



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

ANÁLISIS REGIONAL HIDROLÓGICO COMBINANDO AVENIDA ÍNDICE Y EL ANÁLISIS MULTIVARIADO

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

PRESENTA:
LAURA TATIANA BERBESI PRIETO

TUTOR PRINCIPAL
DR. CARLOS AGUSTÍN ESCALANTE SANDOVAL, FACULTAD DE INGENIERÍA

COMITÉ TUTOR
M.I. VÍCTOR FRANCO (INSTITUTO DE INGENIERÍA)
M.I ALBA BEATRIZ VÁZQUEZ GONZÁLEZ (FACULTAD DE INGENIERÍA)
DRA. LILIA REYES CHÁVEZ (FACULTAD DE INGENIERÍA)
M.I MARCO TULIO MENDOZA ROSAS (FACULTAD DE INGENIERÍA)

Ciudad Universitaria, CD. MX. mayo 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

Presidente: M.I. Víctor Franco
Secretario: M.I. Alba Beatriz Vázquez González
1 er. Vocal: Dr. Carlos Agustín Escalante Sandoval
2 do. Vocal: Dra. Lilia Reyes Chávez
3 er. Vocal: M.I. Marco Tulio Mendoza Rosas

Ciudad de México

TUTOR DE TESIS:

DR. CARLOS AGUSTÍN ESCALANTE SANDOVAL



The image shows a handwritten signature in black ink, which appears to read 'Escalante Sandoval'. The signature is written over a horizontal dashed line. The signature is enclosed within a large, hand-drawn oval shape.

FIRMA

Agradecimientos

A mis padres y mi hermano por su apoyo incondicional durante mi proceso de formación. Por su amor, motivación y comprensión que me dan la fortaleza para mejorar cada día.

A mi tutor el doctor Carlos A. Escalante Sandoval, por su dedicación y orientación en el desarrollo del proyecto. Además de su valioso tiempo dedicado a mi formación durante la maestría. Al comité tutor por sus recomendaciones y sugerencias.

Al Consejo Nacional de Ciencia y Tecnología (Conacyt) por el apoyo económico otorgado mediante la beca nacional durante mis estudios de la maestría.

Índice

1. Planteamiento del Problema	8
1.1. Formulación del problema	8
1.2. Objetivos	8
1.3. Justificación	9
1.4. Alcance	9
2. Marco Teórico	10
2.1. Técnicas de estimación de parámetros	10
2.1.1. Tipo de modelos teóricos	10
2.1.2. Método de los momentos	11
2.1.3. Método de momentos de probabilidad pesada	12
2.1.4. Momentos L	12
2.2. Identificación de puntos atípicos (Outliers)	14
2.2.1. Diagrama de caja (Box Plot)	14
2.2.2. Prueba de Grubbs	15
2.2.3. Tres veces la desviación estándar	15
2.3. Análisis de frecuencias de eventos extremos	15
2.3.1. Periodo de retorno (T)	16
2.3.2. Prueba de homogeneidad	16
2.3.3. Prueba de independencia	21
2.3.4. Distribución Gumbel	23
2.3.5. Modelo logístico bivariado	25
2.3.6. Prueba de bondad de ajuste	26
2.3.7. Análisis de frecuencias de gastos máximos anuales	26
2.4. Análisis regional hidrológico	27

	5
2.4.1. Delimitación de regiones homogéneas	28
2.4.2. Técnica regional avenida índice	32
3. Recolección y análisis de información	35
3.1. Técnica regional avenida índice	41
3.2. Técnica regional avenida índice combinando el análisis multivariado	43
4. Resultados y discusión	45
4.1. Ecuaciones de ajuste a la mediana regional	46
4.2. Análisis para sitios con escasa información	48
4.3. Análisis para sitios con nula información	49
4.4. Discusión	52
5. Conclusiones	64
6. Bibliografía	67

Lista de tablas

Tabla 1. Valores críticos xk – Prueba Pettitt	19
Tabla 2. Valores críticos Tc – Prueba Normal Estándar	20
Tabla 3. Valores críticos Pc – Prueba Buishand	21
Tabla 4. Valores críticos Nc – Prueba Von Neumann	21
Tabla 5. Extensión de los registros rellenados de las estaciones en estudio	37
Tabla 6. Resultados pruebas de homogeneidad e independencia	37
Tabla 7. Periodos de análisis para la técnica regional avenida índice	40
Tabla 8. Parámetros por la distribución Gumbel univariada	41
Tabla 9. Ejemplo de aplicación avenida índice asumiendo la estación 26241 con escasa información	42
Tabla 10. Parámetros estimados por la distribución Gumbel bivariada, estación 26241 de la región homogénea de la RH 26.	43
Tabla 11. Parámetros por la distribución Gumbel bivariada	44
Tabla 12. Comparación EEA [m ³ /s] del análisis puntual entre la distribución Gumbel univariada y bivariada	45
Tabla 13. Estadísticos de las series en el periodo común en análisis	46
Tabla 14. Ecuaciones de ajuste de la mediana regional – Análisis univariado	47
Tabla 15. Ecuaciones de ajuste de la mediana regional – Análisis multivariado	47
Tabla 16. Gasto estimado de 2.33 años [m ³ /s] por la distribución Gumbel univariada y bivariada	48
Tabla 17. EEA [m ³ /s] por la técnica avenida índice para sitios con escasa información	48
Tabla 18. Ecuaciones para determinar el gasto de 2.33 años [m ³ /s] por la distribución Gumbel univariada	49
Tabla 19. Ecuaciones para determinar el gasto de 2.33 años [m ³ /s] por la distribución Gumbel bivariada	50
Tabla 20. Gasto de 2.33 años [m ³ /s] estimados regionalmente para sitios con nula información	50
Tabla 21. EEA [m ³ /s] por la técnica avenida índice para sitios con nula información	51
Tabla 22. Casos con menor EEA [m ³ /s] por la técnica avenida índice combinando el análisis multivariado respecto al univariado	53
Tabla 23. Parámetros por la distribución Gumbel univariada RH 10, periodo completo	54
Tabla 24. Parámetros por la distribución Gumbel bivariada RH 10, periodo completo	54
Tabla 25. Ecuaciones de ajuste de la mediana regional – Análisis univariado RH 10, periodo completo	55
Tabla 26. Ecuaciones de ajuste de la mediana regional – Análisis multivariado RH 10, periodo completo	55
Tabla 27. Gasto estimado de 2.33 años [m ³ /s] por la distribución Gumbel univariada y bivariada, RH 10, periodo completo	56
Tabla 28. EEA [m ³ /s] por la técnica avenida índice para sitios con escasa información RH 10, periodo completo	57
Tabla 29. Ecuaciones para determinar el gasto de 2.33 años [m ³ /s] por la distribución Gumbel univariada RH 10, periodo completo	58

Tabla 30. Ecuaciones para determinar el gasto de 2.33 años [m ³ /s] por la distribución Gumbel bivariada RH 10, periodo completo	58
Tabla 31. Gasto de 2.33 años estimados para sitios con nula información RH 10, periodo completo	59
Tabla 32. EEA [m ³ /s] por la técnica avenida índice para sitios con nula información RH 10, periodo completo	59
Tabla 33. Casos con menor EEA [m ³ /s] por la técnica avenida índice combinando el análisis multivariado RH 10, periodo completo	60
Tabla 34. Ecuaciones para determinar el gasto de 2.33 años [m ³ /s] por la distribución Gumbel univariada RH 10, periodo completo con el área de las cuencas	61
Tabla 35. Ecuaciones para determinar el gasto de 2.33 años [m ³ /s] por la distribución Gumbel bivariada RH 10, periodo completo con el área de las cuencas	62
Tabla 36. EEA [m ³ /s] por la técnica avenida índice para sitios con nula información RH 10, periodo completo utilizando el área para estimar el gasto de 2.33 años [m ³ /s]	62
Tabla 37. Casos con menor EEA [m ³ /s] estimando el gasto de 2.33 años [m ³ /s] por el parámetro de forma y no por el área por análisis bivariado RH 10, periodo completo	63

Lista de figuras

Figura 1. Zonas de estudio	36
Figura 2. Selección de estaciones en la región homogénea RH 10	39
Figura 3. Ajuste mediana regional univariada sin la estación 26241	42

1. Planteamiento del Problema

En el diseño de las obras hidráulicas, es necesario contar con un amplio registro de datos hidrométricos en el punto de interés. Por lo general, el sitio de estudio no suele coincidir con la ubicación de la estación hidrométrica, e incluso, muchas de estas estaciones ya se encuentran suspendidas, limitando así, la información requerida para un adecuado análisis estadístico. Sin embargo, en el análisis hidrológico se han propuesto algunas “técnicas regionales” que hacen uso de la información proveniente de estaciones que se encuentran localizadas en lo que se denomina “Región Hidrológicamente Homogénea”, siendo aplicables a sitios con escasa o nula información hidrométrica. Estas técnicas regionales han sido desarrolladas para reducir la incertidumbre, desde el punto de vista estadístico, de los eventos asociados a diferentes períodos de retorno, evitando afectar los costos y seguridad de las obras hidráulicas.

Con el objetivo de seguir aportando sobre el conocimiento y aplicación de las técnicas de análisis regionales, en este trabajo, se realizará una combinación entre la Técnica de la Avenida Índice y la técnica de estimación multivariada de eventos de diseño, a través de la aplicación del modelo logístico bivariado con marginales Gumbel.

1.1. Formulación del problema

¿Se obtendrán mejores resultados del análisis regional hidrológico por la técnica avenida índice al combinarla con el análisis multivariado respecto al univariado?

1.2. Objetivos

Identificar una región homogénea en la zona de estudio mediante diferentes técnicas con un periodo común consecutivo de mínimo diez años de gastos máximos instantáneos.

Verificar que las estaciones de la región homogénea sean homogéneas e independientes.

Obtener los eventos de diseño para diferentes periodos de retorno por la distribución Gumbel.

Aplicar el análisis regional hidrológico por la técnica avenida índice univariado.

Obtener los eventos de diseño para diferentes periodos de retorno por la distribución Logística Bivariada con marginales Gumbel.

Aplicar el análisis regional hidrológico por la técnica avenida índice combinando los resultados obtenidos mediante el análisis bivariado.

1.3. Justificación

Generalmente el sitio de proyecto no cuenta con un amplio registro de información hidrométrica y, en algunos casos, esta información es nula, razón por la cual es necesario reducir la incertidumbre de los eventos de diseño aplicando las técnicas de análisis regional hidrológico, donde es posible transferir la información de sitios vecinos de la región homogénea al sitio de proyecto.

El objetivo de un proyectista es obtener un gasto de diseño que se acerque a la realidad y esto se consigue al aplicar la técnica regional que mejor se ajuste a la región homogénea. Con base en este objetivo, se combinan la técnica avenida índice con el análisis multivariado para determinar si se alcanzan mejores resultados, en comparación con la aplicación de la técnica tradicional de la Avenida Índice, y así contar con una nueva herramienta para la estimación de gastos de diseño en cuencas no aforadas o con escasa información.

1.4. Alcance

En este trabajo se emplearán estaciones que tengan un registro de información hidrométrica mayor a treinta años dentro de la zona de estudio, seleccionando aquellas donde la delimitación de su cuenca sea confiable, dada su necesidad en la delimitación de regiones homogéneas y el análisis regional; donde se requieren al menos cuatro estaciones. La veracidad de la información será la establecida por cada una de las fuentes consultadas.

2. Marco Teórico

2.1. Técnicas de estimación de parámetros

2.1.1. Tipo de modelos teóricos

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

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

Un modelo paramétrico puede ser considerado como un determinístico en el sentido de que una vez que se estiman los parámetros del modelo, éste siempre genera la misma salida a partir de la información de entrada. Por otro lado, un modelo paramétrico es estocástico en el sentido de que los parámetros estimados dependen de los datos observados y aquellos cambiarán cuando los datos observados también lo hagan.

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

Antes de hacer inferencias de cualquier modelo es importante la estimación de sus parámetros. Cada estimador de un parámetro es una función de los valores de la muestra, las cuales son observaciones de una variable aleatoria. Así, el propio parámetro estimado es una variable aleatoria que tiene su propia distribución muestral. Un estimador que se obtiene a partir de un grupo de valores puede considerarse como un valor observado de una variable aleatoria. Por lo cual, la bondad de un estimador puede ser juzgado a partir de su distribución.

Independientemente de la técnica que se use para la estimación de los parámetros se deben cumplir las siguientes propiedades:

Sesgo nulo. Un estimador $\hat{\theta}$ de un parámetro θ se dice que es no sesgado, si el valor $E(\hat{\theta}) = \theta$. De lo contrario es sesgado. El sesgo se obtiene como $B = E(\hat{\theta}) - \theta$.

Consistencia. Un estimador $\hat{\theta}$ de un parámetro θ se dice que es consistente sí para cualquier número positivo ε $\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \varepsilon) = 0$. Donde n es el tamaño de la muestra.

Eficiencia. Un estimador $\hat{\theta}$ se dice el más eficiente para θ si tiene sesgo nulo y su varianza es al menos tan pequeña como cualquier otro estimador no sesgado para θ .

Suficiencia. $\hat{\theta}$ en un estimador para θ , si $\hat{\theta}$ emplea toda la información relevante contenida en la muestra. (Escalante & Reyes, 2008, págs. 25-26)

2.1.2. Método de los momentos

El método de los momentos es un procedimiento muy sencillo para encontrar un estimador de uno o más parámetros poblacionales. Consiste básicamente en plantear un sistema de ecuaciones, cuyo tamaño depende del número de parámetros a estimar. Esto se hace al igualar los momentos poblacionales con los muestrales.

Los momentos muestrales, también conocidos como estadísticos muestrales, se obtienen con las siguientes expresiones

Media

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

Varianza sesgada

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

Varianza no sesgada

$$S_{insesg}^2 = \frac{1}{n-1} S_{sesg}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3)$$

Coefficiente de asimetría sesgado

$$g_{sesg} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{[S_{sesg}^2]^{3/2}} \quad (4)$$

Coefficiente de asimetría no sesgado

$$g_{insesg} = \frac{n^2}{(n-1)(n-2)} g_{sesg} \quad (5)$$

Coefficiente de curtosis sesgado

$$k_{sesg} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{[S_{sesg}^2]^2} \quad (6)$$

Coefficiente de curtosis no sesgado

$$k_{insesg} = \frac{n^3}{(n-1)(n-2)(n-3)} k_{sesg} \quad (7)$$

Desviación estándar

$$S = \sqrt{S^2} \quad (8)$$

Coefficiente de variación

$$C_v = \frac{S}{\bar{x}} \quad (9)$$

En el análisis hidrológico se recomienda el uso de los estadísticos no sesgados, ya que generalmente se trabaja con muestras relativamente pequeñas. (Escalante & Reyes, 2008, págs. 26-27)

2.1.3. Método de momentos de probabilidad pesada

Greenwood et al. (1979) introdujeron el método y mostraron su utilidad en la estimación de parámetros de distribuciones cuyas formas inversas $x = x(F)$ se definen explícitamente. Se puede obtener un estimador no sesgado $\hat{M}_{(k)}$ y $k =$ entero no negativo si $x_i, i = 1, \dots, n$ son los valores ordenados de mayor a menor como:

$$\hat{M}_{(0)} = \frac{1}{n} \sum_{i=1}^n x_i \quad (10)$$

$$\hat{M}_{(1)} = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} x_i (n-i) \quad (11)$$

$$\hat{M}_{(2)} = \frac{1}{n(n-1)(n-2)} \sum_{i=1}^{n-2} x_i (n-i)(n-i-1) \quad (12)$$

$$\hat{M}_{(3)} = \frac{1}{n(n-1)(n-2)(n-3)} \sum_{i=1}^{n-3} x_i (n-i)(n-i-1)(n-i-2) \quad (13)$$

2.1.4. Momentos L

Los momentos-L son análogos a los momentos convencionales, sin embargo, tienen cierta ventaja sobre ellos, ya que son capaces de caracterizar a un mayor número de

distribuciones, además de estar virtualmente libres de sesgo aún para muestras pequeñas (Hosking, 1990).

Los primeros cuatro momentos-L son

$$\lambda_1 = \beta_{(0)} \quad (14)$$

$$\lambda_2 = 2\beta_{(1)} - \beta_{(0)} \quad (15)$$

$$\lambda_3 = 6\beta_{(2)} - 6\beta_{(1)} + \beta_{(0)} \quad (16)$$

$$\lambda_4 = 20\beta_{(3)} - 30\beta_{(2)} + 12\beta_{(1)} - \beta_{(0)} \quad (17)$$

Donde

$$\beta_r = M_{i,r,0} = \int_0^1 xF^r dF \quad (18)$$

Los estimadores muestrales de los momentos-L, pueden obtenerse al sustituir los estimadores insesgados de las ecuaciones (10) y (13) en las ecuaciones (14) y (17).

Los primeros momentos-L poblacionales se pueden obtener mediante la expresión (18), de la forma:

$$\lambda_1 = \int_0^1 x(F)dF \quad (19)$$

$$\lambda_2 = \int_0^1 x(F)(2F - 1)dF \quad (20)$$

$$\lambda_3 = \int_0^1 x(F)(6F^2 - 6F + 1)dF \quad (21)$$

$$\lambda_4 = \int_0^1 x(F)(20F^3 - 30F^2 + 12F + 1)dF \quad (22)$$

Una vez conocidos λ_1 , λ_2 , λ_3 y λ_4 se pueden obtener las relaciones de los momentos-L

$$\tau_2 = \frac{\lambda_2}{\lambda_1} \quad \text{Coeficiente de variación - L} \quad (23)$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \quad \text{Coeficiente de Sesgo - L} \quad (24)$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} \quad \text{Coeficiente de Curtosis} - L \quad (25)$$

(Escalante & Reyes, 2008, págs. 35-36)

2.2. Identificación de puntos atípicos (Outliers)

De acuerdo con NIST/SEMATECH, un dato atípico es una observación que parece desviarse significativamente de las demás observaciones en una muestra. Identificar posibles puntos atípicos es importante porque pueden ser datos erróneos como, por ejemplo, los causados por mediciones incorrectas. Si se puede determinar que el dato es un error, debe eliminarse o corregirlo si es posible.

En algunos casos, no se puede determinar si un dato atípico es un error, a veces se identifican por variaciones aleatorias o puede indicar algo científicamente interesante, como el pertenecer a una población diferente a la mayoría de los datos.

Se recomienda que la serie de datos tenga una distribución normal antes de aplicar las pruebas de identificación de puntos atípicos. (NIST/SEMATECH)

2.2.1. Diagrama de caja (Box Plot)

El diagrama de caja (Chambers, 1983) es una excelente herramienta para transmitir información de ubicación y variación en una muestra de datos. Agregar límites al diagrama de caja es útil para la identificación de puntos atípicos. Para crear esta variación:

1. Calcular la mediana y los cuartiles inferior y superior (el cuartil superior Q_3 es el percentil del 75% y el cuartil inferior Q_1 es el percentil del 25%).
2. Ubicar un símbolo en la mediana y dibujar una caja en medio del cuartil inferior y superior.
3. Calcular el rango intercuartil (diferencia entre el cuartil superior y el inferior) – IQR.
4. Calcular los siguientes puntos:

$$Q_1 - 1.5 IQR \quad \text{Límite inferior} \quad (26)$$

$$Q_3 + 1.5 IQR \quad \text{Límite superior} \quad (27)$$

Los datos que sean mayores al límite superior o menor que el límite inferior, se consideran datos atípicos. (NIST/SEMATECH)

2.2.2. Prueba de Grubbs

La prueba de Grubbs (Grubbs, 1969 y Stefansky, 1972) es usada para detectar un solo punto atípico en una serie de datos univariados que siguen aproximadamente una distribución normal.

El estadístico se define como

$$G = \frac{\max_{i=1,\dots,N} |Y_i - \bar{Y}|}{S_y} \quad (28)$$

Donde \bar{Y} y S_y son la media y desviación estándar de la muestra.

Si se desea saber si el mínimo valor es un dato atípico:

$$G = \frac{\bar{Y} - Y_{min}}{S_y} \quad (29)$$

Si se desea saber si el máximo valor es un dato atípico:

$$G = \frac{Y_{max} - \bar{Y}}{S_y} \quad (30)$$

La hipótesis de que no hay puntos atípicos es rechazada si

$$G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{t_{\frac{\alpha}{2N}, N-2}^2}{N-2 + t_{\frac{\alpha}{2N}, N-2}^2}} \quad (31)$$

Con $t_{\alpha/2N, (N-2)}$ denota el valor crítico de la distribución t con N - 2 grados de libertad y un valor de significancia de $\alpha/2N$. (NIST/SEMATECH)

2.2.3. Tres veces la desviación estándar

Se asume que la serie tiene una distribución normal y que los datos fuera de la media más o menos tres veces su desviación estándar son datos atípicos.

$$\text{Media} \pm 3S \quad \text{Contiene el 99.7\% de las mediciones} \quad (32)$$

2.3. Análisis de frecuencias de eventos extremos

Las fases de planeación, diseño, construcción y operación de los aprovechamientos hidráulicos están siempre relacionadas con eventos hidrológicos futuros. La complejidad

de los procesos físicos de estos eventos hace casi imposible tener estimaciones confiables de diseño basadas en las leyes de la mecánica o la física, ya sea porque estos métodos son insuficientes o porque el modelo matemático resultante es muy complicado. Una alternativa en el análisis hidrológico es la aplicación de los conceptos de la teoría de probabilidad y estadística.

El análisis de frecuencias de los gastos máximos anuales se emplea para proveer la magnitud de un evento \hat{Q}_T , de cierto periodo de retorno T, para el diseño de una obra hidráulica; el manejo de las llanuras de inundación, y como ayuda en los procesos de planeación y manejo de las cuencas hidrológicas. Sin embargo, el proyectista no solo debe estimar la magnitud del evento de diseño, sino que también debe proporcionar la probabilidad de excedencia, con el fin de fijar la seguridad del funcionamiento de la obra, o bien el riesgo de falla. (Escalante & Reyes, 2008, pág. 129)

2.3.1. Periodo de retorno (T)

En el análisis de frecuencias de una serie Q_i^j la estación j, para un conjunto de datos $i = 1, 2, \dots, n_j$, el primer objetivo es determinar el intervalo de recurrencia o periodo de retorno T en años, de un evento hidrológico dada una magnitud x .

El periodo de retorno se define como el número de años que transcurren en promedio para que un evento de magnitud dada x sea igualado o excedido por lo menos una vez en ese periodo de tiempo.

$$T = \frac{1}{P(X > x)} = \frac{1}{[1 - P(X \leq x)]} \quad (33)$$

(Escalante & Reyes, 2008, pág. 129)

2.3.2. Prueba de homogeneidad

Las características estadísticas de las series hidrológicas, como la media, desviación estándar y los coeficientes de correlación serial, se afectan cuando la serie presenta tendencia en la media o en la varianza, o cuando ocurren saltos negativos a positivos; tales anomalías son producidas por la pérdida de homogeneidad.

En general, la falta de homogeneidad de los datos es inducida por las actividades humanas como la deforestación, aperturas de nuevas áreas al cultivo, rectificación de cauces, construcción de embalses y reforestación. También es producto de los procesos naturales

súbitos, cómo incendios forestales, terremotos, deslizamientos de laderas y erupciones volcánicas.

Las pruebas estadísticas que miden la homogeneidad de una serie de datos presentan una hipótesis nula y una regla para aceptarla o rechazarla. (Escalante & Reyes, 2008, pág. 129)

2.3.2.1. Prueba estadística de Helmert

Esta prueba es sencilla y consiste en analizar el signo de las desviaciones de cada evento Q_i^j de la serie j para $i = 1, 2, \dots, n_j$, con respecto a su valor medio \bar{Q}^j . Si una desviación de un cierto signo es seguida de otra del mismo signo, entonces se dice que se forma una secuencia S, de lo contrario se considera un cambio C.

La serie se considera homogénea si se cumple

$$-\sqrt{n-1} \leq (S - C) \leq \sqrt{n-1} \quad (34)$$

(Escalante & Reyes, 2008, pág. 130)

2.3.2.2. Prueba estadística t de Student

Cuando la causa probable de la pérdida de homogeneidad de la serie sea un cambio abrupto en la media, el estadístico t es muy útil.

Si se considera una serie Q_i^j para $i = 1, 2, \dots, n_j$, del sitio j, la cual se divide en dos conjuntos de tamaño $n_1 = n_2 = \frac{n_j}{2}$, entonces, el estadístico de prueba se define con la expresión

$$\tau_d = \frac{\bar{x}_1 - \bar{x}_2}{\left[\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right]^{1/2}} \quad (35)$$

Donde

\bar{x}_1, s_1^2 son la media y varianza de la primera parte del registro de tamaño n_1 .

\bar{x}_2, s_2^2 son la media y varianza de la segunda parte del registro de tamaño n_2 .

El valor absoluto de t_d se compara con el valor de la distribución t de Student de dos colas y con $\nu = n_1 + n_2 - 2$ grados de libertad y para un nivel $\alpha = 0.05$.

Si y sólo si, el valor de t_d es mayor que aquel de la distribución t de Student, se concluye que la diferencia entre las medias es evidencia de inconsistencia y por lo tanto la serie Q_i^j se considera no homogénea. (Escalante & Reyes, 2008, pág. 130)

Cuando la serie es impar $n_2 = n/2 + 1$.

2.3.2.3. Prueba estadística de Cramer

Esta prueba se utiliza con el propósito de verificar la homogeneidad en el registro Q_i^j de la serie j para $i = 1, 2, \dots, n_j$, y también para determinar si el valor medio no varía significativamente de un periodo de tiempo a otro. Con este propósito se consideran tres bloques, el primero, del tamaño total de la muestra n_j ; el segundo de tamaño n_{60} (60% de los últimos valores de la muestra n_j); y el tercero de tamaño n_{30} (30% de los últimos valores de la muestra n_j).

La prueba compara el valor de \bar{Q}^j del registro total con cada una de las medias de los bloques elegidos \bar{Q}_{60}^j y \bar{Q}_{30}^j . Para que se considere la serie analizada como estacionaria en la media, se deberá cumplir que no existe una diferencia significativa entre las medias de los bloques.

$$\bar{Q}^j = \sum_{i=1}^{n_j} \frac{Q_i^j}{n_j} \quad \text{Para una sola muestra analizada } j = 1 \quad (36)$$

$$S_Q^j = \left[\frac{1}{n_j - 1} \sum_{i=1}^{n_j} (Q_i^j - \bar{Q}^j)^2 \right]^{1/2} \quad (37)$$

$$\bar{Q}_{60}^j = \sum_{i=1}^{n_{60}} \frac{Q_k^j}{n_{60}} \quad (38)$$

$$\bar{Q}_{30}^j = \sum_{i=1}^{n_{30}} \frac{Q_k^j}{n_{30}} \quad (39)$$

$$\tau_{60}^j = \frac{\bar{Q}_{60}^j - \bar{Q}^j}{S_Q^j} \quad (40)$$

$$\tau_{30}^j = \frac{\bar{Q}_{30}^j - \bar{Q}^j}{S_Q^j} \quad (41)$$

$$t_w = \left\{ \frac{n_w(n_j - 2)}{n_j - n_w [1 + (\tau_w^j)^2]} \right\}^{1/2} |\tau_w^j| \quad \text{para } w = 60 \text{ y } w = 30 \quad (42)$$

El estadístico t_w tiene distribución t de Student de dos colas con $\nu = n_1 + n_2 - 2$ grados de libertad y para un nivel $\alpha = 0.05$.

Sí y solo si, el valor absoluto de t_w , para $w = 60$ y $w=30$, es mayor que el de la distribución t de Student, se concluye que la diferencia entre las medias es evidencia de inconsistencia y por lo tanto la serie Q_i^j se considera no homogénea. (Escalante & Reyes, 2008, pág. 131)

2.3.2.4. Prueba de Pettitt

De Wijngaard (2003, págs. 690-691): es una prueba no paramétrica basada en rangos. Los rangos r_1, r_2, \dots, r_n de y_1, y_2, \dots, y_n son usados para calcular:

$$x_k = 2 \left(\sum_{i=1}^k r_i \right) - k(n+1) \quad k = 1, \dots, n \quad (43)$$

Donde r_i son los datos de la serie ordenados de menor a mayor.

Si ocurre un punto de cambio en la medida en el año k , el valor absoluto de x_k alcanza su valor máximo

$$x_k = \max |x_k| \quad 1 \leq k \leq n \quad (44)$$

Los valores de significancia son dados por Pettitt (1979). La Tabla 1, se muestran los valores críticos de x_k :

Tabla 1. Valores críticos x_k – Prueba Pettitt

n	20	30	40	50	70	100
1 %	71	133	208	293	488	841
5 %	57	107	167	235	393	677

Autor: Wijngaard (2003)

Si $x_k \text{ crítico} \geq x_k$ la muestra es homogénea.

2.3.2.5. Prueba Normal Estándar

De Wijngaard (2003, págs. 689-691): Alexandersson (1986) describe un estadístico $T(k)$ compara la media de los primeros k años y los últimos $n-k$ años del registro:

$$T(k) = k\bar{z}_1^2 + (n - k)\bar{z}_2^2 \quad k = 1, \dots, n \quad (45)$$

$$\bar{z}_1 = \frac{1}{k} \sum_{i=1}^n \frac{Q_i - \bar{Q}}{S} \quad (46)$$

$$\bar{z}_2 = \frac{1}{n - k} \sum_{i=k+1}^n \frac{Q_i - \bar{Q}}{S} \quad (47)$$

El estadístico de la prueba T_o es:

$$T_o = \text{máx } T(k) \quad 1 \leq k \leq n \quad (48)$$

Tabla 2. Valores críticos T_c – Prueba Normal Estándar

n	20	30	40	50	70	100
1 %	9.56	10.45	11.01	11.38	11.89	12.32
5 %	6.95	7.65	8.10	8.45	8.80	9.15

Autor: Wijngaard (2003)

Si $T_c \geq T_o$ la muestra es homogénea.

2.3.2.6. Prueba de Buishand

De Wijngaard (2003, págs. 689-691): esta prueba se basa en las desviaciones acumuladas de la media:

$$S_k^* = \sum_{i=1}^k (Q_i - \bar{Q}) \quad k = 1, \dots, n \quad (49)$$

Para un registro homogéneo se puede considerar que los valores S_k^* fluctúan cerca de cero, ya que no existe un patrón sistemático en las desviaciones de los Q_i de su valor medio \bar{Q} . La escala modificada de las sumas parciales se obtiene dividiendo los valores de S_k^* entre la desviación estándar de la muestra.

$$S_k^{**} = \frac{S_k^*}{S} \quad k = 0, \dots, n \quad (50)$$

Un estadístico que es sensible a las desviaciones de la homogeneidad es:

$$P = \max |S_k^{**}| \quad 0 \leq k \leq n \quad (51)$$

Los valores críticos de $\frac{P}{\sqrt{n}}$ son dados por Buishand (1982). La Tabla 3 muestra los valores críticos de P_c :

Tabla 3. Valores críticos P_c – Prueba Buishand

n	10	20	30	40	50	100	∞
1 %	1.29	1.42	1.46	1.50	1.52	1.55	1.62
5 %	1.14	1.22	1.24	1.26	1.27	1.29	1.36
10 %	1.05	1.10	1.12	1.13	1.14	1.17	1.22

Autor: Buishand (1982)

Si $P_c \geq \frac{P}{\sqrt{n}}$ la muestra es homogénea.

2.3.2.7. Prueba de Von Neumann

De Wijngaard (2003, págs. 690-691): es la razón entre la media cuadrática de las diferencias sucesivas y la varianza (Von Neumann, 1941):

$$N = \frac{\sum_{i=1}^{n-1} (Q_i - Q_{i+1})^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2} \quad (52)$$

Cuando la muestra es homogénea el valor esperado es $N = 2$. Si la muestra tiene un salto, entonces el valor de N tiende a ser menor que 2 (Buishand T. , 1981). Si la muestra tiene variaciones rápidas alrededor de la media, entonces los valores de N pueden ser mayores a 2 (Bingham & Nelson, 1981). La Tabla 4 da los valores críticos de N :

Tabla 4. Valores críticos N_c – Prueba Von Neumann

n	20	30	40	50	70	100
1 %	1.04	1.20	1.29	1.36	1.45	1.54
5 %	1.30	1.42	1.49	1.54	1.61	1.67

Autor: Wijngaard (2003)

Si $N_c < N$ la muestra es homogénea.

2.3.3. Prueba de independencia

Para que se pueda llevar a cabo el análisis de frecuencias se requiere que la muestra Q_i^j de la serie j para $i = 1, 2, \dots, n_j$, este compuesta por variables aleatorias. Para probarlo se aplicarán tres pruebas que se explican a continuación:

n Cantidad de datos

$h(m) = \left| \sum p \ln p \right|$ Función de entropía de las permutaciones de orden m

Está asintóticamente $X_{m!-1}^2$ distribuida (chi-cuadrado). La regla de decisión en la aplicación del test $G(m)$ al 100% $(1 - \alpha)$ de nivel confianza es:

Si $0 \leq G(m) \leq X_{m!-1, \alpha}^2$ la serie es independiente.

Se recomienda utilizar el mayor m que satisfaga $5m! \leq n$, para $m = 2, 3, \dots$

2.3.3.3. Prueba de Amigó

Definida por Amigó (2010), sigue la metodología de la prueba de Mantilla, excepto que las ventanas no son consecutivas, el último valor de la primera fila será el primer valor de la segunda fila, y así sucesivamente.

Número de ventanas

$$K = \frac{n - 1}{m - 1} \quad (56)$$

Se aplica una prueba basada en la distribución chi-cuadrado ($X_{m!-1}^2$), con el estadístico

$$X^2 = \frac{m!}{K} \sum \vartheta_j^2 - K \quad (57)$$

Donde ϑ_j son las frecuencias absolutas de las permutaciones asociadas.

Si $0 < X^2 < X_{m!-1, \alpha}^2$ la serie es independiente.

2.3.4. Distribución Gumbel

De Escalante y Reyes (2008, págs. 59-61,77)

$$F(x) = e^{-e^{-\left(\frac{x-v}{\alpha}\right)}} \quad (58)$$

$$f(x) = \frac{1}{\alpha} e^{-\left(\frac{x-v}{\alpha}\right)} e^{-e^{-\left(\frac{x-v}{\alpha}\right)}} \quad \begin{array}{l} -\infty < x < \infty \\ \alpha > 0 \end{array} \quad (59)$$

Donde

v parámetro de ubicación

α parámetro de escala

Además

$$\gamma = 1.1396$$

$$k = 5.4002$$

La variable reducida Gumbel es

$$y_i = \frac{x_i - v}{\alpha} \quad (60)$$

Estimador de eventos de diseño

$$\hat{X}_T = \hat{v} - \hat{\alpha} \ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \quad (61)$$

Estimadores por momentos

$$\hat{v} = \bar{x} - 0.45 S \quad (62)$$

$$\hat{\alpha} = 0.78 S \quad (63)$$

Estimadores por máxima verosimilitud

Considerando la variable reducida (60) se tiene el siguiente proceso iterativo

$$P = n - \sum_{i=1}^n e^{-y_i} \quad (64)$$

$$R = n - \sum_{i=1}^n y_i + \sum_{i=1}^n y_i e^{-y_i} \quad (65)$$

El criterio de convergencia es

$$\frac{P}{\hat{\alpha}} \approx 0 \quad \text{y} \quad \frac{-R}{\hat{\alpha}} \approx 0 \quad (66)$$

Incrementos

$$\delta_{vi} = (1.11P_j - 0.26R_j) \frac{\alpha_i}{n} \quad (67)$$

$$\delta_{\alpha i} = (0.26P_j - 0.61R_j) \frac{\alpha_i}{n} \quad (68)$$

Nuevos valores

$$\hat{v}_{j+1} = \hat{v}_j + \delta_{vj} \quad (69)$$

$$\hat{\alpha}_{j+1} = \hat{\alpha}_j + \delta_{\alpha j} \quad (70)$$

2.3.5. Modelo logístico bivariado

De Escalante y Reyes (2008, págs. 91-92):

Dado que el modelo logístico bivariado tiene mayor versatilidad que el modelo mezclado (amplios rangos del coeficiente de correlación y del índice de dependencia, los cuales permiten alternativas más allá del caso de independencia), es aconsejable (Raynal, 1985) su aplicación a la solución de problemas de análisis de frecuencias en Hidrología.

El modelo tiene la forma

$$F(x, y, m) = \exp \left\{ - \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{1/m} \right\} \quad (71)$$

Para:

$$F(x)F(y) < F(x, y) < \text{mín} [F(x)F(y)] \quad (72)$$

Donde

$F(x)$ y $F(y)$ son las funciones de distribución marginal

m , ($m \geq 1$) es el parámetro de asociación bivariada

Para $m = 1$ la función de distribución bivariada se reduce al caso de independencia

$$F(x, y, m) = F(x)F(y)$$

Gumbel (1962) mostró seis posibles funciones de distribución bivariada, mediante el empleo de las tres asíntotas como funciones marginales dentro del modelo logístico. Raynal (1985) redujo las posibles combinaciones a tres, al utilizar como marginales a las distribuciones Gumbel y GVE. Escalante (1998) propone el empleo de la distribución de valores extremos tipo I de dos poblaciones como marginales del modelo logístico.

Por ejemplo, si $F(x)$ y $F(y)$ son del tipo de la expresión (58), entonces las funciones de distribución acumulada y densidad bivariada (VEB11) tendrán la forma

$$F(x_i, y_i, v_1, \alpha_1, v_2, \alpha_2, m) = \exp \left\{ - \left[e^{-m\left(\frac{x-v_1}{\alpha_1}\right)} + e^{-m\left(\frac{y-v_2}{\alpha_2}\right)} \right]^{1/m} \right\} \quad (73)$$

$$f(x_i, y_i, v_1, \alpha_1, v_2, \alpha_2, m) = \left[\frac{e^{-m\left(\frac{x-v_1}{\alpha_1}\right)} + e^{-m\left(\frac{y-v_2}{\alpha_2}\right)}}{\alpha_1 \alpha_2} \right] \cdot \left[e^{-m\left(\frac{x-v_1}{\alpha_1}\right)} + e^{-m\left(\frac{y-v_2}{\alpha_2}\right)} \right]^{\frac{1}{m}-2} \cdot$$

$$\exp \left\{ - \left[e^{-m\left(\frac{x-v_1}{\alpha_1}\right)} + e^{-m\left(\frac{y-v_2}{\alpha_2}\right)} \right]^{\frac{1}{m}} \right\} \cdot \left\{ (m-1) + \left[e^{-m\left(\frac{x-v_1}{\alpha_1}\right)} + e^{-m\left(\frac{y-v_2}{\alpha_2}\right)} \right]^{\frac{1}{m}} \right\} \quad (74)$$

2.3.6. Prueba de bondad de ajuste

De Escalante y Reyes (2008, pág. 133):

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

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

$$EEA = \left[\frac{\sum_{i=1}^{n_j} (\hat{Q}_T^j - Q_T^j)^2}{n_j - mp} \right]^{1/2} \quad (75)$$

Donde

- Q_T^j Son los eventos Q_i^j ordenados de mayor a menor con un periodo de retorno asignado $T = \frac{n_j+1}{m}$ y una probabilidad de no excedencia $P = 1 - \frac{1}{T}$.
- n_j Longitud en años del registro analizado.
- m Número de orden del registro. $m = 1$ para el evento más grande, $m = n_j$ para el evento más chico.
- \hat{Q}_T^j Eventos estimados por cierta distribución para cada periodo de retorno T asignado a la muestra ordenada Q_i^j .
- mp Número de parámetros de la distribución ajustada, donde $mp = 2$ para la Gumbel.

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

2.3.7. Análisis de frecuencias de gastos máximos anuales

Metodología tomada de Escalante y Reyes (2008, pág. 134):

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

La secuencia de análisis es la siguiente:

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

Paso 2. Verificar la homogeneidad de la serie mediante las pruebas descritas en el capítulo 2.3.2.

Paso 3. Con las expresiones (1) a (9) se obtienen los estadísticos muestrales de la serie Q_i : \bar{x} , S^2 , S , g , k y Cv ; prefiriendo los no sesgados, dado que generalmente se trabaja con muestras pequeñas.

Paso 4. Verificar la independencia de eventos de la serie Q_i según lo descrito en el capítulo 2.3.3.

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

$$T = \frac{n + 1}{m} \quad (\text{Ley empírica de Weibull}) \quad (76)$$

$$P(X \leq x) = 1 - \frac{1}{T} \quad (77)$$

Donde

n Tamaño de la muestra analizada en años.

m número de orden del registro.

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

Paso 6. A la serie Q_i se le ajustan las diferentes distribuciones de probabilidad para el análisis de máximos y se selecciona aquel que proporcione el mínimo error estándar de ajuste EEA.

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

2.4. Análisis regional hidrológico

El análisis de frecuencias se emplea para estimar eventos de cierto periodo de retorno Q_T , en un sitio de proyecto. Si el periodo de retorno T es grande, comparado con la longitud n de la serie analizada, entonces el error del estimador \hat{Q}_T puede ser muy alto e ineficiente

para propósitos de diseño. Más aún, en algunos sitios de interés se carece de información hidrométrica, lo que dificulta la obtención de \hat{Q}_T . Estos inconvenientes han propiciado la generación de modelos de estimación regional, los cuales hacen uso de la información proveniente de cuencas vecinas.

La mayor parte de los modelos regionales propuestos en la literatura requieren que la región considerada en el estudio sea homogénea de alguna manera cuantificable.

Tanto las técnicas de delimitación de regiones homogéneas como los modelos regionales hacen uso de las características fisiográficas de las cuencas. (Escalante & Reyes, 2008, pág. 157)

2.4.1. *Delimitación de regiones homogéneas*

En general, la delimitación regional se ha sustentado en considerar áreas geográficamente continuas, límites políticos o administrativos. Sin embargo, si la variabilidad espacial de las características fisiográficas o hidrológicas es grande, la consideración de homogeneidad no se puede garantizar.

A las técnicas que ubican estaciones en forma iterativa, con base en un algoritmo puramente matemático se le denominan técnicas de racimo. Con esta metodología no es necesario que una cuenca se encuentre dentro de un grupo que sea geográficamente contiguo. Estas regiones podría decirse, serán comunes respecto a sus series de datos en un espacio variable multidimensional antes que en el espacio geográfico. Sin embargo, existen numerosos problemas asociados a la regionalización, y en particular al análisis de grupos. El primer problema común es el relacionado con la selección de las variables para evaluar el grado de similitud entre las diferentes cuencas.

Algunas cuencas son similares con respecto a cierta variable, como la cubierta forestal y la lluvia anual, pero no a otras, como las características fisiográficas. Además, con el análisis por racimos, literalmente cualquier conjunto de variables es capaz de generar grupos. Lo cual es muy útil ya que de esta manera se podrá seleccionar las variables de acuerdo con la importancia del problema.

Si dichas variables son medidas a partir de longitudes de registros grandes, esto contribuirá a obtener una mayor similitud de aquellas características que se obtienen a partir de registros cortos.

Otro problema en el análisis por racimos es la variedad de los diferentes algoritmos que pueden emplearse para formar grupos. Desdichadamente, las diferentes técnicas de agrupación aplicadas a un mismo registro a menudo producen estructuras que son sustancialmente diferentes.

Puede pensarse que la selección de la región es un problema trivial, sin embargo, la inclusión o exclusión de información dentro del modelo regional puede producir estimadores \hat{Q}_T poco confiables. (Escalante & Reyes, 2008, pág. 171)

Se deben utilizar periodos comunes consecutivos para delimitar las regiones homogéneas, estos deben iniciar y finalizar en el mismo año, con un mínimo de diez años de registro. Se deben calcular los estadísticos de las muestras en sus periodos comunes.

2.4.1.1. Método geográfico

Esta técnica se basa en suponer que una región es homogénea, dada la proximidad de sus cuencas, donde la delimitación se puede realizar trazando un radio de 80 km a partir del centro de gravedad del sitio de proyecto y tomando toda la información proveniente de los sitios que quedan dentro de este radio. (Escalante & Reyes, 2008, pág. 171)

2.4.1.2. Método de los trazos multidimensionales

Nathan y McMahon (1990) desarrollaron una técnica de regionalización que resuelve los problemas asociados con la selección de una adecuada técnica de racimos, la definición de la región homogénea y la predicción del grupo de membresía al cual pertenecería una nueva cuenca.

El método emplea la técnica de regresión lineal múltiple para seleccionar las características fisiográficas y climatológicas más adecuadas.

La heterogeneidad de los grupos que se forman de manera preliminar se evalúa mediante una técnica de posicionamiento propuesta por Andrews (1972), en la cual un punto en el espacio multidimensional se representa por una curva en dos dimensiones vía la función

$$f(t) = \frac{x_1}{\sqrt{2}} + x_2 \text{sen}(t) + x_3 \cos(t) + x_4 \text{sen}(2t) + x_5 \cos(2t) + \dots \quad (78)$$

Donde x_1, x_2, \dots Son las características fisiográficas y/o meteorológicas obtenidas del análisis de regresión y la función se evalúa en el rango $-\pi \leq t \leq \pi$.

El hecho de que esta función preserva las distancias, la hace la técnica ideal de comparación visual para la formación de grupos homogéneos. Los racimos de cuencas con comportamientos similares aparecen como una banda de curvas muy próximas unas de otras.

Una característica del método, que no se identifica de forma inmediata, consiste en que los resultados que se obtienen dependen del tipo y del ordenamiento de las variables seleccionadas. Las primeras variables son asociadas con componentes cíclicos de bajas frecuencias y las ultimas de alta frecuencia. Las bajas frecuencias son más fáciles de observar, de esta manera, x_1 representa a aquella que en el análisis de regresión resulta más significativa desde el punto de vista estadístico, x_2 a la segunda y así sucesivamente.

Una vez que se han desarrollado las funciones del tipo $f(t)$ para cada uno de los sitios involucrados, se generan envolventes que forman conjuntos de áreas de comportamiento similar y se derivará una curva media $\bar{f}(t)$ que represente los atributos físicos de cada grupo formado.

Para las curvas donde se tenga duda de que pertenezca o no a un grupo particular se hará la siguiente prueba

$$SS = \sum_{t=-\pi}^{\pi} \{ [f^j(t)] - [\bar{f}(t)]_k \}^2 \quad (79)$$

Donde SS es la suma de las diferencias al cuadrado entre los puntos que forman la curva en el sitio j , la curva característica $\bar{f}(t)$ de cada uno de los k grupos ya formados. La cuenca se localizará en la región k donde se haya obtenido el menor valor de SS . (Escalante & Reyes, 2008, pág. 172)

A cada muestra del periodo común consecutivo se le ajusta la mejor distribución de probabilidad, y se obtiene en cada sitio $i = 1, 2, \dots, n$, para $T = 2, 5, 10, 20, 50, 100, 500, 1000, 5000$ y 10000 años.

Se plantean 10 sistemas de ecuaciones, una para cada periodo de retorno:

$$\begin{aligned} Q_{T=2 \text{ años}}^{i=1} &= a + b_1 x_1^1 + b_2 x_2^1 + \dots \\ Q_{T=2 \text{ años}}^{i=2} &= a + b_1 x_1^2 + b_2 x_2^2 + \dots \\ Q_{T=2 \text{ años}}^{i=3} &= a + b_1 x_1^3 + b_2 x_2^3 + \dots \\ &\dots \\ Q_{T=2 \text{ años}}^{i=n} &= a + b_1 x_1^n + b_2 x_2^n + \dots \\ &\dots \end{aligned} \quad (80)$$

Para n estaciones, habrá máximo n menos una característica fisiográfica. El número de características fisiográficas dependerá de su importancia, entre mayor sea el valor de t , más importante es. Cuando t tienda a cero, no se tiene en cuenta la característica.

2.4.1.3. Método del coeficiente de variación

Esta técnica también hace uso de una comparación gráfica entre atributos estadísticos de los registros analizados.

Para cada una de las series j , por ejemplo, de gastos máximos anuales, se obtendrán con las expresiones (1) y (9) los estadísticos muestrales \bar{x}_j y Cv_j . Posteriormente, se construye una gráfica situando en el eje de las ordenadas los coeficientes Cv_j y en las abscisas las relaciones (\bar{x}_j/A_j) , donde A_j son las áreas drenadas de cada uno de los sitios j . Los grupos se forman por la concentración de puntos alrededor de determinado valor de coeficiente de variación.

Es importante considerar que la técnica requiere que los estadísticos \bar{x}_j y Cv_j se estimen de registros con distribución Normal, de tal manera que ninguna de las dos variables tenga tal influencia que pueda desproporcionar a la otra. Como esta condición no siempre se cumple será necesario transformar las series originales. (Escalante & Reyes, 2008, pág. 173)

2.4.1.4. Distancia euclidiana o región de influencia

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

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

$$D_{jb} = \left[\sum_{i=1}^p (C_j^i - C_b^j)^2 \right]^{1/2} \quad (81)$$

$$C_j^i = \frac{X_j^i}{S(X^i)} \quad (82)$$

$$C_b^i = \frac{X_b^i}{S(X^i)} \quad (83)$$

Donde

p número de atributos considerados

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

X_j^i valores del atributo i del sitio j

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

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

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

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

El próximo paso dentro del proceso de identificación de la región de influencia implica proponer un valor de tolerancia que acote la medida de la distancia. Así, se pudieran excluir todas aquellas estaciones que tengan una distancia más grande que el valor de \bar{D}_{jb} desde el sitio base. (Escalante & Reyes, 2008, págs. 173-174)

2.4.2. Técnica regional avenida índice

Descrito por Campos (1994): el método del índice de crecientes o de avenidas, desarrollado y propuesto por Tate Dalrymple (1960), se ha establecido como un procedimiento formal de la Exploración Geológica Estadounidense (U.S Geological Survey), y por ello en algunas ocasiones así se le designa. El método utiliza las crecientes anuales de ríos no regulados y con los datos de cada sitio i ($i = 1, 2, 3, \dots, m$).

Se sigue el procedimiento indicado por Campos (1994) con algunas variaciones:

Paso 1. A cada muestra i de la región homogénea en un periodo común se le aplica la distribución de probabilidad Gumbel, para obtener sus parámetros $\hat{\nu}$ y $\hat{\alpha}$.

Paso 2. Con los parámetros se estiman los gastos \hat{X}_T^i para los periodos de retorno $T = 1.1, 2, 2.33, 5, 10, 20, 50, 100, 500, 1000, 5000,$ y 10000 años. El periodo de 2.33 años se considera igual a la creciente media anual.

Paso 3. Se calculan los índices de crecientes o cocientes $\hat{Q}_T^i / \hat{Q}_{2.33}^i$ para todo T e i .

Para la finalidad del presente proyecto, que busca determinar si se obtiene una reducción del error estándar de ajuste al combinar la técnica avenida índice con el análisis multivariado, se continuó así: para comparar la técnica en su procedimiento tradicional y combinando el análisis multivariado, se asumió una estación a la vez como si tuviera escasa información, y no se tuvo en cuenta en la aplicación de la técnica con el fin de comparar sus datos medidos con los estimados. Posteriormente, se realizó el procedimiento con todas las estaciones.

Paso 4. Para cada periodo de retorno se obtienen los valores medianos de los índices de crecientes.

Paso 5. Se ajusta una ecuación logarítmica entre los valores medianos de los índices de crecientes y sus respectivos periodos de retorno T .

Paso 6. La serie i se ordena de mayor a menor, se le asigna un periodo de retorno $T = \frac{n+1}{m}$, se ingresa en la ecuación obtenida en el paso 5 y se estiman los valores medianos de los índices de crecientes.

Paso 7. Los valores medianos de los índices de crecientes estimados se multiplican por el $\hat{Q}_{2.33}^i$ de la estación que se quiere comparar con su registro medido, y se obtiene el error estándar de ajuste.

Para el caso donde se asume la estación como si no tuviera información, se ajustaron cinco tipos de curvas: lineal, polinomial de segundo grado, potencial, exponencial y logarítmica para comparar los $\hat{Q}_{2.33}^i$ con las características fisiográficas de las cuencas. Con la característica fisiográfica y el tipo de curva que mejor se ajustan a los $\hat{Q}_{2.33}^i$, se estima el $\hat{Q}_{2.33}^R$ de la estación con nula información, y se ingresa en el paso 7.

El valor de $\hat{Q}_{2.33}$ es mucho más estable y confiable que la media aritmética, ya que esta estimación da un mayor peso a las crecientes medias que a las extremas cuyos errores de muestreo son grandes, y por esta razón no está influenciado por un cambio en la inclusión o exclusión de una gran creciente, como la media aritmética (Dalrymple, 1960).

3. Recolección y análisis de información

De la Comisión Nacional del Agua (CONAGUA), específicamente en el Banco Nacional de Datos de Aguas Superficiales (BANDAS), se descargaron los catálogos de estaciones hidrométricas y de presas, con el fin de ubicar zonas con suficientes estaciones que sus gastos no estuvieran controlados. Posteriormente, de esta misma página se obtuvieron las bases de datos de las estaciones hidrométricas de las zonas de estudio del presente proyecto. La consulta y descarga de información se realizó entre los meses de septiembre y octubre del año 2019 en la dirección: <http://www.conagua.gob.mx/CONAGUA07/Contenido/Documentos/Portada%20BANDAS.htm>

Se requirió de archivos en formato shape para la ubicación de las presas y estaciones hidrométricas, los cuales se obtuvieron del Sistema Nacional de Información del Agua (SINA) y La Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO). Los enlaces respectivos consultados en septiembre del año 2019 fueron:

<http://sina.conagua.gob.mx/sina/tema.php?tema=presasPrincipales&ver=mapa&o=0&n=nacional>
http://www.conabio.gob.mx/informacion/metadatos/gis/esthidgw.xml?_xsl=/db/metadatos/xsl/fgdc_html.xsl&_indent=no

Finalmente, para la delimitación y obtención de las características fisiográficas de las cuencas, se descargaron los archivos Continuo de Elevación Mexicano (CEM) del Instituto Nacional de Estadística y Geografía (INEGI), los cuales tienen una resolución de 15 por 15 metros, y se descargaron en la dirección: <https://www.inegi.org.mx/app/geo2/elevacionesmex/>

En la Figura 1 se pueden observar las tres zonas de estudio elegidas para el proyecto. Dentro de cada zona se eliminaron aquellas estaciones con registros cortos y las que eran gastos controlados según el catálogo del BANDAS. Adicionalmente, tampoco se tuvieron en cuenta para el estudio aquellas estaciones cuya ubicación o delimitación no coincidían con los indicados en el catálogo.

Después de este primer filtro, se obtuvieron las características fisiográficas de las cuencas que tenían al menos 10 años de periodo común consecutivo para delimitar la región homogénea de cada zona, las estaciones seleccionadas se muestran en la Tabla 5.



Figura 1. Zonas de estudio. (Elaborado por la autora)

El relleno de los datos faltantes de las estaciones pertenecientes a cada región homogénea se realizó mensualmente. Luego de normalizar las series, se determinó un periodo común, eliminando aquellos años donde no se tuviera dato en alguna de las estaciones, con el fin de buscar la mejor correlación entre ellas y proceder con su respectivo llenado; cada vez que al rellenar un dato se completaba la serie en ese año, se agregaba al periodo común y se determinaba nuevamente la correlación entre las estaciones. La extensión resultante del proceso de relleno de datos de cada estación se puede ver en la Tabla 5. A estas series se le aplicaron las pruebas de detección de puntos atípicos, para confirmar que se trataban de eventos reales y no de un error en el registro. El criterio para eliminar un “dato atípico” era si dos de las tres técnicas lo identificaban y si no presentaba el mismo comportamiento de las estaciones de la región homogénea.

Posteriormente se aplicaron las pruebas de independencia, para verificar que se trataban de series aleatorias y poder aplicarles el análisis de frecuencias; así mismo, se realizaron las pruebas de homogeneidad, ya que la distribución aplicada en el estudio parte de la hipótesis de que la muestra es estacionaria. Estos resultados se pueden observar en la Tabla 6.

Tabla 5. Extensión de los registros rellenados de las estaciones en estudio

Estación	Nombre	Inicio	Fin	n	Estación	Nombre	Inicio	Fin	n
Región homogénea RH 10					Región homogénea RH 23				
10029	Naranjo	1939	1984	46	23008	Huixtla	1959	1996	38
10031	Guamúchil	1939	1971	33	23012	Novillero	1963	1990	28
10034	Zopilote	1939	1991	53	23015	Cacaluta	1964	2005	42
10036	Jaina	1941	1993	53	23016	Cintalapa	1964	1998	35
10037	Huites	1941	1992	52	23017	Coapa	1964	1998	35
10040	Santa Cruz	1943	1995	53	23019	Despoblado	1964	2005	42
10057	Bamícori	1951	1986	36	23020	Huehuetán	1964	1998	35
10064	Chínipas	1964	2008	45	Región homogénea RH 26				
10065	Ixpalino	1953	2003	51	26241	Ballesmi	1961	2014	54
10066	Choix	1955	2005	51	26263	Santa Rosa	1961	2014	54
10070	Acatitán	1955	2007	53	26267	Gallinas	1961	2014	54
10077	Palo Dulce	1957	1986	30	26285	Micos	1961	2014	54
10078	La Tina	1960	1984	25	26389	Vigas	1961	2014	54
10079	Badiraguato	1959	1999	41	26430	Agua Buena	1973	2014	42
10083	El Quelite	1960	1993	34					
10087	Tamazula	1962	2014	53					
10100	Urique II	1967	2002	36					
10110	Toahayana	1957	1987	31					
10112	Guatenipa	1968	2013	46					
10113	La Huerta	1969	1999	31					
10137	Tecusiapa	1958	2008	51					

Tabla 6. Resultados pruebas de homogeneidad e independencia

Región homogénea RH 10												
	Estación	10029	10031	10034	10036	10037	10040	10057	10064	10065	10066	10070
¿I?	Anderson	Si										
	Mantilla	Si										
	Amigó	Si										
¿H?	Helmert	No	Si	Si	Si	No	No	No	No	Si	No	Si
	T-Student	Si										
	Cramer	Si										
	Pettitt	No	Si	No	Si							
	Normal Estándar	Si	Si	Si	No	Si						
	Buishand	Si										
	Von Neumann	Si	No	Si								

*Continua en la siguiente página

*Continuación Tabla 6. Resultados pruebas de homogeneidad e independencia

	Estación	10077	10078	10079	10083	10087	10100	10110	10112	10113	10137
¿I?	Anderson	Si									
	Mantilla	Si	No	Si							
	Amigó	Si									
¿H?	Helmert	Si	Si	No	No	No	Si	Si	Si	Si	No
	T-Student	Si									
	Cramer	Si									
	Pettitt	Si	Si	No	Si	Si	Si	Si	No	No	Si
	Normal Estándar	Si	No	No	Si						
	Buishand	Si									
	Von Neumann	Si									

Región homogénea RH 23

	Estación	23008	23009	23012	23015	23016	23017	23019	23020
¿I?	Anderson	Si	Si	Si	Si	Si	No	No	No
	Mantilla	Si							
	Amigó	Si							
¿H?	Helmert	No	Si	Si	No	No	No	No	Si
	T-Student	No	No	Si	No	Si	No	No	Si
	Cramer	No	No	Si	No	Si	No	No	Si
	Pettitt	No	No	Si	No	No	No	No	Si
	Normal Estándar	No	No	Si	No	No	No	No	Si
	Buishand	No	No	Si	No	No	No	No	Si
	Von Neumann	No	Si	No	No	No	No	No	Si

Región homogénea RH 26

	Estación	26241	26263	26267	26285	26389	26430
¿I?	Anderson	Si	Si	Si	Si	Si	Si
	Mantilla	Si	Si	Si	Si	Si	Si
	Amigó	Si	Si	Si	Si	Si	Si
¿H?	Helmert	No	No	No	No	No	Si
	T-Student	Si	Si	Si	Si	Si	Si
	Cramer	Si	Si	Si	Si	Si	Si
	Pettitt	Si	Si	Si	Si	No	Si
	Normal Estándar	Si	Si	Si	Si	Si	No
	Buishand	Si	Si	Si	Si	No	No
	Von Neumann	Si	Si	Si	Si	Si	Si

*¿I?: ¿Independiente? - ¿H?: ¿Homogénea?

Las estaciones que no tienen al menos 30 años de registro consecutivo común con su región homogénea se descartaron del análisis regional. Tampoco se tuvieron en cuenta las estaciones a las que no se le ajustó la distribución Gumbel, es decir, aquellas que estimaron valores inferiores a cero, lo que no describe el comportamiento de la variable aleatoria (gastos máximos instantáneos).

Estos filtros redujeron considerablemente la cantidad de estaciones para aplicar la técnica avenida índice. Como se puede observar en la región homogénea de la RH 10 (Figura 2), donde ordenando las estaciones se encontraron dos periodos con más de 30 años de registro en común, en el primer periodo, dos de las seis estaciones no se ajustaron a la distribución Gumbel, quedando cuatro para el análisis, pero en el segundo periodo, tres de los cinco registros tampoco se pudieron estimar por dicha distribución, y por tanto se descartaron todas las estaciones.

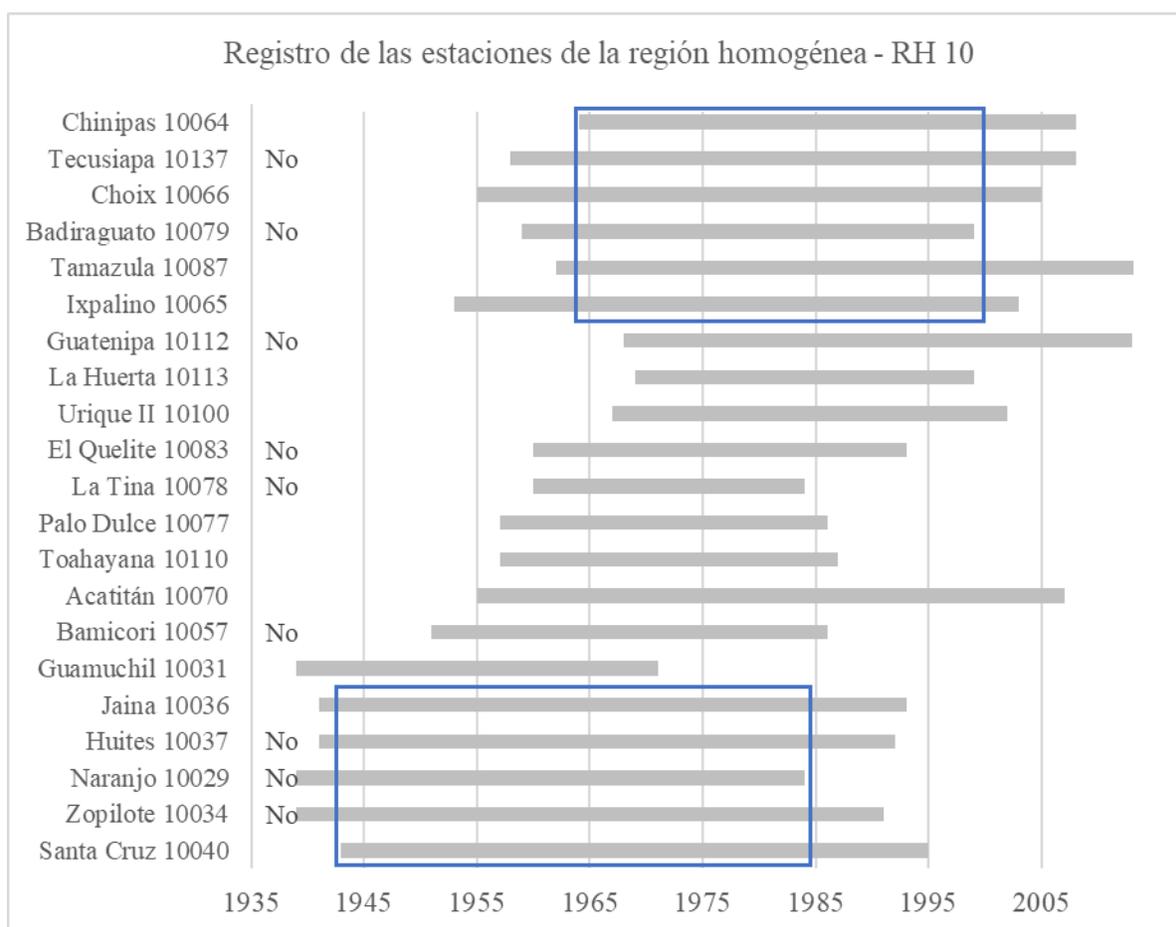


Figura 2. Selección de estaciones en la región homogénea RH 10

Para aplicar la técnica regional de la avenida índice es necesario contar con un periodo común, en el presente proyecto se analizaron tres periodos, los cuales se observan en la Tabla 7 junto con el número total de estaciones por cada región homogénea.

Tabla 7. Periodos de análisis para la técnica regional avenida índice

Región hidrológica	Periodo común		Años n	Número de estaciones
	Inicio	Fin		
Región homogénea RH 10	1964	1999	36	4
Región homogénea RH 23	1964	1996	33	6
Región homogénea RH 26	1973	2014	42	4

Para el método tradicional a cada estación se le ajustó la distribución Gumbel por máxima verosimilitud, y se estimaron los gastos máximos instantáneos para diferentes periodos de retorno de diseño: 1.1, 2, 2.33, 5, 10, 20, 50, 100, 500, 1000, 5000, 10000 años.

En el análisis multivariado, los gastos para diferentes periodos de retorno se estimaron con los parámetros de la distribución Gumbel bivariada que mejor se ajustó a cada serie de datos, este proceso se realizó para cada estación (base) con una estación asociada a la vez, es decir, todas las pertenecientes a la misma región homogénea que compartan el periodo común establecido, sea que se les ajuste la distribución Gumbel o no.

Para comparar los resultados obtenidos se aplicó en ambos procedimientos: sin una estación a la vez y con todas las estaciones del periodo común. Además, se consideraron los dos casos posibles: sitios con escasa y nula información. Esto quiere decir, que una muestra de m_s estaciones se analiza de ocho maneras diferentes, los mismos casos para la avenida índice tradicional y combinando el análisis multivariado:

Caso 1. Gastos estimados con la mediana regional sin una estación a la vez.

Caso 2. Gastos estimados con la mediana regional con todas las estaciones.

Caso 3. Gastos estimados con la mediana regional con el $\hat{Q}_{2.33}^R$ estimado con la función regional – sin una estación a la vez.

Caso 4. Gastos estimados con la mediana regional con el $\hat{Q}_{2.33}^R$ estimado con la función regional - Todas las estaciones.

3.1. Técnica regional avenida índice

En la Tabla 8 se resumen los parámetros obtenidos por la distribución Gumbel para cada estación en su periodo común con su error estándar de ajuste respecto a los gastos medidos.

Tabla 8. Parámetros por la distribución Gumbel univariada

Estación	Nombre	Periodo común		Años n	EEA [m ³ /s]	Parámetro	
		Inicio	Fin			Ubicación	Escala
Región homogénea RH 10							
10064	Chínipas	1964	1999	36	132.26	717.18	417.76
10065	Ixpalino	1964	1999	36	701.27	829.51	552.52
10066	Choix	1964	1999	36	167.73	229.61	143.95
10087	Tamazula	1964	1999	36	200.37	440.96	199.05
Región homogénea RH 23							
23008	Huixtla	1964	1996	33	25.78	177.34	122.61
23009	Pijjiapan	1964	1996	33	72.39	189.34	126.29
23015	Cacaluta	1964	1996	33	19.86	141.58	79.100
23016	Cintalapa	1964	1996	33	59.12	116.30	86.170
23019	Despoblado	1964	1996	33	38.56	192.56	137.87
23020	Huehuetán	1964	1996	33	54.17	336.21	119.99
Región homogénea RH 26							
26241	Ballesmi	1973	2014	42	43.65	208.79	118.48
26267	Gallinas	1973	2014	42	95.42	276.04	135.95
26285	Micos	1973	2014	42	46.68	221.67	110.18
26430	Agua Buena	1973	2014	42	35.67	86.590	48.700

Se ejemplifica el procedimiento con la región homogénea de la RH 26. Primero se estimaron los gastos por la distribución Gumbel con los parámetros indicados en la Tabla 8, estos eventos se modulan en el $\hat{Q}_{2.33}$ de cada estación. Asumiendo que la estación base 26241 tiene escasa información, se determina la mediana entre las tres estaciones restantes (26267, 26285 y 26430), ver Tabla 9. Luego se ajustó una ecuación logarítmica entre los valores de la mediana regional sin la estación 26241 y sus respectivos periodos de retorno (Figura 3).

La serie de la estación 26241 entre los años 1973 y 2014; correspondiente al periodo común en estudio; se ordenó de mayor a menor y se le asignó un periodo de retorno $T = \frac{n+1}{m}$, estos se ingresaron en la ecuación obtenida en la Figura 3 y se estimaron los valores medianos.

Para estimar los gastos regionalmente, se multiplicaron los valores medianos estimados por el $\hat{Q}_{2.33}$ de la estación 26241, con estos últimos se calculó el error estándar de ajuste respecto a la serie medida.

Tabla 9. Ejemplo de aplicación avenida índice asumiendo la estación 26241 con escasa información

Tr años	Eventos estimados \hat{Q}_T^i por la distribución Gumbel			Eventos modulados $\hat{Q}_T^i / \hat{Q}_{2.33}^i$			Mediana regional sin la estación
	26267	26285	26430	26267	26285	26430	26241
1.1	157.14	125.30	43.99	0.443	0.439	0.383	0.439
2	325.87	262.06	104.44	0.919	0.918	0.910	0.918
2.33	354.70	285.42	114.77	1.000	1.000	1.000	1.000
5	479.97	386.94	159.64	1.353	1.356	1.391	1.356
10	581.99	469.63	196.18	1.641	1.645	1.709	1.645
20	679.86	548.94	231.24	1.917	1.923	2.015	1.923
50	806.53	651.60	276.62	2.274	2.283	2.410	2.283
100	901.46	728.54	310.62	2.541	2.552	2.707	2.552
500	1120.81	906.32	389.20	3.160	3.175	3.391	3.175
1000	1215.12	982.74	422.98	3.426	3.443	3.686	3.443
5000	1433.98	1160.12	501.38	4.043	4.065	4.369	4.065
10000	1528.23	1236.50	535.14	4.308	4.332	4.663	4.332

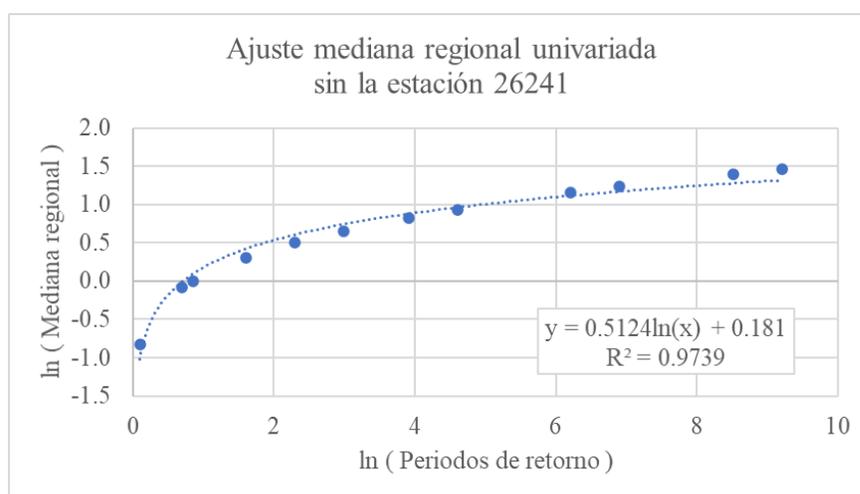


Figura 3. Ajuste mediana regional univariada sin la estación 26241

Este procedimiento se repitió para cada una de las estaciones en estudio, es decir, luego se tomó como base la estación 26267 y se obtuvo la mediana regional con las estaciones 26241, 26285 y 26430, y así sucesivamente. De igual manera, se obtuvo la mediana regional tomando en cuenta todas las estaciones del periodo en estudio.

Para estimar gastos en puntos de proyecto con nula información, es necesario encontrar una ecuación regional que relacione los $\hat{Q}_{2.33}$ de las estaciones en análisis con una característica fisiográfica, para lo cual se tuvo en cuenta cuatro tipos de curvas: lineal, polinomial de segundo grado, potencial y exponencial. En el caso del ejemplo que solo se cuenta con cuatro estaciones, no se realizó el análisis sin una estación a la vez, porque al quedar con tres puntos, cualquier atributo otorgará un coeficiente de determinación igual a uno al ajustar un polinomio de segundo grado. Por este motivo, la estimación regional para casos de nula información sin una estación a la vez, solo se realizó en la región homogénea de la RH 23.

3.2. Técnica regional avenida índice combinando el análisis multivariado

El primer paso consistió en encontrar los mejores parámetros obtenidos por la distribución Gumbel bivariada entre cada estación y el resto de las estaciones de su región homogénea que compartieran el periodo común establecido en Tabla 7, su identificación se hizo por medio del error estándar de ajuste. Nuevamente se ejemplifica con la región homogénea de la RH 26 (Tabla 10), solo se muestra la obtención de los parámetros por la distribución Gumbel bivariada de la estación 26241, el procedimiento se repite para el resto de los casos.

Tabla 10. Parámetros estimados por la distribución Gumbel bivariada, estación 26241 de la región homogénea de la RH 26.

Estación base	Estación asociada	Parámetros			EEA [m ³ /s]
		Ubicación	Escala	m	
26241	26263	210.018	116.244	1.33	45.41
	26267	206.890	111.270	1.37	51.01
	26285	209.102	115.347	1.30	46.44
	26389	205.110	109.761	1.43	53.13
	26430	206.083	113.381	1.26	49.14

De esta manera se calcularon los parámetros por la distribución Gumbel bivariada para cada estación en su periodo común de las tres zonas en estudio, en la Tabla 11 se presentan con

sus respectivos errores estándar de ajuste respecto a los gastos medidos, así como sus estaciones base.

Tabla 11. Parámetros por la distribución Gumbel bivariada

Estación		Periodo común		Años	EEA	Parámetro		
Base	Asociada	Inicio	Fin	n	[m ³ /s]	Ubicación	Escala	m
Región homogénea RH 10								
10064	10065	1964	1999	36	132.40	717.105	417.605	1.00
10065	10079	1964	1999	36	691.74	836.758	564.452	1.00
10066	10137	1964	1999	36	164.59	232.509	148.273	1.09
10087	10064	1964	1999	36	196.28	444.008	205.090	1.10
Región homogénea RH 23								
23008	23009	1964	1996	33	25.67	180.55	120.75	1.89
23009	23019	1964	1996	33	70.23	194.253	130.564	1.66
23015	23008	1964	1996	33	18.59	144.40	81.80	1.88
23016	23019	1964	1996	33	56.83	119.15	89.63	1.41
23019	23009	1964	1996	33	38.23	193.648	136.081	1.66
23020	23019	1964	1996	33	53.73	336.53	120.65	1.21
Región homogénea RH 26								
26241	26263	1973	2014	42	45.41	210.02	116.24	1.33
26267	26241	1973	2014	42	92.55	278.80	139.38	1.37
26285	26241	1973	2014	42	44.56	223.18	112.19	1.30
26430	26241	1973	2014	42	34.83	87.88	49.61	1.26

4. Resultados y discusión

En la Tabla 12 se comparan los resultados del error estándar de ajuste entre los gastos estimados puntualmente por la distribución Gumbel univariada (Tabla 8) y bivariada (Tabla 11) respecto a los gastos máximos instantáneos medidos, en todos los casos el error es menor por la estimación bivariada, excepto en las estaciones 10064 y 26241.

Tabla 12. Comparación EEA [m^3/s] del análisis puntual entre la distribución Gumbel univariada y bivariada

Estación	Nombre	Periodo común		Años n	EEA [m^3/s]	
		Inicio	Fin		Univariada	Bivariada
Región homogénea RH 10						
10064	Chínipas	1964	1999	36	132.26	132.40
10065	Ixpalino	1964	1999	36	701.27	691.74
10066	Choix	1964	1999	36	167.73	164.59
10087	Tamazula	1964	1999	36	200.37	196.28
Región homogénea RH 23						
23008	Huixtla	1964	1996	33	25.78	25.67
23009	Pijjiapan	1964	1996	33	72.39	70.23
23015	Cacaluta	1964	1996	33	19.86	18.59
23016	Cintalapa	1964	1996	33	59.12	56.83
23019	Despoblado	1964	1996	33	38.56	38.23
23020	Huehuetán	1964	1996	33	54.17	53.73
Región homogénea RH 26						
26241	Ballesmi	1973	2014	42	43.65	45.41
26267	Gallinas	1973	2014	42	95.42	92.55
26285	Micos	1973	2014	42	46.68	44.56
26430	Agua Buena	1973	2014	42	35.67	34.83

A pesar de presentar una disminución del error al aplicar el análisis multivariado puntual por la distribución Gumbel bivariada, la diferencia es poca. Además, los valores del error estándar de ajuste son muy elevados en algunas estaciones, ya que estas en realidad no presentan un comportamiento que se ajuste a la distribución Gumbel, la cual se caracteriza por un coeficiente de asimetría de 1.1396 y un coeficiente de curtosis de 5.4. En la Tabla 13 se presentan los estadísticos de las series correspondientes a su respectivo periodo común en análisis. Otra razón que justifica los resultados del error estándar de ajuste es que las regiones de Sinaloa y la Costa de Chiapas, presentan una componente ciclónica al ser zonas costeras y, por tanto, las distribuciones

mezcladas son las que presentan un mejor ajuste a sus series de datos; por el contrario, en la región homogénea de la RH 26, la distribución Gumbel no es la que presenta un mejor ajuste a la serie de datos.

Tabla 13. Estadísticos de las series en el periodo común en análisis

Estación	Número datos	Media	Desviación estándar	Coeficiente		
				Asimetría	Curtosis	Variación
Región homogénea RH 10						
10064	36	969.25	570.76	1.06	4.38	0.59
10065	36	1224.90	1165.60	2.98	13.70	0.95
10066	36	328.28	286.67	3.37	18.64	0.87
10087	36	571.39	368.31	3.18	17.14	0.64
Región homogénea RH 23						
23008	33	245.06	140.25	0.21	2.82	0.57
23009	33	265.05	182.97	2.02	10.92	0.69
23015	33	186.85	100.21	0.96	4.96	0.54
23016	33	169.98	136.24	2.38	12.42	0.80
23019	33	268.07	154.81	0.07	2.49	0.58
23020	33	407.03	168.47	1.72	8.44	0.41
Región homogénea RH 26						
26241	42	281.92	168.63	1.19	4.39	0.60
26267	42	363.81	225.32	2.20	9.83	0.62
26285	42	290.94	163.45	1.37	4.93	0.56
26430	42	118.17	82.02	2.17	8.88	0.69

4.1. Ecuaciones de ajuste a la mediana regional

Como se ejemplificó en la Figura 3, la ecuación es un ajuste entre los periodos de retorno de diseño y sus respectivas medianas regionales. Para aplicar la ecuación de ajuste de la mediana regional se tiene que la x es el valor de los periodos de retorno estimados por la ecuación de Weibull correspondientes al periodo común en análisis de la serie medida y la y son los gastos medianos estimados regionalmente. En la Tabla 14 y la Tabla 15, se presentan las ecuaciones obtenidas por el análisis univariado y bivariado, en dos casos: sin tener en cuenta una estación a la vez y con todas las estaciones de la región homogénea en estudio, adicionalmente se agrega el valor del coeficiente de determinación R^2 entre los periodos de retorno de diseño y las medianas regionales.

Tabla 14. Ecuaciones de ajuste de la mediana regional – Análisis univariado

Sin la estación	Ecuación logarítmica	R ²
Región homogénea RH 10		
10064	$y = \exp (0.59786 * \ln[\ln(x)] + 0.15991)$	0.9892
10065	$y = \exp (0.56874 * \ln[\ln(x)] + 0.16942)$	0.9847
10066	$y = \exp (0.56874 * \ln[\ln(x)] + 0.16942)$	0.9847
10087	$y = \exp (0.59786 * \ln[\ln(x)] + 0.15991)$	0.9892
Todas	$y = \exp (0.58314 * \ln[\ln(x)] + 0.16502)$	0.9870
Región homogénea RH 23		
23008	$y = \exp (0.62437 * \ln[\ln(x)] + 0.14913)$	0.9925
23009	$y = \exp (0.64074 * \ln[\ln(x)] + 0.14147)$	0.9942
23015	$y = \exp (0.64074 * \ln[\ln(x)] + 0.14147)$	0.9942
23016	$y = \exp (0.62437 * \ln[\ln(x)] + 0.14913)$	0.9925
23019	$y = \exp (0.62437 * \ln[\ln(x)] + 0.14913)$	0.9925
23020	$y = \exp (0.64074 * \ln[\ln(x)] + 0.14147)$	0.9942
Todas	$y = \exp (0.63248 * \ln[\ln(x)] + 0.14542)$	0.9934
Región homogénea RH 26		
26241	$y = \exp (0.51241 * \ln[\ln(x)] + 0.18097)$	0.9739
26267	$y = \exp (0.55561 * \ln[\ln(x)] + 0.17292)$	0.9824
26285	$y = \exp (0.55561 * \ln[\ln(x)] + 0.17292)$	0.9824
26430	$y = \exp (0.51241 * \ln[\ln(x)] + 0.18097)$	0.9739
Todas	$y = \exp (0.53380 * \ln[\ln(x)] + 0.17764)$	0.9783

Tabla 15. Ecuaciones de ajuste de la mediana regional – Análisis multivariado

Sin la estación	Ecuación logarítmica	R ²
Región homogénea RH 10		
10064	$y = \exp (0.60497 * \ln[\ln(x)] + 0.15721)$	0.9901
10065	$y = \exp (0.56864 * \ln[\ln(x)] + 0.16945)$	0.9847
10066	$y = \exp (0.56864 * \ln[\ln(x)] + 0.16945)$	0.9847
10087	$y = \exp (0.60497 * \ln[\ln(x)] + 0.15721)$	0.9901
Todas	$y = \exp (0.58656 * \ln[\ln(x)] + 0.16389)$	0.9875
Región homogénea RH 23		
23008	$y = \exp (0.63156 * \ln[\ln(x)] + 0.14585)$	0.9933
23009	$y = \exp (0.62559 * \ln[\ln(x)] + 0.14858)$	0.9926
23015	$y = \exp (0.63156 * \ln[\ln(x)] + 0.14585)$	0.9933
23016	$y = \exp (0.62559 * \ln[\ln(x)] + 0.14858)$	0.9926
23019	$y = \exp (0.62559 * \ln[\ln(x)] + 0.14858)$	0.9926
23020	$y = \exp (0.63156 * \ln[\ln(x)] + 0.14585)$	0.9933
Todas	$y = \exp (0.62857 * \ln[\ln(x)] + 0.14723)$	0.9930
Región homogénea RH 26		
26241	$y = \exp (0.51617 * \ln[\ln(x)] + 0.18048)$	0.9747
26267	$y = \exp (0.54973 * \ln[\ln(x)] + 0.17432)$	0.9813
26285	$y = \exp (0.54973 * \ln[\ln(x)] + 0.17432)$	0.9813
26430	$y = \exp (0.51617 * \ln[\ln(x)] + 0.18048)$	0.9747
Todas	$y = \exp (0.53282 * \ln[\ln(x)] + 0.17782)$	0.9781

4.2. Análisis para sitios con escasa información

Los gastos medianos estimados regionalmente por las ecuaciones de las tablas anteriores se multiplican por el gasto de 2.33 años de la respectiva estación (Tabla 16), los cuales se obtuvieron con los parámetros indicados en la Tabla 8 y la Tabla 11. Los gastos obtenidos regionalmente se comparan con los medidos en cada estación y se calcula el error estándar de ajuste.

Tabla 16. Gasto estimado de 2.33 años [m³/s] por la distribución Gumbel univariada y bivariada

Estación	Nombre	$\hat{Q}_{2.33}$ [m ³ /s]	
		Univariada	Bivariada
Región homogénea RH 10			
10064	Chínipas	958.89	958.73
10065	Ixpalino	1149.19	1163.34
10066	Choix	312.90	318.30
10087	Tamazula	556.13	562.67
Región homogénea RH 23			
23008	Huixtla	248.28	250.42
23009	Pijijiapan	262.41	262.64
23015	Cacaluta	187.34	191.73
23016	Cintalapa	166.16	171.00
23019	Despoblado	272.33	272.38
23020	Huehuetán	405.63	406.34
Región homogénea RH 26			
26241	Ballesmi	277.34	277.28
26267	Gallinas	354.70	359.44
26285	Micos	285.42	288.10
26430	Agua Buena	114.77	116.58

Los resultados del error estándar de ajuste aplicando la técnica avenida índice por medio del análisis univariado y multivariado se observan en la Tabla 17.

Tabla 17. EEA [m³/s] por la técnica avenida índice para sitios con escasa información

Estación	EEA [m ³ /s] por la técnica avenida índice – escasa información			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
Región homogénea RH 10				
10064	107.77	106.65	110.76	109.93
10065	706.63	702.31	697.35	690.66
10066	165.56	164.36	163.37	161.61
10087	183.20	183.05	184.68	184.81

*Continúa en la siguiente página

*Continuación Tabla 17. EEA [m³/s] por la técnica avenida índice para sitios con escasa información

Estación	EEA [m ³ /s] por la técnica avenida índice – escasa información			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
Región homogénea RH 23				
23008	29.44	31.51	30.26	31.22
23009	69.23	69.91	69.59	69.76
23015	27.14	29.45	26.62	29.29
23016	58.84	57.77	58.29	57.55
23019	40.73	40.88	41.59	41.20
23020	105.52	104.29	104.03	103.75
Región homogénea RH 26				
26241	43.47	42.89	40.35	40.48
26267	88.12	89.17	90.75	91.20
26285	39.28	40.48	41.04	41.82
26430	36.88	36.48	35.55	35.46

4.3. Análisis para sitios con nula información

Para estimar gastos en sitios con nula información, se requiere determinar que atributo de las cuencas correspondientes a la región homogénea en estudio tiene una mejor relación con sus respectivos $\hat{Q}_{2.33}$ (Tabla 16). En el presente proyecto, solo se realizó el análisis sin una estación a la vez en la región homogénea 23, como se explicó en el capítulo anterior. Se encontró que el mejor ajuste en todos los casos propuestos se da por una función polinómica de segundo grado. Los resultados se pueden observar en la Tabla 18 y la Tabla 19.

Tabla 18. Ecuaciones para determinar el gasto de 2.33 años [m³/s] por la distribución Gumbel univariada

Estación	R ²	Atributo (x)	Ecuación
Región homogénea RH 10			
Todas	0.9998	A [km ²]	$y = \exp(-0.294 \ln(x)^2 + 5.53950 \ln(x) - 18.9494)$
Región homogénea RH 23			
23008	0.8726	S [%]	$y = \exp(1.328 \ln(x)^2 - 9.18180 \ln(x) + 20.8634)$
23009	0.9922	S [%]	$y = \exp(8.685 \ln(x)^2 - 51.1557 \ln(x) + 80.4051)$
23015	0.7586	S [%]	$y = \exp(4.565 \ln(x)^2 - 27.3489 \ln(x) + 46.2778)$
23016	0.7975	S [%]	$y = \exp(3.449 \ln(x)^2 - 20.8774 \ln(x) + 36.9598)$
23019	0.8001	S [%]	$y = \exp(4.293 \ln(x)^2 - 25.9828 \ln(x) + 44.6191)$

*Continúa en la siguiente página

*Continuación Tabla 18. Ecuaciones para determinar el gasto de 2.33 años [m³/s] por la distribución Gumbel univariada

Estación	R ²	Atributo (x)	Ecuación
23020	0.4542	S [%]	$y = \exp(4.043 \ln(x)^2 - 24.3311 \ln(x) + 41.9057)$
Todas	0.7818	S [%]	$y = \exp(4.564 \ln(x)^2 - 27.4278 \ln(x) + 46.4991)$
Región homogénea RH 26			
Todas	0.99997	A [km ²]	$y = \exp(-0.649 \ln(x)^2 + 9.07350 \ln(x) - 25.7406)$

*A: área en km², S: pendiente media de la cuenca en %.

Tabla 19. Ecuaciones para determinar el gasto de 2.33 años [m³/s] por la distribución Gumbel bivariada

Estación	R ²	Atributo (x)	Ecuación
Región homogénea RH 10			
Todas	0.9997	Área [km ²]	$y = \exp(-0.280 \ln(x)^2 + 5.31040 \ln(x) - 18.0005)$
Región homogénea RH 23			
23008	0.8729	S [%]	$y = \exp(1.545 \ln(x)^2 - 10.3620 \ln(x) + 22.4676)$
23009	0.9950	S [%]	$y = \exp(8.431 \ln(x)^2 - 49.6588 \ln(x) + 78.2202)$
23015	0.7692	S [%]	$y = \exp(4.514 \ln(x)^2 - 27.0219 \ln(x) + 45.7726)$
23016	0.8037	S [%]	$y = \exp(3.479 \ln(x)^2 - 21.0228 \ln(x) + 37.1347)$
23019	0.8139	S [%]	$y = \exp(4.231 \ln(x)^2 - 25.5937 \ln(x) + 44.0241)$
23020	0.4541	S [%]	$y = \exp(3.860 \ln(x)^2 - 23.2189 \ln(x) + 40.2280)$
Todas	0.7912	S [%]	$y = \exp(4.512 \ln(x)^2 - 27.0935 \ln(x) + 45.9754)$
Región homogénea RH 26			
Todas	0.99996	A [km ²]	$y = \exp(-0.645 \ln(x)^2 + 9.01650 \ln(x) - 25.5411)$

*A: área en km², S: pendiente media de la cuenca en %.

En la Tabla 20 se presentan los $\hat{Q}_{2.33}^R$ [m³/s] estimados con las ecuaciones anteriores.

Tabla 20. Gasto de 2.33 años [m³/s] estimados regionalmente para sitios con nula información

Estación	$\hat{Q}_{2.33}^R$ [m ³ /s] por la técnica avenida índice - nula información			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
Región homogénea RH 10				
10064			1012.27	1018.26
10065			1105.96	1114.79
10066			316.75	322.65
10087			540.75	545.39

* Continua en la siguiente página

* Continuación Tabla 20. Gasto de 2.33 años [m³/s] estimados regionalmente para sitios con nula información

Estación	$\hat{Q}_{2.33}^R$ [m ³ /s] por la técnica avenida índice - nula información			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
Región homogénea RH 23				
23008	165.87	172.99	222.21	226.20
23009	164.18	168.14	210.33	212.80
23015	211.36	214.40	204.09	207.56
23016	214.73	216.88	198.41	201.56
23019	311.38	313.02	296.50	297.51
23020	377.03	370.79	399.25	398.35
Región homogénea RH 26				
26241			249.19	251.29
26267			384.25	386.88
26285			280.45	283.48
26430			120.00	121.46

Los $\hat{Q}_{2.33}^R$ indicados en la Tabla 20, se multiplican por los gastos medianos estimados regionalmente por las ecuaciones de las Tabla 14 y Tabla 15. Luego se comparan los gastos obtenidos con los medidos en cada estación y se calcula el error estándar de ajuste (Tabla 21).

Tabla 21. EEA [m³/s] por la técnica avenida índice para sitios con nula información

Estación	EEA [m ³ /s] por la técnica avenida índice - nula información			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
Región homogénea RH 10				
10064			139.36	143.59
10065			713.54	707.87
10066			162.46	160.78
10087			184.92	184.28
Región homogénea RH 23				
23008	89.97	81.85	31.65	28.88
23009	140.81	136.78	95.93	93.97
23015	48.21	51.17	40.42	43.79
23016	74.05	75.65	62.75	64.77

* Continúa en la siguiente página

* Continuación Tabla 21. EEA [m³/s] por la técnica avenida índice para sitios con nula información

Estación	EEA [m ³ /s] por la técnica avenida índice - nula información			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
23019	74.61	76.43	60.12	60.79
23020	93.45	90.59	100.38	99.18
Región homogénea RH 26				
26241			50.93	49.55
26267			98.38	99.74
26285			40.41	40.79
26430			35.50	35.74

Al comparar la Tabla 21 con la Tabla 17 se puede observar que el estimar los gastos máximos instantáneos asumiendo las estaciones como si tuvieran nula información aumentó el error estándar de ajuste en la mayoría de los casos, y esto debido a la incertidumbre que se genera al estimar el gasto $\hat{Q}_{2.33}^R$ a partir de un atributo de la cuenca. Cabe resaltar que, en la mayoría de los casos, la diferencia del error no es mayor a 20 m³/s.

4.4. Discusión

En la Tabla 22 se resumen los errores estándar de ajuste de la Tabla 17 y la Tabla 21, solo se muestran los casos donde el análisis multivariado otorgó una disminución del error estándar de ajuste respecto al análisis univariado de la técnica regional avenida índice.

El combinar el análisis multivariado en la región homogénea de la región hidrológica número 10, logra un menor error estándar de ajuste en diez de los doce escenarios estudiados. Por el contrario, en la región homogénea de la RH 26, solo se logra una reducción en cuatro de los doce escenarios. Y en la región homogénea de la RH 23, nueve de los veinticuatro escenarios presentan un menor error estándar de ajuste con la combinación del análisis multivariado.

La inconsistencia en los resultados puede ser debido a que las series en estudio no presentan el mejor ajuste con la distribución Gumbel, la cual es la requerida por la técnica regional avenida índice, además, en las regiones costeras estudiadas en el presente proyecto se ajusta mejor las distribuciones mezcladas. Otra razón puede ser el limitar la información al periodo común como solicita la técnica.

Tabla 22. Casos con menor EEA [m³/s] por la técnica avenida índice combinando el análisis multivariado respecto al univariado

Estación	Casos con menor EEA [m ³ /s] por la técnica avenida índice combinando el análisis multivariado			
	Sin la estación en mención		Todas las estaciones	
Información:	Escasa	Nula	Escasa	Nula
Región homogénea RH 10				
10064	Si		Si	
10065	Si		Si	Si
10066	Si		Si	Si
10087	Si			Si
Región homogénea RH 23				
23008		Si		Si
23009		Si		Si
23015				
23016	Si		Si	
23019			Si	
23020	Si	Si	Si	Si
Región homogénea RH 26				
26241	Si			Si
26267				
26285				
26430	Si		Si	

Como un análisis adicional, se probó el aplicar la técnica avenida índice con el mismo procedimiento ya antes mencionado al registro completo de las estaciones de la región hidrológica RH 10. Primero se estimaron los parámetros de ubicación y escala por la distribución Gumbel univariada (Tabla 23), y luego por la Gumbel bivariada (Tabla 24), la cual se probó la mejor combinación para cada estación con los registros completos. Con los parámetros se estimaron los gastos máximos instantáneos para diferentes periodos de retorno de diseño, los cuales se modularon en su respectivo $\hat{Q}_{2.33}$ y se obtuvieron las medianas regionales al no tener en cuenta de a una estación a la vez y con todas las estaciones, como lo explicado en la Tabla 9. Con esta información se determinó la ecuación de ajuste de la mediana regional en cada caso, como lo mostrado en la Figura 3, los resultados se observan en la Tabla 25 y Tabla 26.

Tabla 23. Parámetros por la distribución Gumbel univariada RH 10, periodo completo

Estación	Año inicial		Años n	EEA [m ³ /s]	Parámetro	
	Inicio	Fin			Ubicación	Escala
10029	1939	1985	47	285.30	346.19	381.58
10031	1939	1975	37	335.91	372.66	337.98
10034	1939	2014	76	66.39	177.08	183.64
10036	1941	2014	74	614.67	510.68	481.90
10037	1941	1992	52	1712.25	2095.45	1674.62
10040	1943	2014	72	596.73	474.69	526.35
10057	1951	1986	36	71.54	101.51	107.24
10064	1964	2008	45	129.67	711.03	399.33
10065	1953	2003	51	571.80	816.85	514.13
10066	1955	2005	51	154.26	209.58	141.77
10077	1957	1986	30	860.80	664.97	476.69
10083	1960	1993	34	138.73	306.67	276.12
10087	1962	2014	53	804.58	431.16	284.13
10100	1967	2002	36	35.52	253.23	115.19
10110	1957	1987	31	191.74	742.45	465.37
10112	1968	2013	46	583.27	996.98	910.75
10113	1969	1999	31	120.99	627.01	457.73
10137	1958	2008	51	440.59	673.47	558.04

Tabla 24. Parámetros por la distribución Gumbel bivariada RH 10, periodo completo

Estación		Periodo común		Años n	EEA [m ³ /s]	Parámetro		
Base	Asociada	Inicio	Fin			Ubicación	Escala	m
10029	10110	1939	1985	47	282.47	359.86	382.14	1.19
10031	10087	1939	1975	37	334.70	369.87	340.28	1.19
10034	10037	1939	2014	76	61.66	184.61	187.86	1.05
10036	10034	1941	2014	74	590.27	515.52	517.18	1.27
10037	10034	1941	1992	52	1455.92	2263.25	2041.07	1.05
10040	10029	1943	2014	72	564.88	552.52	573.21	1.03
10057	10040	1951	1986	36	67.90	98.93	112.24	1.09
10064	10066	1964	2008	45	109.02	747.30	430.66	1.12
10065	10113	1953	2003	51	550.72	826.27	544.11	1.64
10066	10064	1955	2005	51	148.66	214.00	149.13	1.12
10077	10110	1957	1986	30	823.38	706.73	543.14	1.13
10083	10100	1960	1993	34	137.96	306.01	277.10	1.07
10087	10031	1962	2014	53	801.02	435.68	290.35	1.19
10100	10113	1967	2002	36	34.20	254.95	118.26	1.34
10110	10034	1957	1987	31	184.30	723.17	480.46	1.18
10112	10057	1968	2013	46	520.68	1033.99	982.25	1.07
10113	10036	1969	1999	31	116.36	633.80	471.77	1.03
10137	10065	1958	2008	51	406.82	722.66	593.35	1.07

Tabla 25. Ecuaciones de ajuste de la mediana regional – Análisis univariado RH 10, periodo completo

Sin la estación	Ecuación logarítmica	R ²
10029	$y = \exp(0.71705 * \ln[\ln(x)] + 0.095936)$	0.9985
10031	$y = \exp(0.71705 * \ln[\ln(x)] + 0.095936)$	0.9985
10034	$y = \exp(0.71705 * \ln[\ln(x)] + 0.095936)$	0.9985
10036	$y = \exp(0.71705 * \ln[\ln(x)] + 0.095936)$	0.9985
10037	$y = \exp(0.73971 * \ln[\ln(x)] + 0.079446)$	0.9986
10040	$y = \exp(0.71705 * \ln[\ln(x)] + 0.095936)$	0.9985
10057	$y = \exp(0.71705 * \ln[\ln(x)] + 0.095936)$	0.9985
10064	$y = \exp(0.73971 * \ln[\ln(x)] + 0.079446)$	0.9986
10065	$y = \exp(0.73971 * \ln[\ln(x)] + 0.079446)$	0.9986
10066	$y = \exp(0.73971 * \ln[\ln(x)] + 0.079446)$	0.9986
10077	$y = \exp(0.73971 * \ln[\ln(x)] + 0.079446)$	0.9986
10083	$y = \exp(0.71705 * \ln[\ln(x)] + 0.095936)$	0.9985
10087	$y = \exp(0.73971 * \ln[\ln(x)] + 0.079446)$	0.9986
10100	$y = \exp(0.73971 * \ln[\ln(x)] + 0.079446)$	0.9986
10110	$y = \exp(0.73971 * \ln[\ln(x)] + 0.079446)$	0.9986
10112	$y = \exp(0.71705 * \ln[\ln(x)] + 0.095936)$	0.9985
10113	$y = \exp(0.73971 * \ln[\ln(x)] + 0.079446)$	0.9986
10137	$y = \exp(0.71705 * \ln[\ln(x)] + 0.095936)$	0.9985
Todas	$y = \exp(0.72818 * \ln[\ln(x)] + 0.087998)$	0.9986

Tabla 26. Ecuaciones de ajuste de la mediana regional – Análisis multivariado RH 10, periodo completo

Sin la estación	Ecuación logarítmica	R ²
10029	$y = \exp(0.73381 * \ln[\ln(x)] + 0.08387)$	0.9986
10031	$y = \exp(0.73381 * \ln[\ln(x)] + 0.08387)$	0.9986
10034	$y = \exp(0.73381 * \ln[\ln(x)] + 0.08387)$	0.9986
10036	$y = \exp(0.73381 * \ln[\ln(x)] + 0.08387)$	0.9986
10037	$y = \exp(0.73381 * \ln[\ln(x)] + 0.08387)$	0.9986
10040	$y = \exp(0.73381 * \ln[\ln(x)] + 0.08387)$	0.9986
10057	$y = \exp(0.73381 * \ln[\ln(x)] + 0.08387)$	0.9986
10064	$y = \exp(0.80209 * \ln[\ln(x)] + 0.02771)$	0.9962
10065	$y = \exp(0.80209 * \ln[\ln(x)] + 0.02771)$	0.9962
10066	$y = \exp(0.80209 * \ln[\ln(x)] + 0.02771)$	0.9962
10077	$y = \exp(0.80209 * \ln[\ln(x)] + 0.02771)$	0.9962
10083	$y = \exp(0.73381 * \ln[\ln(x)] + 0.08387)$	0.9986
10087	$y = \exp(0.80209 * \ln[\ln(x)] + 0.02771)$	0.9962
10100	$y = \exp(0.80209 * \ln[\ln(x)] + 0.02771)$	0.9962
10110	$y = \exp(0.80209 * \ln[\ln(x)] + 0.02771)$	0.9962
10112	$y = \exp(0.73381 * \ln[\ln(x)] + 0.08387)$	0.9986
10113	$y = \exp(0.80209 * \ln[\ln(x)] + 0.02771)$	0.9962
10137	$y = \exp(0.80209 * \ln[\ln(x)] + 0.02771)$	0.9962
Todas	$y = \exp(0.76592 * \ln[\ln(x)] + 0.05881)$	0.9981

Para el escenario de escasa información, los gastos medianos estimados regionalmente por las ecuaciones de las tablas anteriores se multiplican por el gasto de 2.33 años de la respectiva estación (Tabla 27). Los gastos obtenidos se comparan con los medidos en cada estación y se calcula el error estándar de ajuste, los cuales se pueden observar en la Tabla 28 y de la cual cabe resaltar que en catorce de las dieciocho estaciones se tiene una disminución del error al combinar la técnica avenida índice con el análisis multivariado, tanto en el caso de no tener en cuenta de a una estación a la vez, como al usar todas las estaciones de la zona en estudio.

Tabla 27. Gasto estimado de 2.33 años [m³/s] por la distribución Gumbel univariada y bivariada, RH 10, periodo completo

Clave	Nombre	$\hat{Q}_{2.33}$ [m ³ /s]	
		Univariada	Bivariada
10029	Naranjo	566.96	580.97
10031	Guamúchil	568.21	566.75
10034	Zopilote	283.34	293.30
10036	Jaina	789.50	814.76
10037	Huites	3064.36	3444.19
10040	Santa Cruz	779.24	884.17
10057	Bamícori	163.56	163.86
10064	Chínipas	942.08	996.48
10065	Ixpalino	1114.32	1141.09
10066	Choix	291.60	300.29
10077	Palo Dulce	940.77	1020.98
10083	El Quelite	466.42	466.34
10087	Tamazula	595.55	603.67
10100	Urique II	319.88	323.37
10110	Toahayana	1011.71	1001.15
10112	Guatenipa	1523.93	1602.31
10113	La Huerta	891.85	906.76
10137	Tecusiapa	996.34	1065.96

Tabla 28. EEA [m³/s] por la técnica avenida índice para sitios con escasa información RH 10, periodo completo

Estación	EEA [m ³ /s] por la técnica avenida índice			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
10029	321.62	307.42	318.74	299.81
10031	338.23	335.61	335.94	330.10
10034	79.12	69.65	77.83	66.22
10036	627.42	608.45	623.43	597.36
10037	1568.46	1394.41	1582.62	1339.92
10040	644.33	588.94	640.29	574.86
10057	79.07	77.82	78.37	76.22
10064	126.94	169.66	122.59	158.30
10065	492.02	465.29	495.76	474.32
10066	134.50	125.74	135.59	128.44
10077	816.40	772.94	819.37	782.18
10083	136.21	133.48	134.30	129.57
10087	773.09	760.44	775.16	766.00
10100	80.97	90.85	79.39	86.17
10110	132.89	162.15	130.12	145.84
10112	584.46	522.95	577.16	500.95
10113	133.17	164.54	127.96	146.28
10137	411.19	333.24	406.32	347.56

Para estimar gastos en sitios con nula información, se determinó qué atributo de las cuencas en estudio tiene una mejor relación con sus respectivos $\hat{Q}_{2.33}$ (Tabla 27). La función que mejor se ajustó en todos los casos para el análisis univariado (Tabla 29) y multivariado (Tabla 30) fue un polinomio de segundo grado con el atributo del parámetro de forma. Se continuó el procedimiento al calcular con dichas ecuaciones los $\hat{Q}_{2.33}^R$ de cada estación y estos se multiplicaron por sus respectivas medianas regionales para obtener los gastos estimados para los periodos de retorno correspondientes a la longitud de la serie original, obteniendo así, el error estándar de ajuste entre los estimados en sitios con nula información y los datos medidos. En el caso sin una estación a la vez dio un menor error al combinar el análisis multivariado en once de las dieciocho estaciones respecto al univariado, mientras que al tener en cuenta todas las estaciones, es mejor en catorce de ellas.

Tabla 29. Ecuaciones para determinar el gasto de 2.33 años [m³/s] por la distribución Gumbel univariada RH 10, periodo completo

Estación	R ²	Atributo (x)	Ecuación
10029	0.8232	Pf [1/km]	$y = \exp(2.376e-01 \ln(x)^2 - 0.0381 \ln(x) + 4.9166)$
10031	0.8231	Pf [1/km]	$y = \exp(2.375e-01 \ln(x)^2 - 0.0384 \ln(x) + 4.9167)$
10034	0.7964	Pf [1/km]	$y = \exp(2.066e-01 \ln(x)^2 - 0.1623 \ln(x) + 4.8221)$
10036	0.8317	Pf [1/km]	$y = \exp(2.142e-01 \ln(x)^2 - 0.1582 \ln(x) + 4.8045)$
10037	0.7629	Pf [1/km]	$y = \exp(-2.596e-02 \ln(x)^2 - 1.1027 \ln(x) + 3.9510)$
10040	0.8327	Pf [1/km]	$y = \exp(2.141e-01 \ln(x)^2 - 0.1593 \ln(x) + 4.8028)$
10057	0.7664	Pf [1/km]	$y = \exp(4.494e-01 \ln(x)^2 + 1.0931 \ln(x) + 6.4028)$
10064	0.8145	Pf [1/km]	$y = \exp(2.084e-01 \ln(x)^2 - 0.1528 \ln(x) + 4.8315)$
10065	0.8118	Pf [1/km]	$y = \exp(2.070e-01 \ln(x)^2 - 0.1552 \ln(x) + 4.8320)$
10066	0.8063	Pf [1/km]	$y = \exp(1.830e-01 \ln(x)^2 - 0.2441 \ln(x) + 4.7850)$
10077	0.8145	Pf [1/km]	$y = \exp(2.058e-01 \ln(x)^2 - 0.1692 \ln(x) + 4.8122)$
10083	0.8326	Pf [1/km]	$y = \exp(2.280e-01 \ln(x)^2 - 0.1175 \ln(x) + 4.7737)$
10087	0.8176	Pf [1/km]	$y = \exp(1.957e-01 \ln(x)^2 - 0.2126 \ln(x) + 4.7730)$
10100	0.8778	Pf [1/km]	$y = \exp(1.211e-01 \ln(x)^2 - 0.5516 \ln(x) + 4.4511)$
10110	0.8132	Pf [1/km]	$y = \exp(2.067e-01 \ln(x)^2 - 0.1654 \ln(x) + 4.8158)$
10112	0.8236	Pf [1/km]	$y = \exp(2.090e-01 \ln(x)^2 - 0.1222 \ln(x) + 4.8805)$
10113	0.8155	Pf [1/km]	$y = \exp(2.048e-01 \ln(x)^2 - 0.1746 \ln(x) + 4.8064)$
10137	0.8403	Pf [1/km]	$y = \exp(2.608e-01 \ln(x)^2 + 0.0811 \ln(x) + 5.0496)$
Todas	0.8169	Pf [1/km]	$y = \exp(2.066e-01 \ln(x)^2 - 0.1632 \ln(x) + 4.8198)$

*Pf: parámetro de forma en 1/km.

Tabla 30. Ecuaciones para determinar el gasto de 2.33 años [m³/s] por la distribución Gumbel bivariada RH 10, periodo completo

Estación	R ²	Atributo (x)	Ecuación
10029	0.8300	Pf [1/km]	$y = \exp(2.679e-01 \ln(x)^2 + 0.0604 \ln(x) + 5.0038)$
10031	0.8281	Pf [1/km]	$y = \exp(2.640e-01 \ln(x)^2 + 0.0450 \ln(x) + 4.9922)$
10034	0.8052	Pf [1/km]	$y = \exp(2.355e-01 \ln(x)^2 - 0.0743 \ln(x) + 4.8911)$
10036	0.8397	Pf [1/km]	$y = \exp(2.439e-01 \ln(x)^2 - 0.0639 \ln(x) + 4.8871)$
10037	0.7697	Pf [1/km]	$y = \exp(-1.019e-02 \ln(x)^2 - 1.0628 \ln(x) + 3.9848)$
10040	0.8322	Pf [1/km]	$y = \exp(2.417e-01 \ln(x)^2 - 0.0662 \ln(x) + 4.8900)$
10057	0.7788	Pf [1/km]	$y = \exp(5.021e-01 \ln(x)^2 + 1.3087 \ln(x) + 6.6398)$
10064	0.8214	Pf [1/km]	$y = \exp(2.381e-01 \ln(x)^2 - 0.0559 \ln(x) + 4.9185)$
10065	0.8189	Pf [1/km]	$y = \exp(2.360e-01 \ln(x)^2 - 0.0639 \ln(x) + 4.9117)$
10066	0.8129	Pf [1/km]	$y = \exp(2.140e-01 \ln(x)^2 - 0.1438 \ln(x) + 4.8714)$
10077	0.8207	Pf [1/km]	$y = \exp(2.356e-01 \ln(x)^2 - 0.0699 \ln(x) + 4.9027)$
10083	0.8370	Pf [1/km]	$y = \exp(2.566e-01 \ln(x)^2 - 0.0247 \ln(x) + 4.8585)$
10087	0.8246	Pf [1/km]	$y = \exp(2.229e-01 \ln(x)^2 - 0.1273 \ln(x) + 4.8485)$
10100	0.8878	Pf [1/km]	$y = \exp(2.826e-02 \ln(x)^2 + 0.1292 \ln(x) + 3.6776)$
10110	0.8222	Pf [1/km]	$y = \exp(2.360e-01 \ln(x)^2 - 0.0757 \ln(x) + 4.8920)$
10112	0.8299	Pf [1/km]	$y = \exp(2.382e-01 \ln(x)^2 - 0.0284 \ln(x) + 4.9640)$
10113	0.8232	Pf [1/km]	$y = \exp(2.330e-01 \ln(x)^2 - 0.0869 \ln(x) + 4.8827)$
10137	0.8500	Pf [1/km]	$y = \exp(2.947e-01 \ln(x)^2 + 0.1965 \ln(x) + 5.1535)$
Todas	0.8236	Pf [1/km]	$y = \exp(2.357e-01 \ln(x)^2 - 0.0693 \ln(x) + 4.9034)$

*Pf: parámetro de forma en 1/km.

Con las ecuaciones presentadas se estimaron los $\hat{Q}_{2.33}^R$ [m³/s] indicados en la Tabla 31, y se obtuvieron sus respectivos errores estándar de ajuste respecto a los gastos medidos (Tabla 32).

Tabla 31. Gasto de 2.33 años estimados para sitios con nula información RH 10, periodo completo

Estación	$\hat{Q}_{2.33}^R$ [m ³ /s] por la técnica avenida índice - nula información			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
10029	439.61	446.28	452.48	459.84
10031	441.29	449.58	454.10	461.53
10034	284.79	285.49	284.54	286.83
10036	1141.59	1209.06	1103.93	1166.42
10037	1681.45	1825.78	2463.87	2734.76
10040	1139.88	1195.71	1101.27	1163.46
10057	312.73	333.60	189.23	192.27
10064	901.57	941.78	904.53	945.76
10065	1016.95	1072.80	1024.56	1078.21
10066	380.23	383.42	367.70	371.78
10077	979.31	1025.08	976.25	1024.76
10083	306.04	309.28	324.70	327.65
10087	657.90	678.52	652.17	671.59
10100	703.57	806.82	656.75	676.51
10110	1048.89	1111.36	1045.72	1101.68
10112	984.43	1035.79	1019.98	1073.13
10113	946.17	993.79	941.93	986.90
10137	613.43	628.82	640.57	659.15

Tabla 32. EEA [m³/s] por la técnica avenida índice para sitios con nula información RH 10, periodo completo

Estación	EEA [m ³ /s] por la técnica avenida índice			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
10029	430.47	422.04	416.71	405.84
10031	426.13	417.64	414.39	405.37
10034	78.00	75.50	76.87	71.52
10036	588.87	604.36	570.72	563.68
10037	2832.82	2681.04	2052.70	1791.99
10040	591.47	599.30	573.99	563.60
10057	148.85	169.55	64.56	59.81
10064	129.21	154.60	123.44	138.32
10065	541.46	498.08	539.48	501.34
10066	136.86	127.14	133.32	128.35
10077	802.36	771.65	806.66	781.11
10083	297.54	293.82	276.45	273.62
10087	756.38	739.42	760.07	748.06
10100	501.64	613.28	446.90	466.09
10110	127.89	159.73	125.76	148.60
10112	1097.89	1040.09	1057.85	997.70
10113	148.01	195.58	142.05	181.23
10137	768.84	750.85	738.59	716.79

En la Tabla 33 se resumen los errores estándar de ajuste de la Tabla 28 y Tabla 32, y solo se muestran los casos donde el análisis multivariado otorgó una disminución del error estándar de ajuste respecto al análisis univariado de la técnica regional avenida índice.

Tabla 33. Casos con menor EEA [m^3/s] por la técnica avenida índice combinando el análisis multivariado RH 10, periodo completo

Estación	Casos con menor EEA [m^3/s] por la técnica avenida índice combinando el análisis multivariado			
	Sin la estación en mención		Todas las estaciones	
	Información:	Escasa	Nula	Escasa
10029	Si	Si	Si	Si
10031	Si	Si	Si	Si
10034	Si	Si	Si	Si
10036	Si		Si	Si
10037	Si	Si	Si	Si
10040	Si		Si	Si
10057	Si		Si	Si
10064				
10065	Si	Si	Si	Si
10066	Si	Si	Si	Si
10077	Si	Si	Si	Si
10083	Si	Si	Si	Si
10087	Si	Si	Si	Si
10100				
10110				
10112	Si	Si	Si	Si
10113				
10137	Si	Si	Si	Si

Estos resultados muestran que el aplicar la técnica regional avenida índice combinada con el análisis multivariado junto con toda la extensión de los registros y no un periodo común, logra reducir el error estándar de ajuste que al realizarlo univariado. Al comparar los EEA obtenidos puntualmente en las estaciones con los obtenidos regionalmente, se puede observar que se logra reducir el error en más casos cuando se aplica en análisis multivariado teniendo en cuenta todas las estaciones como si tuviera escasa información, siendo menor en diez de los dieciocho casos. A pesar de la reducción de la incertidumbre al aplicar la técnica avenida índice con las extensiones completas de los registros, es importante resaltar que varias de las estaciones estimaron datos negativos al ajustar la distribución Gumbel puntualmente, lo cual no refleja el comportamiento de la variable aleatoria, ya que una serie medida no presentará gastos negativos, y por tanto, la técnica

no debería ser utilizada a pesar de la disminución del error, porque el principio es que la distribución de probabilidades refleje el comportamiento de la variable física, la cual no permite valores menores que cero, y no se cumple el concepto de la parte real a la abstracción matemática.

Adicionalmente, se probó el ajustar únicamente el área de las cuencas para obtener los gastos $\hat{Q}_{2.33}^R$ [m³/s] y aplicar el análisis en sitios con nula información, ya que es una práctica muy usual; de lo cual se obtuvo que los coeficientes de determinación R^2 son menores que los obtenidos por el parámetro de forma para el caso en estudio. En la Tabla 34 y la Tabla 35, se muestran las ecuaciones y sus respectivas funciones que mejor se ajustaron al área de drenaje, y los resultados del error estándar de ajuste se presentan en la Tabla 36, que nuevamente otorgaron un mejor resultado al combinar la técnica regional con el análisis multivariado, pero dicho error aumenta en más de la mitad de las estaciones en comparación con el resultado obtenido al estimar el $\hat{Q}_{2.33}^R$ con el parámetro de forma; lo cual demuestra la importancia de obtener el atributo que mejor se ajusta al $\hat{Q}_{2.33}$, y no limitarse a utilizar el área, así como ajustar diferentes tipo de funciones, ya que se observa que el polinomio de segundo grado no fue el que mejor se ajustó en todos los casos, teniendo como opción las funciones exponenciales y potenciales; de haber forzado solamente el uso del polinomio, el error sería aún mayor.

Tabla 34. Ecuaciones para determinar el gasto de 2.33 años [m³/s] por la distribución Gumbel univariada RH 10, periodo completo con el área de las cuencas

Estaciones	R ²	Función	Ecuación
10029	0.7948	Polinomial grado 2	$y = \exp(4.407e-02 \ln(x)^2 - 0.1421 \ln(x) + 4.7232)$
10031	0.8009	Polinomial grado 2	$y = \exp(4.718e-02 \ln(x)^2 - 0.1847 \ln(x) + 4.8528)$
10034	0.7708	Polinomial grado 2	$y = \exp(4.053e-02 \ln(x)^2 - 0.0924 \ln(x) + 4.5662)$
10036	0.8101	Polinomial grado 2	$y = \exp(4.227e-02 \ln(x)^2 - 0.1009 \ln(x) + 4.5387)$
10037	0.7303	Polinomial grado 2	$y = \exp(-1.872e-02 \ln(x)^2 + 0.7694 \ln(x) + 1.5077)$
10040	0.8178	Polinomial grado 2	$y = \exp(4.488e-02 \ln(x)^2 - 0.1361 \ln(x) + 4.6540)$
10057	0.7335	Polinomial grado 2	$y = \exp(9.229e-02 \ln(x)^2 - 0.9662 \ln(x) + 8.2154)$
10064	0.7927	Polinomial grado 2	$y = \exp(4.229e-02 \ln(x)^2 - 0.1200 \ln(x) + 4.6631)$
10065	0.7909	Polinomial grado 2	$y = \exp(4.187e-02 \ln(x)^2 - 0.1162 \ln(x) + 4.6573)$
10066	0.7908	Exponencial	$y = \exp(3.3568 \wedge [0.0818 \ln(x)])$
10077	0.7909	Polinomial grado 2	$y = \exp(4.025e-02 \ln(x)^2 - 0.0852 \ln(x) + 4.5260)$
10083	0.8100	Polinomial grado 2	$y = \exp(4.280e-02 \ln(x)^2 - 0.0982 \ln(x) + 4.4356)$
10087	0.7954	Polinomial grado 2	$y = \exp(4.426e-02 \ln(x)^2 - 0.1456 \ln(x) + 4.7377)$
10100	0.8748	Exponencial	$y = \exp(3.2655 \wedge [0.0858 \ln(x)])$
10110	0.7932	Polinomial grado 2	$y = \exp(4.290e-02 \ln(x)^2 - 0.1308 \ln(x) + 4.7070)$
10112	0.7880	Polinomial grado 2	$y = \exp(3.858e-02 \ln(x)^2 - 0.0738 \ln(x) + 4.5256)$
10113	0.7921	Polinomial grado 2	$y = \exp(3.990e-02 \ln(x)^2 - 0.0788 \ln(x) + 4.4991)$
10137	0.8067	Polinomial grado 2	$y = \exp(4.918e-02 \ln(x)^2 - 0.2267 \ln(x) + 5.0531)$
Todas	0.7937	Polinomial grado 2	$y = \exp(4.036e-02 \ln(x)^2 - 0.0875 \ln(x) + 4.5359)$

Tabla 35. Ecuaciones para determinar el gasto de 2.33 años [m³/s] por la distribución Gumbel bivariada RH 10, periodo completo con el área de las cuencas

Estaciones	R ²	Función	Ecuación
10029	0.8083	Polinomial grado 2	$y = \exp(5.212e-02 \ln(x)^2 - 0.2462 \ln(x) + 5.0649)$
10031	0.8125	Polinomial grado 2	$y = \exp(5.456e-02 \ln(x)^2 - 0.2791 \ln(x) + 5.1625)$
10034	0.7868	Polinomial grado 2	$y = \exp(4.819e-02 \ln(x)^2 - 0.188 \ln(x) + 4.86450)$
10036	0.8250	Polinomial grado 2	$y = \exp(5.025e-02 \ln(x)^2 - 0.2028 \ln(x) + 4.8697)$
10037	0.7438	Potencial	$y = \exp(1.855 \ln(x)^{0.602})$
10040	0.8224	Polinomial grado 2	$y = \exp(5.202e-02 \ln(x)^2 - 0.2295 \ln(x) + 4.9667)$
10057	0.7545	Polinomial grado 2	$y = \exp(1.047e-01 \ln(x)^2 - 1.1436 \ln(x) + 8.8669)$
10064	0.8065	Polinomial grado 2	$y = \exp(5.038e-02 \ln(x)^2 - 0.2252 \ln(x) + 5.0107)$
10065	0.8042	Polinomial grado 2	$y = \exp(4.950e-02 \ln(x)^2 - 0.2131 \ln(x) + 4.9714)$
10066	0.8034	Polinomial grado 2	$y = \exp(3.994e-02 \ln(x)^2 - 0.0747 \ln(x) + 4.5179)$
10077	0.8038	Polinomial grado 2	$y = \exp(4.834e-02 \ln(x)^2 - 0.1913 \ln(x) + 4.8791)$
10083	0.8212	Polinomial grado 2	$y = \exp(5.059e-02 \ln(x)^2 - 0.1989 \ln(x) + 4.7684)$
10087	0.8084	Polinomial grado 2	$y = \exp(5.193e-02 \ln(x)^2 - 0.2439 \ln(x) + 5.0599)$
10100	0.8883	Exponencial	$y = \exp(3.2133 \ln(x)^{0.0884})$
10110	0.8058	Polinomial grado 2	$y = \exp(4.994e-02 \ln(x)^2 - 0.2182 \ln(x) + 4.9845)$
10112	0.8016	Polinomial grado 2	$y = \exp(4.647e-02 \ln(x)^2 - 0.1751 \ln(x) + 4.8567)$
10113	0.8064	Polinomial grado 2	$y = \exp(4.746e-02 \ln(x)^2 - 0.1744 \ln(x) + 4.8079)$
10137	0.8223	Polinomial grado 2	$y = \exp(5.792e-02 \ln(x)^2 - 0.3419 \ln(x) + 5.4370)$
Todas	0.8071	Polinomial grado 2	$y = \exp(4.820e-02 \ln(x)^2 - 0.1884 \ln(x) + 4.8668)$

Tabla 36. EEA [m³/s] por la técnica avenida índice para sitios con nula información RH 10, periodo completo utilizando el área para estimar el gasto de 2.33 años [m³/s]

Estación	EEA [m ³ /s] por la técnica avenida índice			
	Sin la estación en mención		Todas las estaciones	
	Univariada	Multivariada	Univariada	Multivariada
10029	378.78	368.65	370.86	357.53
10031	432.61	425.39	420.73	412.94
10034	72.46	70.14	72.06	66.68
10036	597.28	618.92	575.59	572.77
10037	2890.06	2715.26	2033.39	1747.16
10040	636.73	664.50	600.01	606.69
10057	130.60	145.57	64.57	59.91
10064	172.21	178.53	161.12	154.23
10065	604.24	563.48	596.00	557.66
10066	413.86	161.73	159.06	155.15
10077	810.51	779.78	814.04	787.85
10083	301.56	298.77	279.70	277.73
10087	803.59	791.57	801.95	792.43
10100	349.15	349.04	529.05	553.07
10110	271.42	254.45	256.07	230.82
10112	979.94	906.80	940.81	864.91
10113	146.19	193.61	140.36	179.16
10137	697.99	673.92	674.00	647.58

La Tabla 37 especifica los casos del análisis multivariado para sitios con nula información donde el error estándar de ajuste fue inferior al estimar el $\hat{Q}_{2.33}^R$ por el parámetro de forma y no por el área.

Tabla 37. Casos con menor EEA [m³/s] estimando el gasto de 2.33 años [m³/s] por el parámetro de forma y no por el área por análisis bivariado RH 10, periodo completo

Estación	Casos con menor EEA estimando el $\hat{Q}_{2.33}^R$ por el parámetro de forma y no por el área por análisis multivariado	
	Sin la estación en mención	Todas las estaciones
10029		
10031	Si	Si
10034		
10036	Si	Si
10037	Si	
10040	Si	Si
10057		Si
10064	Si	Si
10065	Si	Si
10066	Si	Si
10077	Si	Si
10083	Si	Si
10087	Si	Si
10100		Si
10110	Si	Si
10112		
10113		
10137		

5. Conclusiones

Es muy importante la correcta ubicación de la estación hidrométrica a estudiar, ya que a partir de ella se delimita su respectiva cuenca y se calculan sus características fisiográficas, las cuales son necesarias para la delimitación de regiones homogéneas, en la técnica de los trazos multidimensionales y la distancia euclidiana, donde se necesitan las características más significativas de cada estación. Además, también son necesarias para el análisis regional hidrológico, donde se necesitan para estimar los datos de una estación no aforada. Por lo cual, es necesario reducir la incertidumbre de las características fisiográficas de la cuenca, debido a que, podría afectar considerablemente los resultados de las técnicas mencionadas.

En el caso de la detección de datos atípicos, se aplicaron tres pruebas: la del rango intercuartil, la prueba de Grubbs y tres veces la desviación estándar. Se pudo observar en el desarrollo del proyecto que la prueba del rango intercuartil es más sensible a identificar datos atípicos, dado que en la mayoría de los casos resaltó datos que las otras dos pruebas no. Es importante mencionar que, para eliminar un dato de un registro, no solo debió ser identificado por al menos dos de las pruebas mencionadas, sino que se verificó que no se tratara de un evento que, si ocurrió, al compararlo con las estaciones vecinas de la región homogénea.

Es fundamental contar con varias herramientas para la toma de decisiones en el análisis de la información, como lo es el definir si una serie de datos corresponde a una variable aleatoria. En el caso de la región homogénea de la RH 23 (Tabla 6), de haber aplicado únicamente la prueba de independencia de Anderson se concluiría que las estaciones 23017, 23019 y 23020 son dependientes y no se hubieran tenido en cuenta en el análisis, pero las pruebas de Mantilla y Amigó muestran que, si son independientes, y por tanto, se concluye que son no dependientes. Es el mismo caso al querer definir si una serie de datos es homogénea o no, contar con más pruebas nos brinda mayor seguridad en el momento de concluir si una serie es independiente o dependiente, homogénea o no.

El aplicar la técnica regional avenida índice tiene dos limitantes, primero que se debe realizar en un periodo común entre las estaciones de la región homogénea, y el segundo es que sus registros deben ajustarse a la distribución Gumbel, esto último limitó considerablemente el número de estaciones utilizadas en el análisis. Se recomienda verificar los gastos estimados al ajustar la distribución Gumbel a la serie original, especialmente para aquellos periodos de retorno inferiores a dos años, los cuales pueden estimar gastos negativos y lo que generaría no tener en cuenta la

serie en la técnica regional. Por esta razón, dentro de los periodos de retorno de diseño, se incluyó el de 1.1 años.

Al aplicar la técnica regional avenida índice combinando el análisis multivariado en la región homogénea de la región hidrológica número 10, logra un menor error estándar de ajuste en diez de los doce escenarios estudiados. Por el contrario, en la región homogénea de la RH 26, solo se logra una reducción en cuatro de los doce escenarios. Y en la región homogénea de la RH 23, nueve de los veinticuatro escenarios presentan un menor error estándar de ajuste con la combinación del análisis multivariado.

Respecto a la inconsistencia en los resultados puede ser debido a que las series en estudio no presentan el mejor ajuste con la distribución Gumbel, la cual es la requerida por la técnica regional avenida índice, además, en las regiones costeras estudiadas en el presente proyecto las distribuciones que mejor se ajustan a las series son las mezcladas, ya que tienen en cuenta la componente ciclónica, y la técnica regional utiliza una distribución unimodal siendo su comportamiento bimodal.

La posible explicación a los resultados obtenidos puede ser el utilizar únicamente la información del periodo común como solicita la técnica, motivo por el cual se repitió el análisis en la región homogénea de la RH 10 con todas las estaciones y sus periodos completos. Los resultados muestran que el aplicar la técnica regional avenida índice combinada con el análisis multivariado junto con toda la extensión de los registros y no un periodo común, logra reducir el error estándar de ajuste que al realizarlo univariado. Al comparar los EEA obtenidos puntualmente en las estaciones con los obtenidos regionalmente, se puede observar que se logra reducir el error en más casos cuando se aplica en análisis multivariado teniendo en cuenta todas las estaciones como si tuviera escasa información, siendo menor en diez de los dieciocho casos. A pesar de la reducción de la incertidumbre al aplicar la técnica avenida índice con las extensiones completas de los registros, es importante resaltar que varias de las estaciones estimaron datos negativos al ajustar la distribución Gumbel puntualmente, lo cual no refleja el comportamiento de la variable aleatoria, ya que una serie medida no presentará gastos negativos, y por tanto, la técnica no debería ser utilizada a pesar de la disminución del error, porque el principio es que la distribución de probabilidades refleje el comportamiento de la variable física, la cual no permite valores menores que cero, y no se cumple el concepto de la parte real a la abstracción matemática.

Se puede observar en la gran mayoría de los casos que el estimar los gastos máximos instantáneos regionales asumiendo las estaciones como si tuvieran nula información aumentó el error estándar de ajuste, esto debido a la incertidumbre que se genera al estimar el gasto $\hat{Q}_{2.33}^R$ a partir de un atributo de la cuenca. Por este motivo, se adicionó el análisis utilizando nada más el área para ajustar la curva a los $\hat{Q}_{2.33}$, ya que es una práctica muy usual. El error estándar de ajuste aumenta en más de la mitad de las estaciones en comparación con el resultado obtenido al estimar el $\hat{Q}_{2.33}^R$ con el parámetro de forma, para el caso del presente proyecto; lo cual demuestra la importancia de obtener el atributo que mejor se ajusta al $\hat{Q}_{2.33}$, y no limitarse a utilizar el área, así como ajustar diferentes tipos de funciones, ya que se observa que el polinomio de segundo grado no fue el que mejor se ajustó en todos los casos, teniendo como opción las funciones exponenciales y potenciales. De haber forzado solamente el uso del polinomio, el error sería aún mayor.

6. Bibliografía

- Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of Climatology*, 6(6), 661-675.
- Amigó, J., Zambrano, S., & Sanjuán, M. (2010). Detecting determinism in time series with ordinal patterns: a comparative study. *International Journal of Bifurcation and Chaos*, 20(9), 2915-2924.
- Andrews, D. F. (1972). Plots of high dimensional data. *Biometrics*, 28(1), 125-136.
- Bingham, C., & Nelson, L. (1981). An approximation for the distribution of the Von Neumann ratio. *Technometrics*, 23(3), 285-288.
- Buishand, T. (1981). The analysis of homogeneity of long-term rainfall records in the Netherlands. Scientific report / Royal Nederlands Meteorological Institute.
- Buishand, T. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, 58, 11-27.
- Campos Aranda, D. (1994). Aplicación del método del índice de crecientes en la región hidrológica número 10, Sinaloa. *Ingeniería hidráulica en México*, IX(3), 41-55.
- Chambers, J., Cleveland, W., Kleiner, B., & Tukey, P. (1983). Graphical methods for data analysis. *Wadsworth*.
- Dalrymple, T. (1960). Flood frequency analysis. *USGS Water Supply Paper*, 1543(A).
- Escalante, C. (1998). Multivariate estimation of extreme flood hydrographs. *Hydrological Science and Technology Journal*, 14(1-4), 19-27.
- Escalante, C., & Reyes, L. (2008). *Técnicas estadísticas en hidrología*. Ciudad de México: UNAM Facultad de ingeniería.
- Greenwood, J., Landwehr, J., & Wallis, J. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5), 1049-1054.
- Grubbs, F. (1969). Procedures for detecting outlying observations in samples. *Technometric*, 11(1), 1-21.
- Gumbel, E. J. (1962). Statistical theory of extreme values (Main results). (Sarhan, & Greenberg, Edits.) 59-93.

- Hosking, J. (1990). L-Moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B*, 52(1), 105-124.
- Kite, G. (1988). Frequency and risk analyses in hydrology. *Water Resources Publications*, 257.
- Matilla García, M., & Ruiz Marín, M. (2008). A non-parametric independence test using permutation entropy. *Journal of Econometrics*, 144(1), 139-155.
- Nathan, R. J., & McMahon, T. A. (1990). Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology*, 121, 217-238.
- NIST/SEMATECH. (s.f.). *e-Handbook of Statistical Methods*. Recuperado el 17 de Diciembre de 2020, de <http://www.itl.nist.gov/div898/handbook/>
- Pettitt, A. (1979). A non-parametric approach to the change-point detection. *Applied Statistics*, 28, 126-135.
- Raynal, J. A. (1985). Bivariate extreme value distributions applied to flood frequency analysis Ph. D. dissertation. *Civil Engineering Department*, 237.
- Salas, J., Delleur, J., & Yevjevich, V. (1988). *Applied modeling of hydrologic time series*. USA: Water Resources Publications.
- Stefansky, W. (1972). Rejecting outliers in factorial designs. *Technometrics*, 14(2), 469-479.
- Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics*, 12(4), 367-395.
- Wijngaard, J. B., Klein Tank, A. M., & Können, G. P. (2003). Homogeneity of 20th century european daily temperatureand precipitation series. *International journal of climatology*, 23(6), 679-692.