.</i> (2013) Sanborn y Bledsoe (2006)

	<i>Basin elongation (E)</i> <i>Diameter of a circle that contains the same magnitude as the area of the basin</i>		
5	Parámetro de forma <i>Shape parameter</i> $B = P/A$	km ⁻¹	Álvarez-Olguín <i>et al.</i> (2011) Gubareva (2012)
6	Parámetro de forma de Horton <i>Horton shape parameter</i> $BH = A/L^2$	-	Brath <i>et al.</i> (2009) Rahman <i>et al.</i> (2013)
7	Longitud de la cuenca (L) <i>Basin length</i>	m	Álvarez-Olguín <i>et al.</i> (2011) Skaugen y Væringstad (2005)
8	Ancho máximo de la cuenca (W_{max}) <i>Maximum width of the basin</i>	m	Álvarez-Olguín <i>et al.</i> (2011)
9	Ancho medio de la cuenca (W_{mean}) <i>Mean width of the basin</i>	m	Álvarez-Olguín <i>et al.</i> (2011)
10	Tiempo de concentración (T_c) <i>Time concentration</i>	s	Brath <i>et al.</i> (2009) Álvarez-Olguín <i>et al.</i> (2011) Requena <i>et al.</i> (2016)
11	Aspecto promedio <i>Average aspect</i>	-	Sanborn y Bledsoe (2006) Núñez <i>et al.</i> (2016)
12	Coseno (Aspecto) <i>Mean cosine (Basin aspect)</i>	-	Ahn y Palmer (2016)
13	Seno (Aspecto) <i>Mean sine (Basin aspect)</i>	-	Ahn y Palmer (2016)
14	Aspecto cara al norte Varía de -1 a 1. El valor de 1 significa que la cuenca hidrográfica está orientada / drenando hacia el norte, valor de -1 significa que la cuenca hidrográfica está enfrentando / drenando hacia el sur <i>Face North</i> <i>Ranges from -1 to 1. Value of 1 means watershed is facing/draining due north, value of -1 means watershed is facing/ draining due south</i>	-	Ahn y Palmer (2016); Kult <i>et al.</i> (2014) Singh <i>et al.</i> (2014)
15	Aspecto cara este Varía de -1 a 1. El valor de 1 significa que la cuenca hidrográfica está orientada / drenando hacia el este, el valor de -1 significa que la cuenca hidrográfica está orientada hacia / hacia el oeste. <i>Face aspect east</i>	o	Ahn y Palmer (2016) Kult <i>et al.</i> (2014) Singh <i>et al.</i> (2014)

	<i>Ranges from -1 to 1. Value of 1 means watershed is facing/draining due east, value of -1 means watershed is facing/draining due west</i>		
CORRIENTES / RED DE DRENAJE			
16	Longitud de la corriente principal (S_{len}) <i>Main river length</i>	km	Acreman y Sinclair (1986) Álvarez–Olguín <i>et al.</i> (2011) Brath <i>et al.</i> (2009) Malekinezhad <i>et al.</i> (2011) Nathan y McMahon (1990) Ouarda <i>et al.</i> (2001) Ouarda <i>et al.</i> (2006) Requena <i>et al.</i> (2016) Yu <i>et al.</i> , 2002
17	Longitud promedio de la corriente (\overline{S}_{len}) <i>Average basin length</i>	m	Singh <i>et al.</i> (2016)
18	Longitud total de la red de drenaje (L_{TOR}) <i>Total length of the drainage network</i>	km	Gubareva (2012)
19	Ancho promedio de las corrientes (\overline{W}_L) <i>Average width stream</i>	m	Singh <i>et al.</i> (2016)
20	Ancho máximo, mínimo y medio (W) <i>Maximun, minimun and mean width stream</i>	m	Paris y Zucarelli (2004)
21	Profundidad máxima, mínimo y medio del río <i>Maximun, minimun and mean river depth</i>	m	Paris y Zucarelli (2004)
22	Orden de corrientes Horton-Strahler (<i>order</i>) <i>Horton-Strahler stream order</i>	-	Brath <i>et al.</i> (2009) Gubareva (2012) Kult <i>et al.</i> (2014)
23	Porcentaje de longitud de cauces que corresponden al orden número 1, de la clasificación de Strahler ($order_{s_1}$) <i>Percentage of length corresponding to order number 1, of the Strahler classification</i>	-	Singh <i>et al.</i> (2014)
24	Porcentaje de longitud de cauces que corresponden al orden número 2, de la clasificación de Strahler ($order_{s_2}$) <i>Percentage of length corresponding to order number 2, of the Strahler classification</i>	-	Singh <i>et al.</i> (2014)

25	<p>Porcentaje de longitud de cauces que corresponden al orden número 4, de la clasificación de Strahler ($order_{s_2}$)</p> <p><i>Percentage of length corresponding to order number 4, of the Strahler classification</i></p>	-	Singh <i>et al.</i> (2014)
26	<p>Máximo orden de corrientes ($Order_{MAX}$)</p> <p><i>Maximun stream order</i></p>	-	Ahn y Palmer (2016)
27	<p>Frecuencia de corrientes ($SFREQ$)</p> <p>Número total de uniones dividida por el área</p> <p><i>Stream Frequency</i></p> <p><i>Total number of junctions divided by area</i></p>	uniones* km ⁻²	Acreman y Sinclair (1986)
28	<p>Densidad de corrientes (D)</p> $D = \frac{L_{TOT}}{A}$ <p><i>Stream density</i></p> <p><i>Streams length (km)/Basin area (km²)</i></p>	km ⁻¹	Ahn y Palmer (2016) Álvarez–Olguín <i>et al.</i> (2011) Kult <i>et al.</i> (2014) Laaha y Blöschl (2006) Nathan y McMahon (1990) Schnier y Cai (2014) Rahman <i>et al.</i> (2013) Sanborn y Bledsoe (2006) Sarhadi y Modarres (2011) Singh <i>et al.</i> (2014)
29	<p>Densidad (frecuencia) ($Sfreq$)</p> <p>Número total de uniones de los tres primeros ordenes dividida por el área</p> <p><i>Stream Frequency</i></p> <p><i>Total number of units of the first three orders divided by area</i></p>	uniones* km ²	Gubareva (2012)
30	<p>Media ponderada de la orientación de los arroyos de los tres principales órdenes (Ar)</p> <p><i>Weighted average of the orientation of the streams of the three main orders</i></p>	°	Gubareva (2012)
31	<p>Altitud del final de la corriente principal (ALT_{MAX})</p> <p><i>Altitude of the end of the main current</i></p>	msnm	Ouarda <i>et al.</i> (2008)
32	<p>Elevación máxima de la corriente principal ($Elev_{MAX}$)</p> <p><i>Maximun elevation of the main stream</i></p>	msnm	Singh <i>et al.</i> (2016)
33	<p>Elevación mínima de la corriente principal ($Elev_{MIN}$)</p> <p><i>Minimun elevation of the main stream</i></p>	msnm	Singh <i>et al.</i> (2016)
34	<p>Pendiente del canal ($S1085$)</p> <p>Pendiente del canal al 10 y 85%</p>	m km ⁻¹	Acreman y Sinclair (1986)

	<i>Channel slope</i> <i>stream slope at 10--85% of the river.</i>		
35	Pendiente promedio del canal principal (S_{prom}) <i>Average slope of the main channel</i>		Álvarez–Olguín <i>et al.</i> (2011) Brath <i>et al.</i> (2009) Ouarda <i>et al.</i> (2001) Ouarda <i>et al.</i> (2006) Ouarda <i>et al.</i> (2008) Rahman <i>et al.</i> (2013) Sanborn y Bledsoe (2006) Sarhadi y Modarres (2011) Singh <i>et al.</i> (2016) Skaugen y Væringstad (2005)
36	Pendiente del cauce principal al 75% desde la parte central (S_{75}) <i>Slope of the main channel to 75% from the central part</i>	m km ⁻¹	Nathan y McMahon (1990)
RELIEVE			
37	Latitud ($LATIT$) Latitud del centroide de la cuenca, grados <i>Latitude</i> <i>Latitude of the centroid of the basin, degrees</i>	°	Dikbas <i>et al.</i> (2013) Kult <i>et al.</i> (2014) Nathan y McMahon (1990) Núñez <i>et al.</i> (2016) Ouarda <i>et al.</i> (2001) Paris y Zucarelli (2004)
38	Media ponderada de la latitud de la cuenca (LAT) <i>Weighted average of the latitude of the basin</i>	°	Gubareva (2012)
39	Latitud de la estación de aforo (LAT_G) <i>Latitude of Gage station</i>		Ouarda <i>et al.</i> (2006)
40	Longitud ($LONGIT_C$) Longitud del centroide de la cuenca <i>Longitude</i> <i>Longitude of centroid of the basin</i>	°	Dikbas <i>et al.</i> (2013) Gubareva (2012) Nathan y McMahon (1990) Nobert <i>et al.</i> (2014) Núñez <i>et al.</i> (2016) Ouarda <i>et al.</i> (2001) Paris y Zucarelli (2004)
41	Longitud de la estación de aforo ($LONG_G$) <i>Longitude of gage station</i>		Ouarda <i>et al.</i> (2006)
42	Altitud promedio (ALT) <i>Average altitude of the basin</i>	msnm	Ahn y Palmer (2016) Álvarez–Olguín <i>et al.</i> (2011) Brath <i>et al.</i> (2009) Gubareva (2012) Lebecherel <i>et al.</i> (2016) Nathan y McMahon (1990) Nobert <i>et al.</i> (2014) Núñez <i>et al.</i> (2016)

			Requena et al. (2016) Sanborn y Bledsoe (2006) Sarhadi y Modarres (2011) Singh <i>et al.</i> (2014) Yu <i>et al.</i> (2002)
43	Altitud del centroide de la cuenca (ALT_C) <i>Altitude of centroid of the basin</i>	msnm	Ouarda <i>et al.</i> (2006)
44	Rango de la altitud (R) <i>Altitude range</i>	m	Álvarez-Olguín <i>et al.</i> (2011) Laaha y Blöschl (2006)
45	Elevación mediana de la cuenca ($ELEV_MEDIAN$) <i>Median elevation of the basin</i>	m	Ahn y Palmer (2016)
46	Elevación máxima de la cuenca ($ELEV_MAX$) <i>Maximum elevation of the basin</i>	m	Ahn y Palmer (2016) Laaha y Blöschl (2006) Sarhadi y Modarres (2011) Requena <i>et al.</i> (2016)
47	Elevación mínima de la cuenca ($ELEV_MIN$) <i>Minimum elevation of the basin</i>	m	Ahn y Palmer (2016) Sanborn y Bledsoe (2006) Sarhadi y Modarres (2011) Requena <i>et al.</i> (2016)
48	Elevación de la estación de aforo ($ELEV_G$) <i>Elevation of the gage station</i>	msnm	Laaha y Blöschl (2006) Schnier y Cai (2014)
49	Relieve relativo de la cuenca ($RRMEAN$) $RRMEAN = \frac{ELEV_MEAN - ELEV_MIN}{ELEV_MAX - ELEV_MIN}$ <i>Basin relative relief</i>	-	Ahn y Palmer (2016)
50	Relieve de la cuenca dividido por la longitud de la cuenca (R_L) <i>Basin relief divided by basin length</i>	-	Sanborn y Bledsoe (2006)
PENDIENTE			
51	Pendiente media de la cuenca ($Slope$) <i>Mean slope of the basin</i>	$m\ m^{-1}$ %	Ahn y Palmer (2016); Gubareva (2012) Kult <i>et al.</i> (2014) Laaha y Blöschl (2006) Nobert <i>et al.</i> (2014) Núñez <i>et al.</i> (2016) Requena <i>et al.</i> (2016) Sanborn y Bledsoe (2006) Sarhadi y Modarres (2011)

			Schnier y Cai (2014) Skaugen y Væringstad (2005) Yu <i>et al.</i> (2002)
52	Porcentaje de superficies con pendiente moderada ($Slop_M$) <i>Percentage of surfaces with moderate slope</i>	%	Laaha y Blöschl (2006)
53	Porcentaje de superficies con pendiente alta ($Slop_H$) <i>Percentage of surfaces with high slope</i>	%	Laaha y Blöschl (2006)

Variables climatológicas

N	VARIABLE	UNIDAD	AUTOR(ES)
PRECIPITACIÓN ASOCIADA A UN PERIODO DE RETORNO			
1	Precipitación máxima asociada a un periodo de retorno (P_{max}) <i>Maximum rainfall associated with a return period</i>	mm	Requena <i>et al.</i> (2016)
2	Lluvia: Máximo 2 días de lluvia en un periodo de retorno de 5 años (P_{2D5T}) <i>Maximum 2 day rainfall of 5 yr return period</i>	mm	Acreman y Sinclair (1986)
3	Precipitación máxima de un día de 5 años período de retorno (P_{1D5T}) <i>Maximum one-day rainfall of 5 yr return period</i>	mm	Acreman y Sinclair (986)
PRECIPITACIÓN ANUAL			
4	Precipitación media anual de la cuenca (PMA) <i>Mean annual precipitation</i>	mm	Acreman y Werritty (1987) Acreman y Sinclair (1986) Ahn y Palmer (2016) Álvarez–Olguín <i>et al.</i> (2011) Arellano-Lara y Escalante-Sandoval (2014) Gubareva (2012) Heřmanovský y Pech (2013) Kult <i>et al.</i> (2014) Laaha y Blöschl (2006) Lebecherel <i>et al.</i> (2016) Malekinezhad <i>et al.</i> (2011) Mamun <i>et al.</i> (2010) Nathan y McMahon (1990) Nobert <i>et al.</i> (2014) Núñez <i>et al.</i> (2016) Ouarda <i>et al.</i> (2001) Ouarda <i>et al.</i> (2006) Ouarda <i>et al.</i> (2008) Requena <i>et al.</i> (2016) Sanborn y Bledsoe (2006) Sarhadi y Modarres (2011) Schnier y Cai (2014) Singh <i>et al.</i> (2014)
5	Varianza de lluvia anual (Var) <i>Variance of the anual rainfall</i>	-	Arellano-Lara y Escalante-Sandoval (2014)
6	Desviación estándar de lluvia anual ($SDAR$) <i>Standard deviation of the annual rainfall</i>	-	Arellano-Lara y Escalante-Sandoval (2014)
7	Coficiente de variación de la lluvia anual ($CVAR$)	-	Arellano-Lara y Escalante-Sandoval (2014)

	<i>Coefficient of variation of the annual rainfall</i>		
8	Promedio máximo anual de lluvia diaria (<i>AAMDR</i>) <i>Average annual maximum of daily rainfall</i>	mm	Arellano-Lara y Escalante-Sandoval (2014)
9	Coeficiente de variación del máximo anual de la lluvia diaria (<i>CVAMDR</i>) <i>Coefficient of variation of the annual maximum of daily rainfall</i>	-	Arellano-Lara y Escalante-Sandoval (2014)
10	Media anual de lluvia máxima en 48 h (<i>MA48MR</i>) <i>Mean annual 48-hour maximum rainfall</i>	mm	Arellano-Lara y Escalante-Sandoval (2014)
11	Coeficiente de variación del máximo anual de lluvia en 48h (<i>CVA48MR</i>) <i>Coefficient of variation of the annual 48-hour maximum rainfall</i>	-	Arellano-Lara y Escalante-Sandoval (2014)
12	Coeficiente de variación del número de días con lluvia por año (<i>CVNDRY</i>) <i>Coefficient of variation of the number of days with rainfall per year</i>	-	Arellano-Lara y Escalante-Sandoval (2014)
PRECIPITACIÓN MENSUAL			
13	Precipitación mensual (<i>Pm</i>) <i>Monthly precipitation</i>	mm	Kult <i>et al.</i> (2014)
14	Precipitación media mensual total (<i>Pmt</i>) <i>Total mean monthly precipitation</i>	mm	Sanborn y Bledsoe (2006)
15	Precipitación en el mes más lluvioso del año (<i>Pmm</i>) <i>Precipitation in the rainiest month of the year</i>	mm	Álvarez-Olguín <i>et al.</i> (2011)
16	Precipitación de Marzo a Mayo (<i>Pm – m</i>) <i>Precipitation from March to May</i>	mm	Nobert <i>et al.</i> (2014)
17	Precipitación de Octubre a Diciembre (<i>Po – d</i>) <i>Precipitation from October to December</i>	mm	Nobert <i>et al.</i> (2014)
18	Relación de precipitación del mes más lluvioso respecto al mes más seco (<i>C_{RD}</i>) <i>Ratio of precipitation in wettest month to that of the driest</i>	-	Sanborn y Bledsoe (2006)
19	Relación de precipitación de los 3 meses más lluvioso respecto a los 3 meses más secos (<i>C_{3RD}</i>)	-	Sanborn y Bledsoe (2006)

	<i>Ratio of precipitation in the wettest 3 months to the driest 3 months</i>		
20	Desviación estándar del número de días con lluvia por año (<i>SDMR#</i>) <i>Standard deviation of the monthly rainfall for each month</i>	-	Arellano-Lara y Escalante-Sandoval (2014)
21	Desviación estándar de la lluvia al mes (<i>SDMR#</i>) <i>Standard deviation of the monthly rainfall for each month</i>	mm	Arellano-Lara y Escalante-Sandoval (2014)
22	Máximo de lluvia diaria en cada mes (<i>MDR#</i>) <i>Maximum daily rainfall for each month</i>	mm	Arellano-Lara y Escalante-Sandoval (2014)
23	Promedio mensual de lluvia diaria (<i>AMDR#</i>) <i>Average monthly of daily rainfall for each month</i>	mm	Arellano-Lara y Escalante-Sandoval (2014)
24	Coefficiente de variación del máximo de lluvia diaria por mes (<i>CVMDR#</i>) <i>Coefficient of variation of maximum daily rainfall for each month</i>	mm	Arellano-Lara y Escalante-Sandoval (2014)
25	Promedio del número de días con lluvia por mes (<i>ANDR#</i>) <i>Average number of days with rainfall for each month</i>	mm	Arellano-Lara y Escalante-Sandoval (2014)
26	Estacionalidad de precipitación (E_p) Rango de 0 a 1. Valor de 0 indica que la precipitación se distribuye exactamente igual en cada mes. Valor 1, toda la precipitación cae en un solo mes. <i>Precipitation Seasonality</i> <i>Range is 0 (precipitation spread out exactly evenly in each month) to 1 (all precipitation falls in a single month)</i>	-	Ahn y Palmer (2016) Kult <i>et al.</i> (2014)
27	Precipitación promedio en verano (<i>PSUM</i>) <i>Summer average precipitation</i>	mm	Laaha y Blöschl (2006) Schnier y Cai, (2014)
28	Precipitación promedio en invierno (<i>PWIN</i>) <i>Winter average precipitation</i>	mm	Laaha y Blöschl, 2006; Schnier y Cai, 2014;
29	Coefficiente de Fournier Indicador de la erosión que se puede producir en una cuenca, debido a factores pluviométricos; se define de la siguiente forma.	-	Álvarez-Olguín <i>et al.</i> (2011)

	$F = \frac{P_{mm}^2}{PMA} \quad (12.6)$ <p>donde: <i>F</i> Coeficiente de Fournier en mm. <i>Pmm</i> Precipitación del mes más lluvioso en mm. <i>PMA</i> Precipitación media anual en mm.</p> <p><i>Fournier's coefficient</i> Indicator of erosion that can occur in a basin, due to rainfall factors; it is defined as the precipitation of the wettest month squared, divided by rainfall and the annual average.</p>		
NIEVE / HELADA			
30	Porcentaje de nieve (<i>S%</i>) <i>Snow percentage</i>	%	Kult <i>et al.</i> (2014)
31	Promedio mensual total de nieve (<i>MATS</i>) <i>Total monthly average of snow</i>	mm	Sanborn y Bledsoe (2006)
32	Promedio anual de nieve (<i>AATS</i>) <i>Mean annual snow</i>	mm	Sanborn y Bledsoe (2006)
33	Media acumulada de nieve en un día específico (<i>IMAS</i>) <i>Accumulated average of snow on a specific day</i>	mm	Ouarda <i>et al.</i> (2006)
34	Relación del total de nieve anual con el total anual de precipitación (<i>TS / TP</i>) <i>Relation: the annual total of snow with the annual total of precipitation</i>	-	Sanborn y Bledsoe (2006)
35	Día de la primera helada (<i>dff</i>) <i>Day of first frost</i>	-	Kult <i>et al.</i> (2014)
36	Día de la última helada (<i>dlf</i>) <i>Day of last frost</i>	-	Kult <i>et al.</i> (2014)
EVAPOTRANSPIRACIÓN			
37	Promedio anual de evapotranspiración (<i>AAE</i>) <i>Annual average evapotranspiration</i>	mm	Sanborn y Bledsoe (2006) Singh <i>et al.</i> (2014)
38	Evapotranspiración potencial (<i>PET</i>) <i>Potential evapotranspiration</i>	mm	Requena <i>et al.</i> (2016)
39	Evapotranspiración potencial media mensual (<i>AAPET</i>) <i>Annual average potential evapotranspiration</i>	mm	Sanborn y Bledsoe (2006)
EVAPORACIÓN			

40	Evaporación potencial media anual (<i>AAPE</i>) <i>Annual average potencial evaporation</i>	mm/año	Heřmanovský y Pech (2013) Lebecherel <i>et al.</i> (2016) Mamun <i>et al.</i> (2010)
41	Relación de la precipitación media anual y la evaporación potencial media anual (<i>PMA / AAPE</i>) <i>Ratio: mean annual precipitation with mean annual potential evaporation</i>	-	Heřmanovský y Pech (2013)
42	Relación de la evaporación media anual con la evaporación media anual potencial (<i>MAE / AAPE</i>) <i>Ratio: mean annual evaporation with mean annual potential evaporation</i>	-	Heřmanovský y Pech (2013)
HUMEDAD			
43	Humedad relativa (<i>R</i>) <i>Relative humidity</i>	%	Kult <i>et al.</i> (2014)
44	Índice de humedad mensual Relación entre la precipitación mensual y la evapotranspiración potencial <i>Monthly wetness index</i> <i>Ratio between monthly precipitation and potential evapotranspiration</i>	-	Kult <i>et al.</i> (2014)
TEMPERATURA ANUAL			
45	Temperatura mínima anual (<i>mAT</i>) <i>Minimun annual temperature</i>	°C	Núñez <i>et al.</i> (2016)
46	Temperatura máxima anual (<i>MAT</i>) <i>Maximun annual temperature</i>	°C	Núñez <i>et al.</i> (2016)
47	Temperatura media anual (<i>AAT</i>) <i>Average annual temperature</i>	°C	Álvarez–Olguín <i>et al.</i> (2011) Kult <i>et al.</i> (2014) Malekinezhad <i>et al.</i> (2011) Schnier y Cai (2014)
TEMPERATURA MENSUAL			
48	Temperatura mensual promedio (<i>AMT</i>) <i>Average monthly temperature</i>	°C	Kult <i>et al.</i> (2014) Sanborn y Bledsoe (2006)
49	Temperatura mensual mínima promedio (<i>AmMT</i>) <i>Average minimun monthly temperature</i>	°C	Sanborn y Bledsoe (2006)
50	Temperatura mínima promedio durante Noviembre a Marzo (<i>AmMTN – M</i>) <i>Average minimun monthly temperature Nov-March</i>	°C	Sanborn y Bledsoe (2006)

51	Temperatura promedio de verano (<i>AST</i>) <i>Average summer temperature</i>	°C	Schnier y Cai (2014)
52	Temperatura promedio de invierno (<i>AWT</i>) <i>Average winter temperature</i>	°C	Schnier y Cai (2014)
RADIACIÓN			
53	Radiación solar mensual (<i>MSR</i>) <i>Monthly solar radiation</i>	KJ/m ² /día	Sanborn y Bledsoe (2006)
54	Radiación solar promedio anual (<i>AASR</i>) <i>Average annual solar radiation</i>	KJ/m ² /día	Sanborn y Bledsoe (2006)

Variables hidrométricas

N	VARIABLE	UNIDAD	AUTOR
1	Gasto específico (q_s) <i>Specific flow</i>	m ³ /skm ²	Merz <i>et al.</i> (2016) Skaugen y Væringstad (2005)
2	Caudal medio anual (AAF) <i>Annual average flow</i>	m ³ /s	Álvarez-Olguín <i>et al.</i> (2011) Lebecherel <i>et al.</i> (2016) Nathan y McMahon (1992) Sanborn y Bledsoe (2006)
3	Desviación estándar del caudal medio anual ($SDAAF$) <i>Standard deviation of the annual average flow</i>	adim	Nathan y McMahon (1992)
4	Caudal máximo anual (MAF) <i>Maximun annual flow</i>	m ³ /s	Dikbas <i>et al.</i> (2013)
5	Caudal máximo medio anual ($AAMF$) <i>Average annual maximum flow</i>	m ³ /s	Paris y Zucarelli (2004)
6	Gasto máximo anual: coeficiente de variación ($CVMAF$) <i>Coefficient of variation and skewness of maximun annual flow</i>	m ³ /s	Dikbas <i>et al.</i> (2013) Dodangeh <i>et al.</i> (2014)
7	Desviación estándar del Caudal mínimo medio ($SDmAF$) <i>Standard deviation of the minimum average flow</i>	-	Dodangeh <i>et al.</i> (2014)
8	Régimen de río (RF) <i>Regime Flow</i>	-	Merz <i>et al.</i> (2016) Nathan y McMahon (1992)
9	Curvas de Duración de Caudal (FDC) <i>Flow duration curve</i>	m ³ /s	Nathan y McMahon (1992)
10	Curva de frecuencia de caudal mínimo (FFC) <i>Flow frequency curve</i>	m ³ /s	Nathan y McMahon (1992)
11	Relación del (caudal medio anual máximo de 1 día) / (Caudal medio durante todo el año) ($R1$) <i>Ratio between mean maximum annual flow in 1 day and yearly median flow</i>	-	Sanborn y Bledsoe (2006)
12	Coficiente de sesgo del gasto máximo anual ($CSMAF$) <i>Coefficient of skewness of annual maximum flow</i>	-	Álvarez-Olguín <i>et al.</i> (2011) Dodangeh <i>et al.</i> (2014) Sanborn y Bledsoe (2006);
13	Coficiente de curtosis del gasto máximo anual ($CKMAF$) <i>Coefficient of kurtosis of annual maximum flow</i>	-	Dodangeh <i>et al.</i> (2014)
14	Medida de discordancia (D)	-	Dodangeh <i>et al.</i> (2014)

	<i>Discordancy measure</i>		
15	Caudal diario (<i>DF</i>) <i>Daily flow</i>	m ³ /s	Yu <i>et al.</i> (2002)
16	Coefficiente de variación del gasto diario (<i>CVDF</i>) <i>Coefficient of variation of daily flow</i>	-	Sanborn y Bledsoe (2006)
17	Caudal medio anual (<i>MeAF</i>) <i>Mean annual flow</i>	m ³ /s	Paris y Zucarelli (2004)
18	Caudal mensual (<i>MnF</i>) <i>Monthly flow</i>	m ³ /s	Sanborn y Bledsoe (2006)
19	Caudal máximo mensual para cada mes <i>Maximun monthly flow for each month</i>	m ³ /s	Sanborn y Bledsoe (2006)
20	Caudal mínimo medio (<i>mMF</i>) <i>Minimun median flow</i>	m ³ /s	Dodangeh <i>et al.</i> (2014)
21	Caudal máximo medio anual (<i>MYMF</i>) <i>Maximun yearly median flow</i>	m ³ /s	Paris y Zucarelli (2004)
22	Caudal mínimo promedio anual en periodos de 1, 3, 7, 30 y 90 días (<i>AAmF</i>) <i>Average annual minimum flow in 1, 3, 7, 30 and 90 days</i>	m ³ /s	Sanborn y Bledsoe (2006)
23	Caudal mínimo de 7 días dividido por el caudal medio para ese año (<i>R7d</i>) <i>7-day minimum flow divided by mean flow for that year</i>	m ³ /s	Sanborn y Bledsoe (2006)
24	Día juliano de ocurrencia del caudal mínimo (<i>JD</i>) <i>Julian date of the minimum flow</i>	-	Sanborn y Bledsoe (2006)
25	Caudal máximo promedio anual en periodos de 1, 3, 7, 30 y 90 días. (<i>AAMAF</i>) <i>Average annual maximum flow in 1, 3, 7, 30 and 90 days</i>	m ³ /s	Sanborn y Bledsoe (2006)
26	Día juliano de ocurrencia del caudal máximo (<i>JDMF</i>) <i>Julian date of the maximum flow</i>	-	Sanborn y Bledsoe (2006)
27	Número de día por año con caudal nulo (caudal = 0) (<i>Q0</i>) <i>Number of days per year with zero flow</i>	-	Sanborn y Bledsoe (2006)

28	Número medio de pulsos bajos, pulso bajo definido como 1 desviación estándar por debajo de la media. (#LP) <i>Average number of low pulses, low pulse defined as 1 standard deviation below the mean</i>	-	Sanborn y Bledsoe (2006)
29	Promedio de la duración de pulsos bajos (ADLP) <i>Average duration of low pulses</i>	-	Sanborn y Bledsoe (2006)
30	Número medio de pulsos altos, pulso bajo definido como 1 desviación estándar por encima de la media. (#HP) <i>Average number of high pulses, high pulse defined as 1 standard deviation above the mean</i>	-	Sanborn y Bledsoe (2006)
31	Promedio de la duración de pulsos altos (ADHP) <i>Average duration of high pulses</i>	-	Sanborn y Bledsoe (2006)
32	Tasa de crecimiento, media de todas las diferencias positivas respecto al caudal medio diario. (RPD) <i>Rise rate mean of all positive differences</i>	m ³ /s / día	Sanborn y Bledsoe (2006)
33	Tasa de decremento, media de todas las diferencias negativas respecto al caudal medio diario (RND) <i>Fall rate mean of all negative differences</i>	m ³ /s / día	Sanborn y Bledsoe (2006)
34	(Media del caudal mensual - mediana del caudal mensual) / (mediana del caudal mensual) (Rm) <i>(Mean monthly flow median monthly flow)/median monthly flow</i>	-	Sanborn y Bledsoe (2006)
35	Media del escurrimiento anual dividida por el área de la cuenca (AAF / A) <i>Mean annual runoff divided by catchment area</i>	cm	Sanborn y Bledsoe (2006)
36	Variabilidad media en los caudales diarios dividida por los caudales diarios medianos para cada año. (VARDF) <i>Average variability in daily flows divided by median daily flows for each year, where variability is calculated as 90th–10th percentile</i>	-	Sanborn y Bledsoe (2006)

37	CV en caudales mínimos mensuales (<i>CVm</i>) <i>CV in minimum monthly flow</i>	-	Sanborn y Bledsoe (2006)
38	Promedio del caudal diario anual más bajo dividido por el promedio anual de caudal diario promedio de todos los años (<i>Rdf</i>) <i>Mean of lowest annual daily flow divided by median annual daily flow averaged across all years</i>	-	Sanborn y Bledsoe (2006)
39	Media del escurrimiento mínimo anual dividida por el área de la cuenca (<i>MeAm / A</i>) <i>Mean annual minimum flows divided by catchment area</i>	m ³ /s / km ²	Sanborn y Bledsoe (2006)
40	Media del percentil 25th de la curva de duración de caudal dividida por la mediana diaria del caudal a través de todos los años de registro (<i>FDC25p</i>) <i>Mean of 25th percentile from the flow duration curve divided by median daily flow across all years</i>	adim	Sanborn y Bledsoe (2006)
41	Número total de eventos de caudal bajo (con un umbral igual al 5% del flujo diario medio) dividido por la longitud de registro en años (<i>If5%</i>) <i>Total number of low flow spells (threshold equal to 5% of mean daily flow) divided by record length in years</i>	-	Sanborn y Bledsoe (2006)
42	Número medio de eventos de inundación discretos por año (<i>FY</i>) <i>Mean number of discrete flood events per year</i>	-	Sanborn y Bledsoe (2006)
43	Media anual mínima de 30 días dividida por la mediana de la descarga (<i>q30m</i>) <i>Mean annual 30-day minimum divided by median discharge</i>	-	Sanborn y Bledsoe (2006)
44	Media anual mínima de 7 días dividida por la mediana de la descarga (<i>q7m</i>) <i>Mean annual 7-day maximum divided by median discharge</i>	-	Sanborn y Bledsoe (2006)
45	Media anual máxima de 30 días dividida por la mediana de la descarga (<i>q30M</i>)	-	Sanborn y Bledsoe (2006)

	<i>Mean annual 30-day maximum divided by median discharge</i>		
46	Máxima proporción del año (número de días / 365) Durante el cual no se han producido nunca inundaciones durante el período de registro (Nf) <i>Max proportion of the year (num days/365) during which no floods have ever occurred over the period of record</i>	-	Sanborn y Bledsoe (2006)
47	Velocidad en el cauce (Vc) <i>Velocity in stream reach</i>	m/s	Álvarez-Olguín <i>et al.</i> (2011)
48	Coficiente estacional de caudales en invierno (CEW) La relación entre el promedio de la suma de los caudales promedio mensuales de enero a marzo y el caudal total anual. <i>Seasonal coefficient of winter streamflow</i> <i>The ratio between the average of the sum of the average monthly streamflows from January to March and the total annual streamflow</i>	-	Matteau <i>et al.</i> (2009)
49	Coficiente estacional de caudales en primavera (CESP) La relación entre el promedio de la suma de los caudales mensuales promedio de abril a junio y el caudal total anual <i>Seasonal coefficient of spring streamflow</i> <i>The ratio between the average of the sum of the average monthly streamflows from April to June and the total annual streamflow</i>	-	Matteau <i>et al.</i> (2009)

50	<p>Coeficiente estacional de caudales en verano (<i>CESM</i>)</p> <p>La relación entre el promedio de la suma de los caudales mensuales promedio de julio a septiembre y el caudal total anual.</p> <p><i>Seasonal coefficient of summer streamflow</i> <i>The ratio between the average of the sum of the average monthly streamflows from July to September and the total annual streamflow.</i></p>	-	Matteau <i>et al.</i> (2009)
51	<p>Coeficiente estacional de caudales en otoño (<i>CEF</i>)</p> <p>La relación entre el promedio de la suma de los caudales promedio mensuales de octubre a diciembre y el caudal total anual</p> <p><i>Seasonal coefficient of fall streamflow</i> <i>The ratio between the average of the sum of the average monthly streamflows from October to December and the total annual streamflow</i></p>	-	Matteau <i>et al.</i> (2009)
52	<p>Relación entre los caudales de invierno y primavera (<i>SWP</i>)</p> <p><i>Ratio between winter streamflows and spring streamflow</i></p>	-	Matteau <i>et al.</i> (2009)
53	<p>Relación entre los caudales de primavera y verano (<i>SPM</i>)</p> <p><i>Ratio between spring streamflows and summer streamflows</i></p>	-	Matteau <i>et al.</i> (2009)
54	<p>Relación entre los caudales de verano y otoño (<i>MF</i>)</p> <p><i>Ratio between summer streamflows and fall streamflows</i></p>	-	Matteau <i>et al.</i> (2009)
55	<p>Relación entre los caudales de otoño e invierno (<i>FW</i>)</p> <p><i>Ratio between fall streamflows and winter streamflow</i></p>	-	Matteau <i>et al.</i> (2009)
56	<p>Coeficiente mensual del máximo caudal mensual promedio (<i>CMax</i>)</p>	%	Matteau <i>et al.</i> (2009)

	<p>La relación entre el flujo máximo promedio mensual y el caudal total anual</p> <p><i>Monthly coefficient of the maximum monthly average streamflow</i></p> <p><i>The ratio between the maximum monthly average streamflow and the total annual streamflow</i></p>		
57	<p>Coeficiente mensual del mínimo caudal mensual promedio (<i>CMin</i>)</p> <p>La relación entre el caudal mínimo mensual promedio y el caudal anual total</p> <p><i>Monthly coefficient of the minimum monthly average streamflow</i></p> <p><i>The ratio between the minimum monthly average streamflow and the total annual streamflow</i></p>	%	Matteau <i>et al.</i> (2009)
58	<p>Coeficiente mensual de inmoderación (<i>CI</i>)</p> <p>La relación entre el caudal mensual máximo y mínimo</p> <p><i>Monthly coefficient of immoderation</i></p> <p><i>The ratio between the maximum and minimum monthly streamflow</i></p>	-	Matteau <i>et al.</i> (2009)
59	<p>Coeficiente de variación (<i>CV – DST</i>)</p> <p>La relación entre la desviación estándar y el flujo promedio mensual</p> <p><i>Coefficient of variation</i></p> <p><i>The ratio between the standard deviation and the average monthly streamflow</i></p>	%	Matteau <i>et al.</i> (2009)
60	<p>Momento del flujo máximo mensual promedio (<i>MAXmF</i>)</p> <p>Promedio de meses de ocurrencia de los caudales promedio máximos mensuales</p> <p><i>Timing of maximum monthly average streamflows</i></p> <p><i>Average month of occurrence of the maximum monthly average streamflows</i></p>	-	Matteau <i>et al.</i> (2009)
61	<p>Medida de tiempo de los caudales del caudal promedio mensual mínimo (<i>Tm</i>)</p> <p>Promedio de meses de ocurrencia de caudales de flujo promedio mensual mínimo</p>	-	Matteau <i>et al.</i> (2009)

	<i>Timing of minimum monthly average streamflows</i> <i>Average month of occurrence of the minimum monthly average streamflows</i>		
62	Velocidad máxima (<i>V max</i>) <i>Maximun velocity</i>	m ³ /s	Paris y Zucarelli (2004)
63	Velocidad mínima (<i>V min</i>) <i>Mínimun velocity</i>	m ³ /s	Paris y Zucarelli (2004)
64	Velocidad media (<i>V m</i>) <i>Mean velocity</i>	m ³ /s	Paris y Zucarelli (2004)

Variables de ingeniería

N	VARIABLE	UNIDAD	AUTOR(ES)
1	Máxima capacidad de almacenamiento del embalse inmediato aguas arriba (<i>Cap _ u</i>) <i>Maximum storage capacity of nearest upstream reservoir</i>	m ³	Matteau <i>et al.</i> , (2009)
2	Máxima capacidad de almacenamiento acumulada de todos los embalses aguas arriba de la presa en cuestión (<i>Cap _ a</i>) <i>Cumulative maximum storage capacity of all upstream reservoirs</i>	m ³	Matteau <i>et al.</i> , (2009)
3	Grado de almacenamiento (<i>DS</i>) <i>Degree of storage</i>	-	Matteau <i>et al.</i> , (2009)
4	Grado de almacenamiento acumulado (<i>DCS</i>) <i>Degree of cumulative storage</i>	-	Matteau <i>et al.</i> , (2009)

Propiedades hidrológicas del suelo

N	VARIABLE	UNIDAD	AUTOR(ES)
1	Déficit medio de la humedad del suelo (<i>SMD</i>) <i>Effective mean soil moisture deficit</i>	mm	Acreman y Sinclair (1986)
2	Tipo de suelo (<i>SOIL</i>) <i>Soil type index</i>	-	Acreman y Sinclair (1986)
3	Permeabilidad promedio de la cuenca (<i>PPC</i>) <i>Average permeability of the basin</i>	mm/h	Ahn y Palmer (2016) Kult <i>et al.</i> (2014) Singh <i>et al.</i> (2014) Singh <i>et al.</i> (2016)
4	Permeabilidad (PPC_{max}, PPC_{min}) <i>Maximunand minimun permeability</i>	mm/h	Brath <i>et al.</i> (2009)
6	Porcentaje del área permeable en la cuenca (<i>PER</i>) <i>Percentage of the permeable área of the basin</i>	%	Requena <i>et al.</i> (2016)
8	Tasa media de infiltración en la cuenca (R_{inf}) <i>Average rate of infiltration in the basin</i>	mm	Requena <i>et al.</i> (2016)
7	índice topográfico de humedad $\ln(\text{área de flujo acumulada}) - \ln[1/\tan(\text{slope})]$ <i>Topographic moisture index</i>		Kult <i>et al.</i> (2014) Sanborn y Bledsoe (2006) Singh <i>et al.</i> (2014)

Variables uso del suelo

N	VARIABLE	UNIDAD	AUTOR(ES)
1	Uso de suelo y vegetación (USV) <i>Soil use and vegetation</i>	%	Acreman y Sinclair (1986) Acreman y Werritty (1987) Ahn y Palmer (2016) Álvarez–Olguín <i>et al.</i> (2011) Brath <i>et al.</i> (2009) Kult <i>et al.</i> (2014) Laaha y Blöschl (2006) Nathan y McMahon (1990) Ouarda <i>et al.</i> (2001) Ouarda <i>et al.</i> (2006) Sanborn y Bledsoe (2006) Sarhadi y Modarres (2011) Schnier y Cai (2014) Singh <i>et al.</i> (2014) Skaugen y Væringstad (2005)
2	Clasificación del suelo en función de la infiltración (<i>HG</i>)	%	Ahn y Palmer (2016) Singh <i>et al.</i> (2014)

	<i>Soil classification based on infiltration</i>		
--	--	--	--

Variables ambientales y físicas

N	VARIABLE	UNIDAD	AUTOR(ES)
1	Índice de acidez (PH)	[H]+	Kirk y Williams (1985)
2	Conductividad específica (C_e)	S/m (siemens por metro) o $\Omega^{-1} \cdot m^{-1}$	Kirk y Williams (1985)
3	Sólidos suspendidos (S_{ST})	mg/litro	Kirk y Williams (1985)
4	Concentración de sulfato (C)	mg/litro	Kirk y Williams (1985)

Variables geológicas

N	VARIABLE	UNIDAD	AUTOR(ES)
1	Porcentaje del tipo de roca ($\%R$) Clasificación del tipo de rocas subyacente al área de drenaje.	%	Laaha y Blöschl (2006) Sanborn y Bledsoe (2006) Yu <i>et al.</i> (2002)

ANEXO II

Aplicación del Análisis de componentes Principales

Se tiene una serie de puntos (mediciones o elementos) en el espacio multidimensional de d variables (X_i, Y_i, \dots, Z_i) , se propone efectuar una transformación en componentes principales (Tabla II-1).

Tabla II-1. Datos estadísticos de 39 estaciones.

N	HEP	DEP	QTE	H2O	PMS	EM2	CTE	GE	GM2	PMG
1	17.6	25	350	47.9	380	596	3.6	28.6	17046	38.9
2	20.2	31	340	50.7	330	522	2.9	34.8	18192	38.3
3	47.1	21	290	45.3	350	502	3.0	28.4	16854	36.3
4	34.3	21	300	46.1	350	515	3.2	30.4	15686	33.9
5	28.8	31	300	50.8	290	533	2.7	32.4	17274	38.0
6	12.9	35	320	56.4	250	552	3.4	39.5	19798	32.7
7	15.7	29	330	49.2	340	568	3.0	27.2	15462	41.7
8	17.7	25	310	45.2	370	557	2.9	28.4	15808	40.9
9	58.9	20	260	44.1	330	365	1.9	33.4	12202	37.1
10	33.8	21	300	43.9	380	517	3.0	27.6	14282	43.1
11	38.3	21	350	47.3	390	474	2.3	31.9	15117	42.4
12	19.5	27	330	50.0	330	483	2.6	34.2	16529	39.5
13	12.3	33	280	54.9	230	610	3.5	30.2	18452	31.7
14	35.7	27	320	47.0	360	497	3.3	36.1	17922	39.9
15	34.1	21	280	41.8	390	497	2.4	27.3	13549	41.1
16	18.1	31	270	50.5	270	602	3.2	29.6	17814	31.1
17	16.5	33	380	54.3	320	435	2.7	41.3	17983	33.6
18	32.7	21	360	45.0	440	634	3.1	22	13938	49.8
19	33.9	25	280	45.2	340	545	2.6	27.5	14998	39.6
20	19.7	33	300	52.6	270	483	3.8	38.1	18423	33.0
21	37.0	21	290	45.3	350	567	2.0	27.3	15512	38.2
22	15.0	33	330	52.4	300	594	3.0	27.7	16484	37.4
23	15.0	35	390	55.7	310	449	3.0	30.4	13675	44.3
24	42.0	19	280	43.7	360	507	3.0	30.3	15360	35.3
25	24.0	31	380	50.7	370	523	2.0	31.8	16662	39.6
26	31.0	25	360	46.7	410	538	3.0	29.1	15676	42.1
27	26.0	20	260	43.2	340	616	4.0	30.7	18904	32.6
28	33.0	31	330	49.2	340	607	3.0	28.5	17306	37.3
29	8.6	40	340	57.6	250	566	4.0	30.9	17507	35.5
30	23.0	29	340	48.6	360	521	3.0	32.9	17142	39.4
31	20.0	28	300	50.8	290	526	3.0	33.2	17451	33.8
32	16.0	31	360	52.9	320	437	3.0	39.9	17448	37.3

Es importante analizar las matrices de correlación y de varianza-covarianza, ya que se debe seleccionar una para efectuar el análisis. Si utilizamos la primera, tendremos los componentes sobre las *observaciones estandarizadas*; de este modo se evita el problema de escala:

CÁLCULO DE LOS ESTADÍSTICOS Y MATRICES

Tabla II.2- Estadísticos descriptivos de las variables medidas.

<i>Variable</i>	<i>Media</i>	<i>Desviación estándar</i>	<i>Varianza</i>
<i>HEP</i>	26.24	10.96	120.11
<i>DEP</i>	27.56	5.39	29.04
<i>QTE</i>	321.54	35.06	1229.15
<i>H2O</i>	49.08	3.99	15.88
<i>PMS</i>	336.67	45.96	2112.28
<i>EM2</i>	531.41	65.84	4335.41
<i>CTE</i>	2.94	0.47	0.22
<i>GE</i>	30.96	4.36	19.03
<i>GM2</i>	16318.08	1833.36	3361198.6
<i>PMG</i>	38.36	4.38	19.23

Resultado:

$$\mu(x_1, y_1) = \begin{bmatrix} 26.24 \\ 27.56 \\ 321.54 \\ 49.08 \\ 336.67 \\ 531.41 \\ 2.94 \\ 30.96 \\ 16318.08 \\ 38.36 \end{bmatrix}$$

Matriz de correlación.

	HEP	DEP	QTE	H2O	PMS	EM2	CTE	GE	GM2	PMG
HEP	1									
DEP	-0.77	1								
QTE	-0.32	0.37	1							
H2O	-0.74	0.92	0.45	1						
PMS	0.55	-0.68	0.23	-0.74	1					
EM2	-0.19	-0.06	-0.13	-0.06	0	1				
CTE	-0.42	0.19	-0.08	0.23	-0.35	0.55	1			
GE	-0.31	0.4	0.1	0.42	-0.44	-0.57	0.1	1		
GM2	-0.52	0.38	-0.09	0.38	-0.53	0.32	0.68	0.55	1	
PMG	0.26	-0.22	0.51	-0.24	0.68	-0.08	-0.45	-0.5	-0.72	1

Matriz de covarianza.

	HEP	DEP	QTE	H2O	PMS	EM2	CTE	GE	GM2	PMG
HEP	1.201									
DEP	-0.454	0.290								
QTE	-1.232	0.696	12.291							
H2O	-0.323	0.197	0.633	0.159						
PMS	2.763	-1.678	3.632	-1.348	21.123					
EM2	-1.394	-0.214	-3.043	-0.148	-0.118	43.354				
CTE	-0.022	0.005	-0.012	0.004	-0.074	0.170	0.002			
GE	-0.148	0.093	0.152	0.073	-0.884	-1.645	0.002	0.190		
GM2	-103.97	37.364	-58.630	28.121	-445.00	380.91	5.819	44.318	33612	
PMG	0.127	-0.051	0.777	-0.042	1.362	-0.232	-0.009	-0.096	-58.093	0.192

Por la fuerte variación en GM2 en la matriz $[Cov_{(x,y)}]$, se utilizará la matriz de correlación.

CÁLCULO DE LOS VALORES PROPIOS DE LA MATRIZ DE CORRELACIÓN:

Con diez variables ($d = 10$), habrá diez valores-propios asociados

$$[Corr_{(x,y)}] = \lambda_1, \lambda_2, \dots, \lambda_{10}$$

$$|[Corr_{(x,y)}] - \lambda I| = \begin{vmatrix} 1 - \lambda_1 & -0.77 & \dots & 0.26 \\ -0.77 & 1 - \lambda_2 & \dots & -0.22 \\ \vdots & \vdots & \ddots & \vdots \\ 0.26 & -0.22 & \dots & 1 - \lambda_p \end{vmatrix} = 0 \quad (12.7)$$

Resolviendo el determinante de la matriz se tiene una ecuación de grado 10 en λ :

$$\lambda^5 + 000\lambda^4 + 000\lambda^3 + 000\lambda^2 + \lambda + 000 = 0 \quad (12.8)$$

Obteniendo los siguientes valores propios:

$$[\lambda_i] = \left[\begin{array}{l} \lambda_1 = 4.424 \quad \lambda_2 = 2.201 \quad \lambda_3 = 1.694 \quad \lambda_4 = 0.856 \quad \lambda_5 = 0.291 \\ \lambda_6 = 0.256 \quad \lambda_7 = 0.139 \quad \lambda_8 = 0.096 \quad \lambda_9 = 0.028 \quad \lambda_{10} = 0.014 \end{array} \right] \quad (12.9)$$

Los valores-propios representan la varianza de los elementos sobre el eje correspondiente. En nuestro ejemplo se tiene una varianza de 4.42 sobre el primer eje, 2.20 sobre el segundo, etc. Esta información no es fácil de interpretar; por lo que se propone expresar cada varianza (cada valor-propio), en un valor relativo (porcentaje) de la varianza total (suma de los valores-propios). En nuestro caso con 5 variables, los porcentajes de variación explicados por cada uno de los ejes principales son respectivamente:

$$\frac{4.42}{10} * 100 = 44.20\%, \quad \frac{2.20}{10} * 100 = 22.00\%, \dots$$

Cabe recordar que los componentes por construcción algebraica no están correlacionados y por lo tanto, la parte de información explicada por dos ejes es igual a la simple suma de la parte que explicada cada uno de ellos.

$$\begin{aligned} 44.20+22.00 &= 66.20\% \text{ por el plano definido por los ejes 1 y 2} \\ 16.90+8.60 &= 25.50\% \text{ por el plano definido por los ejes 3 y 4} \\ 44.20+16.90 &= 61.10\% \text{ por el plano definido por los ejes 1 y 3} \\ 66.20+25.50 &= 91.70\% \text{ por los planos definidos por 1-2 y 3-4} \end{aligned}$$

Es necesario examinar cada uno de los ejes hasta obtener la información que se requiera o se juzgue *suficiente* de cada uno de ellos. De esta forma se tiene que un solo eje explica un poco menos de la mitad de la información inicial (44.2%). Dos ejes explican los 2/3 de esta información (66.2%), mientras que tres ejes explican más de 4/5 (83.1%) y cuatro el (91.7%). Estos resultados se pueden ver el cuadro 1. Según el grado de exigencia, normalmente se retienen los ejes hasta que un 80% de la variación sea explicada (Tabla II-3).

TablaII-3. Valores-propios y porcentaje de la varianza.

Valor-propio λ_i	Total de la varianza (%)	Acumulado del valor-propio	Acumulado total (%)
4.424	44.24	4.424	44.24

2.201	22.012	6.625	66.252
1.694	16.945	8.32	83.197
0.856	8.555	9.175	91.752
0.291	2.911	9.466	94.663
0.256	2.564	9.723	97.227
0.139	1.394	9.862	98.622
0.096	0.96	9.958	99.582
0.028	0.28	9.986	99.862
0.014	0.138	10	100

También es necesario remarcar que dado el orden decreciente de los valores-propios, un mismo porcentaje puede ser obtenido por diferentes ejes. Por ejemplo, un porcentaje mayor al 45% puede ser obtenido en este caso por los ejes 2-3-4 (22.01 + 16.95 + 8.56) o por los ejes 1-5-9 (44.24 + 2.91 + 0.28). En general se recomienda conservar los ejes con un valor-propio entre los cuatro primeros ejes, para este caso. En ocasiones se representan estos porcentajes en gráficas como la que se muestra en la Figura II-1.

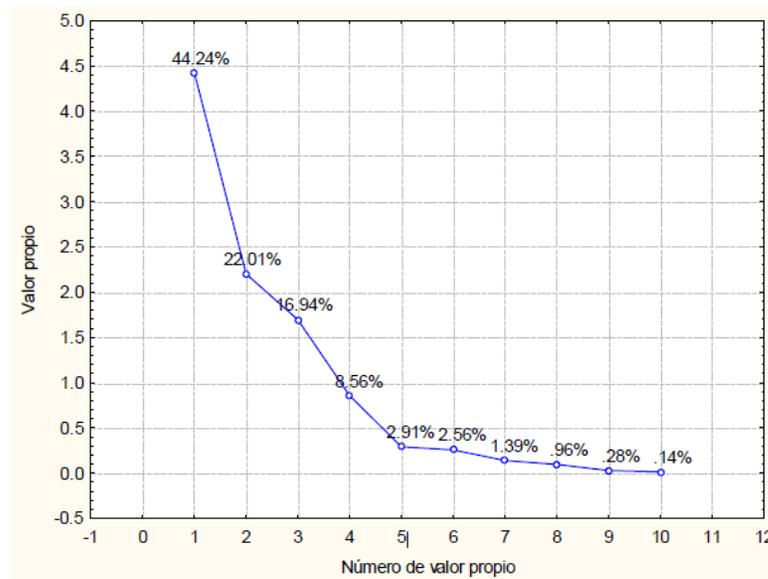


Figura II-1. Figura para las conclusiones

CÁLCULO DE VECTORES PROPIOS

Considerando la ecuación (12.7), donde $\left(\left[Corr_{(x,y)} \right] - \lambda I \right) W = 0$, el vector-propio ϕ_i

es la solución a $\left(\left[Corr_{(x,y)} \right] - \lambda_i I \right) \phi_i = 0$. Esto quiere decir que existe una solución

asociada a cada valor propio. Así se obtienen los vectores-propios para cada valor

asociados de λ_i :

$$\begin{bmatrix} 1-\lambda_i & -0.77 & \cdots & 0.26 \\ -0.77 & 1-\lambda_i & \cdots & -0.22 \\ \vdots & \vdots & \ddots & \vdots \\ 0.26 & -0.22 & \cdots & 1-\lambda_i \end{bmatrix} = 0$$

De manera que los vectores propios quedan de la siguiente forma

$$\phi_1[\lambda_1] = \begin{bmatrix} 0.381 \\ -0.384 \\ -0.048 \\ -0.394 \\ 0.401 \\ -0.043 \\ -0.264 \\ -0.287 \\ -0.373 \\ 0.310 \end{bmatrix} \quad \phi_2[\lambda_2] = \begin{bmatrix} -0.138 \\ 0.305 \\ 0.514 \\ 0.310 \\ 0.059 \\ -0.396 \\ -0.362 \\ 0.139 \\ -0.279 \\ 0.373 \end{bmatrix} \quad \phi_3[\lambda_3] = \begin{bmatrix} 0.265 \\ -0.109 \\ -0.328 \\ -0.113 \\ -0.154 \\ -0.587 \\ -0.295 \\ -0.488 \\ 0.012 \\ -0.324 \end{bmatrix}$$

FORMAR LA MATRIZ DE TRANSFORMACIÓN

Cada uno de los vectores propios forman la matriz de transformación, eigen-vectores.

Matriz de vectores-propios:

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}
HEP	0.381	-0.138	0.265	-0.015	-0.577	0.579	-0.262	-0.166	-0.028	-0.017
DEP	-0.384	0.305	-0.109	0.248	0.06	0.04	-0.458	-0.65	-0.021	-0.221
QTE	-0.048	0.514	-0.328	-0.445	-0.153	0.315	0.393	0.003	0.136	-0.364
H2O	-0.394	0.31	-0.113	0.201	-0.262	0.235	0.109	0.151	-0.18	0.71
PMS	0.401	0.059	-0.154	-0.47	0.293	-0.042	-0.022	-0.48	-0.2	0.482
EM2	-0.043	-0.396	-0.587	0.091	0.191	0.371	-0.078	0.143	-0.509	-0.171
CTE	-0.264	-0.362	-0.295	-0.328	-0.611	-0.453	-0.029	-0.157	-0.02	0.015
GE	-0.287	0.139	0.488	-0.446	0.046	-0.034	-0.201	0.197	-0.598	-0.149
GM2	-0.373	-0.279	0.012	-0.403	0.255	0.338	-0.327	0.133	0.536	0.17
PMG	0.31	0.373	-0.324	-0.044	-0.096	-0.22	-0.633	0.441	0.064	0.026

Los vectores que se acaban de calcular están normalizados. Esto implica que son de longitud 1, por lo que se debe verificar lo siguiente:

$$\phi_{1.1}^2 + \phi_{1.2}^2 + \phi_{1.3}^2 + \phi_{1.4}^2 + \phi_{1.5}^2 + \phi_{1.6}^2 + \phi_{1.7}^2 + \phi_{1.8}^2 + \phi_{1.9}^2 + \phi_{1.10}^2 = 1$$

$$(0.381)^2 + (-0.384)^2 + (-0.048)^2 + (-0.394)^2 + \dots + (-0.287)^2 + (-0.373)^2 + (0.310)^2 = 1$$

Esta matriz W muestra la contribución de cada una de las variables en cada uno de los ejes de los componentes principales. Es decir, que para un cierto eje o componente seleccionado, las variables que tienen los coeficientes mayores (en valor absoluto) son aquellas que contribuyen mayormente a la formación de ese eje (componente principal). Para nuestro ejemplo, las variables PMS, H2O, DEP, HEP y GM2 son las que más contribuyen a la formación del primer componente principal. Para el caso del segundo componente principal, la variable QTE es la que más contribuye a la formación del segundo componente principal. En la práctica no se utilizan mucho los vectores-propios como tales, normalmente se obtiene mayor información si se analiza la correlación de las variables con cada uno de los ejes o componentes principales. Esta correlación se lleva a cabo de con el procedimiento tradicional de correlacionar dos variables cuantitativas. Por ejemplo, se calcula el coeficiente de correlación entre la variable HEP y el primer componente principal Ω_1 . Para representar estas correlaciones se utiliza el llamado círculo de correlaciones, en donde de una manera gráfica, se determinan las variables que están correlacionadas entre si y cuáles no lo están. La interpretación detallada de un círculo de correlaciones se analizará más adelante. Ahora es necesario calcular cada uno de los componentes principales Ω_i .

Para obtener los componentes principales, es necesario recordar que los vectores-propios contienen los coeficientes que afectan a cada una de las variables iniciales y esto permite el cálculo de los componentes principales. Por ejemplo, el primer componente principal Ω_1 se obtiene para la i-ésima medición como:

$$\Omega_1 = 0.381HEP_1 - 0.384DEP_1 - 0.048QTE_1 - 0.394H2O_1 + \dots + 0.310PMG_1$$

Es importante recordar que los valores de cada variable deben introducirse a esta ecuación, dependiendo de la matriz que se haya seleccionado para encontrar los valores-propios. Esto quiere decir que si se utilizó la matriz de varianzas-covarianzas, los datos originales deberán

ser transformados como valores centrados (restando a cada valor su media); mientras que si se utilizó la matriz de correlación los datos deberán ser centrados-estandarizados (restando su media y dividiendo entres su desviación estándar). Para nuestro ejemplo, aplica el segundo caso.

Transformación a valores centrados: $X'_i = X_i - \mu_x$

Transformación a valores centrados-estandarizados: $X''_i = (X_i - \mu_x) / \sigma_x$

Empleando la segunda expresión se obtiene la matriz transformada de datos centrados-estandarizados que será utilizada también para el cálculo del círculo de correlaciones. Matriz de datos originales transformada (datos centrados-estandarizados).

Matriz de datos originales transformada:

	<i>HEP''</i>	<i>DEP'</i>	<i>QTE'</i>	<i>H2O'</i>	<i>PMS'</i>	<i>EM2'</i>	<i>CTE'</i>	<i>GE''</i>	<i>GM2'</i>	<i>PMG'</i>
<i>1</i>	-0.789	-0.476	0.812	-0.297	0.943	0.981	1.418	-0.542	0.397	0.122
<i>2</i>	-0.551	0.638	0.527	0.406	-0.145	-0.143	-0.082	0.879	1.022	-0.015
<i>3</i>	1.903	-1.218	-0.9	-0.949	0.29	-0.447	0.132	-0.588	0.292	-0.471
<i>4</i>	0.735	-1.218	-0.614	-0.748	0.29	-0.249	0.561	-0.129	-0.345	-1.018
<i>5</i>	0.233	0.638	-0.614	0.431	-1.015	0.024	-0.511	0.329	0.521	-0.083
<i>6</i>	-1.218	1.38	-0.044	1.836	-1.886	0.313	0.989	1.957	1.898	-1.292
<i>7</i>	-0.962	0.266	0.241	0.03	0.073	0.556	0.132	-0.863	-0.467	0.761
<i>8</i>	-0.78	-0.476	-0.329	-0.974	0.725	0.389	-0.082	-0.588	-0.278	0.578
<i>9</i>	2.98	-1.404	-1.755	-1.25	-0.145	-2.527	-2.226	0.558	-2.245	-0.288
<i>10</i>	0.689	-1.218	-0.614	-1.3	0.943	-0.219	0.132	-0.771	-1.111	1.08
<i>11</i>	1.1	-1.218	0.812	-0.447	1.16	-0.872	-1.368	0.215	-0.655	0.92
<i>12</i>	-0.615	-0.105	0.241	0.23	-0.145	-0.735	-0.725	0.742	0.115	0.259
<i>13</i>	-1.272	1.009	-1.185	1.46	-2.321	1.194	1.204	-0.175	1.164	-1.52
<i>14</i>	0.863	-0.105	-0.044	-0.522	0.508	-0.523	0.775	1.177	0.875	0.35
<i>15</i>	0.717	-1.218	-1.185	-1.827	1.16	-0.523	-1.154	-0.84	-1.51	0.624
<i>16</i>	-0.743	0.638	-1.47	0.356	-1.451	1.072	0.561	-0.313	0.816	-1.657
<i>17</i>	-0.889	1.009	1.668	1.309	-0.363	-1.464	-0.511	2.37	0.908	-1.086
<i>18</i>	0.589	-1.218	1.097	-1.024	2.248	1.558	0.346	-2.055	-1.298	2.608
<i>19</i>	0.699	-0.476	-1.185	-0.974	0.073	0.206	-0.725	-0.794	-0.72	0.282
<i>20</i>	-	1.009	-0.614	0.883	-1.451	-0.735	1.847	1.636	1.148	-1.223
	0.597*									
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>30</i>	-0.323	0.266	0.527	-0.121	0.508	-0.158	0.132	0.444	0.449	0.236
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>35</i>	0.936	-1.218	-0.044	-0.623	0.725	3.046	1.632	-1.895	0.181	0.054

⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
39	-0.351	1.38	-0.9	0.431	0.29	-0.356	-1.583	-0.657	-1.101	0.419

$$* \frac{19.70 - 26.24}{\sqrt{120.11}} = -0.5967$$

Es interesante mencionar que si se obtiene la matriz de varianza-covarianza y la matriz de correlación de estos datos centrados-reducidos, se obtendrá la misma matriz de correlación que se encontró para los datos originales. Esto es:

$$\left[Corr_{(X,Y,\dots)} \right] = \left[Corr_{(X'',Y'',\dots)} \right] = \left[Cov_{(X'',Y'',\dots)} \right]$$

OBTENER LA MATRIZ DE COMPONENTES PRINCIPALES Ω_i

A continuación con los datos transformados, se procede a calcular el total de los componentes principales Ω_i . Por ejemplo, para el primer término del primer componente principal se tiene:

$$*^1 \Omega_{1,1} = 0.381(-0.789)_1 - 0.384(-0.476)_1 - 0.048(0.812)_1 + \dots + 0.310(-0.122)_1 = -0.034$$

Para el segundo término componente principal, el cálculo será:

$$*^2 \Omega_{1,2} = 0.381(-0.551)_2 - 0.384(0.638)_2 - 0.048(0.527)_2 + \dots + 0.310(-0.015)_2 = -1.308$$

Sucesivamente se van calculando cada uno de los términos de la matriz de componentes principales. Hasta obtener la matriz de componentes principales

Matriz de componentes principales:

	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	Ω_8	Ω_9	Ω_{10}
1	-0.034* ¹	-0.696	-1.828	-1.27	0.042	-0.483	0.474	-0.138	0.025	0.073
2	-1.308* ²	0.576	0.142	-0.707	0.452	0.277	-0.382	0.105	0.126	-0.039
3	1.625	-1.584	1.086	-0.176	-0.772	0.532	-0.052	-0.178	0.665	0.233
4	0.901	-1.409	0.812	-0.291	-0.507	-0.239	0.822	-0.229	-0.059	0.073
5	-0.884	-0.036	0.632	0.809	0.06	0.536	-0.696	0.279	0.097	-0.079
6	-4.416	-0.198	0.544	-0.24	-0.088	0.375	-0.339	0.444	-0.368	0.086

7	0.135	0.379	-1.415	0.491	0.341	-0.616	0.024	0.119	0.063	-0.098
8	1.032	-0.378	-0.729	-0.086	0.91	-0.853	-0.014	0.113	0.06	-0.116
9	3.479	0.263	4.159	1.373	-0.891	0.185	0.107	-0.033	-0.265	-0.136
10	2.595	-0.485	-0.131	-0.064	-0.305	-0.916	-0.179	0.152	0.011	0.006
11	2.357	1.22	0.744	-0.818	0.05	0.554	0.211	0.439	0.004	0.206
12	-0.307	0.961	0.69	-0.227	0.55	-0.212	0.06	0.56	0.063	0.07
13	-3.546	-1.634	-0.501	1.572	-0.16	0.004	0.104	0.318	0.054	0.147
14	0.043	-0.33	0.785	-1.558	-0.541	-0.029	-0.99	-0.091	0.003	-0.018
15	3.308	-0.415	0.756	0.588	0.635	-0.776	0.022	-0.148	-0.038	-0.094
16	-2.102	-1.951	-0.003	1.375	0.388	0.073	0.141	-0.164	-0.005	-0.073
17	-2.623	2.106	1.544	-1.385	0.347	0.526	0.62	-0.153	-0.175	-0.044
18	3.667	0.236	-3.18	-0.71	-0.243	-0.22	-0.271	0.205	-0.208	0.11
19	1.685	-0.772	0.33	1.057	0.261	-0.092	-0.316	0.024	0.005	-0.176
20	-3.247	-0.665	1.152	-0.419	-0.941	-0.738	-0.346	-0.189	-0.06	0.011
21	2.046	-1.05	0.367	0.738	0.499	0.608	0.168	0.364	0	-0.015
22	-1.401	0.15	-1.321	1.051	0.274	0.149	0.192	-0.094	0.045	-0.125
23	-0.71	3.615	-0.788	0.726	-0.671	-0.513	0.346	0.087	0.121	-0.083
24	2.187	-1.411	1.598	0.115	0.049	0.241	0.417	-0.006	-0.058	0.01
25	0.04	1.88	-0.357	-0.576	0.875	1.037	0.229	-0.271	0.099	-0.093
26	1.935	0.923	-0.639	-0.872	0.575	0.546	0.113	-0.21	0.026	0.038
27	-0.196	-3.766	0.043	-0.797	0.363	-0.492	0.043	0.184	0.065	-0.043
28	-0.245	-0.657	-0.956	0.049	-0.219	1.008	-0.375	-0.59	0.001	-0.205
29	-3.96	0.601	-1.315	1.034	-0.665	-0.176	-0.032	-0.356	0.08	-0.038
30	-0.249	0.428	-0.152	-0.874	0.304	-0.094	-0.232	-0.172	0.064	-0.041
31	-1.541	-0.639	0.743	0.385	0.223	-0.067	0.306	0.29	0.055	0.05
32	-2.165	1.553	1.025	-1.418	-0.273	-0.54	0.138	0.152	-0.127	0.028
33	-2.158	-1.883	0.435	-1.441	0.079	-0.043	0.136	-0.446	0.067	-0.004
34	1.451	1.873	-1.703	0.669	-0.954	0.52	-0.232	0.538	0.093	0.012
35	1.292	-2.762	-2.855	-0.25	-0.691	0.807	0.214	0.047	-0.391	-0.003
36	-1.869	1.249	-0.304	1.084	0.082	-0.064	0.529	0.219	0.089	0.028
37	2.232	2.31	-0.467	-0.094	-0.814	-0.373	0.213	-0.55	0.007	-0.044
38	0.465	1.154	0.893	-0.748	0.436	-0.245	-0.623	0.187	-0.001	-0.047
39	0.489	1.243	0.165	1.902	0.942	-0.199	-0.551	-0.808	-0.235	0.432

Si se obtienen los estadísticos básicos de esta matriz se obtendrá media cero y varianzas igual a los valores-propios. También es interesante comprobar las consideraciones comentadas anteriormente en el sentido de que los nuevos ejes no están correlacionados y que la matriz de varianza-covarianza contiene a los valores-propios.

Para ratificar lo anterior, a continuación, se muestran ambas matrices para su análisis.

Matriz de Correlación $[Corr_{(\Omega_i)}]$

	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	Ω_8	Ω_9	Ω_{10}
Ω_1	1									
Ω_2	0	1								
Ω_3	0	0	1							
Ω_4	0	0	0	1						
Ω_5	0	0	0	0	1					
Ω_6	0	0	0	0	0	1				
Ω_7	0	0	0	0	0	0	1			
Ω_8	0	0	0	0	0	0	0	1		
Ω_9	0	0	0	0	0	0	0	0	1	
Ω_{10}	0	0	0	0	0	0	0	0	0	1

Matriz de Covarianza $[Cov_{(\Omega_i)}]$

	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	Ω_8	Ω_9	Ω_{10}
Ω_1	4.424									
Ω_2	0	2.201								
Ω_3	0	0	1.694							
Ω_4	0	0	0	0.856						
Ω_5	0	0	0	0	0.291					
Ω_6	0	0	0	0	0	0.256				
Ω_7	0	0	0	0	0	0	0.139			
Ω_8	0	0	0	0	0	0	0	0.096		
Ω_9	0	0	0	0	0	0	0	0	0.028	
Ω_{10}	0	0	0	0	0	0	0	0	0	0.014

De estos resultados, se concluye que no existe correlación entre las variables transformadas $(\Omega_1, \Omega_2, \dots, \Omega_d)$, mientras que las variables iniciales (X_i, Y_i, \dots, Z_i) , están fuertemente correlacionadas (ver matriz de correlación inicial). Asimismo, la transformación aplicada ha tenido el efecto de maximizar la varianza. La varianza en el primer eje principal Ω_1 es 4.424; mayor que en X_1 . Además, no existe ningún otro eje en el que haya una varianza mayor.

RELACIÓN DE LAS MEDICIONES CON LOS COMPONENTES PRINCIPALES

De la misma forma como se realizó para los vectores-propios y las variables; ahora también es posible conocer la aportación de cada elemento, a la formación de cada componente principal. Es decir, cuál de los elementos contribuye más en la varianza de cada uno de los ejes o componentes principales. Por ejemplo, si observamos el primer componente Ω_1 , podemos concluir que los elementos que tienen el valor más grande (en valor absoluto) son los que mayormente contribuyen a la formación del primer componente; estos son los puntos 6, 9, 13, 15, 18 y 29. Si se grafica en un plano coordenado, por ejemplo, los dos primeros componentes; se tiene una representación de la correlación de cada elemento con Ω_1 y Ω_2 . Es interesante observar en la figura, la posición de los puntos mencionados (de mayor contribución a la varianza), con respecto a los componentes principales.

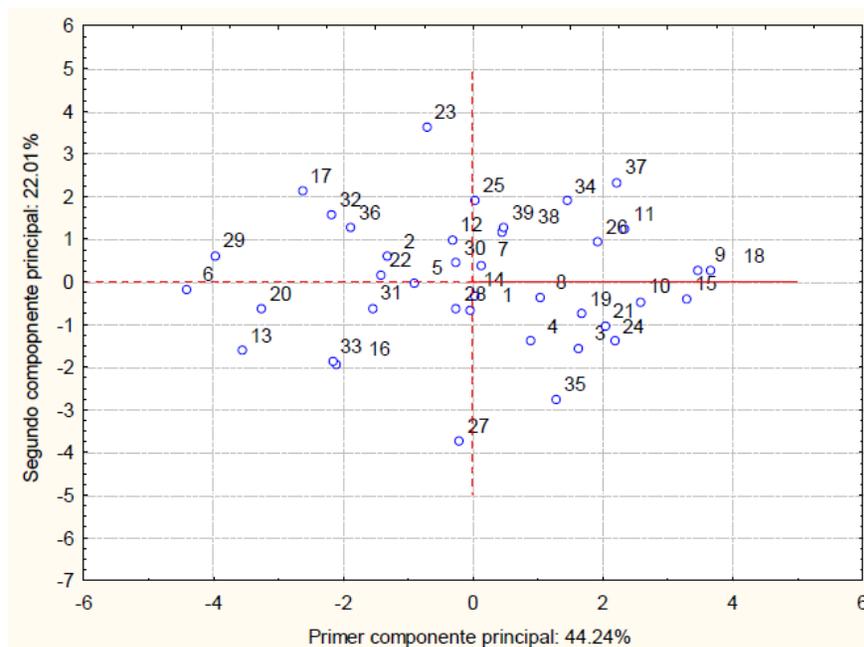


Figura II-2. Relación entre los componentes principales 1-2 y los elementos.

El procedimiento siguiente es realizado con el software STATISTICA. La Figura II-2, representa de manera gráfica, la contribución de cada uno de los elementos en la formación de los componentes principales 1 y 2. En total, este plano 1-2 representa el 66.25 % de la variación total de la muestra. De la misma forma, se puede graficar la posición de los puntos con los componentes principales 1 y 3; esto explicaría el 61.18 % de la variación total de la

muestra (Figura II-3). Posteriormente, se analizará con detalle la interpretación de un plano de componentes principales y de un *círculo de correlaciones*.

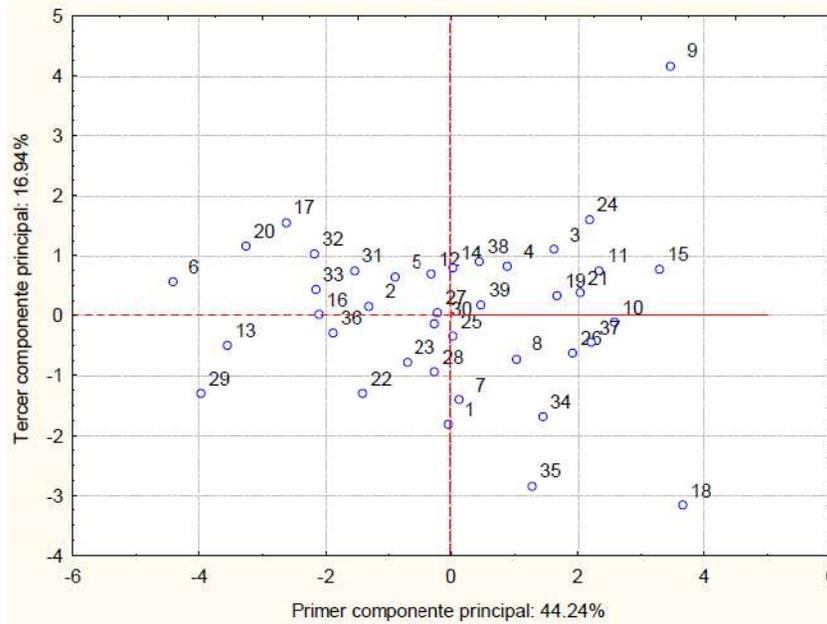


Figura II-3. Relación entre los componentes principales 1-3 y los elementos.

RELACIÓN DE LAS VARIABLES CON LOS COMPONENTES PRINCIPALES

Las variables PMS, H2O, DEP, HEP y GM2 son las que más contribuyen a la formación del primer componente principal y QTE al segundo. Sin embargo, es interesante representar esta contribución de cada una de las variables, en una gráfica similar a la que representa la contribución de cada elemento, con los componentes principales. A esta gráfica se le llama *círculo de correlaciones* y representa la correlación entre los componentes principales y cada una de las variables. El análisis de estas correlaciones nos permite determinar, entre otras cosas, cómo están estructuradas las variables. El cálculo e interpretación de las correlaciones, se lleva a cabo de la manera tradicional: por lo que el coeficiente de correlación se calcula como el cociente de la covarianza, entre el producto de las desviaciones estándar. Por facilidad se recomienda formar la matriz de varianzas-covarianzas.

Como ejemplo tomaremos la matriz de varianzas-covarianzas $[Cov_{(X_1, \Omega_1)}]$ estimada entre la primera variable (HEP) y el primer componente principal esto es:

$$\left[Cov_{(X_1'', \Omega_1)} \right] = \begin{bmatrix} 1.000 & 1.688 \\ 1.688 & 4.424 \end{bmatrix}$$

De esta forma, el coeficiente de correlación es:

$$\rho(X_1'', \Omega_1) = \frac{Cov_{(X_1'', \Omega_1)}}{\sigma(X_1'', X_1'')\sigma(\Omega_1, \Omega_1)} = \frac{1.688}{\sqrt{1.0}\sqrt{4.424}} = 0.8024^*$$

De una forma similar se calcula cada uno de los términos de la matriz de correlación.

Matriz de correlación $\rho(X_1'', \Omega_1) = \left[Corr_{(X_1'', \Omega_1)} \right]$

Matriz de Covarianza $\left[Cov_{(\Omega_i)} \right]$

	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	Ω_8	Ω_9	Ω_{10}
HEP	0.8024*	-0.205	0.345	-0.014	-0.311	0.293	-0.098	-0.051	-0.005	-0.002
DEP	-0.807	0.452	-0.141	0.229	0.032	0.020	-0.171	-0.201	-0.003	-0.026
QTE	-0.101	0.763	-0.426	-0.412	-0.083	0.159	0.147	0.001	0.023	-0.043
H2O	-0.828	0.460	-0.148	0.186	-0.141	0.119	0.041	0.047	-0.030	0.084
PMS	0.843	0.087	-0.201	-0.435	0.158	-0.022	-0.008	-0.149	-0.033	0.057
EM2	-0.091	-0.587	-0.764	0.085	0.103	0.188	-0.029	0.044	-0.085	-0.020
CTE	-0.556	-0.537	-0.384	-0.304	-0.330	-0.229	-0.011	-0.049	-0.003	0.002
GE	-0.603	0.206	0.635	-0.412	0.025	-0.017	-0.075	0.061	-0.100	-0.018
GM2	-0.785	-0.414	0.015	-0.372	0.137	0.171	-0.122	0.041	0.090	0.020
PMG	0.652	0.554	-0.421	-0.041	-0.052	-0.111	0.236	0.137	0.011	0.003

Una vez que se obtiene esta matriz de correlación se representa gráficamente como un círculo unitario, en donde se localizan las variables en un plano formado por los componentes principales que se desean observar. La ilustración 4 muestra el círculo de correlaciones entre los componentes principales Ω_1 y Ω_2 y las variables estudiadas. En el siguiente apartado se estudiará a detalle la interpretación de esta figura.

Es importante señalar que por tratarse de un planteamiento de correlación lineal, es común elevar al cuadrado el coeficiente de correlación, tal y como se hace en un análisis de regresión lineal. Esta operación, se realiza con el objeto de conocer la parte de la variable, que explica

la variación sobre el eje. Para que una variable se considere bien representada en un plano tendrá que tener un coeficiente alto en los componentes principales que forman el plano. En general, para obtener la parte de variación de una variable explicada por k ejes; estará dada por la suma de los cuadrados de los coeficientes de esos k ejes.

Por ejemplo, 68% ($0.6438+0.0422=0.6860$) de la información de la variable HEP está representada sobre un plano formado por los componentes principales Ω_1 y Ω_2 (Figura II-4)

Matriz de correlación $\rho^2(X_i, \Omega_1) = [Corr_{(X_i, \Omega_1)}]^2$:

	Ω_1	Ω_2	Ω_3	Ω_4	Ω_5	Ω_6	Ω_7	Ω_8	Ω_9	Ω_{10}
HEP	0.644	0.042	0.119	0.000	0.097	0.086	0.01	0.003	0.000	0.000
DEP	0.651	0.204	0.02	0.052	0.001	0.000	0.029	0.041	0.000	0.001
QTE	0.010	0.582	0.182	0.17	0.007	0.025	0.022	0.000	0.001	0.002
H2O	0.686	0.212	0.022	0.035	0.02	0.014	0.002	0.002	0.001	0.007
PMS	0.711	0.008	0.04	0.189	0.025	0.000	0.000	0.022	0.001	0.003
EM2	0.008	0.345	0.584	0.007	0.011	0.035	0.001	0.002	0.007	0.000
CTE	0.309	0.288	0.147	0.092	0.109	0.053	0.000	0.002	0.000	0.000
GE	0.364	0.042	0.403	0.17	0.001	0.000	0.006	0.004	0.010	0.000
GM2	0.616	0.172	0.000	0.139	0.019	0.029	0.015	0.002	0.008	0.000
PMG	0.425	0.306	0.177	0.002	0.003	0.012	0.056	0.019	0.000	0.000

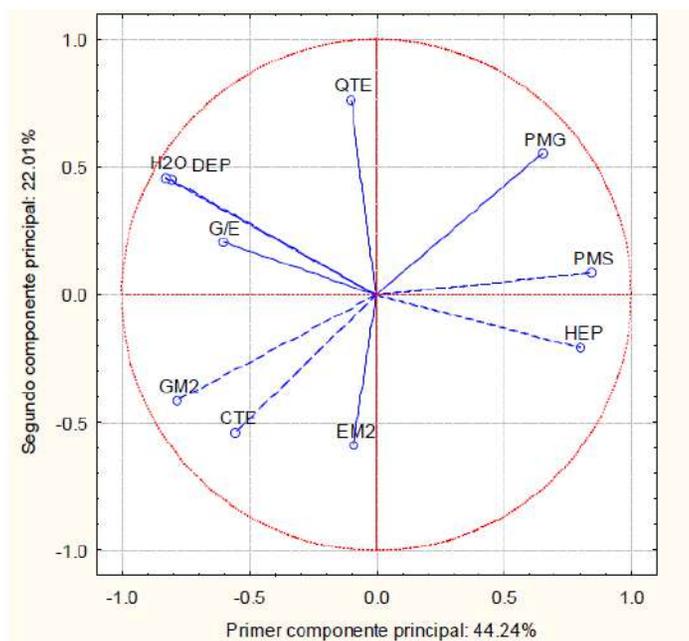


Figura II-4. Circulo de correlaciones entre los componentes principales 1-2.

Otro ejemplo de variables *bien representadas* en el plano (1-2) formado por los componentes principales Ω_1 y Ω_2 , son: DEP y H2O ($\sum \rho^2 > 0.80$). Las variables HEP, PMS, GM" Y PMG también están *bien representadas* en este plano ($0.65 < \sum \rho^2 < 0.80$). Un ejemplo de variables medianamente representadas son: QTE y CTE ya que $\sum \rho^2 = 0.59$. Finalmente se consideran variables pobremente representadas a EM y GE pues $\sum \rho^2 \leq 0.40$.

Antes de pasar al apartado de interpretación, se puede mencionar que, en general, una variable estará bien representada sobre un plano $\Omega_i - \Omega_j$; si se encuentra cerca del círculo de correlaciones.

REPRESENTACIÓN DE LOS UN PLANO DE COMPONENTES PRINCIPALES

La representación gráfica de los resultados de un ACP se lleva a cabo con dos tipos de ilustraciones. La primera, es la representación de los elementos, sobre un plano de ejes coordenados, formados por cada uno de los componentes principales. La segunda, es lo que se conoce como círculo de correlaciones y se forma con los valores del coeficiente de correlación entre los componentes principales y las variables de medición.

Representación de los elementos

Para analizar los resultados de un ACP sobre los elementos, es suficiente construir un plano tomando los componentes principales más importantes, como ejes coordenados (ilustraciones 2 y 3). Lógicamente, el plano más importante será el formado por los componentes 1 y 2; ya que por definición y por construcción, es el plano sobre el cual existe la mayor cantidad de información visible (generalmente se le da el nombre de plano principal). Según la calidad de información que se presenta en este plano, será necesario o no examinar alguna otra combinación de ejes, por ejemplo el plano 3-4. Cabe mencionar que no es suficiente observar cómo se distribuyen estos elementos sobre el plano; también es importante conocer el significado de las variables medidas, ya que éstas juegan un papel importante en la construcción de los componentes principales y por lo tanto, son las que explican la posición de cada uno de los elementos.

La primera relación que se debe observar es la proximidad en el plano de los elementos, dos individuos serán similares mientras más cerca se encuentre uno del otro. En este aspecto es importante analizar si estos individuos también son análogos o presentan alguna relación de similitud en la realidad. Esto puede ocurrir con dos o más elementos, si este fuera el caso, se presume que existe o que se forman grupos de individuos. Esta es justamente una de las aplicaciones del ACP, la formación de grupos homogéneos. Estos resultados son utilizados con frecuencia para la definición de regiones hidrológicamente homogéneas y por lo regular se complementan con un análisis hiperjerárquico de variables y elementos (dendogramas).

Por otro lado, si uno o más elementos se localizan totalmente fuera del resto de los elementos, pueden suceder dos casos. El primero, que los datos originales sean erróneos o que exista algún error de escritura en los valores de la matriz inicial. Si este fuera el caso, se recomienda revisar la muestra y en su caso corregir o retirar la medición y volver a calcular el ACP. El segundo caso, después de verificar los posibles errores y que el elemento continuara fuera del resto de los elementos, se podría tomar la decisión de retirarlo de la muestra y volver a comenzar al ACP; o conservar el elemento en la muestra, ignorándolo y utilizando sólo los elementos restantes.

Círculo de correlaciones

La primera aproximación será observar la matriz de $\rho^2(X_1, \Omega_1)$. Este valor nos indica la *calidad de la representación* de cada una de las variables.

De la misma forma que con los elementos, las variables pueden formar grupos y en este caso se recomienda estudiar por separado la inter-relación de estas variables y en su caso realizar un ACP para cada uno de los grupos de variables formados.

Si dos variables se sitúan en una misma dirección pero con sentido opuesto, significa que mientras una aumenta la otra disminuye.

RECOMENDACIONES

Antes de pretender interpretar los resultados gráficos de un ACP, es importante analizar los estadísticos descriptivos de la muestra de datos (media, varianza, correlación, etc.) Este análisis permite tener una idea general de la magnitud de las variables y de llevar a cabo una rápida verificación de la calidad de las mediciones. También se recomienda observar los

valores-propios y el porcentaje de varianza de cada uno de ellos; esto con la idea de saber cuántos componentes principales se examinarán a detalle.

Debido a que los resultados de un ACP son influenciados por la magnitud de medición de las variables, se recomienda transformar los datos originales para que todos tengan el mismo peso. Esta transformación se lleva a cabo restando a cada valor de medición, el valor de la media (datos centrados) o restando la media y dividiendo entre su desviación estándar (datos centrados-estandarizados). Es importante esta transformación ya que dependiendo de ello, se utilizará para la estimación de los componentes principales; la matriz de correlación o la matriz de varianza-covarianza. La recomendación de utilizar una u otra matriz radica en el origen de las mediciones, es decir, cuando las variables tienen las mismas unidades se recomienda utilizar la matriz de varianza-covarianza, mientras que cuando las variables se miden en diferentes unidades se propone utilizar la matriz de correlación.

Cuando se analiza la correlación entre dos componentes principales y los elementos, a través de una figura de ejes coordenados que forma un plano, se observa la contribución a la varianza de ese plano de cada uno de los elementos. Sin embargo algunos programas de cómputo ofrecen la opción de graficar en 3D, lo que nos ayuda a mejorar la interpretación de la relación entre elementos. Es decir, dos elementos pueden estar cercanos en un plano de componentes principales 1-2, pero muy alejados en un espacio de componentes principales 1-2-3.

ANEXO III

Ejercicios DRH

Método de Ward

A continuación, se muestra el desarrollo de un ejemplo de aplicación con el método de agrupamiento de Ward para cinco sitios (Tabla III-1 y Figura III-1), a los cuales se les asigna el mismo peso, de esta forma se tiene un peso de $1/5$ para cinco sitios, sumando un 1 el total del peso de la muestra.

Tabla III-1. Datos para ejemplo de Ward

PUNTO	x_i	y_i	Peso de la variable
A	3	6	$1/5$
B	8	7	$1/5$
C	4	6	$1/5$
D	4	4	$1/5$
E	7	6	$1/5$

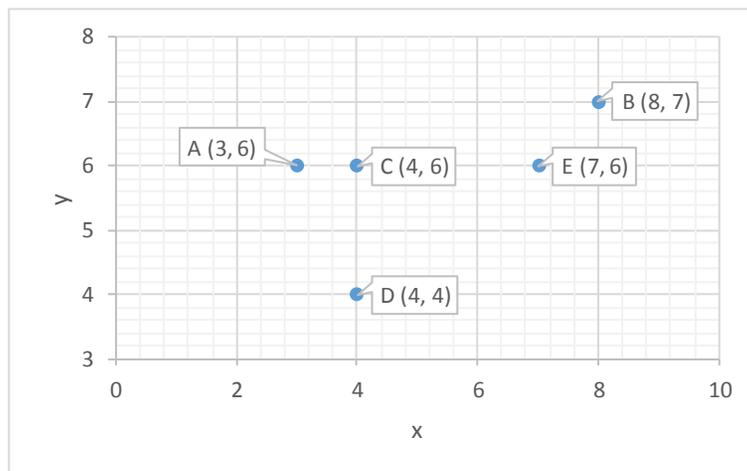


Figura III-1. Distribución espacial de los sitios a agrupar por método de Ward.

Paso 1. Obtener la matriz de distancia euclidiana entre todos los sitios

La distancia euclidiana es calculada entre todos los puntos del sitio a analizar, los datos se colocan en una matriz de datos en la diagonal inferior

$$d_{i,i'}^2 = (x_{i'} - x_i)^2 + (y_{i'} - y_i)^2$$

A continuación, se desarrolla el cálculo para la estimación de la distancia euclidiana de los puntos B, C, D y E respecto al punto A, de la misma forma se calcula para todos los demás puntos, quedando la matriz de distancias formada por los siguientes elementos:

$$d_{AB}^2 = (x_A - x_B)^2 + (y_A - y_B)^2 = (3-8)^2 + (6-7)^2 = 26$$

$$d_{AC}^2 = (x_A - x_C)^2 + (y_A - y_C)^2 = (3-4)^2 + (6-6)^2 = 1$$

$$d_{AD}^2 = (x_A - x_D)^2 + (y_A - y_D)^2 = (3-4)^2 + (6-4)^2 = 5$$

$$d_{AE}^2 = (x_A - x_E)^2 + (y_A - y_E)^2 = (3-7)^2 + (6-6)^2 = 16$$

Se selecciona el par de puntos que mantienen la menor distancia entre sí, en este caso el punto A y C.

	A	B	C	D	E
A	0				
B	26	0			
C	1	17	0		
D	5	25	4	0	
E	16	2	9	13	0

Paso 2. Obtener la matriz de pérdida de inercia mínima

Calcular la inercia inicial con la siguiente ecuación

$$\Delta I_{min} = (h, h') = \frac{m_h m_{h'}}{m_h + m_{h'}} d^2(g_h, g_{h'})$$

Donde m son las masas de cada elemento h y h' ; g son los centros de gravedad de cada elemento h y h' ; d^2 es el cuadrado de la distancia euclidiana entre los centros de gravedad.

Este criterio representa el momento de inercia $M_g [g_h, g_{h'}]$ entre dos puntos (elementos) h y h' con sus respectivas masas m_h y $m_{h'}$, con relación a sus centros de gravedad g_h y $g_{h'}$ de la siguiente forma:

$$g = \frac{g_h m_h + g_{h'} m_{h'}}{m_h + m_{h'}}$$

Se muestra el cálculo para el elemento A-B:

$$\Delta I_{A-B}^{min} = \frac{(1/5)(1/5)}{1/5+1/5}(26) = \frac{(0.2)(0.2)}{0.2+0.2}(26) = 2.6$$

$$\Delta I_{A-B}^{min} = (0.1)(26) = 2.6$$

De forma análoga se calcula para los demás elementos:

$$\Delta I_{A-C} = (0.1)(1) = 0.1$$

$$\Delta I_{B-C} = (0.1)(17) = 1.7$$

$$\Delta I_{C-D} = (0.1)(4) = 0.4$$

$$\Delta I_{A-D} = (0.1)(5) = 0.5$$

$$\Delta I_{B-D} = (0.1)(25) = 2.5$$

$$\Delta I_{C-E} = (0.1)(9) = 0.9$$

$$\Delta I_{A-E} = (0.1)(16) = 1.6$$

$$\Delta I_{B-E} = (0.1)(2) = 0.2$$

$$\Delta I_{D-E} = (0.1)(13) = 1.3$$

Se obtiene la siguiente matriz de pérdida inercia mínima y se selecciona el menor valor, es este caso el punto A-C con valor de 0.1.

Tabla III-2. Matriz de pérdida de inercia mínima inicial

	A	B	C	D	E
A	0.0				
B	2.6	0.0			
C	0.1	1.7	0.0		
D	0.5	2.5	0.4	0.0	
E	1.6	0.2	0.9	1.3	0.0

A partir la primera agregación jerárquica de elementos se desarrolla el cálculo de la inercia entre puntos, considerando el agrupamiento de elementos, por lo cual la masa se incrementa, la nueva notación será la siguiente, donde $i \cup i'$ forma el nuevo elemento j , de la siguiente forma:

$$\Delta I_{j,r} = \frac{1}{m_j + m_r} \left[(m_i + m_r) \Delta I_{ir} + (m_{i'} + m_r) \Delta I_{i'r} - m_r \Delta I_{ii'} \right]$$

donde $j = i + i'$, es el nuevo elemento, r es el elemento objetivo, con respecto al cual se calcula la inercia, de la misma forma m_j será el total del peso de la variable debido a la suma de las masas $m_i + m_{i'}$ y m_r es el peso del elemento r .

Se unen los elementos A-C para formar el punto 6, y se forman la matriz de distancia euclidiana al cuadrado y la matriz de inercia:

$$j = i + i'$$

$$j = 6 \quad i = A \quad i' = C$$

$$m_6 = m_A + m_C = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}$$

Con el planteamiento anterior se calcula la matriz de inercia del elemento $j = 6$ a los elementos r , es decir, cuando $r = B$, $r = D$ y cuando $r = E$ e ingresar el resultado en las celdas correspondientes a 6B, 6D y 6E.

Desarrollo:

Inercia de B a 6

$$j = 6$$

$$r = B$$

$$\Delta I_{6,B} = \frac{1}{m_6 + m_B} \left[(m_A + m_B) \Delta I_{AB} + (m_C + m_B) \Delta I_{CB} - (m_B) \Delta I_{AC} \right]$$

$$\Delta I_{6,B} = \frac{1}{0.4 + 0.2} \left[(0.2 + 0.2) 2.6 + (0.2 + 0.2) 1.7 - (0.2) 0.1 \right]$$

$$\Delta I_{6,B} = 2.83$$

Inercia de D a 6:

$$j = 6$$

$$r = D$$

$$\Delta I_{6,D} = \frac{1}{m_6 + m_D} \left[(m_A + m_D) \Delta I_{AD} + (m_C + m_D) \Delta I_{CD} - (m_D) \Delta I_{AC} \right]$$

$$\Delta I_{6,D} = \frac{1}{0.4+0.2} [(0.2+0.2)0.5 + (0.2+0.2)0.4 - (0.2)0.1]$$

$$\Delta I_{6,D} = 0.57$$

Inercia de E a 6:

$$j = 6$$

$$r = E$$

$$\Delta I_{6,E} = \frac{1}{m_6 + m_E} [(m_A + m_E)\Delta I_{AE} + (m_C + m_E)\Delta I_{CE} - (m_E)\Delta I_{AC}]$$

$$\Delta I_{6,E} = \frac{1}{0.4+0.2} [(0.2+0.2)1.6 + (0.2+0.2)0.9 - (0.2)0.1]$$

$$\Delta I_{6,E} = 1.63$$

Con los valores de inercia calculados para el elemento 6, se completa la matriz de inercia, se representan en color azul los valores calculados, mientras que el resto son los valores originales de la inercia inicial y se elige el mínimo valor, siendo 6-D, por lo que estos elementos se agrupan formando el elemento 7

	6	B	D	E
6	0			
B	2.83	0		
D	0.57	2.50	0	
E	1.63	0.20	1.30	0

El siguiente paso consiste en calcular la inercia entre los puntos $j = 7$ a los elementos r , es decir, cuando $r = 6$, $r = E$ y cuando $r = D$ e ingresar el resultado en las celdas correspondientes a 7D y 76.

$$j = 7 \quad r = D \quad i = B \quad i' = E$$

$$m_7 = m_B + m_E = 0.2 + 0.2 = 0.4$$

$$\Delta I_{7,D} = \frac{1}{m_7 + m_D} [(m_B + m_D)\Delta I_{BD} + (m_E + m_D)\Delta I_{ED} - (m_D)\Delta I_{BE}]$$

$$\Delta I_{6,B} = \frac{1}{0.4+0.2} [(0.2+0.2)2.5 + (0.2+0.2)1.3 - (0.2)0.2]$$

$$\Delta I_{6,B} = 2.47$$

Inercia de 7 a 6:

$$j = 7 \quad r = 6 \quad i = B \quad i' = E$$

$$m_j = m_7 = m_B + m_E = 0.2 + 0.2 = 0.4$$

$$m_r = m_6 = m_A + m_C = 0.2 + 0.2 = 0.4$$

$$\Delta I_{7,D} = \frac{1}{m_7 + m_6} \left[(m_B + m_6) \Delta I_{B6} + (m_E + m_6) \Delta I_{E6} - (m_6) \Delta I_{BE} \right]$$

$$\Delta I_{7,D} = \frac{1}{0.4 + 0.4} \left[(0.2 + 0.4) 2.83 + (0.2 + 0.4) 1.63 - (0.4) 0.2 \right]$$

$$\Delta I_{7,D} = 3.25$$

Resultando la siguiente matriz con los valores de inercia nuevos respecto a la nueva agrupación, por lo que se elige el menor valor, el cual es 6-D con inercia de 0.57, el nuevo elemento se denomina 8.

	6	7	D
6	0		
B	3.25	0	
D	0.57	2.47	0

Con lo anterior es necesario calcular la inercia entre el elemento 8 y el previo, el elemento 7, de la siguiente forma:

$$r = 7; \quad j = 8; \quad i = D; \quad i' = 6$$

$$m_j = m_8 = m_D + m_6 = 0.2 + 0.4 = 0.6$$

$$m_r = m_7 = m_B + m_E = 0.2 + 0.2 = 0.4$$

$$\Delta I_{8,7} = \frac{1}{m_8 + m_7} \left[(m_D + m_7) \Delta I_{D7} + (m_6 + m_7) \Delta I_{67} - (m_7) \Delta I_{D6} \right]$$

$$\Delta I_{8,7} = \frac{1}{0.6 + 0.4} \left[(0.2 + 0.4) 2.47 + (0.4 + 0.4) 3.25 - (0.4) 0.57 \right]$$

$$\Delta I_{8,7} = 3.85$$

Resultando la matriz siguiente:

	7	8
7	0	
8	3.85	0

Mediante un dendograma (Figura III-2) es factible representar el procedimiento de agrupamiento entre las variables.

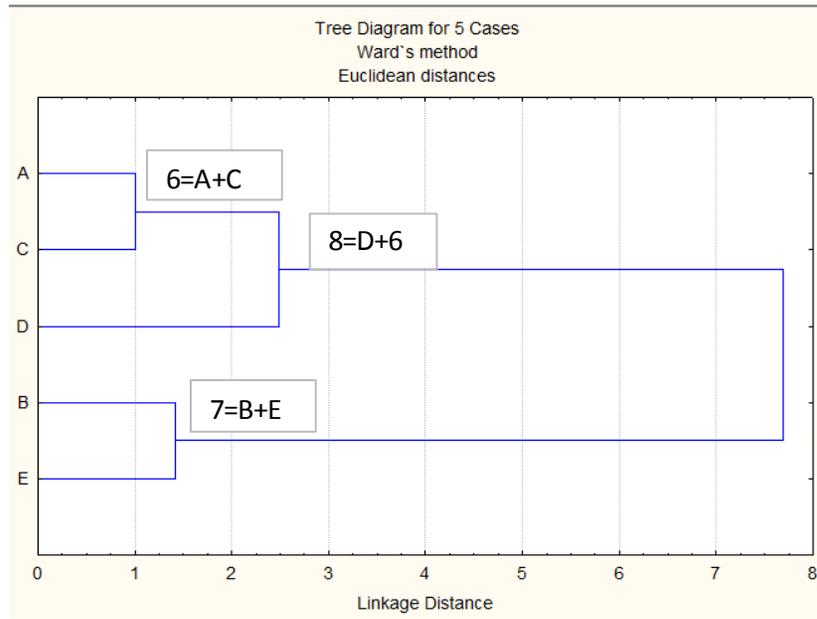


Figura III-2. Dendograma de agrupación por método de Ward.

A manera de resumen se muestran los distintos valores de inercia que fueron calculados en el proceso (Tabla III-3). Es posible observar que la variable j será el agrupamiento formado por medio de la inercia mínima previa, mientras que la variable r corresponderá al elemento con respecto a la cual se calculará la nueva inercia.

Tabla III-3. Resumen de valores de inercia calculados.

AGRUPACIÓN	6			7		8
r	B	D	E	6	D	7
j	6	6	6	7	7	8
i	A	A	A	B	B	D
i'	C	C	C	E	E	6
m_r	0.2	0.2	0.2	0.4	0.2	0.4
m_j	0.4	0.4	0.4	0.4	0.4	0.6

Método de Ward aplicado a datos de precipitación

De forma similar, pero ahora para 46 elementos se realiza un ejemplo de agrupación de elementos homogéneos mediante el método de Ward para 46 estaciones climatológicas con el objetivo de caracterizar la precipitación media anual en la RH10, denominada Sinaloa. Esta aplicación se desarrolló para la caracterización del sitio de estudio de esta investigación.

La República Mexicana está dividida en 13 Regiones Hidrológico-Administrativas y en 37 regiones hidrológicas. La zona de estudio seleccionada corresponde a la Región Hidrológica número 10 (RH10) y pertenece a la Región Hidrológico-Administrativa III, denominada Sinaloa y Pacífico Norte, respectivamente.

La RH10 (ver Figura III-3) se encuentra localizada al noroeste de México, tiene una superficie de 103055,56 km², pertenecen a esta zona la totalidad del estado de Sinaloa y parte de los estados de Durango, Chihuahua y Sonora. Las principales corrientes superficiales las representan los ríos: Fuerte, Sinaloa, Mocorito, Culiacán, San Lorenzo, Elota, Piaxtla, Quelite, Presidio, Baluarte, Cañas, Acaponeta y San Pedro.

Mediante la metodología expuesta anteriormente se realizó la extracción de datos de precipitación mensual y anual por medio del extractor rápido de información climatológica, ERIC, creado por la Comisión Nacional del Agua (IMTA, 1988; Ouarda et al., 2008). Se consideraron todas las estaciones que cumplían con un periodo de registro mínimo de 10 años, además de que se eliminaron todas aquellas estaciones que contenían datos faltantes, por lo que un total de 46 estaciones climatológicas cumplen con el periodo de información definido como aceptable.

Para la caracterización fisiográfica de la zona de estudio se extrajeron datos del Instituto Nacional de Estadística y Geografía, INEGI. En la Tabla III-4 se muestran las variables empleadas y sus principales características, a diferencia del primer ejemplo que contaba solo con la ubicación de latitud y longitud del elemento, contando con solo 2 características.

Tabla III-4. Variables empleadas en la delimitación de regiones homogéneas.

Variable	Características de la variable			
	Mínimo	Máximo	Unidades	Fuente
Promedio anual de precipitación	298.57	1458.78	mm	SMN*

Latitud	23.03	27.40	m	INEGI
Longitud	-105.75	-109.18	m	INEGI
Altitud de la estación de aforo	2.00	2700.00	m	INEGI
Distancia al océano	1.00	224.00	km	INEGI*

(*) Variable que adicionalmente incluye procedimiento de análisis de datos de la fuente consultada.

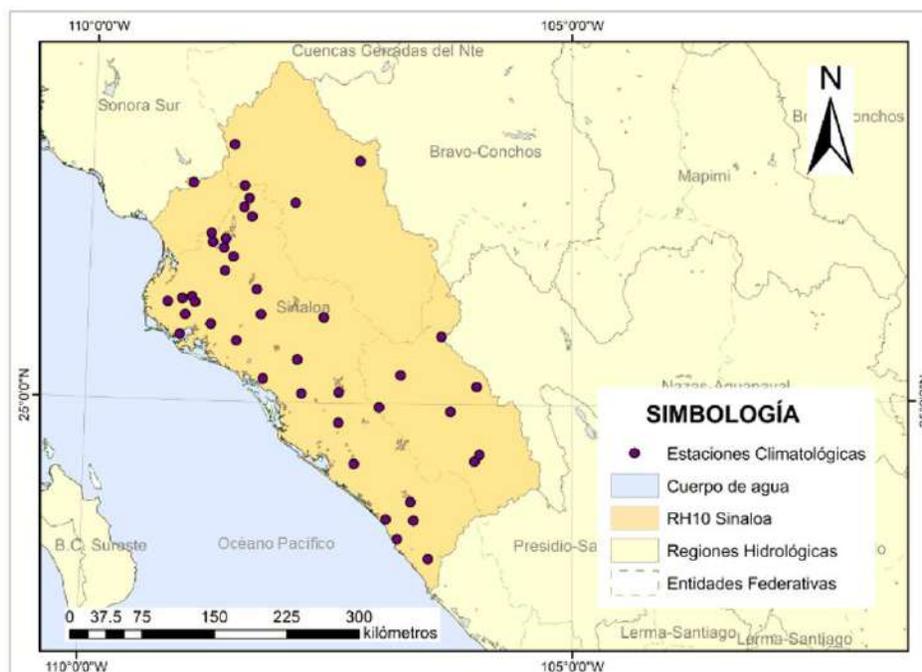


Figura III-3. Mapa de ubicación de la Región Hidrológica 10, Sinaloa, México.

Para cada una de las estaciones climatológicas se aplicaron las ecuaciones descritas en la metodología, en la Tabla III-5 se muestra el nombre y clave de las estaciones, así como sus características descriptivas: latitud, longitud, altitud de la estación de aforo, distancia de la estación al océano y la precipitación media anual. En la Tabla III-6 se muestran los datos estadísticos de las estaciones climatológicas.

Tabla III-5. Estaciones climatológicas de la Región Hidrológica 10 empleadas

Clave	Nombre de la estación	Latitud	Longitud	Altitud	n
25001	Acatitán. San Ignacio	24.083	-106.65	130	24
25003	Ahome. Ahome	25.917	-109.18	33	22
26002	Alamos. Alamos	27.033	-108.95	389	20
25007	Bamícori. El Fuerte	26.35	-108.52	152	25
25002	Agua caliente. Choix	26.817	-108.42	180	16
10018	El Cantil. Papasquiario	24.933	-106.25	2100	32
26017	Casante. Álamos	26.566	-108.75	137	15
8167	Chínipas. Chínipas	27.4	-108.53	480	18

25017	Choix pueblo. Choix	26.73	-108.33	270	13
25011	Concordia	23.266	-106.05	138	26
25080	Ruíz Cortines. Guasave	25.72	-108.73	35	19
25050	La Cruz. Elota	23.92	-106.9	23	17
25015	Culiacán. Culiacán	24.817	-107.4	40	24
25021	Dimas. San Ignacio	23.733	-106.78	12	19
25023	El Fuerte. El Fuerte	26.433	-108.67	84	23
25038	Guasave. Guasave	25.567	-108.47	20	13
25058	Las Habitas. Rosario	23.033	-105.75	80	16
10031	Huahuapan. San Dimas	24.53	-105.95	1150	18
25044	Huites. Choix	26.90	-108.37	268	26
25045	Ixpalino. San Ignacio	23.91	-106.62	80	25
10071	San Miguel de Lobos	25.17	-105.98	2410	22
25025	El Mahome. El Fuerte	26.52	-108.6	120	25
25065	Mochicahui. El Fuerte	25.97	-108.93	22	19
25116	Los Mochis. Ahome	25.80	-109	14	18
8124	San Ignacio. Batopilas	26.87	-107.88	355	19
25070	Pánuco. Sinaloa	23.42	-105.93	450	18
25068	Palo dulce. Choix	27.02	-108.47	237	24
25071	Pericos. Moco rito	25.08	-107.78	35	24
25030	El Playón. Sinaloa	25.22	-108.18	2	18
25031	El Quelite. Mazatlán	23.55	-106.47	60	25
25076	Quila. Culiacán	24.43	-107.23	45	15
25077	Rosa Morada. Moco rito	25.40	-107.83	300	14
25032	El Sabino(El Sabinito)	26.48	-108.73	123	17
10079	Sianori. Topia	25.27	-106.777	850	18
25090	Sinaloa de Leyva. Sin.	25.82	-108.22	80	12
8138	Siquirichic. Guachochic	27.27	-107.22	2120	19
25086	San Miguel Zapotitlán	25.95	-109.03	38	19
25088	Santa Rosa. El Fuerte	25.92	-108.9	28	19
25093	Surutato. Badiraguato	25.8	-107.57	1400	23
10082	Tamazula. Tamazula	24.967	-106.98	250	38
10026	El Tarahumar. Tepehuanes	25.633	-106.35	2560	20
25053	La Tina. El Fuerte	26.217	-108.6	180	23
25098	Topolobampo. Ahome	25.617	-109.05	34	14
10042	Las Truchas. San Dimas	24.467	-106	2700	24
25054	La Vainilla. Sinaloa	26.05	-108.27	185	23
25033	El Varejonal. Badiraguato	25.1	-107.4	160	25

Tabla III-6. Datos estadísticos de las estaciones climatológicas de la Región Hidrológica

Clave	D	\bar{x}	S	C_s	K	C_v
25001	30	796.77	205.86	0.7725	3.8815	0.2584
25003	25	342.42	119.50	0.5522	3.7616	0.349
26002	64	683.32	176.68	1.2248	7.4759	0.2586
25007	76	714.08	171.69	0.4689	3.3369	0.2404
25002	90	734.81	213.33	-0.4286	8.9917	0.2903
10018	133	1458.78	292.33	0.6181	4.2344	0.2004
26017	52	658.32	145.53	-0.4873	4.2215	0.2211

8167	121	979.93	297.23	0.1536	3.4432	0.3033
25017	95	814.96	236.65	1.1208	7.9508	0.2904
25011	37	832.31	201.93	0.601	4.1683	0.2426
25080	36	390.01	130.59	1.9545	8.0673	0.3348
25050	4	499.05	239.78	1.2339	5.8757	0.4805
25015	52	677.65	170.13	1.2882	5.1121	0.2511
25021	4	672.73	298.81	0.7563	4.8743	0.4442
25023	64	638.67	185.54	0.8903	5.5981	0.2905
25038	30	492.74	176.70	1.1828	6.0855	0.3586
25058	35	1159.96	224.29	0.9072	5.0147	0.1934
10031	122	781.94	180.76	0.6743	3.6502	0.2312
25044	103	828.39	230.66	1.1944	5.1989	0.2784
25045	33	774.77	184.25	0.7508	3.042	0.2378
10071	167	875.19	208.43	0.7987	4.4669	0.2382
25025	65	617.62	159.96	-0.1276	3.4431	0.259
25065	47	298.57	81.01	-0.4773	3.9372	0.2713
25116	21	379.02	140.36	0.781	7.0741	0.3703
8124	144	636.86	144.92	0.7022	3.275	0.2276
25070	51	1402.99	342.56	-0.2841	5.6865	0.2442
25068	105	960.99	244.98	0.7706	3.3661	0.2549
25071	50	655.23	178.64	0.7968	3.6685	0.2726
25030	7	611.78	145.50	0.3116	4.7857	0.2378
25031	13	709.17	155.96	0.6913	5.2359	0.2199
25076	27	557.05	150.99	2.2623	9.0787	0.271
25077	55	750.89	177.35	1.3174	7.9823	0.2362
25032	56	555.50	185.77	0.1096	3.2424	0.3344
10079	129	1131.53	258.87	0.943	5.2502	0.2288
25090	68	554.60	215.51	1.1663	9.6503	0.3886
8138	224	493.19	154.29	1.2354	8.4041	0.3128
25086	35	436.58	129.69	1.4047	8.5482	0.2971
25088	52	415.93	134.77	1.7358	11.204	0.324
25093	110	1263.09	231.47	0.8192	4.0171	0.1833
10082	90	989.43	248.95	0.2373	2.9651	0.2516
10026	188	895.82	219.94	0.9247	4.7695	0.2455
25053	68	636.27	210.32	0.9522	3.5252	0.3306
25098	1	310.47	137.26	0.8954	10.0046	0.4421
10042	116	1197.76	226.13	0.6365	2.991	0.1888
25054	90	793.18	152.01	0.2115	2.9852	0.1916
25033	78	908.89	250.47	0.4592	3.6623	0.2756

Donde n: Longitud de registro, D: distancia al océano.

Resultando tres regiones homogéneas (Figura III-4) por medio de la clasificación jerárquica ascendente, donde el dendograma de clasificación se muestra en la Figura III-5:

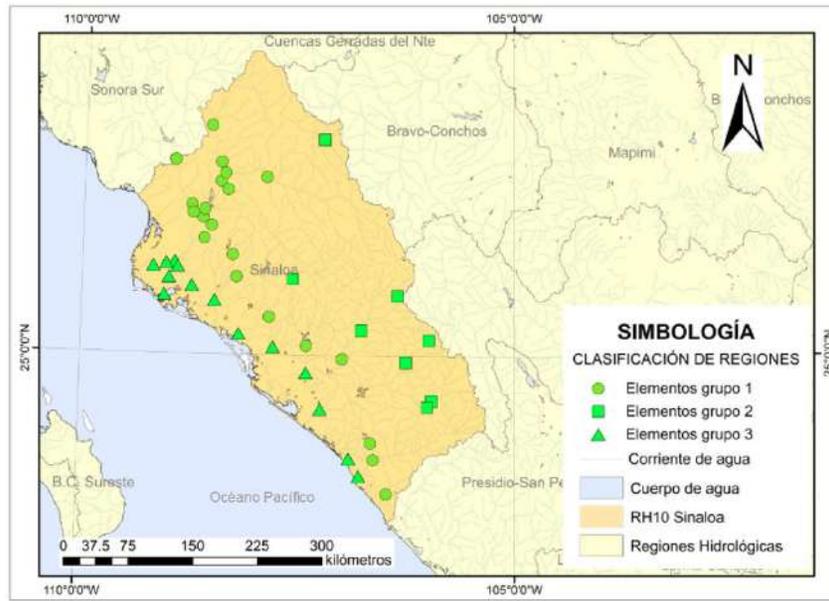


Figura III-4. Agrupación de regiones homogéneas en la Región Hidrológica 10, Sinaloa

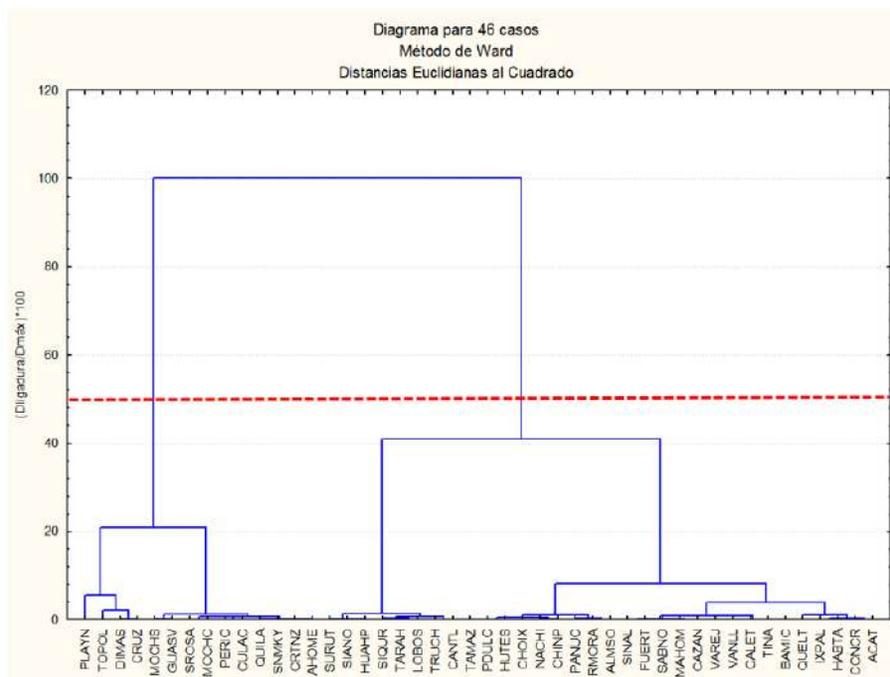


Figura III-5. Agrupación jerárquica de las estaciones climatológicas mediante el método de Ward

Se definieron tres agrupaciones homogéneas para la región hidrológica (Tabla III-7), la característica de este método es que permite al investigador definir qué cantidad de grupos requiere en función de la distancia de ligadura, para lo cual en este desarrollo se definió una distancia de 50 km, donde el grupo con mayor cantidad de elementos fue el formado para la

región 1, que se localiza a una altitud intermedia con respecto a las demás estaciones (Figura III-6), además, las estaciones cumplen con el criterio de homogeneidad que establece que la relación regional entre la desviación estándar con el coeficiente de variación medio no deben exceder un valor de 0.3, por lo que se verifica la homogeneidad en cada zona (Tabla III-7).

Tabla III-7. Resultados de la clasificación por el método de Ward.

N	Región homogénea 1		Región homogénea 2	Región homogénea 3
1	MAHOM	ACAT	CANTL	AHOME
2	NACHI	ALMSO	HUAHP	CRTNZ
3	PANUC	BAMIC	LOBOS	CRUZ
4	PDULC	CALET	SIANO	CULAC
5	QUELT	CAZAN	SIQUR	DIMAS
6	RMORA	CHINP	SURUT	GUASV
7	SABNO	CHOIX	TARAH	MOCHC
8	SINAL	CONCR	TRUCH	MOCHS
9	TAMAZ	FUERT		PERIC
10	TINA	HABTA		PLAYN
11	VANLL	HUTES		QUILA
12	VAREJ	IXPAL		SNMKY
13				SROSA
14				TOPOL
N	24		8	14
CV	0.263		0.229	0.336
σ_{cv}	0.044		0.039	0.074
CC	0.169		0.169	0.220

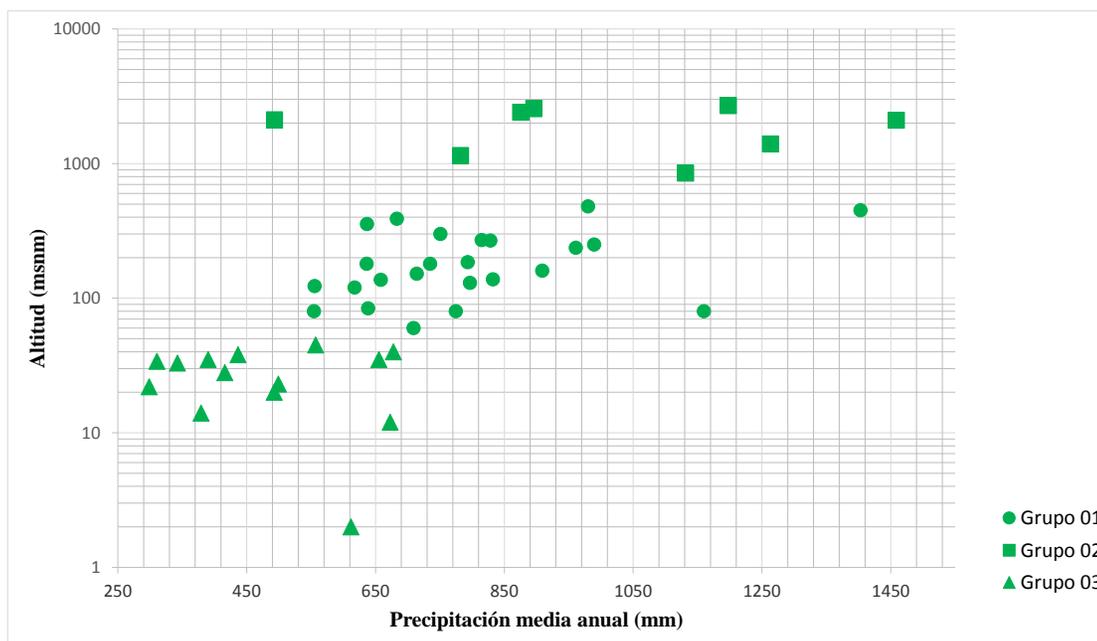


Figura III-6. Gráfico de agrupación de regiones de altitud respecto a la precipitación media anual.

Ejercicio de Método de K-Medias

Ejemplo caso: en total se tienen 10 puntos (ver Tabla III-8 y Figura III-7) el objetivo consiste en formar dos grupos homogéneos, es decir, $k=2$.

Tabla III-8. Coeficientes Momentos-L Datos estadísticos

ELEMENTO	x_i	y_i	$e_i(x_i, y_i)$
1	3	4	3,4
2	2	3	2,3
3	6	3	6,3
4	1	4	1,4
5	4	3	4,3
6	5	4	5,4
7	2	5	2,5
8	6	1	6,1
9	1	2	1,2
10	5	2	5,2

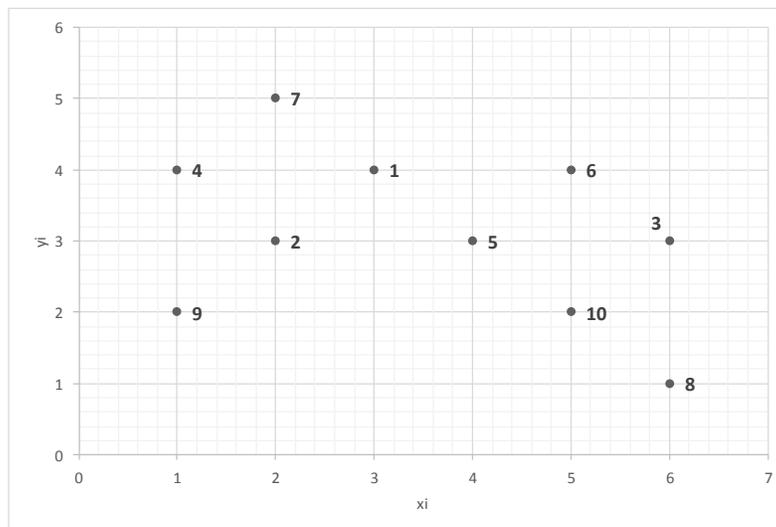


Figura III-7. Ejemplo K-medias

Breve descripción del método. A partir de un conjunto de elementos, con ciertas características, específicamente de ubicación para este ejemplo, se parte de un número arbitrario de formación de grupos homogéneos, es decir, el parámetro K , donde se elegirán K centros de grupos, la calidad de la partición, $Q(P)$, se evalúa en cada iteración, las iteraciones se concluyen hasta que se obtenga el menor valor de calidad de partición.

La nomenclatura a emplear es la siguiente:

$$[\text{centro de grupo}] C_{\substack{\text{número de partición} \\ \text{número de k-grupo}}}$$

PASO 01. El primer paso consiste en obtener la matriz de distancia euclidiana con la siguiente ecuación:

$$d_{i,i'} = \sqrt{(x_{i'} - x_i)^2 + (y_{i'} - y_i)^2}$$

A manera de ejemplo, se muestra el desarrollo para calcular la distancia del punto 4 al punto 3.

$$d_{4,3} = \sqrt{(x_3 - x_4)^2 + (y_3 - y_4)^2}$$

$$d_{4,3} = \sqrt{(6-1)^2 + (3-4)^2}$$

$$d_{4,3} = 26$$

Matriz de distancia euclidiana:

i\i'	1	2	3	4	5	6	7	8	9	10
1	0	2	10	4	2	4	2	18	8	8
2	2	0	16	2	4	10	4	20	2	10
3	10	16	0	26	4	2	20	4	26	2
4	4	2	26	0	10	16	2	34	4	20
5	2	4	4	10	0	2	8	8	10	2
6	4	10	2	16	2	0	10	10	20	4
7	2	4	20	2	8	10	0	32	10	18
8	18	20	4	34	8	10	32	0	26	2
9	8	2	26	4	10	20	10	26	0	16
10	8	10	2	20	2	4	18	2	16	0

PASO 02. Para los k grupos a formar, es decir, los 2 grupos a formar, se elige arbitrariamente el elemento 7 y 8 como centros de grupo. Se selecciona cada elemento uno a uno hasta formar los elementos de los k-grupos de la siguiente forma:

$$\sum_{i \in E} d^2(i, 7) = \min_{i' \in E} \sum_{i \in E} d^2(i, i')$$

$$\sum_{\substack{i \in E \\ i \neq 7}} d^2(i, 8) = \min_{i' \in E} \sum_{\substack{i \in E \\ i \neq 7}} d^2(i, i')$$

Desarrollando lo anterior se

$$[7]C_1^1 = [e_1 \ e_2 \ e_4 \ e_5 \ e_6 \ e_7 \ e_9] = [2 \ 4 \ 2 \ 8 \ 10 \ 0 \ 10] = 2+4+2+8+10+0+10 = 36$$

$$[8]C_2^1 = [e_3 \ e_8 \ e_{10}] = [4 \ 0 \ 2] = 4+0+2 = 6$$

$$Q(P_1) = [7]C_1^1 + [8]C_2^1 = 36+6$$

$$Q(P_1) = 42$$

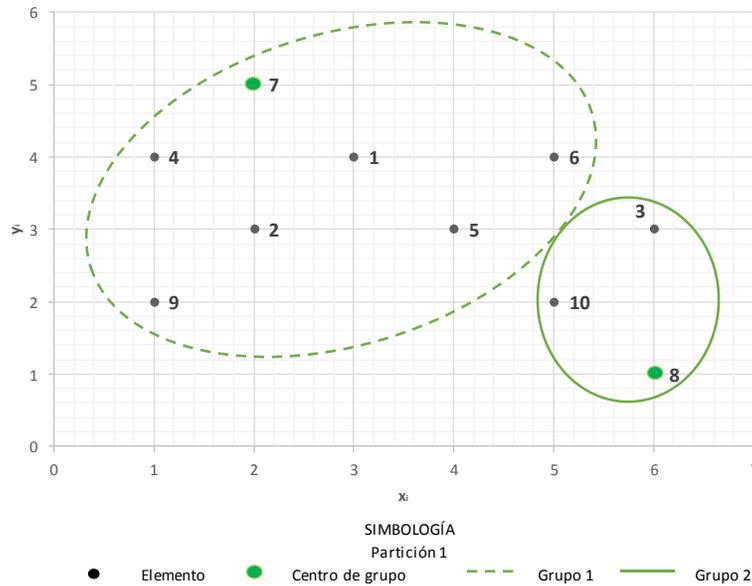


Figura III-8. Partición 1, dos grupos homogéneos

La calidad de la partición, $Q(P)$, representa un coeficiente aproximado a la varianza intra-grupo; considerando que R_r es el centro de gravedad del grupo r .

Una partición P es mejor, mientras menor sea el valor de Q . Esto es, grupos homogéneos entre sus elementos y diferentes entre sus grupos.

Ahora se procede a obtener una segunda partición, conservando el centro de grupo con menor valor de partición, el centro de grupo 8, y del grupo formado con mayor valor de partición se elige un nuevo centro de grupo, en este caso al elemento 4 para representar al grupo a formar, quedando de la siguiente forma:

$$[8]C_1^2 = [e_3 \ e_5 \ e_6 \ e_8 \ e_{10}] = [4 \ 8 \ 10 \ 0 \ 2] = 4+8+10+0+2 = 24$$

$$[4]C_2^2 = [e_1 \ e_2 \ e_4 \ e_7 \ e_9] = [4 \ 2 \ 0 \ 2 \ 4] = 4+2+0+2+4 = 12$$

$$Q(P_2) = [8]C_1^2 + [4]C_2^2 = 24+12$$

$$Q(P_2) = 36$$

$$Q(P_2) < Q(P_1)$$

$$36 < 42$$

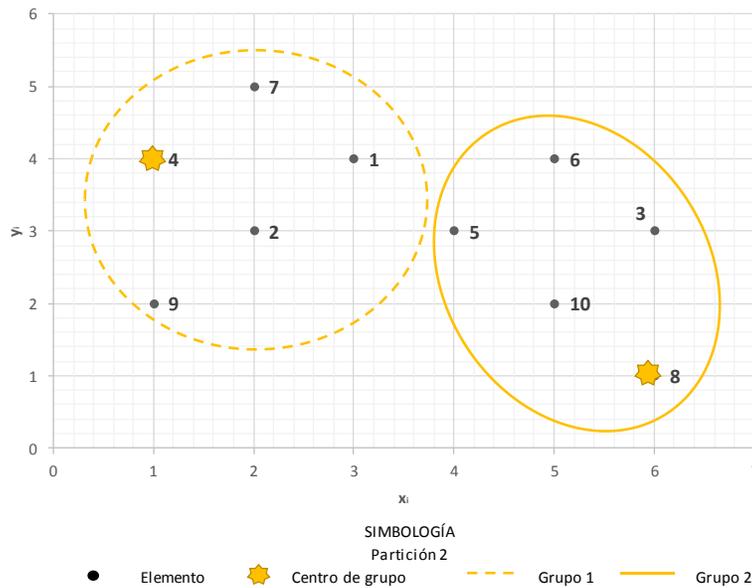


Figura III-9. Partición 2, dos grupos homogéneos.

Debido a que la calidad de la partición dos es mejor que la uno, a partir de P_2 conservamos el centro de grupo formado por el elemento 4 y se elige al elemento 5 como centro de grupo para la partición P_3 :

$$[4]C_1^3 = [e_1 \ e_2 \ e_4 \ e_7 \ e_9] = [4 \ 2 \ 0 \ 2 \ 4] = 4+2+0+2+4 = 12$$

$$[10]C_2^3 = [e_3 \ e_5 \ e_6 \ e_8 \ e_{10}] = [2 \ 2 \ 4 \ 2 \ 0] = 2+2+4+2+0 = 10$$

$$Q(P_3) = [4]C_1^4 + [10]C_2^4 = 12+10$$

$$Q(P_3) = 22$$

$$Q(P_3) < Q(P_2)$$

$$22 < 36$$

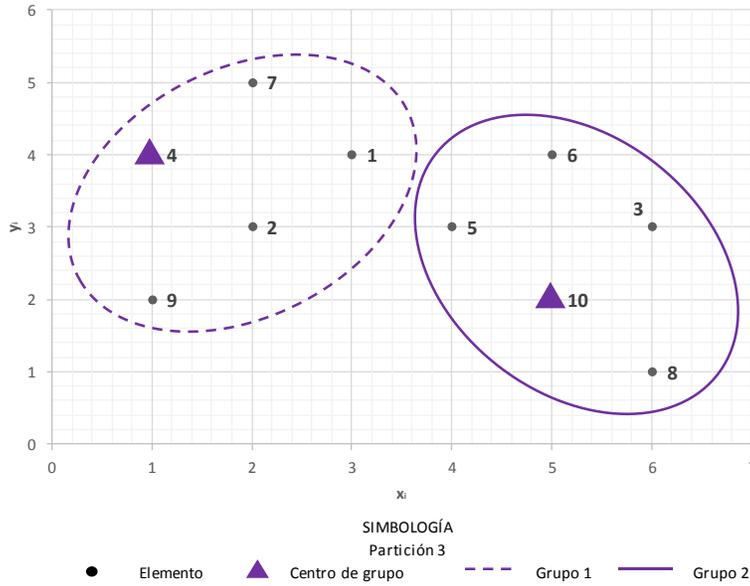


Figura III-10. Partición 3, dos grupos homogéneos.

La calidad de la partición tres es mejor que las anteriores, por lo que a partir de P_3 se conserva el centro de grupo formado por el elemento 4 y se elige al elemento 10 como centro de grupo para la partición P_4 :

$$[10]C_1^4 = [e_3 \ e_5 \ e_6 \ e_8 \ e_{10}] = [2 \ 2 \ 4 \ 2 \ 0] = 2 + 2 + 4 + 2 + 0 = 10$$

$$[2]C_2^4 = [e_1 \ e_2 \ e_4 \ e_7 \ e_9] = [2 \ 0 \ 2 \ 4 \ 2] = 2 + 0 + 2 + 4 + 2 = 10$$

$$Q(P_4) = [10]C_1^4 + [2]C_2^4 = 10 + 10$$

$$Q(P_4) = 20$$

$$Q(P_4) < Q(P_3)$$

$$20 < 22$$

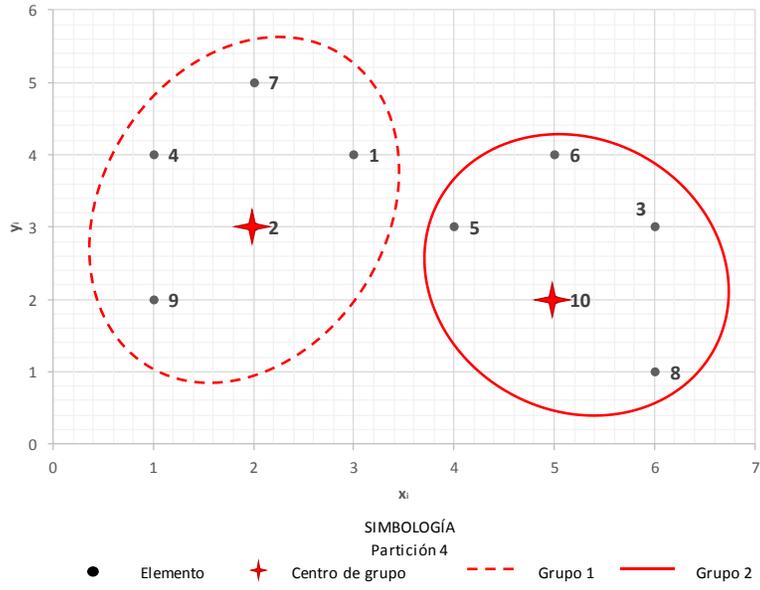


Figura III-11. Partición 4, dos grupos homogéneos

A manera de resumen se muestran las cuatro particiones desarrolladas anteriormente, así como la calidad de la partición (Figura III-12 y Tabla III-9).

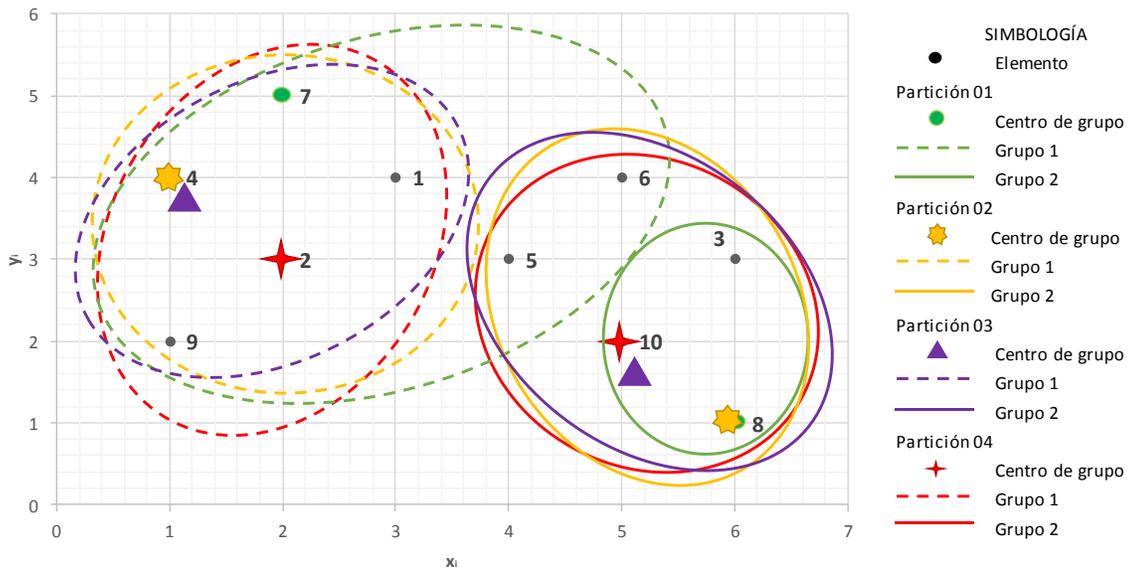


Figura III-12. Resumen de particiones, dos grupos homogéneos.

Tabla III-9. Coeficientes Momentos-L Datos estadísticos

Partición	Centro de grupo 1	Elementos del grupo 1	Centro de grupo 2	Elementos del grupo 2	$Q(P)$
P_1	[7]	$(e_1, e_2, e_4, e_5, e_6, e_7, e_9)$	[8]	(e_3, e_8, e_{10})	42
P_2	[4]	$(e_1, e_2, e_4, e_7, e_9)$	[8]	$(e_3, e_5, e_6, e_8, e_{10})$	36
P_3	[4]	$(e_1, e_2, e_4, e_7, e_9)$	[10]	$(e_3, e_5, e_6, e_8, e_{10})$	22
P_4	[2]	$(e_1, e_2, e_4, e_7, e_9)$	[10]	$(e_3, e_5, e_6, e_8, e_{10})$	20

Con el fin de revisar que el centro de los k grupos haya sido el óptimo, se realizaron verificaciones como se muestra en la Tabla III-10 donde es posible observar que los elementos difieren poco en cuanto a la cantidad de elementos de cada grupo, más bien, la calidad de la partición mejora al seleccionar los centros de grupo que minimizan la función objetivo. Nota. Es importante estandarizar los datos, es decir, media 0 y desviación de 1.

Tabla III-10. Verificación de la calidad de partición óptima con otros elementos.

Centro de grupo 1	Elementos del grupo 1	Centro de grupo 2	Elementos del grupo 2	$Q(P)$
[4]	$(e_1, e_2, e_4, e_7, e_9)$	[10]	$(e_3, e_5, e_6, e_8, e_{10})$	22
[4]	(e_2, e_4, e_7, e_9)	[5]	$(e_1, e_3, e_5, e_6, e_8, e_{10})$	26
[4]	$(e_1, e_2, e_4, e_7, e_9)$	[3]	$(e_3, e_5, e_6, e_8, e_{10})$	24
[2]	$(e_1, e_2, e_4, e_7, e_9)$	[6]	$(e_3, e_5, e_6, e_8, e_{10})$	30
[1]	$(e_1, e_2, e_4, e_5, e_6, e_7, e_9)$	[10]	(e_3, e_8, e_{10})	26
[7]	$(e_1, e_2, e_4, e_7, e_9)$	[10]	$(e_3, e_5, e_6, e_8, e_{10})$	28
[9]	$(e_1, e_2, e_4, e_7, e_9)$	[10]	$(e_3, e_5, e_6, e_8, e_{10})$	34

Método del árbol de regresión

En este apartado se muestra la aplicación del método de minería de datos a través del software STATISTICA: El software STATISTICA en la página de soporte, contiene un conjunto de tutoriales que permiten seguir los procedimientos paso a paso. Para efecto de aplicación se tomó un ejercicio de los propuestos en la mencionada página.

Objetivo del ejercicio: analizar cuáles variables predicen el porcentaje de pobreza en un condado.

Abrir el archivo con los datos: Los archivos que contienen los datos para los ejemplos se encuentran en Abrir\File-Open Examples\Datasets Folder. Abrir el archivo “Poverty.sta”

Especificaciones de las variables Es factible revisar y modificar, según se requiera las características de las variables en estudio. En Data Menu\ All variable specs. Se despliega un cuadro de Variable Specifications Editor

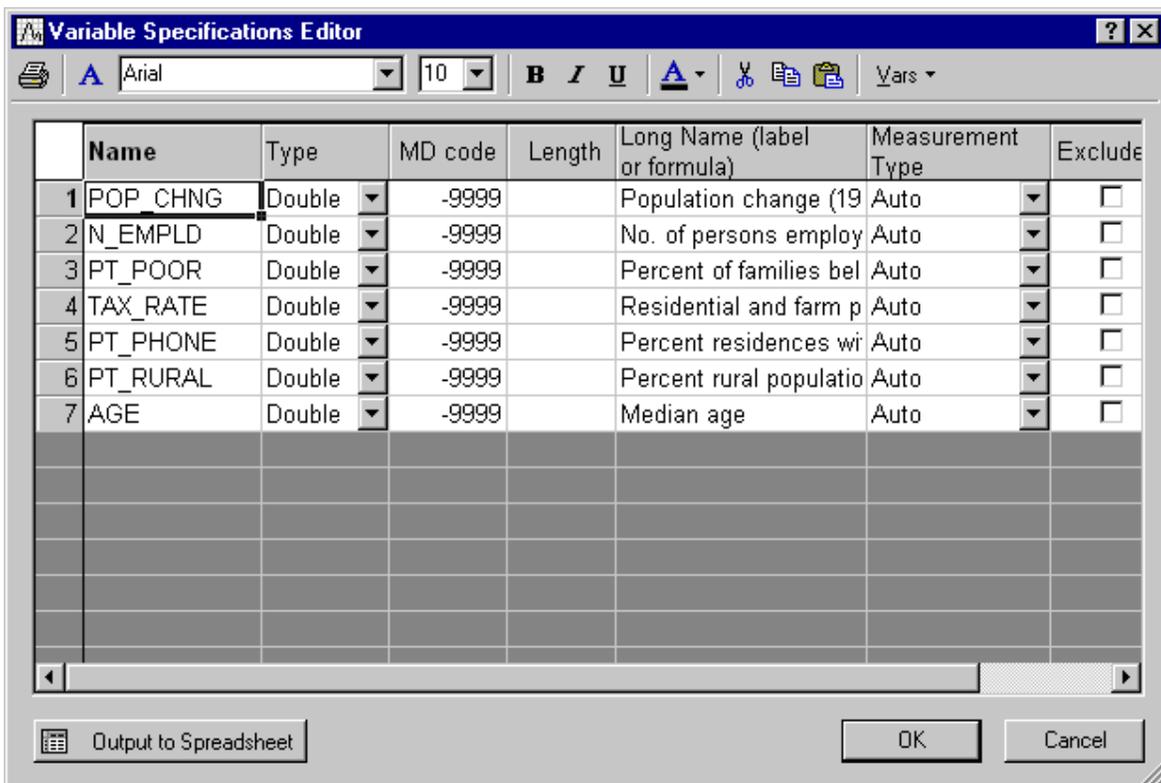


Figura III-13. Variables de entrada

Análisis: En el menú Statistics\Data Mining\Startup Panel\ seleccionar: General Classification and Regression trees Models .

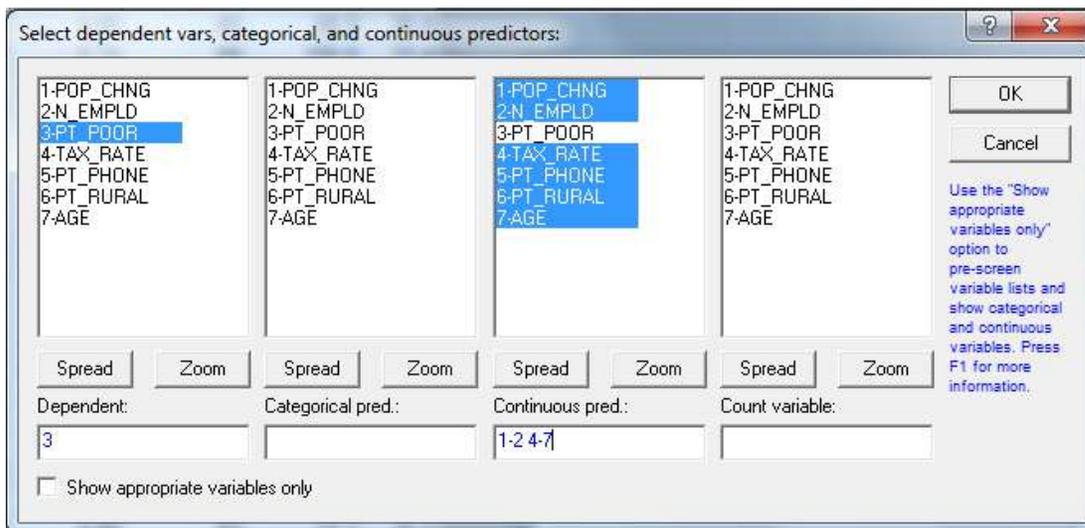
En la Figura III-14 (a-g) se muestra en procedimiento de agrupación.

En la ventana de diálogo dar click en “OK”

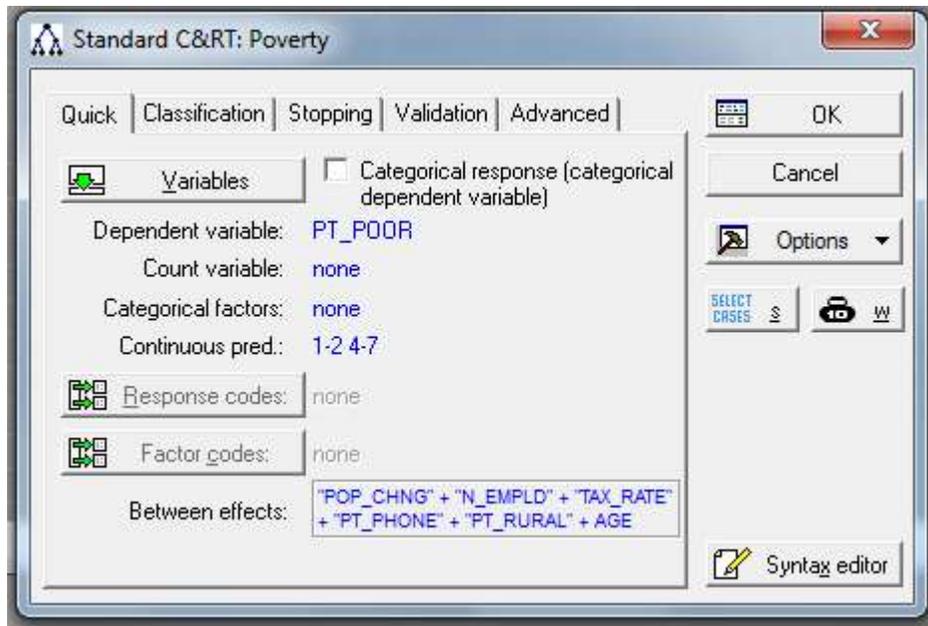


a)

En la pestaña “Quick” seleccionar las variables dependientes e independientes, la variable PT_POOR será la variable dependiente, mientras que las variables independientes, serán de tipo continuo

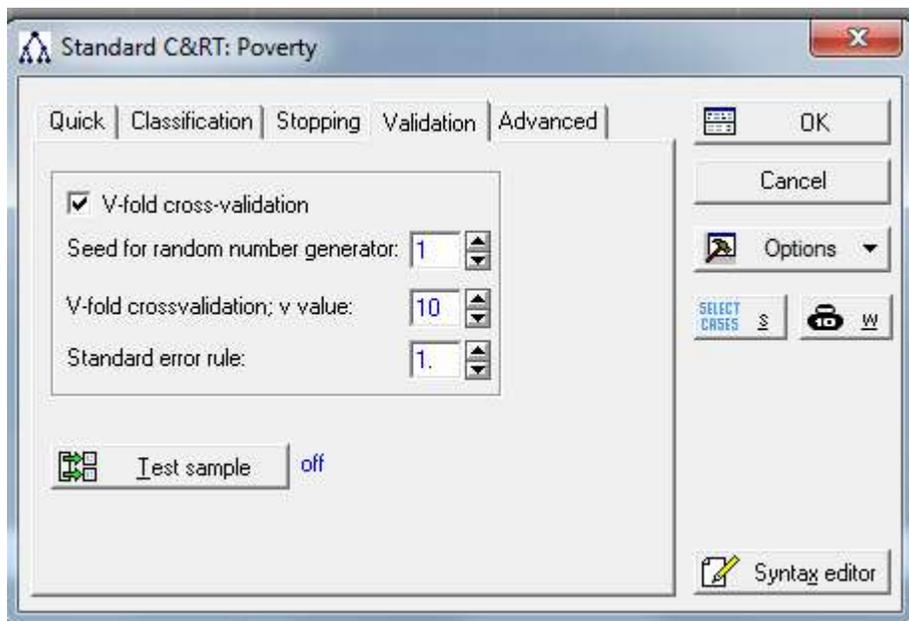


b)



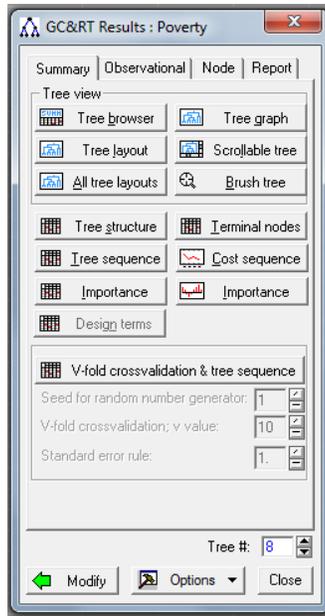
c)

En la pestaña de validación se activa la casilla de “Validation cross-fold” con un valor de 10



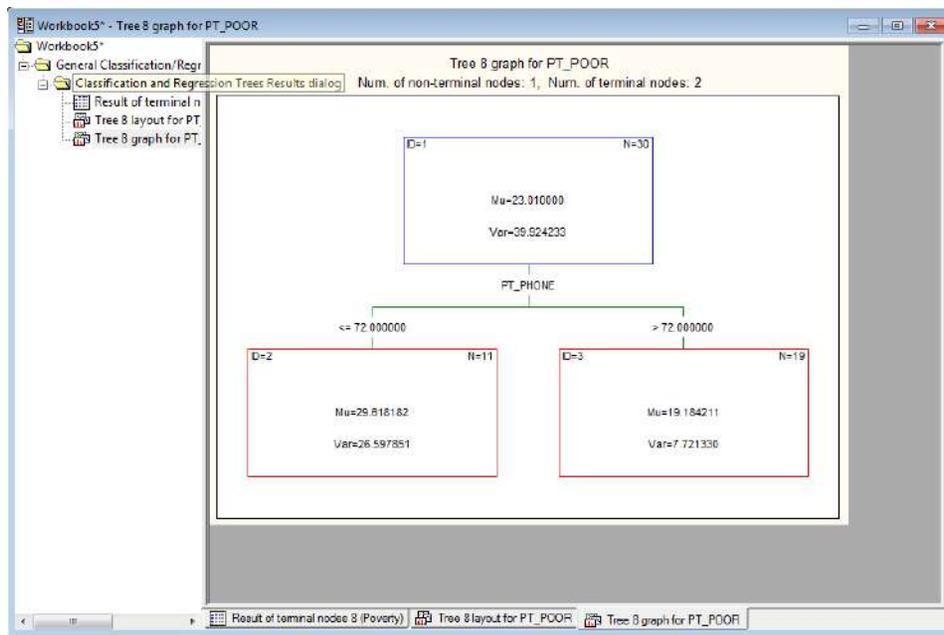
d)

Dar click en “OK”, esperar a que el programa despliegue los resultados.



e)

Se obtiene una gráfica como la siguiente, que permite visualizar la cantidad óptima de grupos formados así como la variable de decisión de la división, que para este ejemplo es la variable “PT_PHONE(teléfono):



f)

Figura III-14. Figuras a-g, procedimiento de delimitación