

UNIVERSIDAD DE COSTA RICA  
SISTEMA DE ESTUDIOS DE POSGRADO

COMPARACIÓN DE LA PROBABILIDAD DE NO EXCEDENCIA DEL  
CAUDAL PICO Y VOLUMEN DE CRECIENTES OBTENIDA  
MEDIANTE UN ANÁLISIS DE FRECUENCIA BIDIMENSIONAL CON  
LA METODOLOGÍA DE CÓPULAS Y UN ANÁLISIS DE  
FRECUENCIA UNIDIMENSIONAL

Tesis sometida a la consideración de la Comisión del Programa de  
Estudios de Posgrado Ingeniería Civil para optar al grado y título de  
Maestría Académica en Ingeniería Hidráulica

DAVID JIMÉNEZ GONZÁLEZ

Ciudad Universitaria Rodrigo Facio, Costa Rica

2022

# **Agradecimientos**

A mi familia de sangre y de la vida, gracias por existir, sin ustedes sería nada.

Esta Tesis fue aceptada por la Comisión del Programa de Estudios de Postgrado en Ingeniería Civil de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Ingeniería Hidráulica

---

PhD. Rafael Murillo Muñoz

**Representante de la Decana del Sistema de Estudios de Postgrado**

---

PhD. Alberto Serrano Pacheco

**Profesor Guía**

---

MSc. Juan José Leitón Montero

**Lector**

---

PhD. Luis Alberto Barboza Chinchilla

**Lector**

---

PhD. Hugo Hidalgo León

**Representante del Director Programa de Postgrado en Ingeniería Civil**

---

David Jiménez González

**Sustentante**

# Índice general

Agradecimientos . . . . .	ii
Hoja de Aprobación . . . . .	iii
Índice General . . . . .	iv
Resumen . . . . .	vii
Índice de Figuras . . . . .	viii
Índice de Cuadros . . . . .	xi
Citas . . . . .	xii
<b>1. Introducción</b>	<b>1</b>
1.1. Justificación . . . . .	1
1.2. Antecedentes . . . . .	3
1.3. Objetivos . . . . .	4
1.3.1. Objetivo General . . . . .	4
1.3.2. Objetivos específicos . . . . .	4
1.4. Alcances . . . . .	5
1.5. Limitaciones . . . . .	6
<b>2. Marco teórico</b>	<b>7</b>
2.1. Análisis de frecuencia de eventos extremos en hidrología . . . . .	7
2.1.1. Estimación de los parámetros de la distribución . . . . .	10
2.2. Probabilidad conjunta y cópulas . . . . .	11
2.3. Dependencia . . . . .	12



2.3.1. Cópula empírica y pseudo-observaciones . . . . .	14
2.3.2. Dependencia en colas . . . . .	14
2.4. Familias de cópulas . . . . .	15
2.5. Pruebas de características de la cópula . . . . .	16
2.5.1. Independencia . . . . .	16
2.5.2. Intercambiabilidad . . . . .	17
2.5.3. Simetría Radial . . . . .	17
2.5.4. Valor extremo . . . . .	18
2.6. Pruebas de bondad de ajuste y calidad de modelo . . . . .	19
2.7. Medida de calidad de modelo . . . . .	19
2.8. Período de retorno en el caso de análisis multivariado . . . . .	20
2.9. Medición de caudal . . . . .	21
2.10. Separación de flujo base . . . . .	22
2.11. Definición de volumen y caudal pico de una creciente . . . . .	25
2.12. Análisis de sensibilidad . . . . .	25
<b>3. Metodología</b>	<b>27</b>
3.1. Herramientas . . . . .	28
<b>4. Desarrollo</b>	<b>30</b>
4.1. Revisión general de la información . . . . .	30
4.2. Obtención de eventos . . . . .	32
4.2.1. Método de refinamiento de separación de eventos . . . . .	36
4.3. Análisis univariado de series de valores máximos . . . . .	38
4.4. Análisis multivariado de series de valores máximos . . . . .	42
4.4.1. Análisis exploratorio . . . . .	42
4.4.2. Estimación de parámetros . . . . .	51
4.4.3. Comprobación visual . . . . .	53

4.4.4. Estimación de probabilidad de excedencia y comparación con análisis univariado . . . . .	57
4.5. Análisis de sensibilidad . . . . .	62
4.6. Discusión general de resultados . . . . .	68
<b>5. Cierre</b>	<b>71</b>
5.1. Conclusiones . . . . .	71
5.2. Recomendaciones . . . . .	72
<b>6. Referencias bibliográficas</b>	<b>74</b>
<b>A. Crecientes máximas empleadas en los análisis</b>	<b>83</b>

# Resumen

**[ES]** Las crecientes pueden estar caracterizadas por más de una variable, e.g. caudal pico, volumen, duración, etc. En este estudio se emplea la metodología de cópulas para calcular probabilidades de no excedencia y definir la estructura de dependencia entre el caudal pico y el volumen de eventos extremos en tres estaciones de medición de caudal ubicadas en la cuenca del río General. Se muestra cómo la dependencia entre dichas características de crecienta pueden variar según los parámetros que se empleen en el método de segmentación de eventos. Además, se indica la diferencia que puede existir entre la estimación de la probabilidad de excedencia en una variable y en múltiples variables de probabilidad de excedencia.

**[EN]** Floods can be characterized by more than one variable, e.g., peak discharge, volume, duration, etc. In this study copula models are used to estimate the non-exceedance probabilities and the dependence structure associated with the peak discharge and volume of extreme events in three discharge gauging stations in the General River. It is shown how the dependence can vary depending on the parameters used in the event segmentation procedure. Furthermore, it is demonstrated that significant differences in the exceedance probability can be observed between univariate and multivariate methodologies estimations.

# Índice de figuras

1.1. Estaciones propuestas para ser analizadas. . . . .	5
2.1. Distribuciones de probabilidad (a) y densidades (b) de las familias Fréchet (punteada), Gumbel (continua) y Weibull (guiones) pertenecientes a la GEV. McNeil , <i>et al</i> (2005). . . . .	9
2.2. Representación de la cópulas $W(u, v)$ , $\prod(u, v) = uv$ y $M(u, v)$ . Fuente:McNeil , <i>et al</i> (2005) . . . . .	13
2.3. Separación de flujo mediante métodos gráficos. Fuente:Chow , <i>et al</i> (1996) . . . . .	23
4.1. Registro completo para estaciones 3103, 3104 y 3105 . . . . .	31
4.2. Porcentaje de vacíos por mes en estaciones 3103, 3104 y 3105 . . . . .	31
4.3. Muestra del registro para verificación visual del modelo para separación de flujo. . . . .	33
4.4. Resultado de separación para distintos parámetros $\alpha$ . . . . .	34
4.5. Resultado de separación para distintos parámetros $\alpha$ en evento de mayor magnitud del 2021 . . . . .	35
4.6. Separación de tormentas para parámetros $u_e = 0.15$ , $u_b = 0.15$ y $w = 72$ . . . . .	38
4.7. Series de valores máximos anuales . . . . .	39
4.8. Distribución GEV ajustada a máximos anuales . . . . .	41
4.9. Distribución de máximos anuales . . . . .	42
4.10.Relación de pares $(Q, V)$ . . . . .	43
4.11.Relación de pares $(Q, V)$ . . . . .	45

4.12. Comparación entre valores de $\widehat{\theta}_{MPL}$ . . . . .	52
4.13. Comparación directa entre cópula estimada y empírica. . . . .	54
4.14. Comparación de los contornos entre cópula empírica y la modelada. . . . .	56
4.15. Comparación de la función de Pickand de la cópula empírica y la modelada. . . . .	57
4.16. Comparación de $1 - U_V$ y $1 - U_Q$ con $P_{Q,V}^V$ . . . . .	59
4.17. Comparación de $1 - U_V$ y $1 - U_Q$ con $P_{Q,V}^A$ . . . . .	60
4.18. Comparación de $1/(1 - U_V)$ y $1/(1 - U_Q)$ con $1/P_{Q,V}^V$ . . . . .	61
4.19. Comparación de $1/(1 - U_V)$ y $1/(1 - U_Q)$ con $1/P_{Q,V}^A$ . . . . .	62
4.20. Impacto en series de máximos asociado a variación en parámetros de separación de flujo y segmentación se crecientes . . . . .	65
4.21. Impacto en series de máximos asociado a variación en parámetros de separación de flujo y segmentación se crecientes empleando modelos lineales. Flujo Base. . . . .	66
4.22. Impacto en estructura de dependencia de la variación en parámetros de separación de flujo y segmentación se crecientes . . . . .	67
4.23. Impacto en estructura de dependencia de la variación en parámetros de separación de flujo y segmentación se crecientes empleando modelos lineales . . . . .	67
A.1. Crecientes estación 3103 $\{V_b, \max\{Q_b\}\}$ . . . . .	84
A.2. Crecientes estación 3103 $\{V_e, \max\{Q_e\}\}$ . . . . .	85
A.3. Crecientes estación 3103 $\{\max\{V_b\}, Q_b\}$ . . . . .	86
A.4. Crecientes estación 3103 $\{\max\{V_e\}, Q_e\}$ . . . . .	87
A.5. Crecientes estación 3104 $\{V_b, \max\{Q_b\}\}$ . . . . .	88
A.6. Crecientes estación 3104 $\{V_e, \max\{Q_e\}\}$ . . . . .	89
A.7. Crecientes estación 3104 $\{\max\{V_b\}, Q_b\}$ . . . . .	90
A.8. Crecientes estación 3104 $\{\max\{V_e\}, Q_e\}$ . . . . .	91
A.9. Crecientes estación 3105 $\{V_b, \max\{Q_b\}\}$ . . . . .	92
A.10. Crecientes estación 3105 $\{V_e, \max\{Q_e\}\}$ . . . . .	93

A.11.Crecientes estación 3105 $\{\max\{V_b\}, Q_b\}$ . . . . .	94
A.12.Crecientes estación 3105 $\{\max\{V_e\}, Q_e\}$ . . . . .	95

# Índice de cuadros

3.1. Paquetes de R a ser empleados en el análisis. . . . .	29
4.1. Medidas de dependencia para cada serie y estación . . . . .	46
4.2. Pruebas estadísticas para cada cópula. . . . .	48
4.3. Pruebas de bondad de ajuste para familias principales de cópulas. . .	49
4.4. Prueba de calidad de modelo $n\widehat{xv}_n$ para familias principales de cópulas.	51
4.5. Estimación de cópula Gumbel. . . . .	53

*“Knowing what we don’t know is better than thinking we know what we don’t.”*

— Philip Tetlock, *Superforecasting: The Art and Science of Prediction*





UNIVERSIDAD DE  
COSTA RICA

SEP Sistema de  
Estudios de Posgrado

**Autorización para digitalización y comunicación pública de Trabajos Finales de Graduación del Sistema de Estudios de Posgrado en el Repositorio Institucional de la Universidad de Costa Rica.**

Yo, David Jiménez González, con cédula de identidad 800810489, en mi condición de autor del TFG titulado Comparación de la probabilidad de no excedencia del caudal pico y volumen de crecientes obtenida mediante un análisis de frecuencia bidimensional con la metodología de cópulas y un análisis de frecuencia unidimensional.

Autorizo a la Universidad de Costa Rica para digitalizar y hacer divulgación pública de forma gratuita de dicho TFG a través del Repositorio Institucional u otro medio electrónico, para ser puesto a disposición del público según lo que establezca el Sistema de Estudios de Posgrado. SI  NO \*

\*En caso de la negativa favor indicar el tiempo de restricción: \_\_\_\_\_ año (s).

Este Trabajo Final de Graduación será publicado en formato PDF, o en el formato que en el momento se establezca, de tal forma que el acceso al mismo sea libre, con el fin de permitir la consulta e impresión, pero no su modificación.

Manifiesto que mi Trabajo Final de Graduación fue debidamente subido al sistema digital Kerwá y su contenido corresponde al documento original que sirvió para la obtención de mi título, y que su información no infringe ni violenta ningún derecho a terceros. El TFG además cuenta con el visto bueno de mi Director (a) de Tesis o Tutor (a) y cumplió con lo establecido en la revisión del Formato por parte del Sistema de Estudios de Posgrado.

**FIRMA ESTUDIANTE**

Nota: El presente documento constituye una declaración jurada, cuyos alcances aseguran a la Universidad, que su contenido sea tomado como cierto. Su importancia radica en que permite abreviar procedimientos administrativos, y al mismo tiempo genera una responsabilidad legal para que quien declare contrario a la verdad de lo que manifiesta, puede como consecuencia, enfrentar un proceso penal por delito de perjurio, tipificado en el artículo 318 de nuestro Código Penal. Lo anterior implica que el estudiante se vea forzado a realizar su mayor esfuerzo para que no sólo incluya información veraz en la Licencia de Publicación, sino que también realice diligentemente la gestión de subir el documento correcto en la plataforma digital Kerwá.

# Capítulo 1

## Introducción

### 1.1. Justificación

En el estado de la práctica de la ingeniería civil es común el uso de eventos hidrometeorológicos extremos como insumo para el análisis hidráulico en el diseño de obras de paso, dimensionamiento de sistemas de drenaje, definición de zonas de inundación, prevención de desastres naturales y otros. Como se indica en Mays (1999) y Chow, *et al* (1988), se requiere para la gran mayoría de aplicaciones que los eventos empleados en el modelado hidráulico estén asociados a una probabilidad de no excedencia que permita explícita o tácitamente tomar en consideración la naturaleza estocástica de dichos fenómenos. Por ejemplo, en Costa Rica parte de la normativa empleada para el diseño de obras hidráulicas establece el período de retorno como un parámetro base (AyA (2017), MOPT (s.f.), Municipalidad de Heredia (2015)).

Para obtener este período de retorno o, más precisamente, la probabilidad de no excedencia, habitualmente se emplea inferencia estadística para obtener los parámetros de alguna distribución de probabilidad que caracterice alguna dimensión relevante de las crecientes, por ejemplo el caudal pico o (infrecuentemente) el volumen.

En el análisis de eventos extremos aplicado al diseño hidráulico, rara vez se evalúa la relación que existe entre las diferentes dimensiones que caracterizan el fenómeno como el caudal pico, el volumen y la duración. Para la mayoría de obras hidráulicas, la variable más significativa en el análisis es el caudal pico, ya que el concepto de riesgo en estas se fundamenta en el balance entre la capacidad hidráulica de la obra y el caudal que deben transitar. La ubicuidad del uso de estos métodos

puede estar asociada a la facilidad de emplear técnicas de estimación hidrológica que solamente proporcionan el caudal pico, y también a que para la mayoría de los análisis hidráulicos se emplean métodos que sólo requieren o emplean el caudal pico y no todo el hidrograma para determinar las dimensiones de la obra.

Actualmente, hay evidencia de que esta tendencia está cambiando. Un ejemplo de ello es el uso, cada vez más común, de métodos de análisis hidráulico en dos dimensiones. Estos métodos pueden considerar apropiadamente los efectos de almacenamiento en el cauce en zonas donde el flujo no sigue patrones unidimensionales, tales como en las llanuras de inundación seguidas de contracciones o expansiones. En este escenario, la utilización solamente de caudales pico podría generar resultados alejados de la realidad, pues no tomarían en consideración la capacidad de amortiguamiento que tiene el sistema fluvial. Se hace necesario entonces la incorporación del volumen y la duración como variables importantes dentro del análisis.

El uso del volumen es de particular interés en el caso de las presas, ya que, en algunas el tamaño del embalse relativo a la cuenca es suficientemente grande como para amortiguar parte de la creciente y con ello el caudal necesario para ser manejado por las estructuras hidráulicas disminuye. En estos casos es común tomar una de las variables de interés (usualmente caudal) para realizar el análisis de frecuencias y escoger el hidrograma que generó dicho valor extremo, o uno similar, para obtener las restantes variables. Como se indica en Brunner , *et al* (2016) el problema con esta práctica es que no necesariamente se está asignando una probabilidad de excedencia correspondiente a todo el fenómeno; por ejemplo, se podría estar estimando adecuadamente qué tan inusual o común es el caudal escogido, pero no necesariamente qué tan inusual o común es el hidrograma en términos de las demás variables como el volumen o la duración.

En los últimos 20 años, para suplir la necesidad de tomar en consideración la multidimensionalidad de los fenómenos extremos, como se explica en la sección 1.2, se ha adoptado en diversos ámbitos de la hidrología el uso de la teoría de cópulas . Las cópulas permiten obtener la probabilidad conjunta de las variables de interés de los fenómenos extremos y caracterizar la dependencia que existe entre ellas. Por ejemplo, dicha estructura permite determinar qué tan inusuales son las crecientes tomando en consideración tanto el caudal pico como el volumen; y no solamente uno de ellos.

A pesar de la importancia de este tipo de análisis multidimensional de crecientes y de la existencia de una metodología específica para su exploración, no se encontraron referencias de trabajos de investigación realizados en Costa Rica enfocados en su estudio. Tampoco se encontraron referencias a estudios empleando cópulas para la caracterización de otros eventos hidrometeorológicos extremos.

Se propone comenzar la exploración del tema mediante la obtención de relaciones de dependencia que existen entre el caudal pico y volumen en crecientes en el país y obtener su probabilidad de no excedencia conjunta. Para ello se propone el estudio de tres subcuencas del río General en las que se cuenta con estaciones hidrológicas.

## 1.2. Antecedentes

Según indica Nelsen (2006), la aparición en literatura científica del concepto de cópula para describir el análisis de frecuencias empleando las distribuciones marginales de un fenómeno multidimensional es relativamente reciente. A pesar de ello, ha recibido mucha atención en los últimos 20 años debido a su utilidad en diversos ámbitos de aplicación que van desde las finanzas hasta la hidrología.

Dos de las primeras apariciones en el campo de la hidrología de la estructura de cópulas para analizar la dependencia entre las características de eventos extremos son Favre , *et al* (2004) y Salvadori y De Michele (2004). Poco después les siguieron artículos explorando la aplicación en casos específicos de ingeniería como la estimación de la avenida de diseño o de revisión como el caso de De Michele , *et al* (2005).

Desde entonces, el uso de cópulas se ha extendido a la mayoría de los dominios de la hidrometeorología con particular énfasis en los análisis de eventos extremos. Algunos de los usos más frecuentes de la metodología, como describe Chen y Guo (2018), se da en el análisis de relaciones espaciales y temporales de fenómenos como escorrentía, precipitación y sequías.

Algunos de los múltiples ejemplos del uso del uso de cópulas para análisis bivariados incluyen: Wen , *et al* (2022) quienes emplearon cópulas de Gumbel-Hougaard para predecir la correlación hidrológica entre lluvia y caudal, Stamatatou , *et al* (2018) quien realizó un análisis de frecuencia de crecientes en Chipre empleando volúmenes y caudales como variables de interés; Requena (2015) detalla el procedimiento a seguir para el análisis de crecientes univariado local, multivariado local, univariado regional y

multivariado regional; por su parte Ávila y Mono (2018) realizó un análisis de cópulas para la estimación de riesgo hidrológico. Previamente, en el ámbito latinoamericano, Zegpi (2008) realizó un análisis de cópulas bivariadas para el análisis de sistemas de drenaje urbanos en Chile.

Como se indicó en la sección 1.1, no se encontraron ejemplos de aplicación en Costa Rica de cópulas para estimación de la probabilidad de crecientes empleando volúmenes y caudales. Incluso, a pesar de su uso generalizado en la práctica profesional asociada a represamientos, no ha habido en los últimos años muchos ejemplos de investigación específicamente sobre los procedimientos para el análisis de frecuencia de crecientes. Se encontraron ejemplos de dicho enfoque en Quirós (1979), Rudín (1983), Zúñiga (1983). Además, se observa una continuidad hasta la actualidad de investigación respecto a la estimación de frecuencia de otros fenómenos hidrometeorológicos como sequías y precipitaciones intensas, como se evidencia en los estudios de García (2015), Muñoz (2017), Murillo (1994, 2006), Rojas (2011a, 2011b).

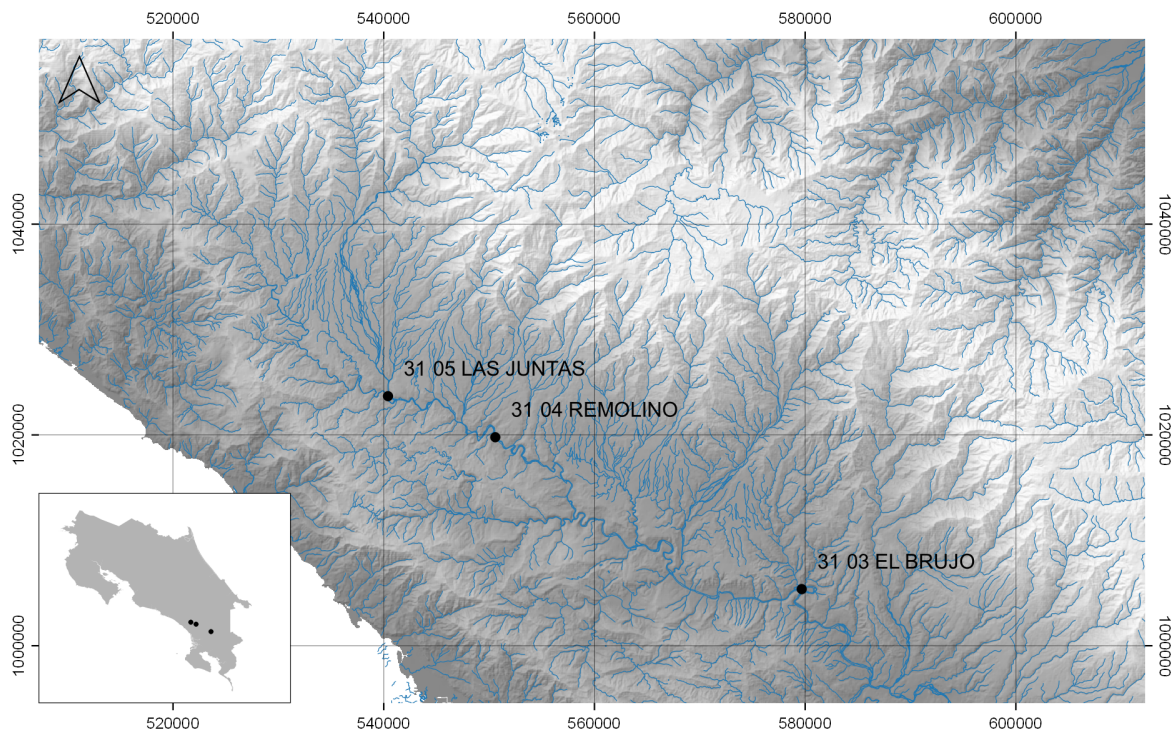
## 1.3. Objetivos

### 1.3.1. Objetivo General

Comparar la probabilidad de no excedencia del caudal pico y volumen de crecientes del río General obtenida mediante un análisis de frecuencia bidimensional con la metodología de cópulas y un análisis de frecuencia unidimensional empleando muestreo de bloque máximo.

### 1.3.2. Objetivos específicos

- Obtener una serie de máximos anuales de volumen y caudal pico para cada una de las estaciones analizadas en la cuenca del río General.
- Realizar un análisis univariado de frecuencias en las series anuales de caudal pico y volumen para cada una de las estaciones.
- Realizar un análisis multivariado de frecuencias empleando cópulas a las series de caudal pico y volumen para cada una de las estaciones.



**Figura 1.1:** Estaciones propuestas para ser analizadas.

- Desarrollar un análisis de sensibilidad del modelo respecto a la metodología de obtención de la series de máximos.

## 1.4. Alcances

Se realiza un análisis de frecuencia univariados y multivariados de crecientes empleando cópulas para las variables de caudal pico y volumen efectuando un muestreo de bloque máximo.

Se realiza el estudio empleando las estaciones de medición de caudal 3103 El Brujo, 3104 Remolino y 3105 las Juntas de la cuenca del río General, en el período desde 1969 hasta 2018, llegando así a 60 años de registro. Se escogieron estas estaciones y cuencas debido, principalmente, a la extensión, coincidencia y densidad del registro de datos que presentan. Además, dichas estaciones se encuentran ubicadas en tres puntos subsecuentes con áreas de cuenca relativamente diferentes entre sí. La ubicación de las estaciones se observa en la Figura 1.1.

## 1.5. Limitaciones

No se realiza dentro del estudio un análisis de validación y verificación de la información suministrada por el Instituto Costarricense de Electricidad (ICE), dado que para ello se requiere de la disponibilidad de información histórica adicional de la que no se dispone.

Como se indica en la sección 2.9, la medición de caudal es un proceso complejo que por su naturaleza está asociado a errores aleatorios y sistemáticos significativos; sin embargo, no se consideran esas fuentes de error en el análisis debido a que ello sobrepasa el alcance de esta investigación.

Para la separación de flujo excedente del flujo total y para la segmentación de eventos hidrológicos se emplean métodos cuya calibración se fundamentó en criterios asociados a la apreciación subjetiva del autor. Para garantizar que estos procedimientos sean objetivos se requiere de datos de precipitación, análisis isotópicos, modelos hidrológicos previamente calibrados o una base de datos etiquetada con extensiones de eventos. No se dispone para este estudio de dichas fuentes de información. Si bien esta es una fuente de error, se considera su potencial impacto en el apartado de análisis de sensibilidad (2.12, 4.5).

Algunos de los supuestos requeridos por la teoría de valores extremos potencialmente no se cumplen en este análisis. Uno de los más importantes es la estacionaridad de la serie de valores extremos, ya que se conoce a priori que la cuenca en la que están ubicadas las estaciones han sufrido modificaciones antrópicas, aunque se desconoce en qué medida dichas modificaciones afectaron el comportamiento hidrológico medido en las estaciones de interés. Además, existen efectos de variabilidad y cambio climático que podrían generar condiciones no estacionarias en la serie. Se considera que la evaluación de dichas hipótesis consisten en investigaciones con suficiente complejidad por sí solas, por lo que no se consideran en el trabajo.



# Capítulo 2

## Marco teórico

Para el desarrollo del análisis que se pretende realizar, se requieren abordar algunos elementos básicos de la teoría subyacente de estadística para así aclarar algunos términos y conceptos. Cada uno de los temas tratados a continuación serán de alguna forma mencionados en uno de los subprocesos de la metodología.

### 2.1. Análisis de frecuencia de eventos extremos en hidrología

Como indica Hamed y Rao (1999), el objetivo principal de un análisis de frecuencia es determinar la relación que existe entre la magnitud de un evento extremo y la frecuencia de ocurrencia del mismo empleando distribuciones de probabilidad. Para ello se recurren a una serie de supuestos que no necesariamente son verificables o siquiera válidos; a pesar de lo anterior, su uso es ubicuo en la ingeniería hidráulica e hidrológica para el diseño de elementos que deben manejar escorrentía.

Quizá uno de los términos más empleados en este ámbito es el período de retorno. Como indica Hamed y Rao (1999), el mismo está definido como el tiempo promedio entre arribos de eventos asociados a una determinada magnitud. Se puede decir entonces que la probabilidad de excedencia asociada a un caudal definido  $Q_T$  por un período de retorno  $T$  puede ser entendido como  $P(Q_T > q) = 1/T$ , pues será en promedio excedido en el período solamente una vez, y por consiguiente la probabilidad de no excedencia viene dada por una distribución de probabilidad acumulada  $F(Q_T) = P(Q_T < q) = 1 - 1/T$ .



Para construir dicha distribución es necesario obtener una muestra  $\mathcal{Q}$  de la población de caudales tal que permita, con cierto nivel de incertidumbre, estimar los parámetros que definen la distribución  $\hat{F}(\mathcal{Q})$ . Para ello existen varias formas de muestrear la serie de caudales: el máximo valor anual de cada año (llamada serie anual) o todos los caudales que superen cierto umbral (serie parcial). Es importante notar que, según indica Hamed y Rao (1999), para hacer inferencia estadística empleando una serie parcial o anual se requiere que los eventos sean idénticamente distribuidos e independientes entre sí. McNeil , *et al* (2005) indica que el concepto de independencia puede ser relajado requiriendo que al menos la serie de máximos sea estrictamente estacionaria; es decir, que sus propiedades estadísticas se mantengan en el tiempo.

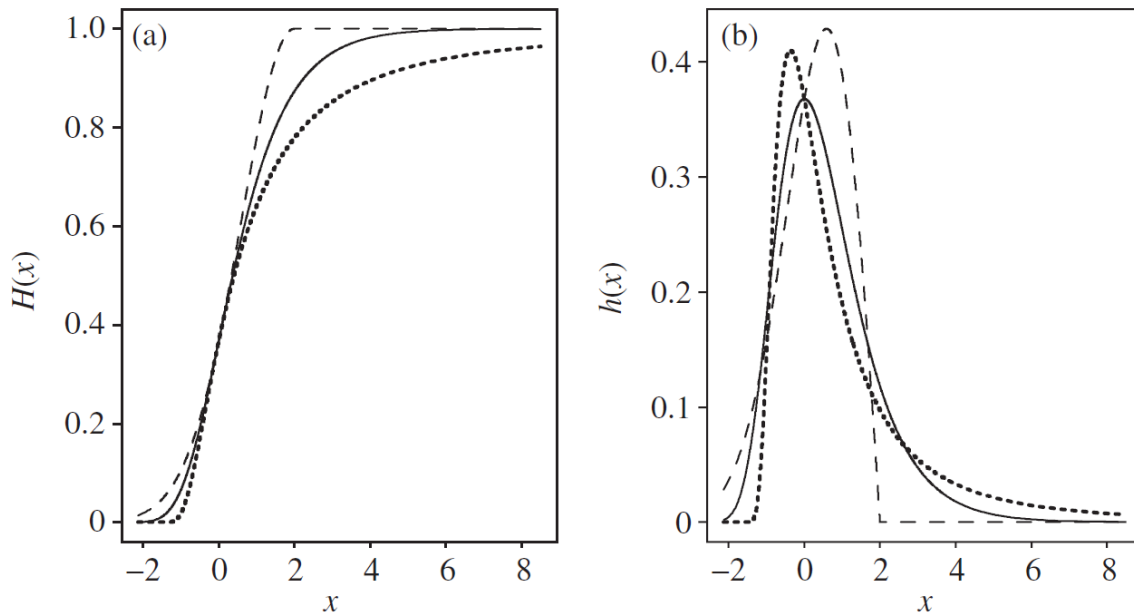
En cualquiera de los casos previamente mencionados, se están tomando muestras que representan la cola superior de la distribución, es decir, se está trabajando con distribuciones de valor extremo. A continuación se explorará la base teórica del análisis de máximos enfocándose en un muestreo de serie anual de los datos.

En el caso del Teorema del Límite Central en donde la distribución de las sumas normalizadas de  $n$  valores de una variable aleatoria independiente e idénticamente distribuída (iid) converge, cuando  $n$  es suficientemente grande, a una distribución normal estándar. Según McNeil , *et al* (2005), homológamente en el análisis de valores extremos existe una familia de distribuciones de probabilidad a la que convergen todos los máximos normalizados estandarizados. Dicha familia se conoce como GEV por sus siglas en inglés indicando Valor Extremo Generalizado y su distribución estándar de probabilidad es:

$$H_{\xi}(x) = \begin{cases} \exp\left(-\left(1 + \xi x\right)_+^{-1/\xi}\right), & \xi \neq 0 \\ \exp\left(-e^{-x}\right), & \xi = 0 \end{cases}, 1 + \xi > 0 \quad (2.1)$$

Dónde  $x = (y - \mu)/\sigma$  es la variable  $y$  estandarizada mediante los parámetros de localización  $\mu \in \mathbb{R}$  y dispersión  $\sigma > 0$ ; mientras que  $\xi$  es el parámetro de forma. Como se indica en McNeil , *et al* (2005) el tipo de distribución es definido por  $\xi$  y puede ser clasificado en tres dominios de atracción: Fréchet, Weibull y Gumbel.

La distribución Fréchet se da cuando  $\xi > 0$ , y además de no tener una cota superior su cola es pesada. La distribución Gumbel se da cuando  $\xi = 0$  y si bien no tiene cota superior ni inferior su cola es menos pesada que la de Fréchet. Por último, cuando  $\xi < 0$  la distribución es Weibull y tiene acotada la cola inferior en un valor definido.



**Figura 2.1:** Distribuciones de probabilidad (a) y densidades (b) de las familias Fréchet (punteada), Gumbel (continua) y Weibull (guiones) pertenecientes a la GEV. McNeil , *et al* (2005).

Dichas distribuciones pueden ser observadas en la figura 2.1 para parámetros  $\xi = (0.5, 0.0, -0.5)$ , respectivamente.

Supóngase que la distribución de los máximos de una variable iid ( $M_n$ ) converge dada una apropiada normalización con las constantes  $d_n$  y  $c_n$ ,  $c_n > 0, \forall n$ . Entonces se puede decir que:  $\lim_{n \rightarrow \infty} P\left(\frac{M_n - d_n}{c_n} \leq x\right) = \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x)$ . Se puede indicar que  $F \in MDA(H)$ , es decir, que pertenece al dominio de atracción de  $H$  y además, se sabe que dicha función debe ser GEV por el Teorema de Fisher-Tippett, Gnedenko, Gnedenko (1943).

**Teorema de Fisher-Tippett, Gnedenko:** Si  $F \in MDA(H)$  para alguna función de distribución no degenerada  $H$ , entonces  $H$  es un distribución de tipo  $H_\xi$ , es decir, una distribución GEV

Esto implica que sabemos, a priori, cuál es la familia de distribuciones de probabilidad de la que provienen las series de máximos de caudal y de volumen; sin embargo, no conocemos a priori cuáles son los parámetros de definen la distribución propiamente. Para ello podemos recurrir a distintos métodos, como será explorado en el siguiente apartado.

### 2.1.1. Estimación de los parámetros de la distribución

Uno de los métodos más usuales para la estimación de los parámetros de una distribución es empleando el método de momentos. En este método se calculan los  $r$  momentos de una distribución, y luego se calculan sus contrapartes empíricas usando la muestra. Por último, los mismos se igualan para así obtener, con un sistema de  $r$  ecuaciones y  $r$  incógnitas, los parámetros de la distribución. Si definimos  $x(p)$  como la función inversa de  $F(x)$  y  $u = F(x)$ , la ecuación general de los momentos puede ser descrita, según Hosking y Wallis (1997), como  $\mathbb{E}(X^r) = \int_0^1 (x(u))^r du$ . La media es un ejemplo de momentos ordinarios cuando  $r = 1$ .

Según Hosking y Wallis (1997), el uso de la metodología de momentos ordinarios, para el cálculo de los parámetros de distribuciones de probabilidad de eventos extremos resulta en algunos casos problemática. Esto porque, cuando el tamaño de muestra es pequeño, la inclusión de valores en extremo altos o bajos dentro de la serie varía significativamente el valor de los momentos y, por tanto, de los parámetros.

Hosking y Wallis (1997), indica que una alternativa al uso de los momentos ordinarios para el cálculo de los parámetros son los momentos lineales o momentos ponderados. Una diferencia marcada entre ambas metodologías es que en el caso de los momentos ordinarios la función integrada de  $x$  crece a la potencia  $r$  por lo que un valor lo suficientemente diferente a los demás (lejos de la media) puede generar un gran cambio, ya que dicha diferencia está elevada a números cada vez más grandes. En el caso de los momentos lineales los datos de la muestra no se elevan a potencias cada vez más grandes; es la función de ponderación la que aumenta de grado conforme se requieren momentos de mayor orden, por ejemplo,  $\lambda_r = \int_0^1 x(u)P_{r-1}^*(u)du$ , donde  $P_{r-1}^*(u)$  son polinomios de Legendre desplazados. Hosking y Wallis (1997) indica que los momentos lineales de la muestra pueden ser obtenidos mediante combinaciones lineales de la muestra, y mediante estos se pueden obtener los parámetros de la distribución.

Según indica Hamed y Rao (1999), otro de los métodos empleados para el ajuste de las distribuciones de probabilidad es el de máxima verosimilitud. Si se supone que la muestra de una distribución es una variable aleatoria independiente, entonces la probabilidad conjunta de obtener dicha muestra puede ser obtenida mediante la multiplicación de la probabilidad asociada a obtener cada uno de los valores de esta (véase 2.2 para más información sobre probabilidad conjunta). En el método de máxima verosimilitud se definen los parámetros de la distribución, a la que se supone

la muestra pertenece, de forma tal que la probabilidad de que la muestra provenga de ella se maximice. Es decir, si  $\alpha_1, \alpha_2, \dots, \alpha_k$  son los parámetros de la distribución de densidad  $f$  y  $x_1, x_2, \dots, x_n$  es la muestra de dicha distribución, entonces, la función de verosimilitud se define como  $L(\alpha_1, \alpha_2, \dots, \alpha_k) = \prod_{i=1}^n f(x_i; \alpha_1, \alpha_2, \dots, \alpha_k)$ . Maximizar  $L$  conlleva entonces a obtener los parámetros de la distribución.

## 2.2. Probabilidad conjunta y cópulas

Sean  $(X, Y)$  un par ordenado de variables continuas aleatorias iid de las cuales se conocen adicionalmente  $F(x)$  y  $G(y)$ , las distribuciones de probabilidad que las caracterizan.

Se desea conocer  $P(X < x, Y < y) = H(x, y)$ ; como se indica en Blanco , *et al* (2012), si  $X$  y  $Y$  son independientes, entonces  $H(x, y) = P(X < x)P(Y < y) = F(x)G(y)$ . Este podría ser identificado como un caso particular en el que la dependencia entre las variables es nula.

Para encontrar la probabilidad conjunta ( $H(x, y)$ ) en el caso general donde las variables no necesariamente son independientes se debe construir una distribución de probabilidad bidimensional que tome en consideración esta relación, o bien, se podría analizar mediante una cópula. Como se describe en Nelsen (2006), cópula es una función que permite, a partir de las distribuciones de probabilidad de las distribuciones marginales que caracterizan las variables que definen el fenómeno, encontrar el valor de la probabilidad conjunta, es decir, es una función que mapea desde el codominio de las funciones de probabilidad al codominio de la distribución conjunta. En el ejemplo antes citado la cópula está definida como  $H(x, y) = C(F(x), G(y)) = C(u, v)$ , donde claramente  $F(x) = u$ ,  $G(y) = v$ . Más formalmente, se puede hacer referencia al teorema de Sklar (1959):

**Teorema de Sklar:** *Sea  $H$  una distribución de probabilidad conjunta con marginales  $F$  y  $G$ . Entonces existe una copula  $C$  tal que  $\forall x, y \in \mathbb{R}$ :*

$$H(x, y) = C(F(x), G(y)) \quad (2.2)$$

*Donde si  $F$  y  $G$  son continuas,  $C$  es única, caso contrario  $C$  está determinada únicamente en  $\text{Ran}F \times \text{Ran}G$ .*

Las cópulas bidimensionales están definidas con un Dominio  $\text{Dom}C = \mathbf{I}^2$  y un rango  $\text{Ran}C = \mathbf{I}$ . Además, las cópulas deben ser doblemente crecientes. Esto implica que un incremento en cualquiera de las distribuciones marginales genera un cambio no decreciente en la probabilidad conjunta.

$$\Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1) \geq 0, \quad (2.3)$$

$$\forall (x_1, y_1), (x_2, y_2) \in \text{Dom}H$$

Finalmente, una cópula se encuentra definida entre los límites de Fréchet-Hoeffding definidos como:

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) = M(u, v) \quad (2.4)$$

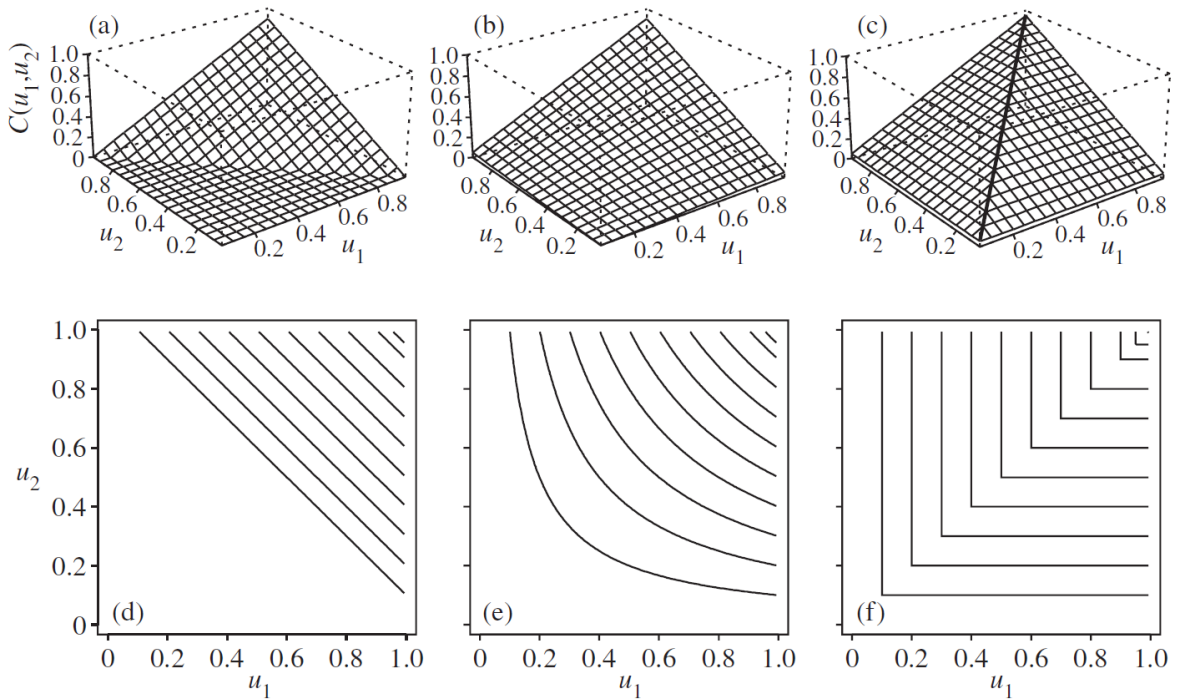
Es importante notar que, tanto  $W(u, v)$  como  $M(u, v)$ , son a su vez cópulas que indican cuando  $Y$  es una función decreciente de  $X$  y cuando  $Y$  es una función creciente de  $X$  respectivamente. Estos límites se dan cuando la relación de dependencia entre las variables es funcional, el otro caso límite se da cuando las variables son independientes, donde la relación entre estas pueden ser definidas mediante la cópula independiente  $\prod(u, v) = uv$ . En la figura 2.2 se observan diferentes representaciones de estos casos particulares.

## 2.3. Dependencia

Como indica Hofert (2018), las cópulas no solamente pueden ser utilizadas para determinar la probabilidad conjunta de dos variables aleatorias, sino también pueden ser empleadas como medio para caracterizar funcionalmente la dependencia que existe entre las variables; de hecho, de ahí viene su particular nombre, por la acción de asociar las variables transformadas por medio de las distribuciones marginales entre sí.

Existen, además de las cópulas, otros medios para entender dicha relación que, como es de esperarse están a su vez asociados a la estructura matemática de la cópula. A continuación se indican algunos de ellos.

Supongamos que se conoce una muestra  $(x_1, y_1), \dots, (x_n, y_n)$  de  $X$  y  $Y$ , es decir  $(\hat{X}, \hat{Y})$ . Como indica Hofert (2018) y Genest y Favre (2007) no es suficiente obtener



**Figura 2.2:** Representación de la cópulas  $W(u, v)$ ,  $\prod(u, v) = uv$  y  $M(u, v)$ .  
 Fuente: McNeil, et al (2005)

el coeficiente de correlación de Pearson entre ellas, ya que las distribuciones de probabilidad que caracterizan a las variables pueden distorsionar dicha medida.

$$r_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{2.5}$$

Según Genest y Favre (2007), existen varios estadísticos de orden que brindan mayor información sobre esta relación independientemente de cuales sean  $F(x)$  y  $G(y)$ . Dos ejemplos de ellos son  $\rho$  de Spearman y la  $\tau$  de Kendal. Sean las coordenadas del par  $(R_i, S_i)$  los órdenes que representan los valores  $X_i$  e  $Y_i$  dentro de sus respectivos rangos, entonces:

$$\rho_n = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (S_i - \bar{S})^2}} = \frac{12}{n(n-1)(n+1)} \sum_{i=1}^n R_i S_i - 3 \frac{n+1}{n-1} \tag{2.6}$$

donde  $\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i = \frac{n+1}{2} = \sum_{i=1}^n S_i$ . Nótese que la estructura matemática es muy similar a  $r$  con la diferencia que emplea los órdenes en vez de los valores para obtener la dependencia.

Para obtener el  $\tau_n$  de Kendal se requiere obtener el número de pares concordantes  $P_n$  y discordantes  $Q_n$ . Un par concordante es aquel donde  $(X_i - X_j)(Y_i - Y_j) > 0$  y uno discordante  $(X_i - X_j)(Y_i - Y_j) < 0$ . Entonces:

$$\tau_n = \frac{P_n - Q_n}{\binom{n}{2}} = \frac{4}{n(n-1)} P_n - 1 \quad (2.7)$$

Como se indica en Genest y Favre (2007), estas medidas de dependencia pueden ser empleadas para obtener el o los  $n$  parámetros de la cópula usualmente denotados  $\theta_n$ . Así, existe una función  $g$  que permite para cada familia de cópulas obtener el valor de  $\theta_n$ .

### 2.3.1. Cópula empírica y pseudo-observaciones

Así como se pueden obtener distribuciones empíricas para datos univariados, existe el concepto de cópula empírica. Para poder construirla se emplean pseudo-observaciones que se pueden entender como los pares (o tuplas) de posiciones de graficación de cada una de las variables. En este caso, si denotamos  $R_X$  y  $R_Y$  como los órdenes (*rankings*) y  $n$  como el número total de observaciones, entonces las pseudo observaciones se definen como  $\{U = \frac{R_X}{1+n}, V = \frac{R_Y}{1+n}\}$ . La cópula empírica entonces, según Nelsen (2006), puede ser descrita como la función acumulativa de probabilidad multivariada empírica que puede ser denotada como:

$$C_n(U_i, V_i) = \frac{1}{n} \sum_{j=1}^n \mathbb{I}(U_j \leq U_i, V_j \leq V_i) \quad (2.8)$$

Donde:  $U_i, V_i$  son los pares ordenados asociados a una pseudo-observación e  $\mathbb{I}$  es la función indicatriz.

### 2.3.2. Dependencia en colas

Como indica Serinaldi (2008), para realizar análisis en las colas de la cópula resulta útil tener medidas que nos permitan determinar qué tan bien ajusta a los datos. Hacer

inferencia en dicha zona es particularmente interesante en el análisis de eventos extremos. Una forma usual de entender este fenómeno es mediante el coeficiente de dependencia que se puede definir de la siguiente forma para la cola inferior:

$$\lambda_L(w) = P(F(x) < w | G(y) < w) = \frac{P(F(x) < w, G(y) < w)}{P(G(y) < w)} \quad (2.9)$$

Donde  $w$  es un umbral arbitrario debajo o encima del cual se define la cola. En el caso de que se desee determinar el coeficiente de dependencia en la cola superior se emplea un valor alto de  $w$  y se invierte la dirección de la desigualdad de la siguiente forma:

$$\lambda_U(w) = P(F(x) > w | G(y) > w) = \frac{P(F(x) > w, G(y) > w)}{P(G(y) > w)} \quad (2.10)$$

Es usual definir la dependencia de las colas de las cópulas en los casos límite de las funciones anteriormente descritas, es decir:  $\lambda_U = \lim_{w \rightarrow 1^-} \lambda_U(w)$  y  $\lambda_L = \lim_{w \rightarrow 1^+} \lambda_L(w)$ .

## 2.4. Familias de cópulas

Así como existen diferentes distribuciones de probabilidad que responden a las características estadísticas de la población o de la muestra, así mismo existen diferentes familias de cópulas para describir el comportamiento de la dependencia entre las variables analizadas. De todas las familias disponibles para modelar Joe (2014) indica que las más comunes son las Arquimedianas, las Elípticas y las de valor extremo. A continuación se presenta una breve descripción de las características de cada una de ellas.

Según Hofert (2018), las cópulas elípticas se emplean para describir dependencia entre distribuciones marginales elípticas, de esta forma, unas de las más empleadas cópulas en ese sentido son la cópula normal o Gaussiana y la cópula de  $t$  o Student. Como indica su nombre y su estructura dichas cópulas tienen a estar concentradas en áreas o volúmenes de forma esférica o elíptica.

Las cópulas Arquimedianas son, como indica Chen y Guo (2018), la familia más empleada en hidrología por su flexibilidad para describir la estructura de dependencia



de diversas variables que caracterizan los eventos extremos, en especial cuando alguna de ellas no es en sí parte de la cola de la distribución.

Como indica Hofert (2018), dichas cópulas comparten la forma matemática para su definición  $C(\mathbf{u}) = \psi(\psi^{-1}(u_1) + \dots + \psi^{-1}(u_d))$ ,  $\mathbf{u} \in [0, 1]^d$  y la diferencia entre ellas está en el la función  $\psi$  o generador que se escoja. Dentro de esta familia se encuentran la cópula Clayton, Frank, Gumbel-Hougaard y Joe.

Finalmente, las cópulas de valor extremo, según Hofert (2018), son una generalización multidimensional del análisis de eventos extremos empleando la GEV, de esta forma, las cópulas de valor extremo definen las estructuras de dependencia de fenómenos en los que ambas marginales provienen de dicha distribución. Según McNeil, *et al* (2005), si ambas variables son obtenidas mediante el procedimiento de serie anual, la cópula que define su dependencia pertenece al dominio de atracción de la cópula de valor extremo, o a uno de los casos límite anteriormente mencionados ( $W(u, v)$ ,  $M(u, v)$  y  $\prod(u, v) = uv$ ).

## 2.5. Pruebas de características de la cópula

Una vez obtenida la cópula empírica, es posible realizar pruebas que permitan entender las características de la estructura de los datos para así escoger mejor las familias de cópulas como candidatas para ser probadas.

Algunas de estas características son: independencia, intercambiabilidad, simetría radial y dependencia de tipo valor extremo.

### 2.5.1. Independencia

Según Hofert (2018) uno de las características más importantes para ser probadas antes de continuar con el modelado mediante cópulas de la estructura de dependencia de los datos, es probar si esta es realmente diferente a la cópula independiente  $\prod$ . Es decir  $H_0 : C = \prod$  y  $H_1 : C \neq \prod$ .

Si bien  $\tau = 0$  no implica que  $C = \prod$ , Hofert (2018) indica que para casos prácticos es apropiado emplear dicha prueba estadística y complementariamente visualizar la relación entre las pseudo observaciones en un gráfico para descartar otro tipo de

relaciones que podría llevar a un falsa falta de evidencia para rechazar la hipótesis nula.

## 2.5.2. Intercambiabilidad

Como se indica en Nelsen (2006), esta propiedad implica que las características de la cópula se mantienen iguales si se intercambian los vectores que la componen el uno por el otro, es decir, para el caso bivariado  $C(u_1, u_2) = C(u_2, u_1)$ . Visualmente esto puede ser interpretado como una simetría respecto a una línea 1 : 1 (recta con una pendiente de 1 y un intercepto de 0, en el dominio  $\mathbf{I}^2$ ).

Hofert (2018) indica que un estadístico natural para evaluar esta característica de la cópula es:

$$S_n^{exc} = \int_{[0,1]^2} n(C_n(u_1, u_2) - C_n(u_2, u_1))^2 dC_n(u) \quad (2.11)$$

Donde:  $n$  := número de observaciones,  $C_n$  := cópula empírica asociada a la  $n$ -ésima observación. Esta prueba estadística fue estudiada en Genest , *et al* (2012) e implementada en el paquete computacional indicado en Hofert, Kojadinovic , *et al* (2018).

Como se indica en Hofert (2018) bajo el supuesto de que la cópula provenga de una distribución extrema se puede realizar una prueba con mayor poder empleando el hecho de que la intercambiabilidad es equivalente a simetría respecto a un eje formado por la línea vertical  $t = 1/2$  en la función de dependencia de Pickarnds.

$$S_n^{exA} = \int n(A_n(t) - A_t(1 - t))^2 dt \quad (2.12)$$

Donde  $A_n$  es un estimador no paramétrico de  $A$ . Esta prueba estadística fue estudiado en Kojadinovic y Yan (2012) e implementada en el paquete computacional descrita en Hofert, Kojadinovic , *et al* (2018).

## 2.5.3. Simetría Radial

La simetría radial implica más bien que la cópula es idénticamente igual a la cópula sobreviviente, es decir  $C = \bar{C}$ . Visualmente esto puede interpretarse como una

simetría con respecto a una línea  $1 : -1$  (recta con una pendiente de  $-1$  y un intercepto de  $1$ , en el dominio  $I^2$ ). Equivalentemente al caso de la intercambiabilidad se puede obtener un estadístico para esta característica:

$$S_n^{sym} = \int_{[0,1]^2} n(C_n(u) - \bar{C}_n(u))^2 dC_n(u) \quad (2.13)$$

Donde  $\bar{C}_n(u) = C_n(1 - u)$ . Este estadístico fue estudiado en Genest y Nešlehová (2014) e implementado en el paquete computacional indicado en Hofert, Kojadinovic , *et al* (2018).

#### 2.5.4. Valor extremo

Si bien en muchas aplicaciones se conoce a priori si la distribución de la que provienen los datos es extrema, las cópulas extremas pueden ser empleadas para describir la estructura de dependencia de vectores que no necesariamente tienen dicho origen. De ahí que, según Hofert (2018), sea útil poner a prueba la hipótesis  $H_0 : C \in \mathbb{E}$  con respecto a la hipótesis alternativa  $H_1 : C \notin \mathbb{E}$ . Donde  $\mathbb{E}$  es el conjunto de familias de cópulas extremas.

Para ello se puede emplear parte de las características de las cópulas extremas. Según Hofert (2018) una cópula  $C(u)$  es extrema si y sólo si es max estable. Esto implica que para cualquier  $u \in [1, 0]^d$  y  $r \in \mathbb{N}$ ,  $C(u) = C(u^{1/r})^r$

Según Kojadinovic , *et al* (2011) un estadístico para medir esta característica en los datos es:

$$S_n^{evc} = T_{3,n} + T_{4,n} + T_{5,n} \quad (2.14)$$

Donde:

$$T_{r,n} = \int_{[0,1]^2} n((C_n(u_1^{1/r}, u_2^{1/r})^r - C_n(u))^2 dC_n(u) \quad (2.15)$$

Esta prueba estadística se encuentra implementada en el paquete computacional descrito en Hofert, Kojadinovic , *et al* (2018).

## 2.6. Pruebas de bondad de ajuste y calidad de modelo

Como indica Genest , *et al* (2009) las pruebas de bondad de ajuste, en el caso de las cópulas, buscan encontrar si la estructura de dependencia multivariada está adecuadamente representada por una parametrización específica de cópulas  $C_\theta$ , es decir, se está tratando de probar la hipótesis  $H_0 : C \in C_\theta$ . Genest , *et al* (2009) indica que para realizar pruebas de bondad de ajuste es preferible emplear el método de pseudo-máxima verosimilitud. En dicho artículo se realizan pruebas de poder para rechazar la hipótesis nula y la prueba con mayor poder resultó ser la Cramér-von Mises.

$$S_n = \int_{[0,1]^d} n (C_n(u) - C_{\theta_n}(u))^2 dC_n(u) = \sum_{i=1}^n (C_n(U_{i,n}) - C_{\theta_n}(U_{i,n}))^2 \quad (2.16)$$

Donde  $C_{\theta_n} :=$  cópula parametrizada. Con base en lo indicado por Genest , *et al* (2009), Hofert (2018) implementó la ecuación 2.16 y el procedimiento de "bootstrap"paramétrico para determinar el  $p$  – valor de la prueba, ambos en el paquete de  $R$  descrito en Hofert, Kojadinovic , *et al* (2018). Una vez realizadas las pruebas de bondad de ajuste, se pueden descartar algunas de las distribuciones o cópulas que inicialmente se consideraron, pero que es poco probable que describan la estructura de los datos observados.

## 2.7. Medida de calidad de modelo

Con el grupo de modelos que no pueden ser descartados como distribuciones de las que proviene la muestra, se pueden emplear criterios de calidad para así escoger entre ellos. Grønneberg y Hjort (2014) desarrollaron el Criterio de Información de Cópulas que es un homólogo del Criterio de Información de Akaike en el caso de ajustes de cópulas en dos pasos como un ajuste por pseudo-verosimilitud.

Una forma práctica, según los autores, es emplear una metodología análoga a la validación cruzada dejando una muestra afuera. Esto puede ser escrito como:

$$\widehat{xv}_n = \frac{1}{n} \sum_{i=1}^n \log C_{\theta_{n,-i}}(F_{n,-i}(X_i)) \quad (2.17)$$

Donde  $C_{\theta_{n,-i}}$  es el conjunto de parámetros de la cópula estimados a partir de todas las pseudo observaciones menos  $i$ ,  $F_{n,-i}$  es la distribución empírica marginal calculada sin la observación  $i$  y  $X_i$  es la observación  $i$ , El procedimiento computacional para desarrollar este análisis está descrito en Grønneberg y Hjort (2014) y se encuentra implementado en el paquete Hofert, Kojadinovic, et al (2018) en conjunto con una versión similar a la indicada pero empleando validación cruzada con  $k$  pliegues.

## 2.8. Período de retorno en el caso de análisis multivariado

A diferencia del caso univariado, donde para un valor existe solamente una probabilidad asociada determinada por la función de distribución de probabilidad, en el caso bivariado existen diferentes posibilidades para visualizar el problema.

Por ejemplo, una distribución conjunta de probabilidad (o cópula) permite analizar la probabilidad de que ocurra un evento tal que al menos tenga una de las características en exceso de determinados umbrales, o bien, la probabilidad de que un evento que tenga ambas características simultáneamente en exceso. Salvadori y De Michele (2004), refiriéndose a dichas posibilidades, indica la nomenclatura  $P_{u,v}^{\vee} = P(U > u \vee V > v) = 1 - C(u, v)$  y  $P_{u,v}^{\wedge} = P(U > u \wedge V > v) = 1 - u - v + C(u, v)$  respectivamente. Además, si entendemos que la cópula sobreviviente está definida como  $\hat{C} = u + v - 1 + C(1 - u, 1 - v)$ , entonces, podemos simplificar la expresión del caso  $\wedge$  como  $P_{u,v}^{\wedge} = P(U > u \wedge V > v) = 1 - \hat{C}(u, v)$ . Con estas probabilidades se pueden calcular los períodos de retorno para ambos casos de la forma  $T_{u,v}^{\wedge} = \mu_T / P_{u,v}^{\wedge}$  y  $T_{u,v}^{\vee} = \mu_T / P_{u,v}^{\vee}$  donde  $\mu_T$  es la duración de bloque analizado.

Como se indica en Chen y Guo (2018) si bien para una combinación de  $u$  y  $v$  se puede obtener una probabilidad  $P_{u,v}^{\vee}$  o  $P_{u,v}^{\wedge}$ , dicha relación no es única para la función inversa, ya que varias combinaciones pueden generar la misma probabilidad. De ahí que no sólo para la representación, sino también para el análisis, varios autores emplean las iso-líneas o líneas de igual probabilidad para representar esta característica de la cópula, en especial si es una bidimensional. Para ello se establece

una relación funcional explícita o implícita para obtener  $u$  a partir de  $v$  y  $P$  o vice versa.

Stamatatou , *et al* (2018), Requena (2015) y Chen y Guo (2018) indican que se puede obtener el valor de la función de densidad en cada uno de los puntos de dicha curva sencillamente derivando la función de probabilidad por ambas variables. En este sentido, si bien la probabilidad de excedencia para todos dichos eventos es la misma, existen entre ellos eventos más probables y menos probables. Este conocimiento resulta útil para escoger conjuntos probables de eventos asociados a un determinado período de retorno.

## 2.9. Medición de caudal

En los apartados anteriores se parte de el hecho de tener una serie de eventos máximos de caudales y volumen; sin embargo, ese no es el caso en esta investigación. Es por ello que es útil mencionar el proceso mediante el cual se mide la variable de interés y cuáles podrían ser sus fuentes de incertidumbre asociada.

El procedimiento para la medición de caudal de forma continua específicamente en las estaciones estudiadas está fundamentado en la obtención de un nivel de agua (escala  $h$ ) respecto a algún punto de referencia en el cauce (cero de la escala) usando un sensor de presión. Para pasar de la escala a un caudal  $q$  se realizan aforos mediante molinete en los cuales se encuentra la velocidad en diferentes verticales de la sección transversal donde se mide la escala y luego, multiplicados por el área aproximada de la vertical correspondiente se suman para encontrar el caudal total en la sección.

Luego de tener suficientes puntos asociando  $h$  con el caudal  $q$  se estiman los parámetros de una función potencial desplazada (2.18) donde se emplean los coeficientes  $\{a, b, h_0\}$  para estimar, mediante regresión, los valores de caudal. En algunos casos incluso se desarrollan ecuaciones para distintos rangos de  $h$  para modelar los cambios en la forma de la sección.

$$\hat{q} = a(h - h_0)^b \quad (2.18)$$

Como indica World Meteorological Organization (2008) esta es una de las formas más usuales de medición continua de caudal; sin embargo, por la naturaleza cambiantes

del río, es necesario realizar un esfuerzo constante por actualizar estas curvas. Además, existen cambios morfológicos en crecientes extraordinarias donde la relación entre  $h$  y  $q$  pueden cambiar en cuestión de minutos y el cambio puede mantenerse en el tiempo.

Según Steinbakk , *et al* (2016) la incertidumbre asociada a la estimación de de caudal es muy significativa en la estimación de las características estadísticas de eventos extremos, incluso respecto a otras fuentes de incertidumbre en la estimacion tales como el tamaño de muestra.

Cabe resaltar que además de las fuentes de incertidumbre mencionadas por estos autores, en algunos casos las decisiones de los expertos en hidrometría implican cierto nivel de subjetividad. Al darse cambios de personal encargado, podría darse una discontinuidad en el criterio y, por consiguiente, en la estimación.

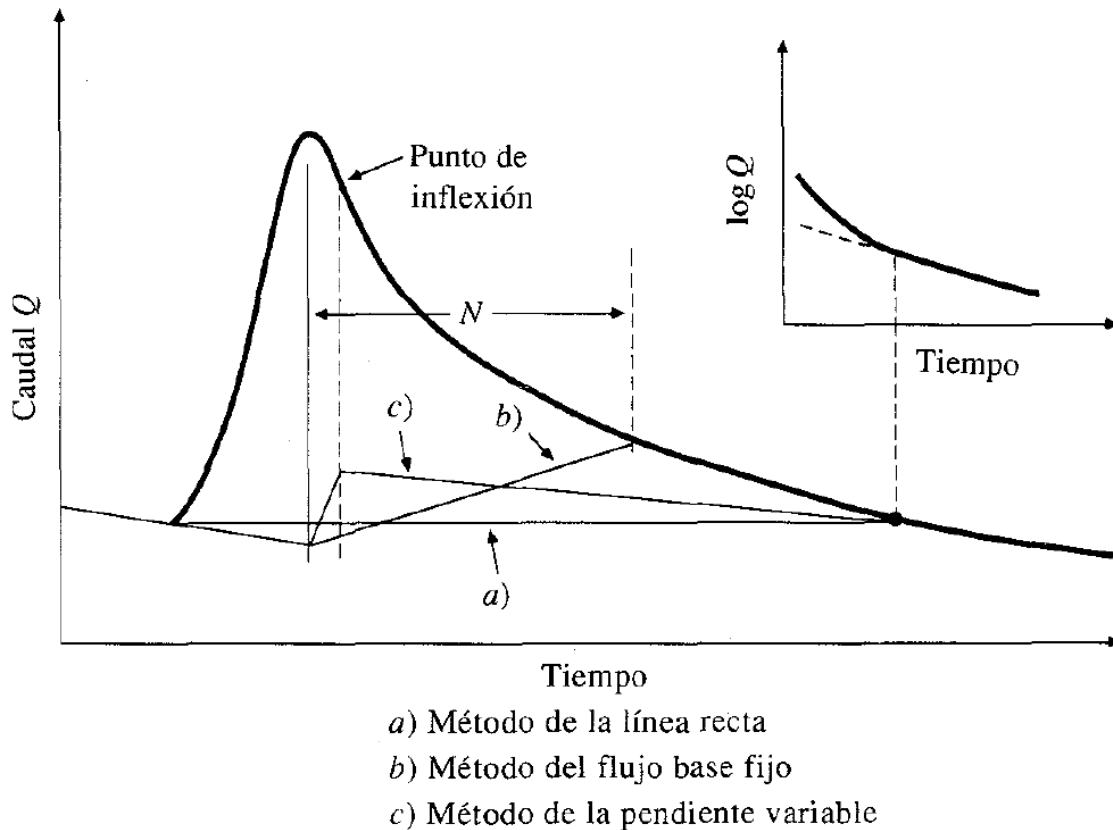
## 2.10. Separación de flujo base

Uno de los problemas con que se enfrenta la hidrología es que el caudal que se observa en un fenómeno de creciente es la suma de los efectos de eventos de precipitación pasados (llámese caudal base  $q_b$ ) y el efecto del evento actual (llámese flujo excedente  $q_e$ ). Esta subdivisión es relevante en el contexto de análisis de crecientes ya que, usualmente se requiere el cálculo del volumen de la creciente que está asociado un evento de precipitación específico como se indica en el siguiente apartado.

El caudal base y el excedente pueden ser también entendidos como diferentes mecanismos físicos que llevan agua desde la zona donde cae la precipitación hasta el cuerpo de agua analizado. Algunos mecanismos rápidos como la escorrentía superficial generan una señal en el río que corresponde directamente con el evento de precipitación que se está observando o se ha observado recientemente. Otros mecanismos, como la infiltración, percolación y eventualmente exfiltración, generan una señal lenta que se asocian mayormente a eventos pasados y en menor medida al evento que se está observando.

Para poder separar ambos flujos existen diversas metodologías que van desde métodos gráficos hasta aprendizaje estadístico e hidrológica isotópica. Algunos de los ejemplos más clásicos de separación del flujo en eventos definidos se encuentra plasmado en Chow , *et al* (1988). Estos métodos dan una idea general de cómo

deberían "verse" los eventos una vez ya separados  $q_e$  y  $q_b$ . Como indican Tan , *et al* (2009) Blume , *et al* (2007) uno de los principales problemas con los métodos de este tipo es que es muy complejo automatizar su aplicación a una serie continua de caudales y el uso en crecientes con más de una cresta es complejo cuanto menos. Además, su carácter subjetivo hace complejo, sino imposible reproducir los resultados.



**Figura 2.3:** Separación de flujo mediante métodos gráficos.

Fuente: Chow , *et al* (1996)



Una alternativa empleada extensamente en análisis hidrológicos son los filtros digitales auto-regresivos. Uno de los más usados y también de los primeros es el recomendado por Lyne y Hollick (1979), que emplea únicamente un parámetro  $\alpha$  y tiene la siguiente forma:

$$q_{e,k} = \alpha q_{e,k-1} + \frac{1 + \alpha}{2} (q_{b,k} - q_{b,k-1}) \quad (2.19)$$

Donde  $k$  es el  $k$ -ésimo instante muestreado del flujo. Empleando esta metodología en distintas cuencas en Australia, Ladson (2013) recomienda varios procedimientos asociados al método para distintas frecuencias de registro y además indica aproximadamente el rango en que se encuentra  $\alpha$  luego de la calibración. Es importante también indicar que este fue el mismo procedimiento empleado por Nikoletta (2017) para separar el flujo base en su análisis de eventos extremos y cópulas.

Cabe indicar que existen filtros que incluyen más parámetros o tienen una formulación diferente a la desarrollada por Lyne y Hollick (1979), como por ejemplo la recomendada en Eckhardt (2005) y revisada en Eckhardt (2012), o bien la indicada en Chapman (1999). Existen metodologías que emplean adicionalmente precipitación para el análisis (Mei y Anagnostou (2015), Shao , *et al* (2020)) o trazadores químicos (Foks , *et al* (2019)). La mayoría de los autores citados indican que el método con mayor probabilidad de acercarse a la realidad es el uso de isótopos para determinar cuál es el funcionamiento de la cuenca e indican que los valores de los filtros o algoritmos requieren calibración con base en ellos.

Si bien, una separación adecuada de  $q_e$  y  $q_b$  debería conllevar por sí misma a una separación de los eventos, no siempre las metodologías llegan a tener un resultado de  $q_e = 0$  en un tiempo que se considere razonable y congruente con los supuestos empleados. Esto ha conllevado a que algunos autores se enfoquen directamente en la segmentación de eventos de creciente sin necesariamente hacer previamente una separación de flujo base. Un ejemplo de este tipo de enfoques es presentado en Thiesen , *et al* (2019) y Oppel y Mewes (2020) donde se emplean metodologías de aprendizaje estadístico para la clasificación binaria de los períodos correspondientes a crecientes. Ambos autores emplean para ello además de datos de precipitación y caudal, una base de datos de eventos previamente segmentados y etiquetados.

Es importante indicar que no se cuenta para este estudio con datos de precipitación, análisis isotópicos o conductividad, modelos hidrológicos previamente calibrados ni una base de datos etiquetada con extensiones de eventos.

## 2.11. Definición de volumen y caudal pico de una creciente

Dado que el caudal es una variable continua en el tiempo, y, al menos en el caso de los cuerpos de agua perennes, siempre mayor que 0, la definición de volumen no es trivial. Usualmente se asocia esta variable a la cantidad de agua escurrida directamente en el cauce como causa de un evento de precipitación ( $EP$ ) definido en el intervalo de tiempo  $t \in [t_{ep1}, t_{ep2}]$ . También puede interpretarse el volumen de una creciente como la integral del caudal  $q_e$  en un periodo de tiempo definido por la duración de esta; es decir, si la creciente ( $EC$ ) está definida por el intervalo de tiempo  $t \in [t_{ec1}, t_{ec2}]$  entonces su volumen (o volumen excedente) es:  $V_{e,EC} = \int_{t_{ec1}}^{t_{ec2}} q_e(t) dt$ . Nótese que el intervalo en que está definido el evento de  $EP$  es diferente al intervalo de tiempo en que está definida  $EC$  por las características de respuesta que tiene una cuenca tal y como se indicó en el apartado anterior.

Cabe indicar que en algunos casos desde el punto de vista práctico, y usualmente asociado al diseño de infraestructura, interesa también conocer el volumen total que escurre en el cuerpo de agua en un período de tiempo definido por una creciente particular. En este caso no sólo se está cuantificando el agua asociada a la tormenta  $EP$ , sino también el agua asociada a tormentas pasadas que todavía tienen un efecto en el caudal ( $q_b$ ) observado durante  $EC$ . Para ello se define el caudal total como  $q(t) = q_b(t) + q_e(t)$ , con lo que volumen total sería  $V_{e+b,EC} = \int_{t_{ec1}}^{t_{ec2}} q(t) dt$ .

El caudal pico de una creciente puede estar definido para su análisis, al igual que el volumen, de dos formas. Si  $t \in [t_{ec1}, t_{ec2}]$  el caudal pico excedente es  $Q_{e,EC} = \text{máx}(q_e(t))$ , y el caudal pico total es  $Q_{e+b,EC} = \text{máx}(q(t))$ .

## 2.12. Análisis de sensibilidad

Debido a que el modelo de separación de flujo y segmentación de eventos está asociado a un alto grado de incertidumbre en los casos en que no se dispone de información adicional, es útil emplear métodos para determinar qué tanto los resultados del análisis podrían variar para diferentes elecciones de los parámetros empleados. Algunos ejemplos de este tipo de análisis se presentan en Eckhardt (2005) y Eckhardt (2012) aplicados exclusivamente a la metodología de separación de flujo.

Como se indica en Saltelli , *et al* (2008) existen muchos métodos para evaluar de alguna forma la sensibilidad en los resultados de un modelo a la variación de los parámetros de entrada del mismo, sin embargo uno de los más sencillos de emplear e interpretar es el uso de simulaciones de Montecarlo y modelos lineales múltiples para evaluar cuál variación del parámetro genera mayores cambios en la salida.

Si no se tuviera a priori una distribución para los parámetros  $\{z_1, z_2, \dots, z_m\}$ , se puede indicar un intervalo de valores en que podrían encontrarse  $\{z_1 \in [a_1, b_1], z_2 \in [a_2, b_2], \dots, z_m \in [a_m, b_m]\}$ . Luego, se realiza un muestreo de aleatorio de  $n$  tuplas en el espacio definido por los intervalos mencionados y se corre el modelo para cada una de ellas.

Defínase  $Y$  como el vector columna de todas salidas del modelo  $(y_1, \dots, y_n)$  y  $Z$  como la matriz  $(n \cdot m)$  definida por todos los parámetros del modelo donde  $Z_{i,\cdot} = (z_{i,1}, \dots, z_{i,m})$ . Entonces, se puede obtener mediante cualquiera de los métodos usuales de regresión lineal múltiple un meta-modelo tal que  $\hat{Y} = \hat{W} \cdot Z^T + \hat{w}_0$ , donde  $\hat{W} = (\hat{w}_1, \dots, \hat{w}_m)$  y  $\hat{w}_0$  son meta-parámetros que se encuentran mediante la minimización de alguna función de costo tal como errores cuadrados (e.g.  $\arg \min_{W, w_0} \sum_{i=1}^n \frac{1}{2} (y_i - \hat{w}_i z_i^T + \hat{w}_0)^2$ ).

Saltelli , *et al* (2008) indica que apropiadamente estandarizados, los meta-parámetros  $W$  y  $w_0$  pueden indicar cuáles de los parámetros explican la mayor parte de la varianza en los resultados de las simulaciones de Montecarlo. Adicionalmente, los estadísticos  $t$  asociados a cada coeficiente puede indicar la significancia de cada uno de los meta-parámetros permitiendo así considerar la certeza con que se conoce que estos no son cero.

# Capítulo 3

## Metodología

Los datos horarios de las estaciones analizadas en cada cuenca fueron facilitados por el Instituto Costarricense de Electricidad (ICE).

Se segmentó ese registro en eventos de creciente. Para ello, se empleó el filtro autorregresivo de Lyne Hollick según Ladson (2013). Una vez realizada dicha separación se empleó un umbral mínimo de la razón de flujo excedente entre el flujo total y otros indicadores para identificar puntos de segmentación como se indica en 2.10. Debido a que la escogencia de los parámetros de este procedimiento podría variar las distribuciones marginales y en las cópulas, se realizó un análisis de sensibilidad para describir el impacto de dichas decisiones en los resultados finales del estudio.

Se escogieron las tormentas con máximos anuales de caudal pico y volumen; luego se registraron los valores de volumen y caudal pico que respectivamente caracterizan dichas crecientes. Esto implica que se trabaja con dos series de valores extremos en cada una de las estaciones escogidas para el caso donde se remueve el caudal base y otras dos series en el caso en que se toma el caudal total. Es importante recalcar que es posible que alguna creciente no solamente contenga el valor máximo de caudal pico, sino también de volumen. En dicho caso la misma observación existirá en ambas series.

Con las series indicadas anteriormente se obtiene preliminarmente la distribuciones de frecuencia para el caudal y el volumen. Para la característica máxima anual, se emplea la distribución GEV. Para la obtención de los parámetros de las distribuciones de volumen y caudal, se emplean las metodologías de momentos lineales según se indica en Hosking y Wallis (1997) y las recomendaciones de Requena (2015). Debido a factores particulares de la serie de datos con la que se dispone, se considera

inapropiado emplear las distribuciones paramétricas marginales encontradas para los datos. Se prosigue el estudio empleando las distribuciones empíricas de estos. Para mayor detalle refiérase al apartado 2.3.1 y 4.3.

Con las pseudo observaciones se ajustaron diferentes cópulas (mayormente arquimedias) mediante tres procedimientos a manera de comparación. Primeramente se determinó mediante el uso del  $K$  de Kendall y la  $\rho$  de Spearman el parámetro de la cópula. Luego se empleó el procedimiento de máxima pseudo-verosimilitud para obtener dicho parámetro. Se realizará esto con la idea de entender cuáles son las diferencias generadas por cada uno de los métodos de estimación.

Se emplea la prueba de bondad de ajuste de Cramér Von Mises adaptada para Cópulas según se indica en Genest, *et al* (2009) y que es implementada por Hofert (2018) para descartar familias que no ajustan de forma adecuada los datos.

A las familias de cópulas que no se descartaron con los procedimientos previos se les calcula el criterio información de cópulas según Grønneberg y Hjort (2014). Por último, el ajuste se verifica mediante inspección visual.

Con la cópula escogida de cada una de las series de datos se determina cuál es el valor de  $P_{u,v}^{\vee}$  y  $P_{u,v}^{\wedge}$  y se compara con el valor de  $P(u)$  y  $P(v)$  obtenido de las marginales de cada una de las series de datos. Con ello, se compara los resultados obtenidos para las estaciones 3103, 3104 y 3105 de la cuenca del río General.

Finalmente, se emplea un análisis de sensibilidad según lo indicado en el apartado 2.12 para determinar qué tan sensibles son los resultados obtenidos a cambios en los parámetros de la metodología de separación de flujo y segmentación de eventos.

### 3.1. Herramientas

Para llevar a cabo lo anteriormente mencionado se empleó el lenguaje de programación  $R$  (R Core Team (2019)) en conjunto con algunos de paquetes del mismo ecosistema. A continuación, se citan algunos de los paquetes que se emplearon ligados a su uso específico.

**Cuadro 3.1:** Paquetes de R a ser empleados en el análisis.

Paquete	Referencia	Uso
<i>copula</i>	Hofert, Kojadinovic , <i>et al</i> (2018)	Obtención de copulas a partir de marginales, graficación de cópulas, pruebas de bondad de ajuste, generación de vectores aleatorios.
<i>qrmtools</i>	Hofert, Hornik , <i>et al</i> (2018)	Análisis de frecuencias.
<i>ggplot2</i>	Wickham (2016)	Graficación.
<i>tidyverse</i>	Wickham (2017)	Manejo General de Datos.
<i>Ecohydrology</i>	DR , <i>et al</i> (2018)	Filtro Lyne Hollick.
<i>lmom</i>	Hosking (2019)	Momentos lineales
<i>snow</i>	Tierney , <i>et al</i> (2018)	Paralelización
<i>patchwork</i>	Pedersen (2020)	Composición de Gráficos

# Capítulo 4

## Desarrollo

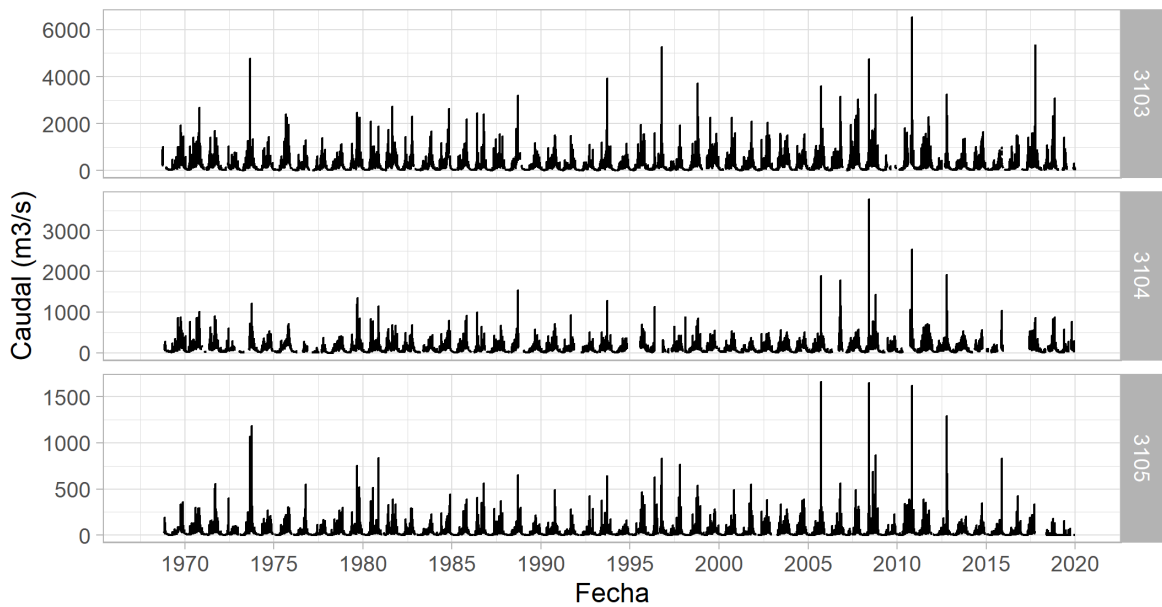
### 4.1. Revisión general de la información

En el análisis de series de tiempo, un primer paso útil consiste en revisar, cómo la variable de interés (en este caso caudal), cambia en el transcurrir del período evaluado. En la figura 4.1 se muestran los registros de las estaciones 3103, 3104 y 3105 evidenciando que existen períodos vacíos y, al menos en apariencia, una diferencia entre la primera parte del registro y la última.

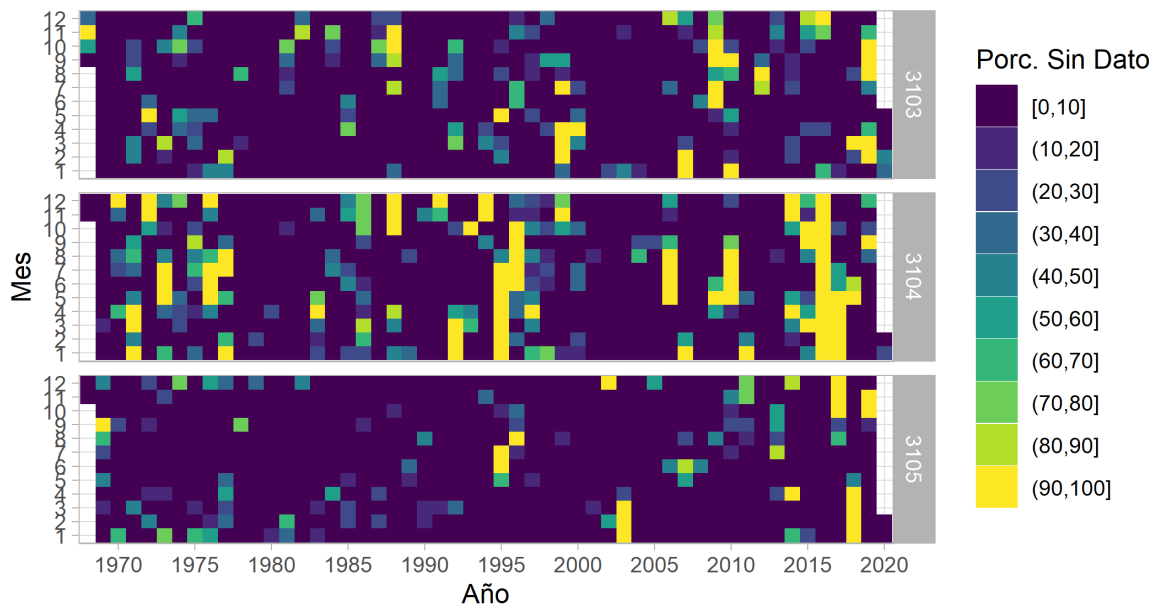
Este tipo de diferencias podría estar ocasionada por cambios en las curvas de descarga que no necesariamente respondan a fenómenos físicos tal y como se indica en la sección 2.9. Cabe resaltar que los efectos de estos errores se propagan a través de todos los resultados y se considera una de las potenciales fuentes de incertidumbre no cuantificadas más grandes en este estudio.

Otro elemento que se puede observar de la figura 4.1 es que si bien hay cierto nivel de asociación entre las estaciones, no necesariamente el orden (*ranking*) de los eventos coincide. Este fenómeno podría estar asociado tanto a particulares de la medición y procesamiento de los datos (apartado 2.9) como a la variabilidad espacial de la precipitación y a la respuesta de la cuenca. Una recomendación para futuras investigaciones en esta temática es el uso de cópulas para entender la asociación entre eventos extremos de distintas estaciones de la misma red de drenaje.

El hecho de que el registro cuente con vacíos es relevante, ya que, si algunos de esos vacíos corresponden con tormentas extremas puede generar errores en la estimación de parámetros tanto de las distribuciones marginales como de la cópula, en especial



**Figura 4.1:** Registro completo para estaciones 3103, 3104 y 3105



**Figura 4.2:** Porcentaje de vacíos por mes en estaciones 3103, 3104 y 3105

debido a que se emplea la metodología de bloques (máximo anual) para definir el conjunto de valores con los que se realizará el análisis.

En la figura 4.2 se aprecia cómo, existen períodos sin registros de meses y en algunos casos años. En la estación 3105 es donde se encuentran menos vacíos mientras



que en la estación 3104 es donde hay más períodos sin datos. En algunos casos no se cuenta con información en períodos lluviosos lo que implica la posibilidad de que se tomen valores máximos para el año menores que el valor máximo que realmente ocurrió.

Ejemplos de este tipo de ausencias son los eventos Joan, Otto, Nate y Cesar. Se encontró que para estos fenómenos al menos en dos de las estaciones estudiadas se tiene un faltante de información de más de 50% del fenómeno usualmente concentrado en la fase de aumento del caudal medido. Es importante recalcar que estos eventos son sólo menciones importantes, mas, no exhaustivas de vacíos que tiene la base de datos con que se está trabajando.

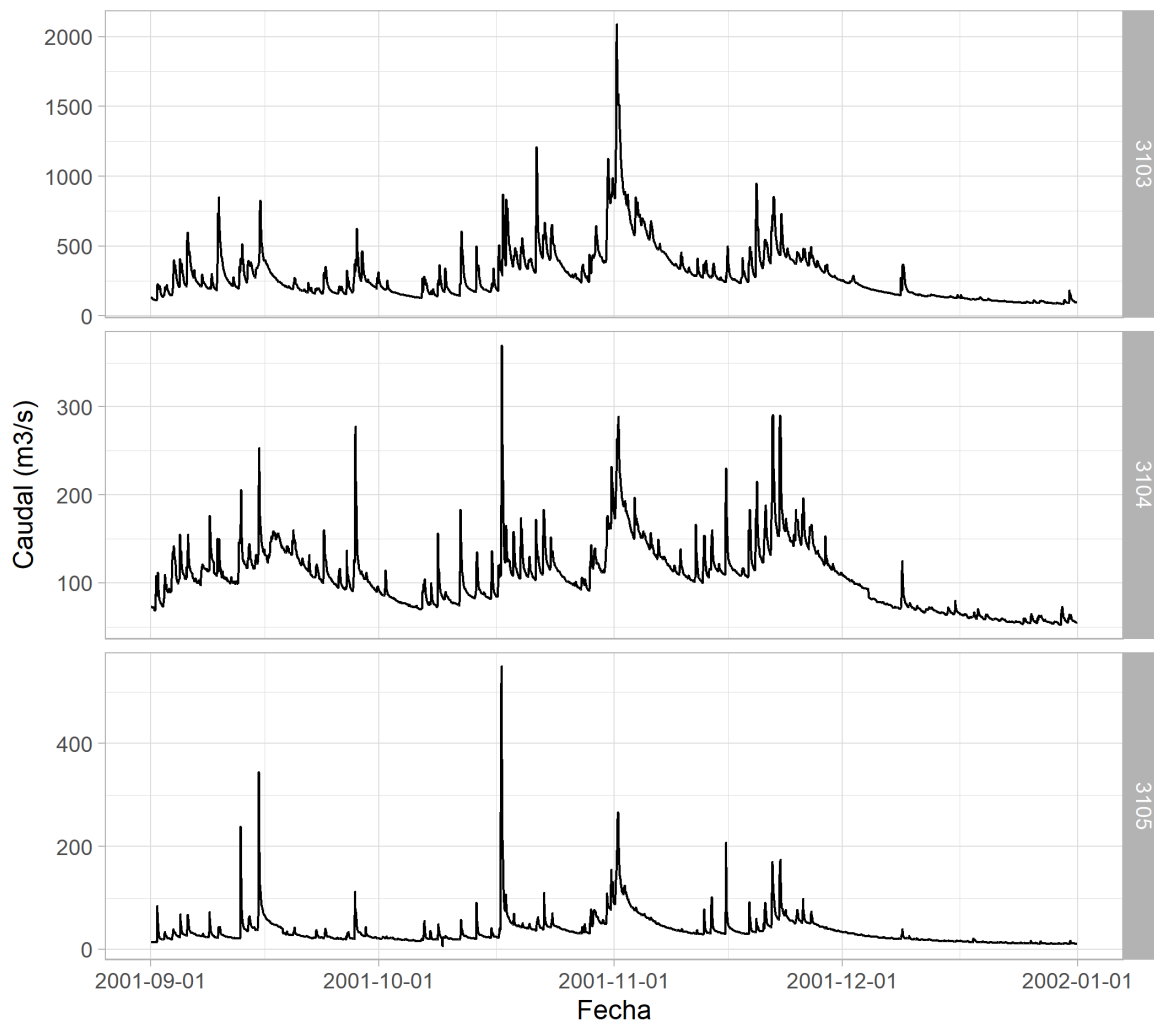
## 4.2. Obtención de eventos

Como se indicó en el apartado 2.10 usualmente los métodos de separación de flujo requieren de información experimental de concentración de trazadores químicos o isótopos para determinar el correcto valor de los parámetros. Diferentes valores de  $\alpha$  en el caso del uso del filtro Lyne Hollick pueden llevar a resultados dispares de separación y eventualmente de segmentación de los eventos.

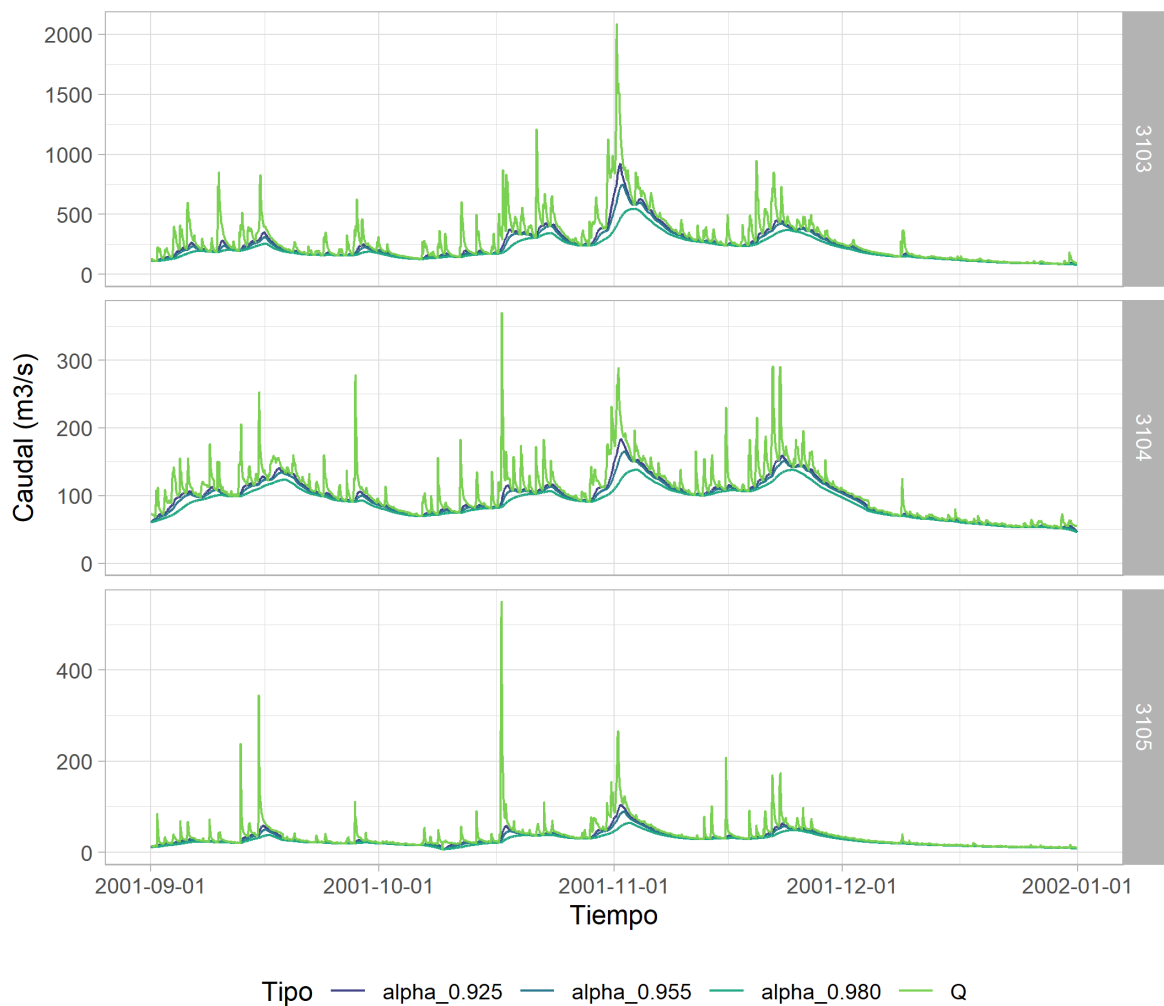
Para al menos realizar un control visual de los resultados obtenidos se definió un período lluvioso (agosto a diciembre) de análisis común entre las estaciones analizadas de forma tal que no se encontraran vacíos de datos. Se encontró que el período desde 2001-09-01 hasta 2002-01-01 es el mayor registro con esas características. En la figura 4.3 se muestran los registros para este periodo.

Como indica Ladson (2013), valores usuales de  $\alpha$  están entre 0.90 y 0.98, siendo un valor de 0.925 uno de los más usados en la literatura técnica y 0.980 el recomendado por Ladson (2013) según la calibración realizada en cuencas australianas. Se presenta en la figura 4.4 esos valores y un valor intermedio para mostrar las diferencias que se pueden dar.

Mientras mayor el valor de  $\alpha$ , mayor es la *memoria* del filtro y por tanto los resultados son mucho más suavizados. Esto implica que para valores de  $\alpha$  cercanos a 1 se generan valores de flujo base menor y para valores menores se generan valores de flujo base mayores en el transcurso de un evento.



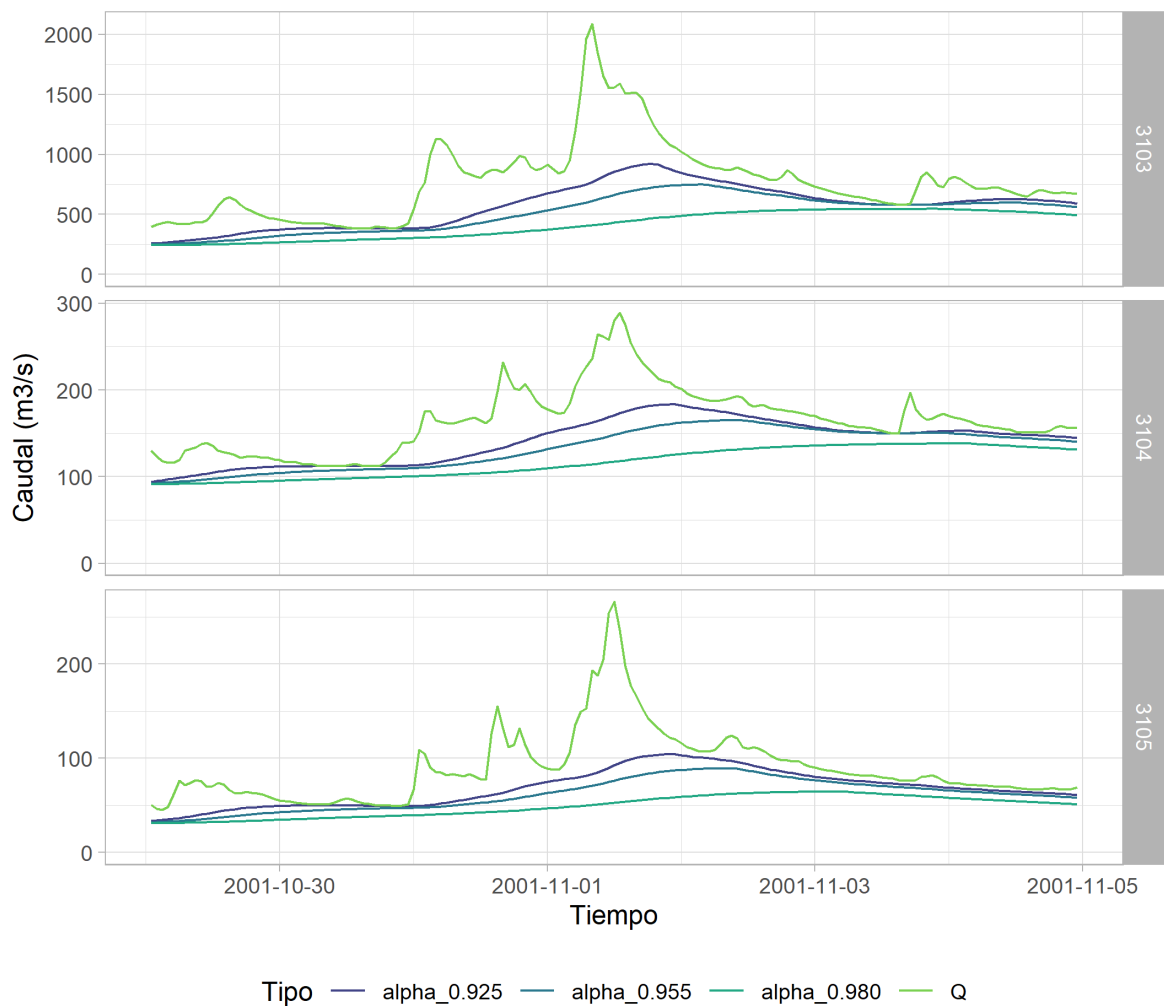
**Figura 4.3:** Muestra del registro para verificación visual del modelo para separación de flujo.



**Figura 4.4:** Resultado de separación para distintos parámetros  $\alpha$ .

Además, obsérvese que para valores de  $\alpha$  mayores, la cantidad de puntos donde el caudal excedente es cercano a cero son menores, generándose así "eventos" de mayor duración llegando en casos extremos a semanas como el evento con un máximo en 2002-11-01 que se observa más detalladamente en la figura 4.5.

La figura 4.4 y 4.5 ponen en evidencia la razón por la que la definición de evento o creciente podría estar sujeta a ambigüedad en casos donde no se tienen datos para separar el caudal excedente del base en cuencas donde la precipitación ocurre continuamente en la época lluviosa. Obsérvese a finales de octubre cómo una serie de crecientes generaron un incremento en el caudal que se mantuvo hasta mediados de noviembre. Si se tomara un valor de  $\alpha = 0.980$  la creciente estaría definida en una duración en el orden de 15 días, mientras que para  $\alpha = 0.925$  en dicho periodo se



**Figura 4.5:** Resultado de separación para distintos parámetros  $\alpha$  en evento de mayor magnitud del 2021

observan varias crecientes cuya duración está en el orden de días u horas. Además, la diferencia en la separación de flujo base y flujo excedente entre ambos coeficientes es apreciable.

Se emplea entonces un valor de  $\alpha = 0.925$  ya que en apariencia logra captar adecuadamente la separación entre los eventos de precipitación y además es de los valores más empleados en la literatura según Ladson (2013).

En el caso de épocas de estiaje extendidas donde el caudal excedente no es cero y para evitar inicios anticipados, se empleó un método alternativo para identificar rangos visualmente razonables de duración de los eventos empleando la información del filtro como base. Este método se inspira en el trabajo de Blume, *et al* (2007)

quien propone un método para limitar el final del estiaje en los hidrogramas con base en la primera derivada del caudal total respecto al tiempo.

### 4.2.1. Método de refinamiento de separación de eventos

Es necesario aclarar que la razón por la que se necesita definir eventos de creciente  $EC$  es que se requiere identificar aquellos  $EC$  que contengan un máximo de caudal pico ( $Q_{e,EC}$  o  $Q_{e+b,EC}$ ) y volumen ( $V_{e,EC}$  o  $V_{e+b,EC}$ ), que, de aquí en adelante, se denotarán por facilidad de lectura como ( $Q_e$  y  $Q_b$ ) y ( $V_e$ ,  $V_b$ ) correspondientemente.

Uno de los mayores retos para cumplir con ese objetivo es que en épocas de estiaje extendidas, si no se acota la duración del evento la magnitud de  $V_b$  podría incluir mucha información (caudal) no asociada al evento de creciente  $EC$ , influenciando así la relación que existe entre el caudal pico y el volumen. Para solucionar este problema se propone el uso de una metodología alternativa para la segmentación de series de tiempo de caudal en eventos de creciente. En los próximos párrafos se describirá en qué consiste ese método.

Se conoce que cuando  $q_e$  comienza a aumentar desde un valor *relativamente* bajo, entonces es común que un evento esté comenzando y que cuando  $q_e$  disminuye hasta acercarse a 0, entonces el evento está terminando; además entre el inicio y el final del evento se encuentra el mayor valor de  $q_e$  que llamaremos caudal pico excedente  $Q_e$ .

- Sea  $H$  una serie discreta de caudales horarios donde  $q_t$  indica el caudal medido de dicha serie en el tiempo  $t$ , donde  $q_{e,t}$  y  $q_{b,t}$  son los caudales excedentes y base asociados a  $q_t$ .
- Sea  $EC_i$  la creciente  $i$  observada en  $H$  y  $EC$  el conjunto de todas las crecientes  $EC_i$ .
- Sea  $t_{EC}$  el conjunto de tiempos de  $H$  que están asociados a alguna de las crecientes  $EC_i$ .
- Sea  $w \in \mathbb{N}$  el tamaño (número de casillas) de una ventana móvil centrada en  $t$  que calcula  $Q_{e,t,w} = \max\{q_{e,t-\frac{w-1}{2}}, \dots, q_{e,t}, \dots, q_{e,t+\frac{w-1}{2}}\}$  si  $w$  es impar y en el caso de ser par, calcula  $Q_{e,t,w} = \max\{q_{e,t+1-\frac{w}{2}}, \dots, q_{e,t}, \dots, q_{e,t+\frac{w}{2}}\}$ .
- Se considera entonces que que  $t \in t_{EC}$  si  $1 = \mathbb{I}(u_e \leq \frac{q_{e,t}}{Q_{e,t,w}}, u_b \leq \frac{q_{e,t}}{q_{b,t}})$ . Donde  $u_b, u_e > 0$  son escalares y  $\mathbb{I}$  es la función indicadora.

Se define un evento de creciente  $EC_i$  como el intervalo de tiempo compuesto por el conjunto de los tiempos subsecuentes donde  $t \in t_{EC}$ . Este criterio genera una serie de crecientes disjuntas y separadas por intervalos de tiempos  $t \notin t_{EC}$ .

Dado que las crecientes de interés son aquellas que terminarán formando parte del conjunto de máximos anuales, es de esperarse que para un  $w$  suficientemente grande pero mucho menor que un año  $Q_{e,t,w} = Q_e$  para los caudales iniciales y finales de los eventos analizados. Es decir, es de esperar que el máximo en la ventana incluya el caudal pico del evento para el final y el inicio de este. Por lo que  $u_e$  puede ser interpretado como el umbral que define la magnitud relativa que debe tener el caudal excedente respecto al caudal pico excedente para pertenecer a la creciente; y  $u_b$  puede ser interpretado como el umbral que define qué tan grande tiene que ser un caudal excedente respecto al caudal base para ser considerado una creciente.

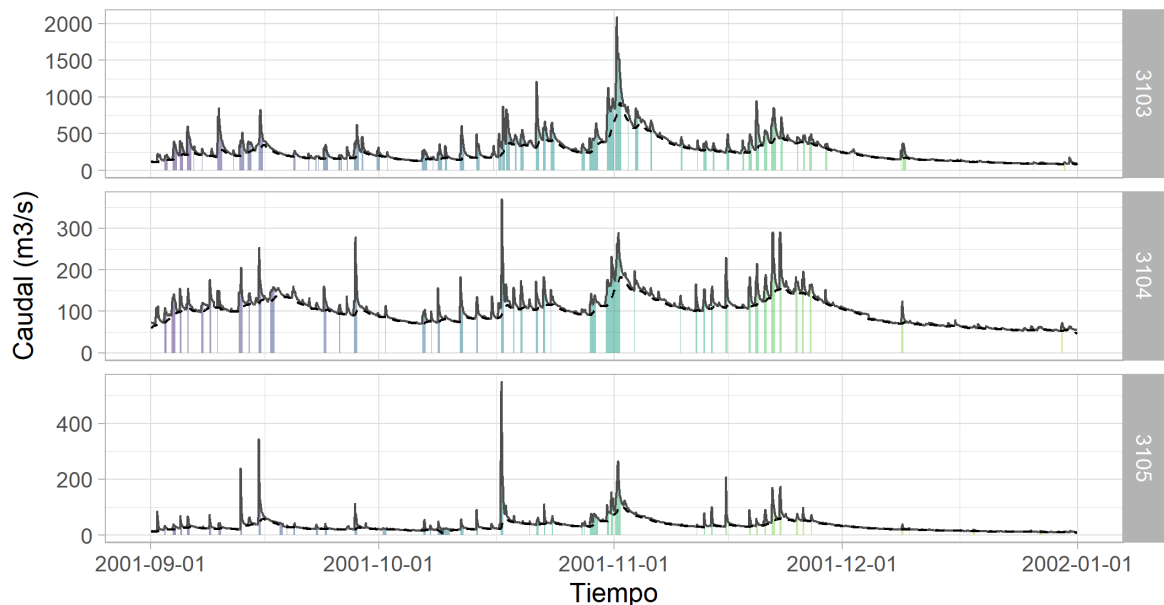
Desde el punto de vista práctico, ambos filtros permiten disminuir duraciones poco realista de crecientes (i.e. excesivamente largas) y por tanto acotan los valores del volumen total. Además, al clasificar fenómenos relativamente pequeños como no pertenecientes a crecientes, disminuye la carga computacional para los posteriores análisis.

La calibración de los parámetros  $u_e, u_b, w$  se realizó iterativamente y buscando que visualmente la separación de las crecientes se pareciera a los criterios indicados por Chow, et al (1988). Además, se realizó un análisis de sensibilidad mostrado al final del apartado de desarrollo donde se indica cuáles son las consecuencias de la elección realizada. En la figura 4.6 se observan los resultados obtenidos en el periodo de prueba con los parámetros calibrados.

Obsérvese que el método hace que la división entre los eventos de las tres estaciones no sea equivalente tanto en tiempo como en el número de eventos considerados creciente por lo que, en el caso de desearse emplear una estrategia similar para realizar análisis de cópulas entre estaciones se requeriría otro enfoque y elección de parámetros. Nótese que en este caso se emplearon parámetros iguales para todas las cuencas; sin embargo, si se tuviese información para calibrar podría ser apropiado el uso de distintos valores para cada una de ellas.

Una vez realizado este análisis se procedió con los mismos parámetros a obtener las tormentas para todas las series continuas de caudal y para todas las estaciones.

Cabe indicar que se consideró el uso de Detección Bayesiana de Puntos de Cambio para la segmentación de crecientes y se obtuvieron resultados alentadores en un



**Figura 4.6:** Separación de tormentas para parámetros  $u_e = 0.15$ ,  $u_b = 0.15$  y  $w = 72$ .

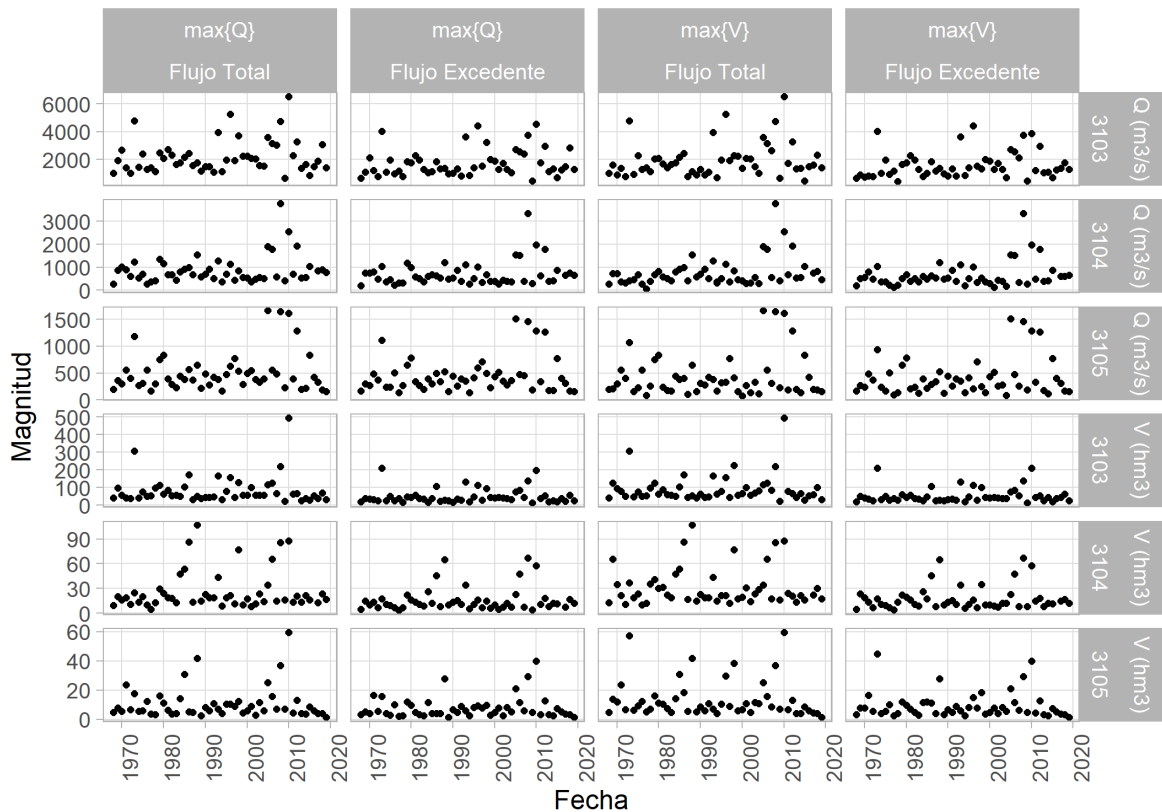
análisis preliminar; sin embargo se considera apropiada la evaluación de esa metodología en conjunto con otras de las mencionadas en el apartado 2.10 en un estudio aparte empleando idealmente otras fuentes de información como la precipitación en las cuencas analizadas.

### 4.3. Análisis univariado de series de valores máximos

Una vez segmentadas las crecientes de toda las estaciones se procedió a obtener para cada una de ellas los valores de  $V_e$ ,  $V_b$ ,  $Q_e$  y  $Q_b$ . Luego, se generaron cuatro series de máximos anuales:

- $\{V_b, \text{máx}\{Q_b\}\}$  : Volúmenes totales asociados y máximos caudales pico totales.
- $\{V_e, \text{máx}\{Q_e\}\}$  : Volúmenes excedentes asociados y máximos caudales pico excedentes.
- $\{\text{máx}\{V_b\}, Q_b\}$  : Máximos volúmenes totales y caudales pico total asociados.
- $\{\text{máx}\{V_e\}, Q_e\}$  : Máximos volúmenes excedentes y caudales pico excedentes asociados.

Las crecientes asociadas a cada una de las series anteriormente indicadas pueden ser observadas en el apéndice A y en la figura 4.7 se puede observar la variación a través del tiempo de las magnitudes.



**Figura 4.7:** Series de valores máximos anuales

Observando los valores de estas series en el tiempo, como se muestra en la figura 4.7, es poco realista pensar que la serie sea estacionaria. Todo lo contrario, en algunos casos parecen haber grupos de valores conglomerados entre sí e incluso se observan tendencias (e.g. cuadrante (1, 3) donde se observa una  $U$  en los años 80). Además, la varianza de los datos no se está manteniendo en todo el registro.

Este comportamiento pone en duda la validez del análisis debido a que los supuestos indicados en el apartado 2.1 no necesariamente se cumplen en este caso. Algunas explicaciones plausibles de este fenómeno son: 1. La metodología de medición del caudal está asociada a altos niveles de incertidumbre y a sesgos sistemáticos que varían en el tiempo según se indicó en el apartado 2.9; 2. La ausencia de algunas crecientes máximas hace que se pierdan las propiedades estadísticas de la serie de máximos según se indicó en el apartado 4.1; 3. Existen procesos hidroclimáticos



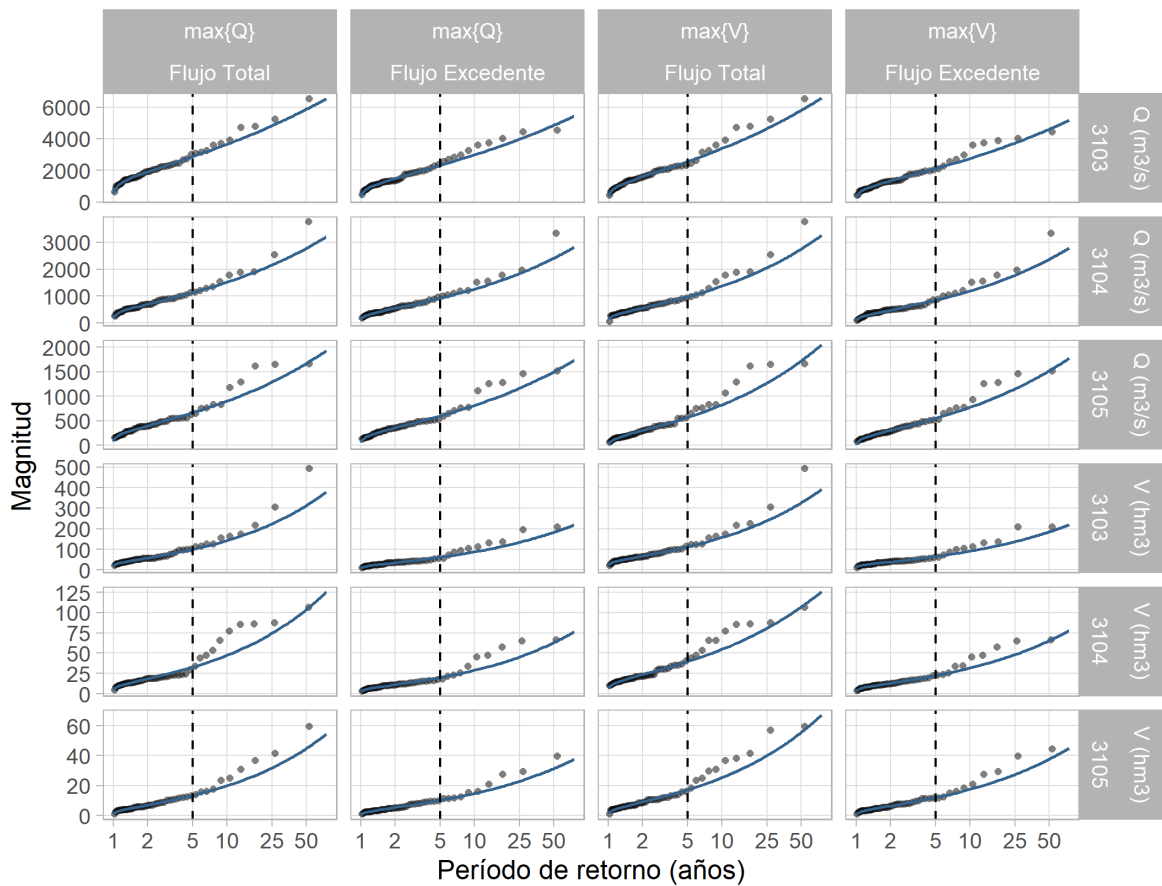
que no necesariamente cumplen con los supuestos requeridos para el análisis de máximos.

En el caso de que la explicación de lo observado resida en el punto 1. o el 2. habría que reprocesar la información y evaluar la posibilidad de reconstruir los fenómenos extremos faltantes. Es del conocimiento del autor que esto se realiza de forma esporádica para análisis de eventos extremos en el Área de Hidrología del ICE, sin embargo se desconoce si estas series de caudal horario en particular tiene dicho control de calidad y se considera que ese proceso excede los alcances de este trabajo.

En el caso 3. se podría evaluar si otro tipo de modelos se podrían aplicar a los datos que no requieran el cumplimiento de esos supuestos como ejemplo modelos bayesianos o mezclas de distribuciones de valores extremos. Se considera la exploración de estas metodologías un paso necesario para futuras investigaciones en esta área una vez que se tenga la certeza de que los datos con los que se cuenta han sido rellenados apropiadamente y se hayan validado la coherencia de las series en el tiempo. Cabe resaltar que de emplearse un muestreo del tipo máximos por encima de un umbral y una Distribución Generalizada de Pareto podría disminuir el impacto de la ausencia de algunos eventos en la serie.

Si se aplica la metodología de momentos lineales a los datos sin ningún tipo de tratamiento adicional se distingue que la distribución está asociada a más de una familia de datos como se observa en la figura 4.8. Nótese que después de períodos de retorno de cinco años se observa en muchas de las series ajustadas una desviación del ajuste obtenido mediante momentos lineales. En la figura 4.9 se muestran los valores por encima de cinco años de período de retorno.

Esta evidencia podría indicar que existen varios procesos de generación de eventos siendo captados por la misma serie de datos. Si ese fuera el caso, metodologías como el uso de mezclas podría ser apropiado para el análisis de extremos univariado siempre y cuando se pueda tener algún nivel de certidumbre que el procesamiento y completitud de los datos. Sería interesante también, conocer si una división entre diferentes tipos de eventos (huracanes, vaguadas, tormentas tropicales, etc.) o magnitud de señales sinópticas (ENOS, AMO, PDO, TNA, etc.) puede, en cierta medida, clasificar a cuál proceso de generación de eventos extremos está asociada cada creciente.



**Figura 4.8:** Distribución GEV ajustada a máximos anuales

Para evitar incluir en posteriores análisis errores estructurales asociados a la escogencia del modelo estadístico para representar las distribuciones marginales, acorde a Joe (2014) es preferible emplear un método de estimación del parámetro de la cópula que emple estadísticos de orden  $\tau$  o  $\rho_s$  o bien las pseudo observaciones en el caso de la máxima pseudo-verosimilitud. Cabe resaltar que este procedimiento no solventa los problemas asociados a los puntos indicados previamente (4.3) pero al menos disminuye las potenciales distorsiones asociadas al ajuste de un modelo no apropiado a las distribuciones marginales.

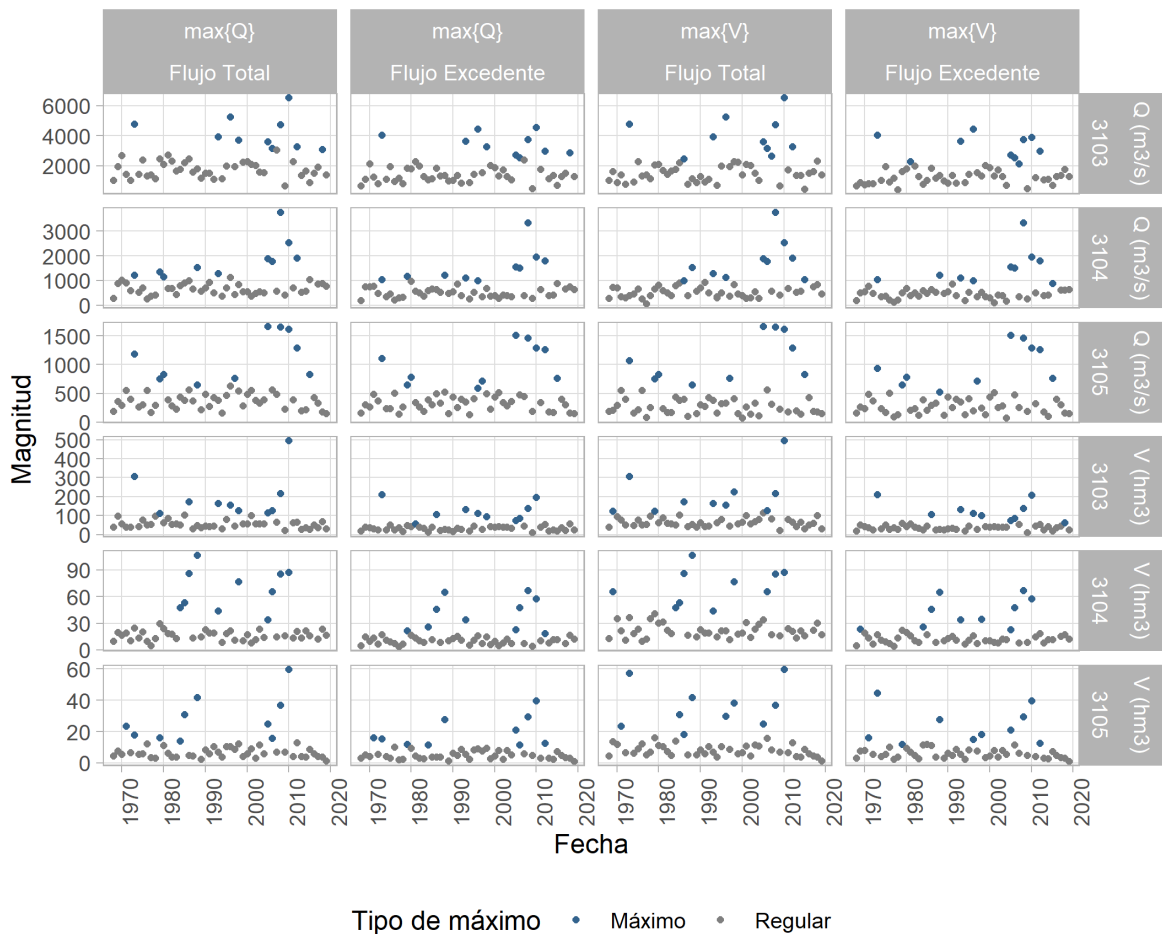


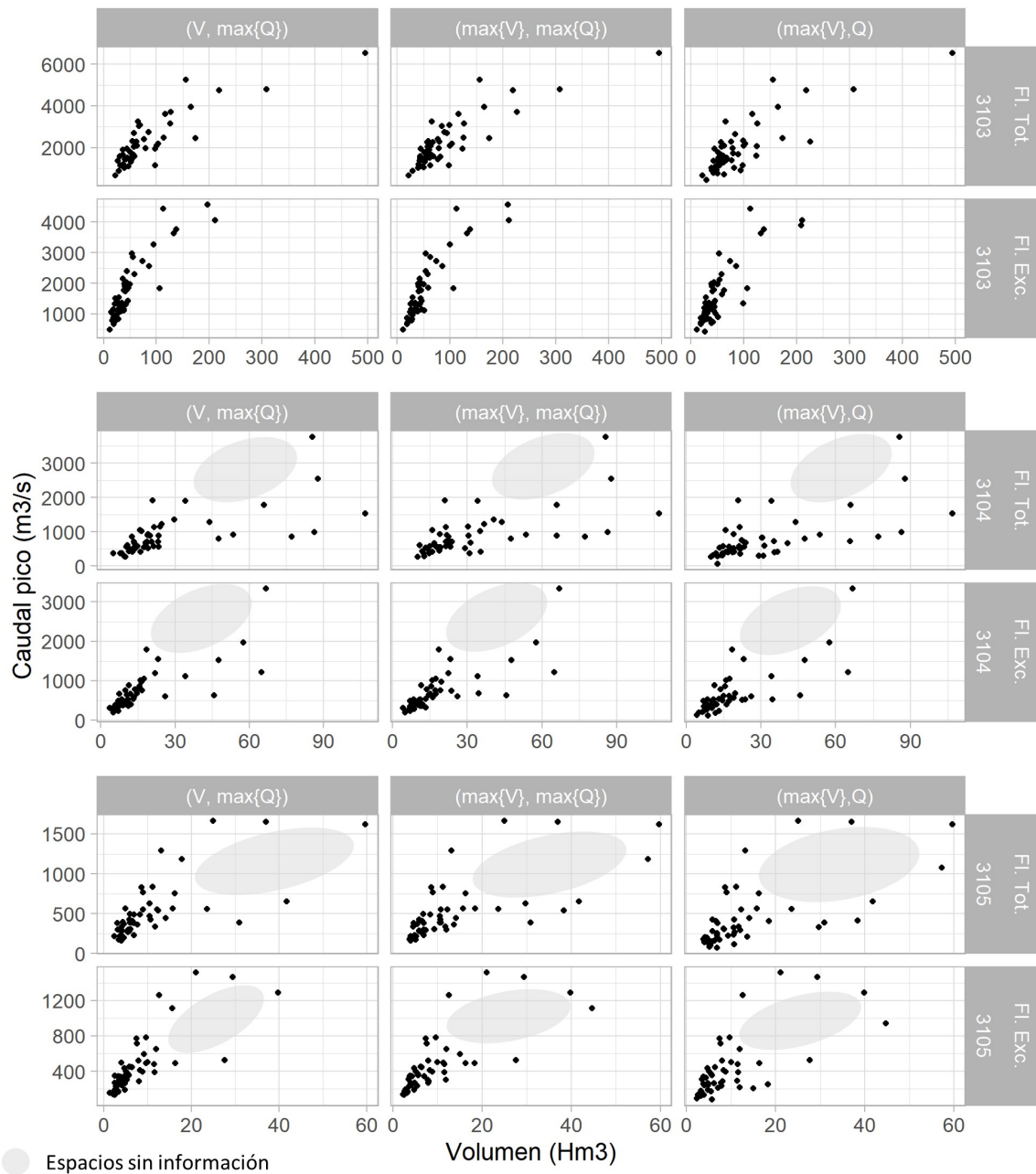
Figura 4.9: Distribución de máximos anuales

## 4.4. Análisis multivariado de series de valores máximos

### 4.4.1. Análisis exploratorio

Uno de los primeros pasos para la construcción de modelos de cópulas es evaluar visualmente cuál es la dependencia que existen entre los vectores aleatorios como se muestra en la figura 4.10. Obsérvese que las variables en general tienen una relación creciente y que existen *vacíos* en el espacio de la muestra y valores atípicos que podrían estar asociados a lo indicado en el apartado anterior.

Si se toman exclusivamente los máximos anuales de cada una de las variables, se puede llegar a dos series adicionales:  $\{\max\{V_b\}, \max\{Q_b\}\}$  y  $\{\max\{V_e\}, \max\{Q_e\}\}$ .



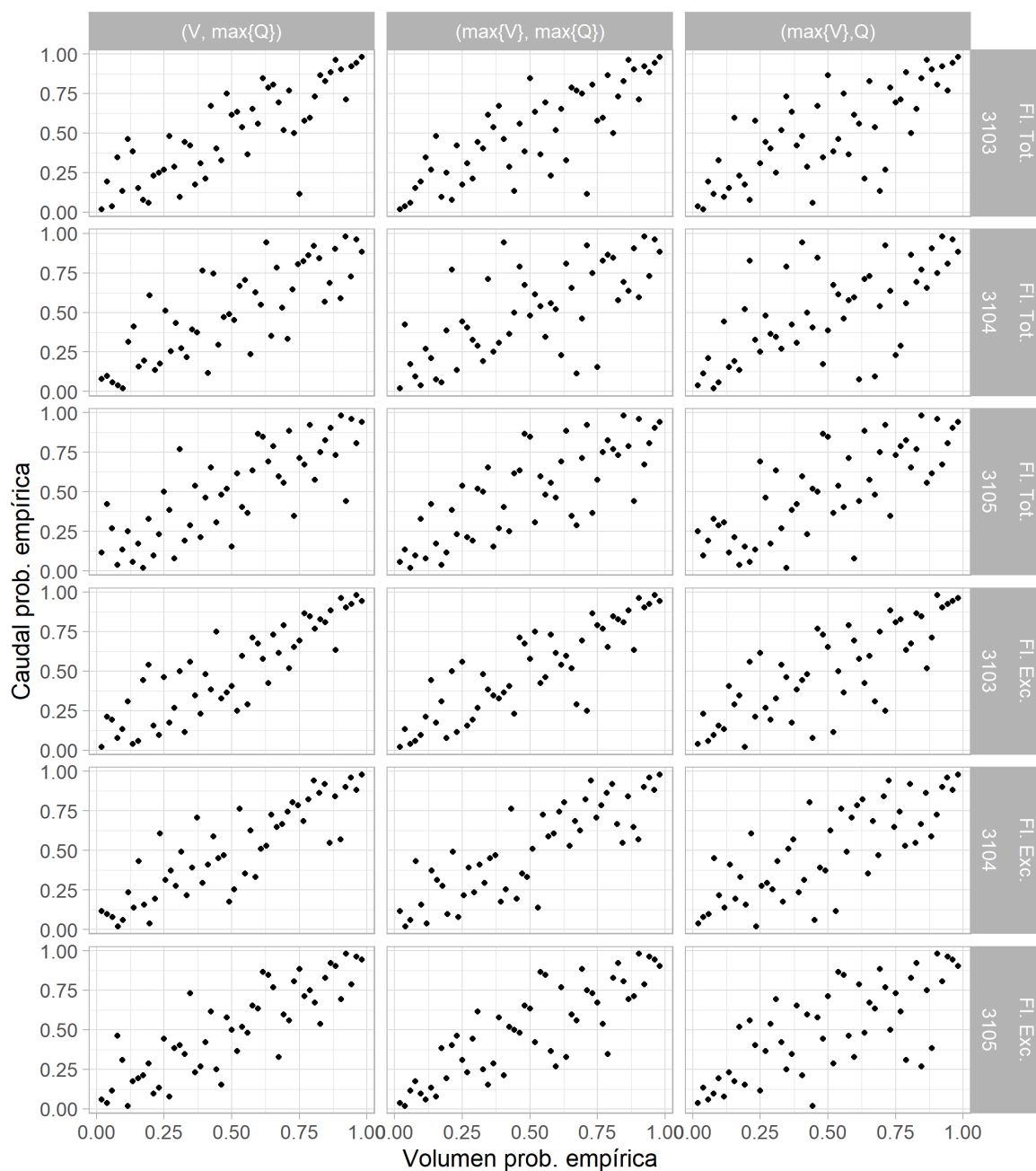
**Figura 4.10:** Relación de pares  $(Q, V)$

Para ello es importante repasar cuál es la interpretación de las cópulas que se pretenden desarrollar con estas series de datos. Si tomamos  $(V_i, Q_i)$  como un dato puntual cualquiera de volumen y caudal, en el sitio correspondiente (3103, 3104 o 3105) y del tipo correspondiente (excedente o total), entonces:

- $C_{\theta, (V, \max\{Q\})}(V_i, Q_i)$  : Probabilidad de que en un año dado se dé un caudal pico máximo menor que  $Q_i$  con un volumen asociado, para la misma creciente, menor que  $V_i$ .
- $C_{\theta, (\max\{V\}, Q)}(V_i, Q_i)$  : Probabilidad de que en un año dado se dé un volumen máximo menor que  $V_i$  con un caudal pico asociado, para la misma creciente, menor que  $Q_i$ .
- $C_{\theta, (\max\{V\}, \max\{Q\})}(V_i, Q_i)$  : Probabilidad de que en un año dado se dé un caudal pico máximo menor que  $Q_i$  y que en el mismo periodo volumen máximo sea menor que  $V_i$  no necesariamente para la misma creciente.

Empleando las pseudo-observaciones se puede visualizar la estructura de la cópula empírica para cada uno de los pares de interés. En la figura 4.11 se muestra cómo la relación entre  $U_Q$  y  $U_V$  muestra una forma elipsoidal y en algunos casos se observa una dependencia de la cola superior marcada  $\lambda_U$  (e.g. páneces (1, 1), (3, 1), (3, 2), (4, 1)). Al igual que en el caso de las variables originales se observan ciertos vacíos e inclusive algunos eventos o grupos de eventos cuya relación entre  $U_Q$  y  $U_V$  es atípica o distante de los demás valores.

En general se puede decir, mediante esta simple inspección, que los caudales y los volúmenes de las crecientes presentan una estructura de dependencia clara. Además, se puede inferir que es raro que una creciente con uno de los valores más altos de caudal no contenga al mismo tiempo uno de los valores más altos de volumen y vice-versa. También es relativamente raro encontrar que los valores más bajos de caudal no estén en una creciente que contenga valores bajos de volumen. Estas aseveraciones generales pueden observarse para todas las cópulas obtenidas en este estudio independientemente del tipo de muestreo, separación de flujo o estación.



**Figura 4.11:** Relación de pares  $(Q, V)$

**Cuadro 4.1:** Medidas de dependencia para cada serie y estación

Tipo Flujo	Muestra	id	$\rho$	$\rho_s$	$\tau$	$\lambda_U$	$\lambda_L$
Total	(V, max{Q})	3103	0.844	0.823	0.648	0.562	0.499
Total	(V, max{Q})	3104	0.680	0.803	0.613	0.271	0.244
Total	(V, max{Q})	3105	0.718	0.794	0.596	0.322	0.009
Total	(max{V}, max{Q})	3103	0.850	0.800	0.626	0.533	0.688
Total	(max{V}, max{Q})	3104	0.650	0.689	0.509	0.278	0.354
Total	(max{V}, max{Q})	3105	0.692	0.780	0.591	0.166	0.359
Total	(max{V},Q)	3103	0.822	0.738	0.557	0.533	0.551
Total	(max{V},Q)	3104	0.652	0.648	0.495	0.278	0.364
Total	(max{V},Q)	3105	0.667	0.749	0.540	0.166	0.000
Excedente	(V, max{Q})	3103	0.879	0.873	0.697	0.535	0.368
Excedente	(V, max{Q})	3104	0.789	0.864	0.696	0.503	0.102
Excedente	(V, max{Q})	3105	0.788	0.852	0.658	0.496	0.377
Excedente	(max{V}, max{Q})	3103	0.877	0.861	0.693	0.535	0.562
Excedente	(max{V}, max{Q})	3104	0.775	0.838	0.639	0.503	0.346
Excedente	(max{V}, max{Q})	3105	0.764	0.835	0.644	0.290	0.552
Excedente	(max{V},Q)	3103	0.836	0.785	0.600	0.488	0.354
Excedente	(max{V},Q)	3104	0.763	0.782	0.589	0.503	0.326
Excedente	(max{V},Q)	3105	0.702	0.718	0.539	0.290	0.354

En el cuadro 4.1, se aprecian algunos de los resultados que corroboran las observaciones efectuadas previamente acerca de las características de la dependencia entre las variables. Otro detalle a notar en dicha tabla es que en general la dependencia es mayor para máximos anuales de caudal que para máximos anuales de volumen; es mayor para flujo excedente que para flujo total; y es mayor en la 3103 que en la 3104 y 3105. Además, con pocas excepciones (mayormente presentes en la estación 3105) se presenta una mayor dependencia en la cola superior que en la cola inferior de la distribución.

Posteriormente a este análisis se realizaron algunas pruebas estadísticas para determinar las características generales de las cópulas. En el cuadro 4.2 se observan los resultados obtenidos representados en el valor  $p$  de cada prueba. En general se puede decir que:

- Independencia (Ind): hay mucha evidencia para rechazar la hipótesis de que las cópulas sean independientes.
- Intercambiabilidad (Int): existe poca evidencia para indicar que las cópulas no son intercambiables.
- Intercambiabilidad dado que es Extrema (IntEv): dado que las cópulas sean extremas existe muy poca evidencia para indicar que no sean intercambiables.
- Simetría Radial (SimR) : existe muy poca evidencia para indicar que las cópulas no sean simétricas.
- Valor Extremo (VE):
  - Para la estación 3103 hay poca evidencia indicando que alguna de las cópulas podría no ser extrema.
  - Para la estación 3104 , a excepción del caso de máximos de caudal pico totales, no hay suficiente evidencia de que la cópula no sea extrema.
  - Para la estación 3105 existe evidencia de que la cópula podría no ser extrema en el caso de máximos de caudal pico y volumen tanto total como excedente. Además, no se puede descartar la posibilidad de que la cópula de máximos de volumen excedente no sea extrema.

Es importante indicar que estos resultados deben de considerarse sólo como información adicional para la caracterización de la estructura de dependencia de las series. Por el corto registro y las potenciales fuentes de error mencionadas con anterioridad los resultados indicados en estos análisis no deben considerarse como definitivos. Además, debido a la gran cantidad de pruebas de hipótesis no necesariamente independientes que se realizan simultáneamente, la probabilidad de encontrar un falso positivo para algún nivel de significancia sencillamente por el azar se vuelve no despreciable como se indica en Bonferroni ([1936](#)).



**Cuadro 4.2:** Pruebas estadísticas para cada cópula.

Flujo	Muestra	id	Ind	Int	IntVE	SimR	VE
Tot.	(V, max{Q})	3103	0	1.000	0.508	0.647	0.907
Tot.	(V, max{Q})	3104	0	0.302	0.826	0.998	0.055
Tot.	(V, max{Q})	3105	0	0.996	0.750	0.999	0.304
Tot.	(max{V}, max{Q})	3103	0	0.902	0.264	0.955	0.670
Tot.	(max{V}, max{Q})	3104	0	0.686	0.850	0.993	0.158
Tot.	(max{V}, max{Q})	3105	0	0.904	0.587	0.550	0.003
Tot.	(max{V},Q)	3103	0	1.000	0.487	0.995	0.800
Tot.	(max{V},Q)	3104	0	0.869	0.699	0.999	0.198
Tot.	(max{V},Q)	3105	0	0.999	0.785	1.000	0.508
Exc.	(V, max{Q})	3103	0	0.673	0.867	0.460	0.572
Exc.	(V, max{Q})	3104	0	0.191	0.581	0.951	0.535
Exc.	(V, max{Q})	3105	0	0.568	0.997	1.000	0.453
Exc.	(max{V}, max{Q})	3103	0	0.634	0.629	0.999	0.474
Exc.	(max{V}, max{Q})	3104	0	0.161	0.842	0.997	0.654
Exc.	(max{V}, max{Q})	3105	0	0.479	0.997	0.408	0.001
Exc.	(max{V},Q)	3103	0	0.431	0.754	0.987	0.967
Exc.	(max{V},Q)	3104	0	0.843	0.772	0.996	0.601
Exc.	(max{V},Q)	3105	0	0.276	0.344	0.749	0.067

$H_0$  :

Ind: Independencia

Int: Intercambiabilidad

IntVE: Intercambiabilidad dado que cópula es extrema

SimR: Simetría Radial

EV: de Valor Extremo

Habiendo descrito las características principales de las cópulas, se busca información para saber si alguna familia en particular es poco apropiada para modelar la estructura de dependencia que estas presentan. Para ello se realizan pruebas de bondad de ajuste para tres familias de cópulas Gumbel, la t de valor extremo (tEV), Clayton, Normal y t. Debido al tamaño de la muestra es de esperarse que no haya suficiente

evidencia para rechazar las familias; en el cuadro 4.3 se muestra que solamente se puede rechazar mediante esta prueba a la cópula Clayton.

**Cuadro 4.3:** Pruebas de bondad de ajuste para familias principales de cópulas.

Flujo	Muestra	id	Gumbel	tEV	Clayton	Normal	t
Tot.	(V, max{Q})	3103	0.583	0.544	0.000	0.302	0.317
Tot.	(V, max{Q})	3104	0.201	0.151	0.002	0.600	0.236
Tot.	(V, max{Q})	3105	0.164	0.124	0.000	0.293	0.068
Tot.	(max{V}, max{Q})	3103	0.544	0.514	0.005	0.379	0.375
Tot.	(max{V}, max{Q})	3104	0.351	0.354	0.035	0.624	0.580
Tot.	(max{V}, max{Q})	3105	0.095	0.051	0.014	0.432	0.111
Tot.	(max{V},Q)	3103	0.631	0.649	0.004	0.334	0.425
Tot.	(max{V},Q)	3104	0.344	0.338	0.008	0.478	0.503
Tot.	(max{V},Q)	3105	0.112	0.092	0.000	0.189	0.050
Exc.	(V, max{Q})	3103	0.340	0.267	0.001	0.115	0.056
Exc.	(V, max{Q})	3104	0.170	0.151	0.000	0.082	0.040
Exc.	(V, max{Q})	3105	0.325	0.269	0.000	0.429	0.174
Exc.	(max{V}, max{Q})	3103	0.462	0.440	0.003	0.376	0.327
Exc.	(max{V}, max{Q})	3104	0.267	0.223	0.000	0.188	0.113
Exc.	(max{V}, max{Q})	3105	0.166	0.127	0.006	0.571	0.355
Exc.	(max{V},Q)	3103	0.591	0.497	0.002	0.280	0.198
Exc.	(max{V},Q)	3104	0.366	0.271	0.000	0.284	0.183
Exc.	(max{V},Q)	3105	0.286	0.244	0.032	0.584	0.537

Para poder medir, de las familias que no se pudieron rechazar, cuál de ellas modela mejor los datos disponibles se aplica una prueba de calidad de modelo  $\widehat{nxv}_n$  (según Grønneberg y Hjort (2014)). Los resultados obtenidos muestran que en general la estructura de dependencia de la estación 3103 está mejor representada por las cópula Gumbel y tEV mientras que para el caso de las estaciones 3104 y 3105 en general la cópula Normal y t obtienen mejores resultados. Este es un resultado realmente inesperado debido a que las series de datos provienen de máximos.

Al igual que en el caso de las pruebas de bondad de ajuste, es importante resaltar que la mayoría de resultados de calidad de modelo no difirieron mucho unos de otros

y que por el tamaño de la muestra algunos de los conjuntos de puntos atípicos y potenciales errores de las series podrían estar afectando el resultado.

Un potencial problema con la elección de la cópula normal para la descripción de la estructura de dependencia es que en la parte superior del régimen, que es la que más interesa, podría sobrestimar la probabilidad de no excedencia de algunos fenómenos debido a que  $\lambda_U = 0$  en el caso límite en la cópula normal mientras que para la cópula Gumbel  $\lambda_U = 2 - 2^{1/\theta}$ . Este resultado sería contraintuitivo puesto que físicamente se espera que crecientes con altos caudales pico contengan volúmenes altos al mismo tiempo. No son plausibles casos en los que tendiendo a valores muy altos no haya asociación entre estas variables.

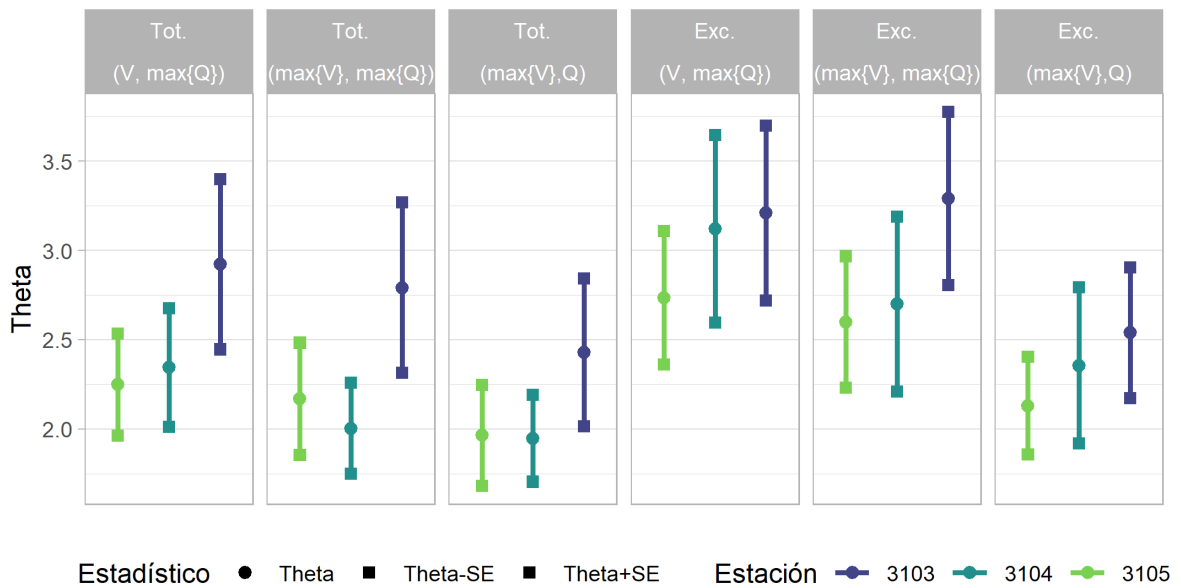
Por las razones anteriormente mencionadas y procurando escoger un modelo parsimonioso, se decide emplear la cópula Gumbel para todas las series analizadas. El análisis anterior pone en evidencia que si bien, los recursos de pruebas estadísticas y de calidad de modelo pueden guiar decisiones, también es necesario considerar otros aspectos no explícitamente incluidos en los datos disponibles.

**Cuadro 4.4:** Prueba de calidad de modelo  $n\widehat{xv}_n$  para familias principales de cópulas.

Flujo	Muestra	id	Gumbel	tEV	Normal	t
Tot.	(V, max{Q})	3103	32.0	30.6	27.4	28.2
Tot.	(V, max{Q})	3104	20.0	19.8	24.5	24.5
Tot.	(V, max{Q})	3105	19.1	19.1	20.8	20.8
Tot.	(max{V}, max{Q})	3103	30.5	30.3	29.2	29.1
Tot.	(max{V}, max{Q})	3104	14.3	13.2	15.8	16.0
Tot.	(max{V}, max{Q})	3105	17.8	18.3	23.1	23.1
Tot.	(max{V},Q)	3103	24.2	24.4	22.6	23.1
Tot.	(max{V},Q)	3104	13.3	11.1	13.7	14.7
Tot.	(max{V},Q)	3105	14.1	14.9	15.7	15.7
Exc.	(V, max{Q})	3103	35.5	35.8	36.2	36.2
Exc.	(V, max{Q})	3104	32.9	31.9	32.4	32.6
Exc.	(V, max{Q})	3105	27.6	27.7	30.1	30.1
Exc.	(max{V}, max{Q})	3103	35.6	35.4	35.4	37.3
Exc.	(max{V}, max{Q})	3104	27.7	27.7	28.4	28.4
Exc.	(max{V}, max{Q})	3105	25.7	25.6	31.6	31.6
Exc.	(max{V},Q)	3103	24.9	25.1	24.6	24.6
Exc.	(max{V},Q)	3104	22.6	23.4	22.3	22.3
Exc.	(max{V},Q)	3105	17.2	17.3	18.6	18.2

#### 4.4.2. Estimación de parámetros

Dado que se escoge la cópula Gumbel-Hougaard como la familia que se empleará para caracterizar la dependencia, se procede a emplear distintos métodos para la estimación del parámetro  $\hat{\theta}$  que define el modelo. Como se indicó anteriormente, se usarán como referencia los métodos asociados a  $\tau$ ,  $\rho_s$  y máxima pseudo-verosimilitud (MPL). En el cuadro 4.5 se muestran los valores de  $\hat{\theta}$  para cada uno de los métodos de estimación. Además, se muestra el error estándar para cada una de las estimaciones. Nótese que el error es relativamente bajo respecto a la magnitud del parámetro.



**Figura 4.12:** Comparación entre valores de  $\widehat{\theta}_{MPL}$

Distintas formas de estimación de los valores en general llevan resultados del parámetro y del error estándar similares entre sí. Además, como se observa en la 4.12, el patrón mostrado por dichos parámetros es afín a lo indicado previamente respecto al cuadro 4.1. Kojadinovic y Yan (2010) indican que, comparando los métodos de momentos ( $\rho_s$  o  $\tau$ ) y el método de MPL, el último presenta mejores resultados para un número limitado de datos por sus características asintóticas. Por ello, se emplearán los valores obtenidos a partir de esa metodología en los siguientes procedimientos.

**Cuadro 4.5:** Estimación de cópula Gumbel.

Flujo	Muestra	id	$\widehat{\theta}_{\text{MPL}}$	$\widehat{\theta}_{\tau}$	$\widehat{\theta}_{\rho}$	$\widehat{\sigma}_{\theta_{\text{MPL}}}$	$\widehat{\sigma}_{\theta_{\tau}}$	$\widehat{\sigma}_{\theta_{\rho}}$
Tot.	(V, max{Q})	3103	2.923	2.842	2.750	0.477	0.495	0.464
Tot.	(V, max{Q})	3104	2.344	2.584	2.599	0.331	0.369	0.402
Tot.	(V, max{Q})	3105	2.249	2.474	2.541	0.285	0.323	0.353
Tot.	(max{V}, max{Q})	3103	2.791	2.675	2.580	0.477	0.476	0.427
Tot.	(max{V}, max{Q})	3104	2.003	2.035	2.021	0.255	0.318	0.319
Tot.	(max{V}, max{Q})	3105	2.169	2.445	2.451	0.314	0.335	0.392
Tot.	(max{V},Q)	3103	2.429	2.255	2.230	0.414	0.375	0.344
Tot.	(max{V},Q)	3104	1.947	1.979	1.886	0.243	0.331	0.290
Tot.	(max{V},Q)	3105	1.965	2.173	2.280	0.283	0.250	0.269
Exc.	(V, max{Q})	3103	3.209	3.303	3.295	0.489	0.527	0.483
Exc.	(V, max{Q})	3104	3.121	3.293	3.167	0.525	0.571	0.529
Exc.	(V, max{Q})	3105	2.734	2.924	3.037	0.373	0.417	0.411
Exc.	(max{V}, max{Q})	3103	3.291	3.253	3.132	0.485	0.614	0.521
Exc.	(max{V}, max{Q})	3104	2.700	2.771	2.892	0.489	0.415	0.406
Exc.	(max{V}, max{Q})	3105	2.600	2.808	2.867	0.368	0.408	0.548
Exc.	(max{V},Q)	3103	2.539	2.500	2.482	0.366	0.404	0.361
Exc.	(max{V},Q)	3104	2.356	2.431	2.459	0.438	0.337	0.308
Exc.	(max{V},Q)	3105	2.130	2.168	2.140	0.272	0.340	0.352

### 4.4.3. Comprobación visual

Con la cópula modelada para cada una de las series de máximos es valioso comprobar visualmente el resultado obtenido. Una de las formas más fáciles y evidentes de realizar dicha comprobación es mediante un gráfico donde se superpongan las pseudo observaciones y una simulación de la cópula obtenida. En la figura 4.13 se muestran 1000 datos simulados. A partir de este resultado no es difícil apreciar que ambos conjuntos de datos muestran similitudes, una de las más importantes siendo una concentración marcada en la cola superior del registro.

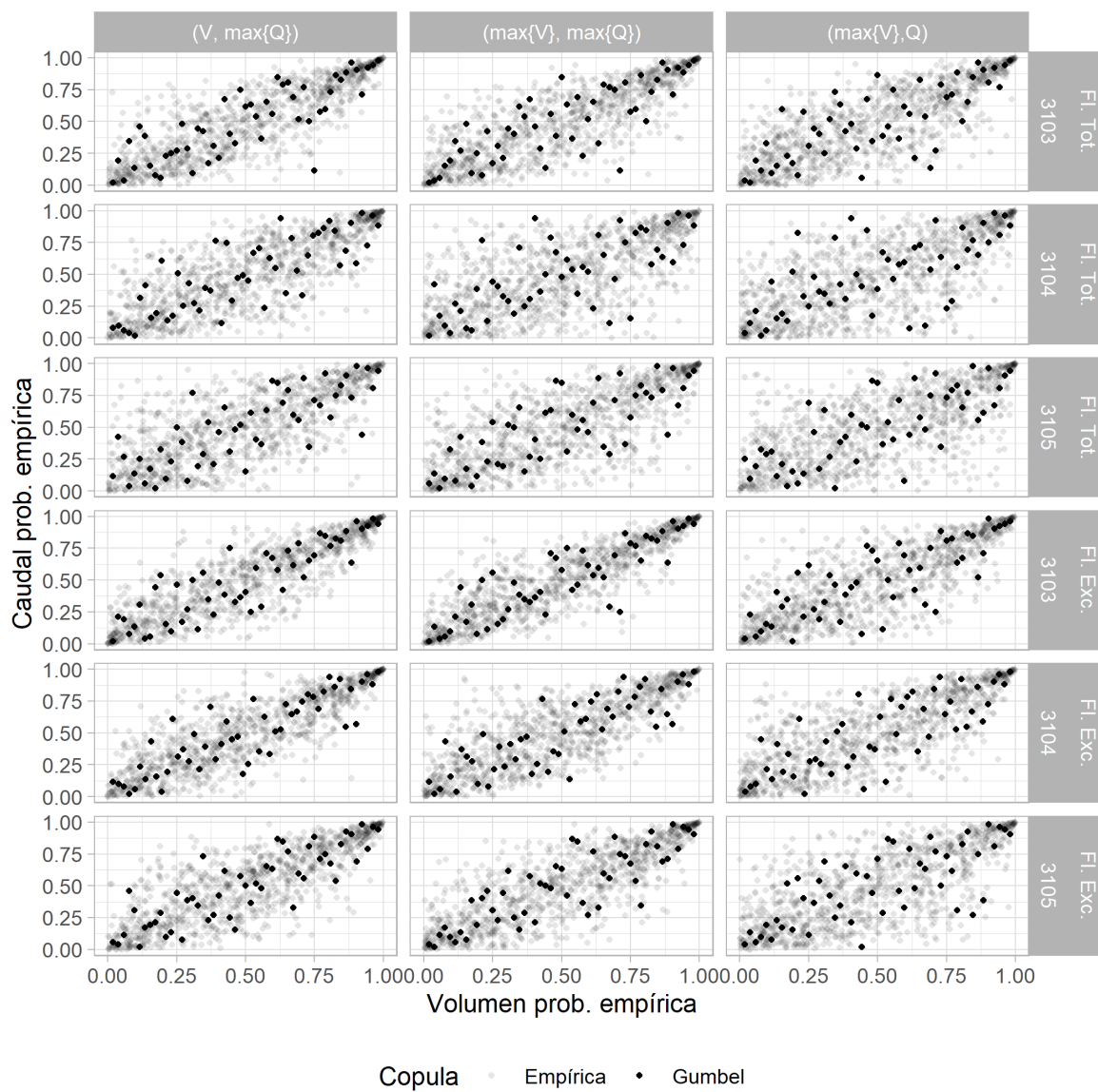
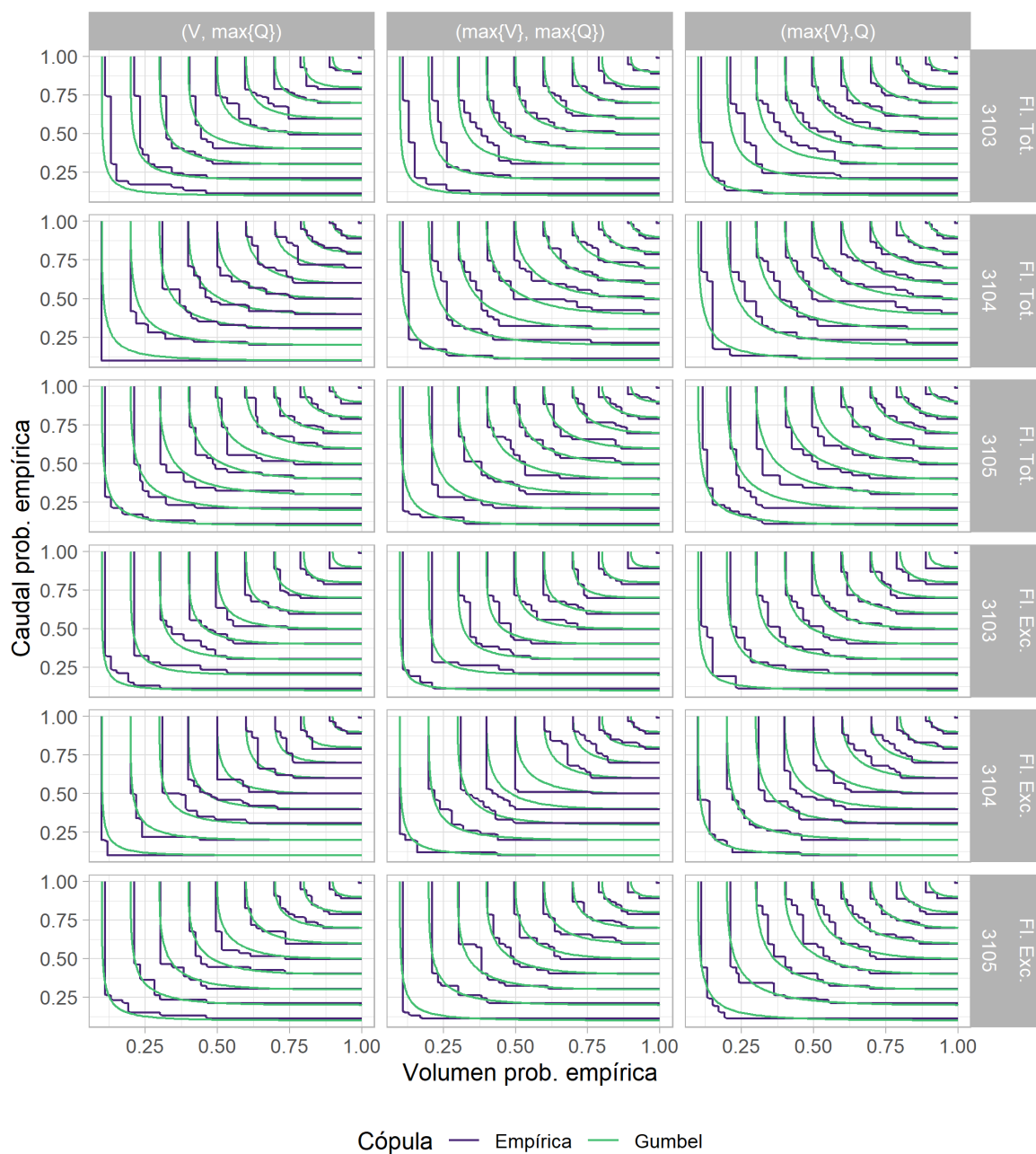


Figura 4.13: Comparación directa entre cópula estimada y empírica.

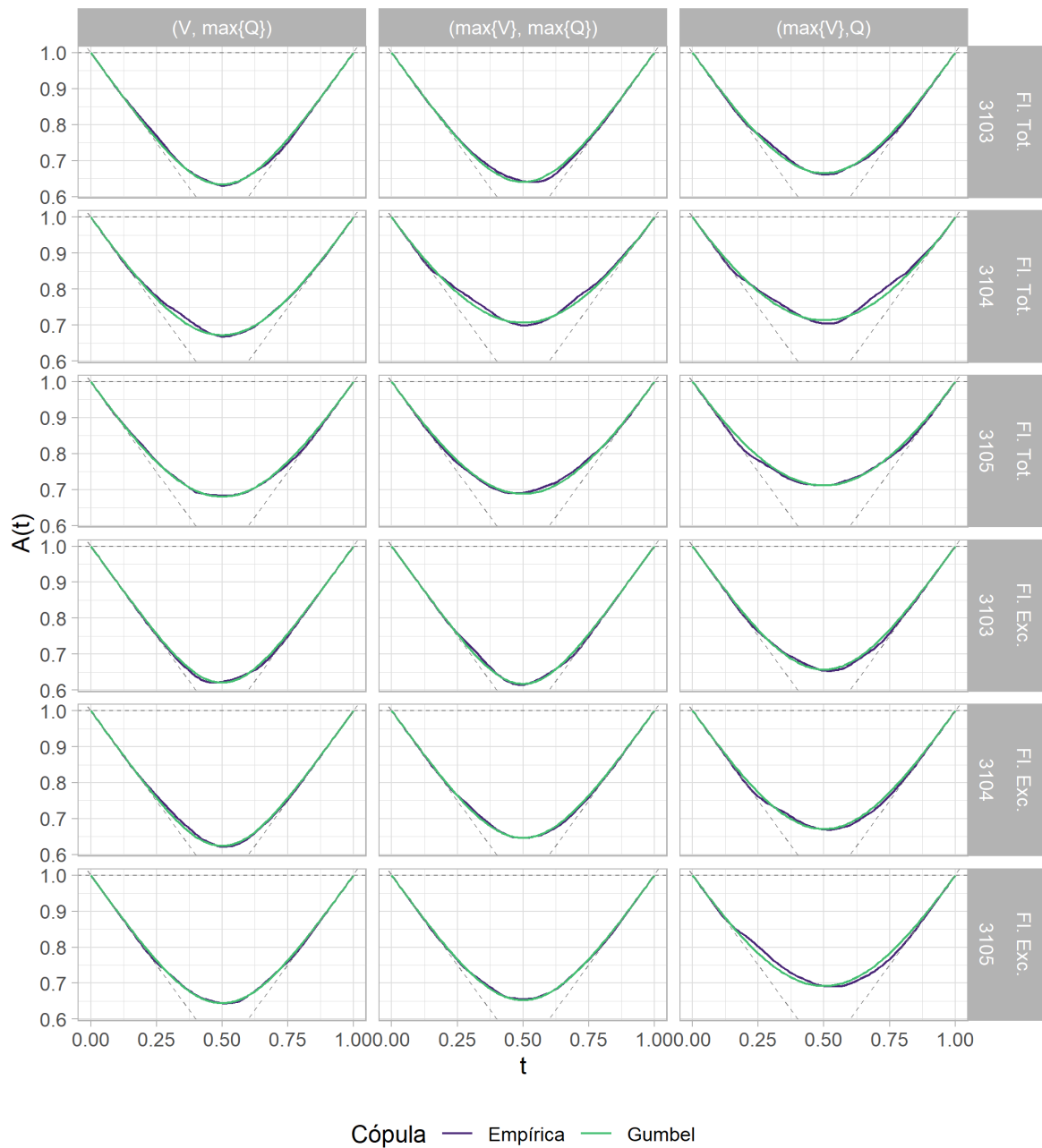
También se pueden generar contornos que muestren en qué zonas la cópula modelada difiere más de la cópula empírica. En la figura 4.14 se muestra que si bien en la zona media pueden darse discrepancias (que parcialmente podrían estar asociadas a las fuentes de error anteriormente mencionadas y al tamaño de la muestra) estas son menores y usualmente no se encuentran en todo el espacio analizado.

Finalmente, se comparan las funciones de Pickand de la cópula empírica y la modelada para todas las series analizadas. En la figura 4.15 se muestra, nuevamente que la cópula modelada sigue, con bastante cercanía a la cópula empírica. En este gráfico es más evidente que en el caso del flujo total para la estación 3104 existen diferencias entre las funciones (páneos (2, 1), (2, 2) y (2, 3)), lo mismo ocurre para volúmenes máximos de excedencia (pánel (6,3)). Este resultado es congruente con la información obtenida previamente en las pruebas estadísticas realizadas a dichos conjuntos de datos. En general, se considera que la copula modelada estima adecuadamente la estructura de dependencia mostrada por los datos.





**Figura 4.14:** Comparación de los contornos entre cópula empírica y la modelada.



**Figura 4.15:** Comparación de la función de Pickand de la cópula empírica y la modelada.

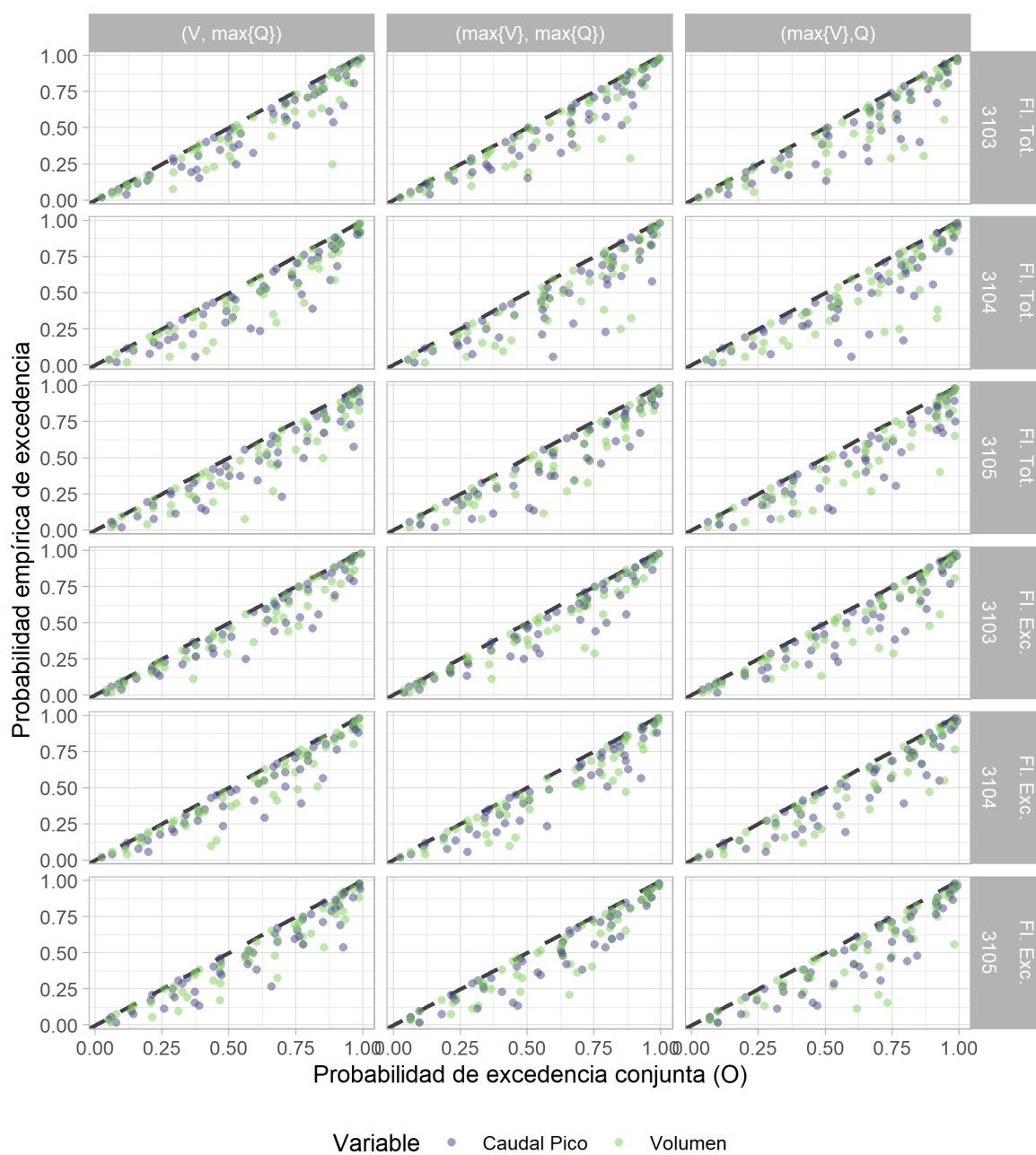
#### 4.4.4. Estimación de probabilidad de excedencia y comparación con análisis univariado

Al tener ya definidas la cópula para cada una de las series de datos, se puede estimar para cada punto la probabilidad de no excedencia conjunta  $P_{Q,V}^V$  (figura 4.16)

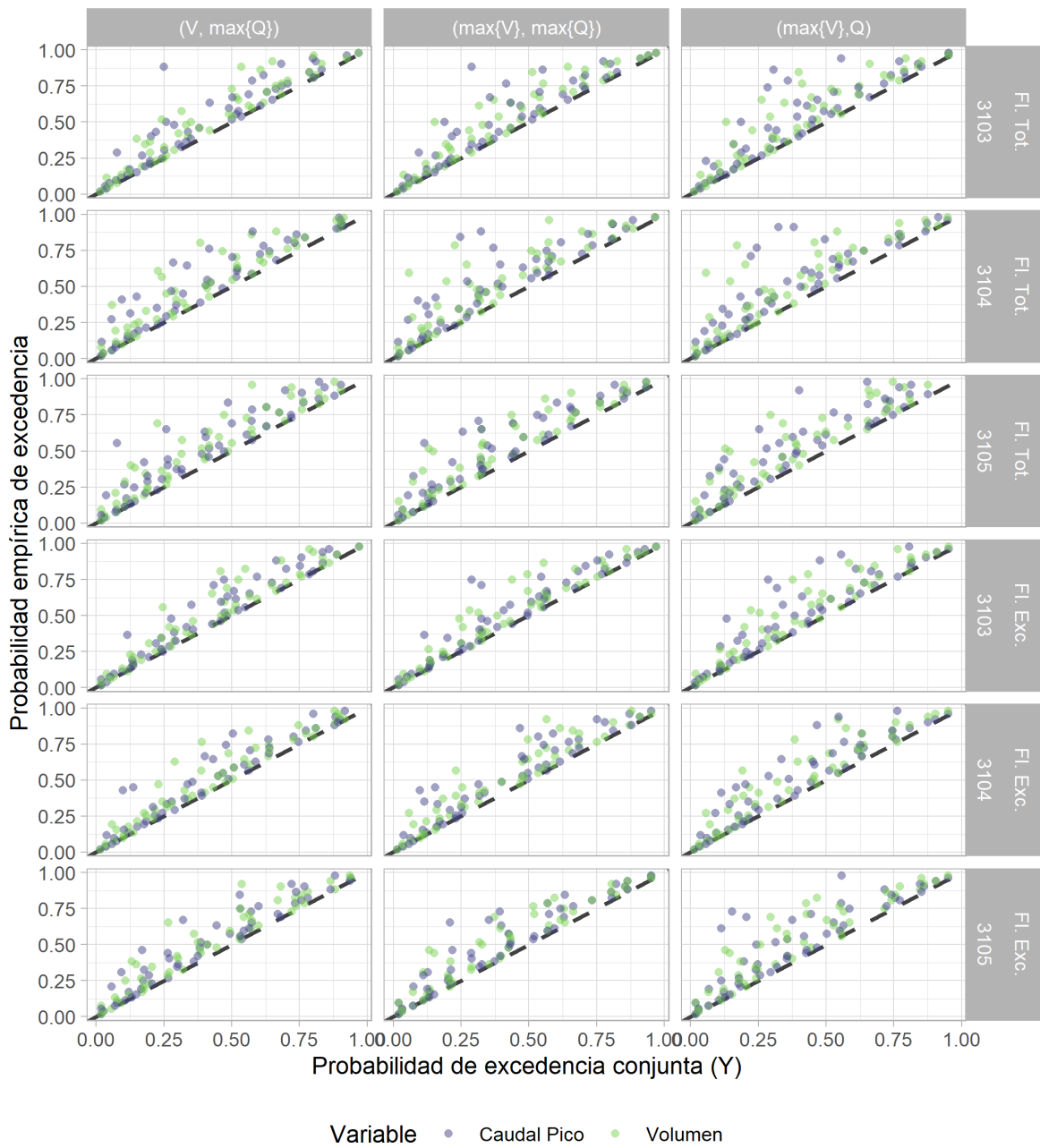
en la que se estima la probabilidad de que al menos uno de los componentes del fenómeno ocurra en exceso y  $P_{Q,V}^{\wedge}$  (figura 4.17) en la que se estima la probabilidad de que ambos componentes ocurran en exceso.

Lo que indica la figura 4.16 es que en general es menos probable que alguno de los valores esté por encima de un umbral (cada uno por aparte) que al menos uno de ellos se encuentre por encima de este. En el caso de la figura 4.17 lo que se indica es que es más probable que alguno de los valores esté por encima de un umbral que ambos lo estén. Lo anterior es verdadero por definición según indica Salvadori y De Michele (2004).

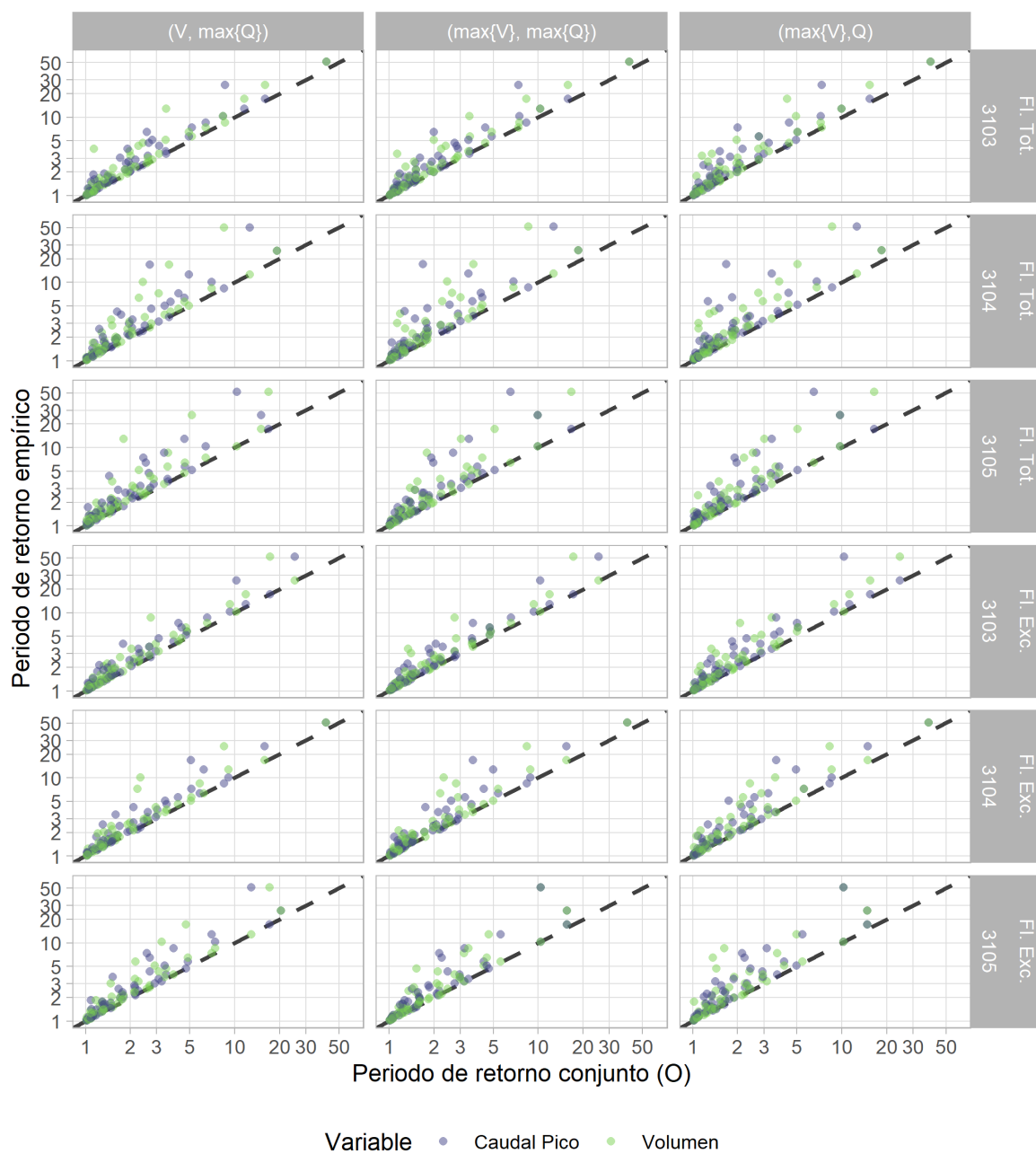
Si bien el concepto de período de retorno no necesariamente sería aplicable en un caso en el que los datos no son estacionarios, será empleado aquí para ilustrar las posibles consecuencias prácticas de esta diferencia en el cálculo. En el caso de  $P_{Q,V}^{\vee}$  las estimaciones univariadas tienden a asignarle un período de retorno mayor que las estimaciones multivariadas a la misma tormenta, es decir, se es **no** conservador al emplear el método univariado (figura 4.18). En el caso  $P_{Q,V}^{\wedge}$ , las estimaciones univariadas tienden a asignarle un período de retorno menor que mas multivariadas a la misma tormenta, es decir, se es conservador al emplear el método univariado (figura 4.19).



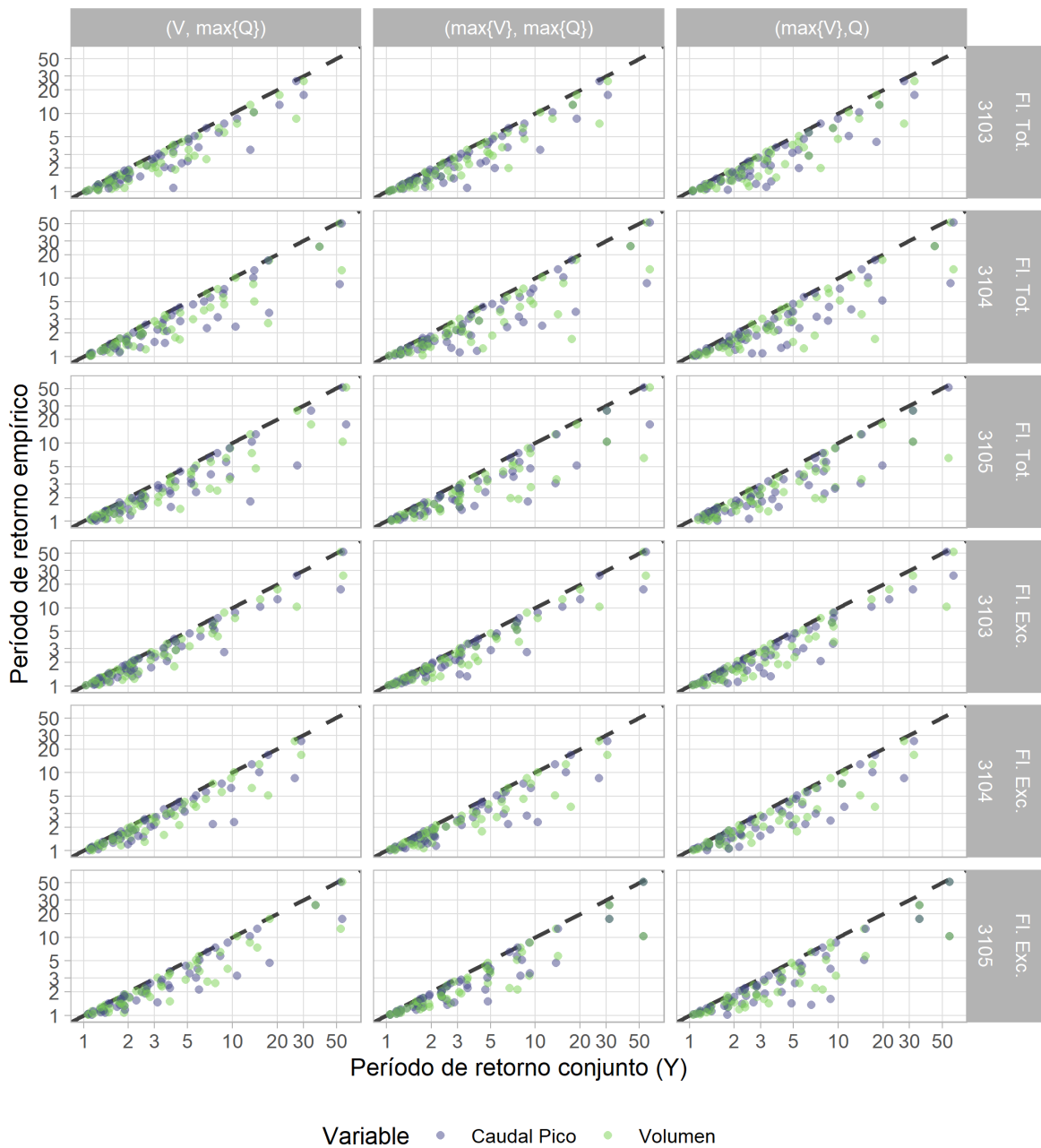
**Figura 4.16:** Comparación de  $1 - U_V$  y  $1 - U_Q$  con  $P_{Q,V}^V$ .



**Figura 4.17:** Comparación de  $1 - U_V$  y  $1 - U_Q$  con  $P_{Q,V}^\wedge$ .



**Figura 4.18:** Comparación de  $1/(1 - U_V)$  y  $1/(1 - U_Q)$  con  $1/P_{Q,V}^V$ .



**Figura 4.19:** Comparación de  $1/(1 - U_V)$  y  $1/(1 - U_Q)$  con  $1/P_{Q,V}^\wedge$ .

## 4.5. Análisis de sensibilidad

Hasta este apartado todos los análisis se han realizado empleando como supuesto que los parámetros de separación de flujo base y excedentes, y segmentación de

crecientes representan adecuadamente los fenómenos asociados al proceso de generación estocástico que se investiga.

Como se indicó en 4.2.1 algunos de estos valores fueron escogidos mediante verificación visual, por lo que son estimaciones gruesas. Adicionalmente, dado que no se tienen datos de referencia con los que realizar validación o verificación del método, no se puede explorar cuál podría ser el error asociado a la elección del método de separación de flujo, la elección del método de segmentación de crecientes; y la escogencia de los parámetros empleados en ellos.

Se pretende entonces evaluar cuál es el impacto en las series de máximos anuales y en la estimación de los parámetros de las cópulas para cada una de ellas mediante el método indicado en el apartado 2.12. Para ello, primeramente se generaron 10000 tuplas para los parámetros del modelo de separación de caudal base y segmentación de tormentas; es decir, se muestreó aleatoriamente del espacio generado por los intervalos  $\{\alpha \in [0.90, 0.99], w \in [3, 240], u_e \in [0.025, 0.250], u_b \in [0.0250, 0.250]\}$ . Para mayor explicación de estos parámetros se puede referir a los apartados 2.10 y 4.2.1.

Es importante indicar que en este análisis la distribución que supuso, por la carencia de información de campo para validación, que la distribución entre los parámetros evaluados es independiente; sin embargo, este supuesto no necesariamente es válido y con información de hidrología isotópica acerca del comportamiento de las crecientes de varias cuencas podría encontrarse que relación entre ellos no es independiente.

Luego, para cada una de estas tuplas se corrió el modelo de separación de flujo y de segmentación de tormentas para quedarse con las tormentas con caudal pico máximo anual o volumen máximo anual, ya sea excedente o total; adicionalmente se les calculó la duración a las tormentas escogidas.

Para cada tupla y estación se generaron las series:  $\{V_b, \text{máx}\{Q_b\}\}$ ,  $\{V_e, \text{máx}\{Q_e\}\}$ ,  $\{\text{máx}\{V_b\}, Q_b\}$ ,  $\{\text{máx}\{V_e\}, Q_e\}$  y con dichas series se calculó  $\rho_s$ ,  $\widehat{\theta}_{MPL}$ ,  $\widehat{\sigma}_{\theta_{MPL}}$  y  $n\widehat{xv}_n$ . Cada cada uno de esos objetos pueden interpretarse como una salida del modelo ( $o_i$ ) para configuraciones distintas de parámetros de separación de flujo y segmentación de eventos  $\alpha$ ,  $u_e$ ,  $u_v$  y  $w$ .

Tal y como indica Saltelli , *et al* (2008), para evaluar la sensibilidad de las salidas  $o_i$  a variaciones en estos parámetros se puede generar meta-modelo lineal múltiple donde estos serán tomados como predictores ( $X$ ) y las salidas  $o_i$  como variables de respuesta. Nótese que en el caso de las series de caudal, volumen o duración para



cada tupla de predictores se tienen 60 resultados (uno por cada año de registro). Para estos casos se busca determinar si los predictores tienen alguna influencia en el valor esperado de las series.

En el caso de los parámetros de las cópulas sólo se tiene un resultado para cada tupla, por lo que se desea encontrar cómo los predictores influyen las características de las cópulas.

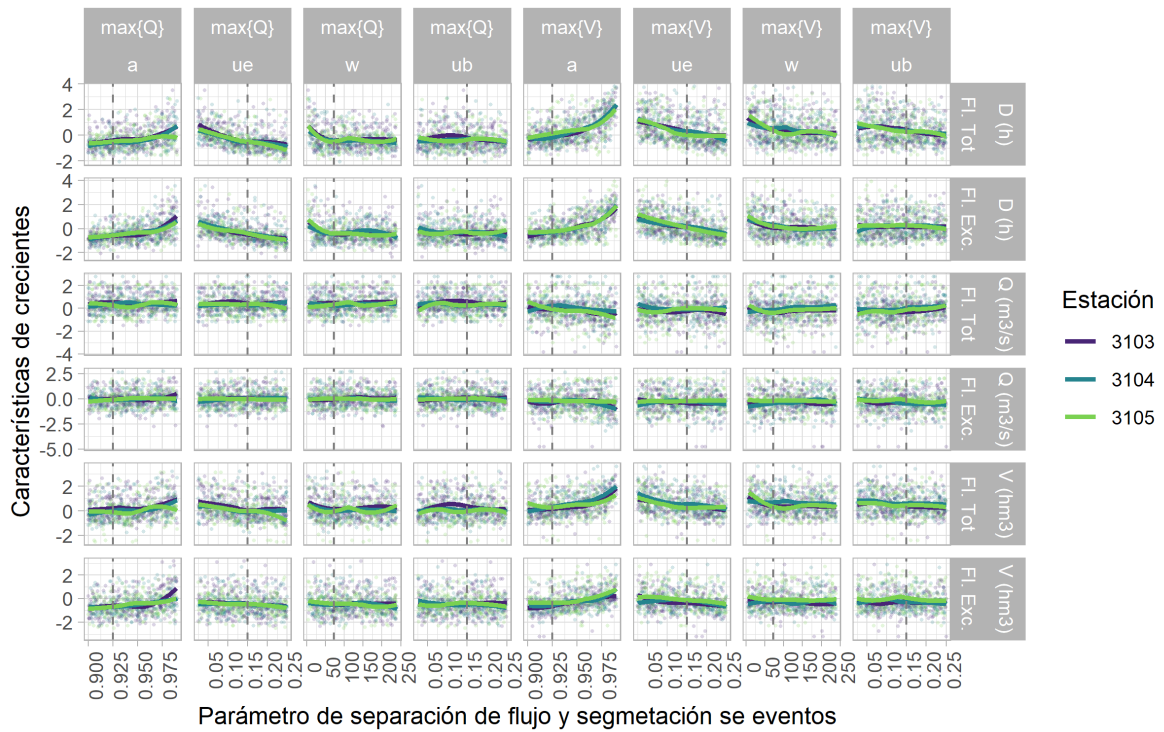
Es importante indicar que en estos meta-modelos se puede considerar no sólo el efecto principal, sino las interacciones de primer orden entre los predictores, es decir:

$$\hat{o}_i = \hat{\beta} \cdot X_i^T + \hat{\beta}_0 = (\hat{\beta}_1 \cdots \hat{\beta}_{10}) \cdot (\alpha w u_e u_b \alpha w \alpha u_e \alpha u_b w u_e w u_b u_e u_b)^T + \hat{\beta}_0 \quad (4.1)$$

Debido a que se están incluyendo los términos de interacciones de primer orden, y para garantizar que no se estén incluyendo variables irrelevantes en el análisis, se emplea para cada uno de los modelos un proceso de selección de variables hacia atrás según se indica en Venables y Ripley (2002). Esto implica que no necesariamente todas las variables se encuentran presentes para todos los meta-modelos, pero que las variables que se encuentran aportan a la explicación de la varianza de la variable estimada al menos con los datos disponibles y los supuestos estructurales asociados a modelos lineales múltiples.

Se realizaron algunos ajustes a los valores de las variables de respuesta  $o_i$  para facilitar su modelado e interpretación de los coeficientes del modelo  $\beta$ . Primero debido a su asimetría a los valores caudal, volumen y duración, se les realizó una transformación logarítmica ( $o'_i = \log(o_i)$ ). Luego, se estandarizaron los valores de las respuestas  $o_i$  separadamente por cada estación.

En la figura 4.20 se visualiza en el eje vertical el valor transformado de las características de las tormentas (e.g. caudal pico ( $Q$ ), volumen ( $V$ ), duración ( $D$ )) y en el eje horizontal el valor de distintos parámetros de separación de flujo y segmentación de crecientes. Además, se representa separadamente las series de flujo excedente (Fl. Exc.) y flujo total (Fl. Tot.). En diferentes colores se observan en el mismo panel los valores asociados a las distintas estaciones analizadas. Se agrega sólo para efectos de visualización una representación suavizada del valor central de las características de las tormentas.

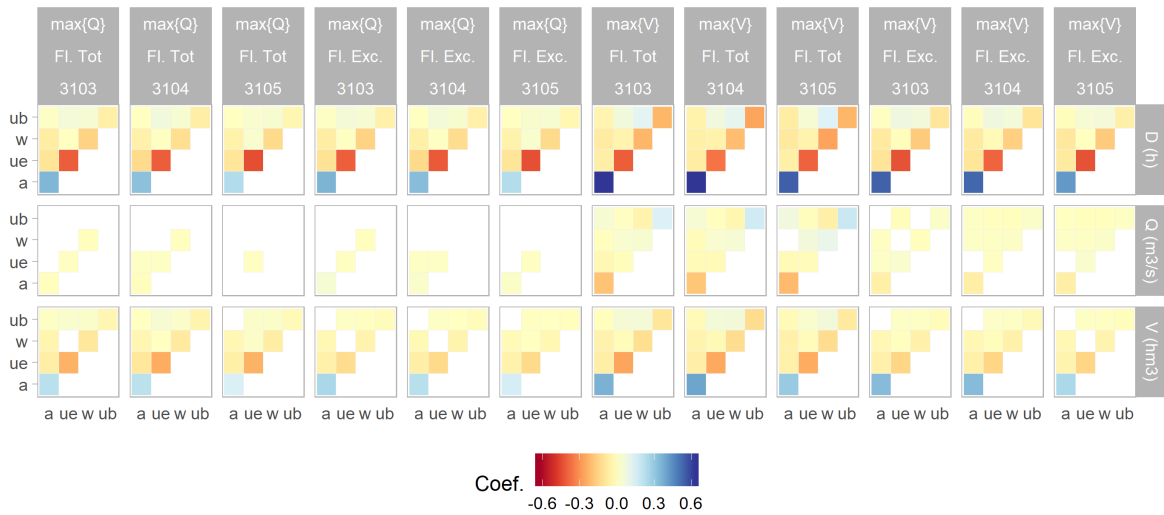


**Figura 4.20:** Impacto en series de máximos asociado a variación en parámetros de separación de flujo y segmentación se crecientes

Es interesante observar que en general el efecto de la variación de los parámetros es muy similar entre las estaciones. Se aprecia que la afectación en el caudal pico es relativamente baja respecto a la varianza asociada a la serie mientras que, en el caso de la duración y el volumen se observa una variación significativa especialmente en las series  $\{\text{máx}\{V_b\}, Q_b\}$ ,  $\{\text{máx}\{V_e\}, Q_e\}$ . Claramente la duración y el volumen están por definición asociados por lo que no es de sorprenderse que ambas variables tengan el mismo comportamiento respecto a la variación de los predictores.

En general, conforme  $\{\alpha \uparrow, u_e \downarrow, u_b \downarrow, w \downarrow\}$ , mayores son las duraciones y volúmenes. Estos resultados coinciden con las intuiciones previas sobre el funcionamiento del modelo según se indica en el apartado 4.2.1. Obsérvese que en el caso de  $w$  el efecto parece disminuir para valores mayores que cierto umbral, lo cual coincide con la intuición de que cuando el valor es suficientemente grande, permite que el pico de la tormenta se incluya dentro de la ventana para toda la duración de la tormenta. Si bien los efectos en las series no es necesariamente lineal, este aparentemente muestra un comportamiento aproximadamente monótonico.

Al aplicar el modelo lineal refuerza la validez de las apreciaciones anteriormente indicadas. Además, se observa que en general para los términos de interacciones la magnitud y la significancia no son tan altos relativamente, es decir, los efectos principales explican la mayor parte de la varianza como se muestra en las figura 4.21.

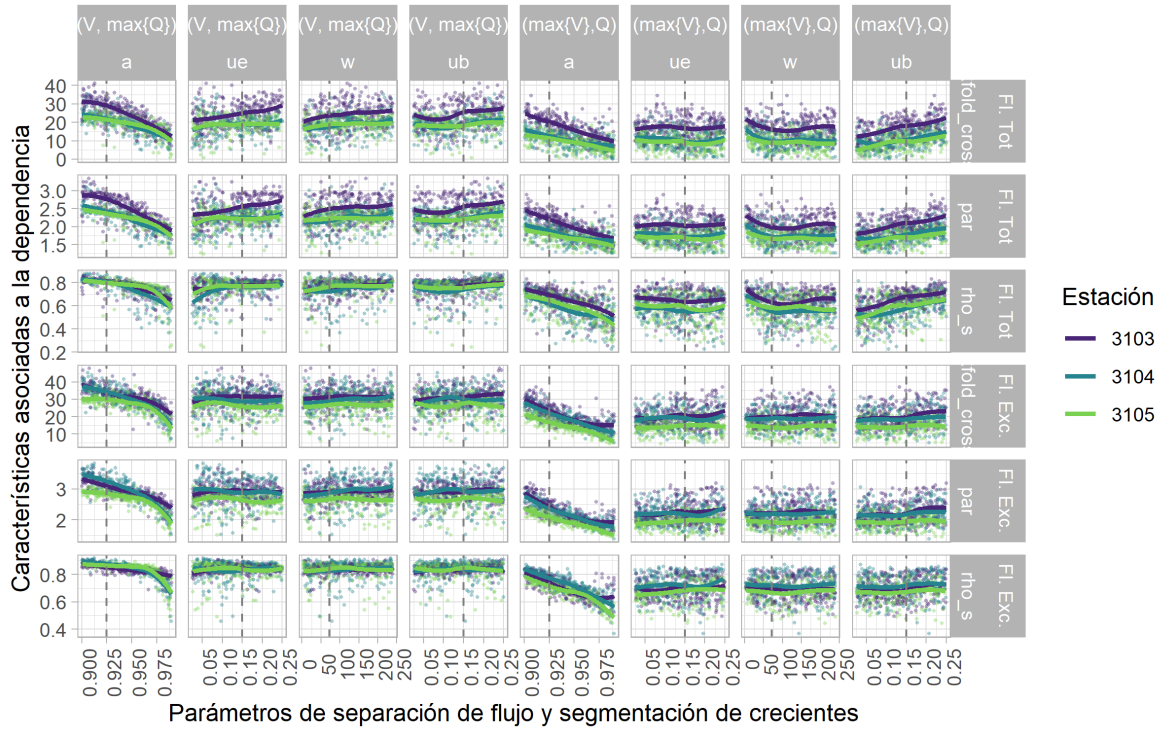


**Figura 4.21:** Impacto en series de máximos asociado a variación en parámetros de separación de flujo y segmentación se crecientes empleando modelos lineales. Flujo Base.

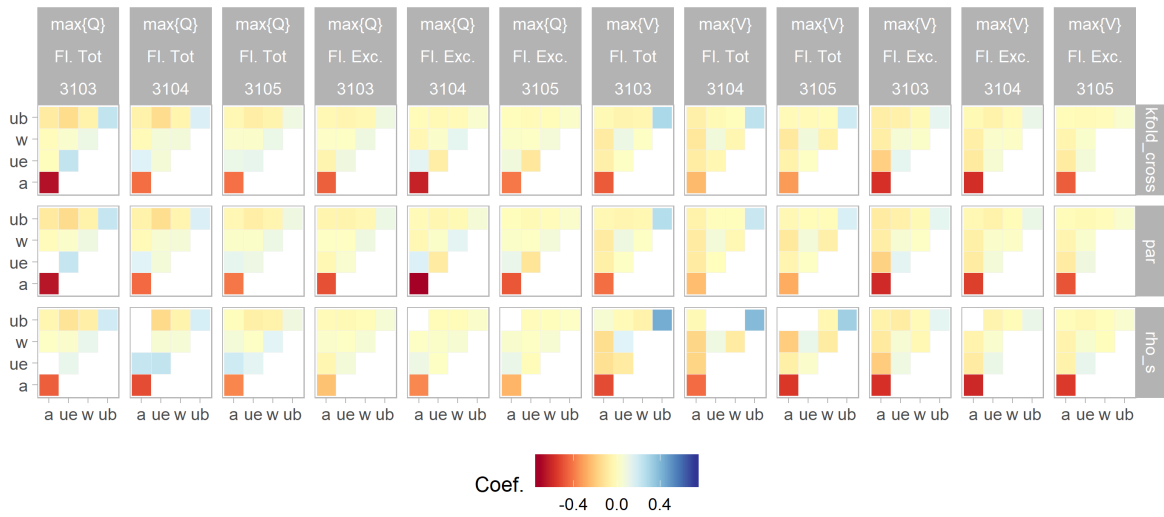
Posteriormente, se realiza el mismo procedimiento anteriormente mencionado para  $\rho_s$ ,  $\widehat{\theta}_{MPL}$ , y  $n\widehat{xv}_n$ , con la excepción de que la estandarización se realiza agrupando únicamente por estación y no se les aplicó una transformación logarítmica. Los resultados de los efectos de cada una de las variables se muestra en la figura 4.22.

En ella se observan efectos muy significativos en las mediciones paramétricas y no paramétricas de dependencia así como en la medida de calidad del modelo. En general, no es tan evidente cuál es el efecto de las demás variables, sin embargo, es claro que si  $\{\alpha\}$  la dependencia encontrada entre caudal pico y volumen de las crecientes disminuye al igual que la calidad de los modelos.

Los mismos resultados indicados anteriormente se obtienen cuando se emplean modelos lineales. En general, el efecto de  $\alpha$  es mucho mayor que el de las demás variables. Después,  $u_b$  muestra un efecto relevante y opuesto a  $\alpha$ , donde a mayores valores se encuentra que la dependencia es mayor. Cabe resaltar que existen interacciones que muestran mayor relevancia que algunos efectos principales, estas interacciones están a su vez asociadas principalmente a  $\alpha$  y  $u_b$



**Figura 4.22:** Impacto en estructura de dependencia de la variación en parámetros de separación de flujo y segmentación se crecientes



**Figura 4.23:** Impacto en estructura de dependencia de la variación en parámetros de separación de flujo y segmentación se crecientes empleando modelos lineales

Los resultados de este análisis implican que muchos de los resultados a los que se llegan en los apartados anteriores hay que tomarlos con cautela, ya que la selección de los parámetros del modelo de separación de flujo y segmentación de tormentas los puede afectar significativamente.

Por ejemplo, una de las decisiones acerca de los parámetros de dichos modelos fue tomar los mismos valores para las tres estaciones de medición. Es complejo identificar qué parte de las similitudes entre los resultados de las estaciones está relacionado específicamente con esta decisión y que parte está relacionada con una diferencia física de las cuencas.

## 4.6. Discusión general de resultados

Debido a los numerosos ejemplos de aplicación de cópulas en análisis de crecientes, no es de sorprenderse que se hayan podido modelar la relación de dependencia entre el caudal pico y volumen en las estaciones estudiadas. Dadas las facilidades que ofrecen ecosistemas como el desarrollado alrededor de R, no es obvio inicialmente el por qué estos métodos no han sido empleados de forma más generalizada; sin embargo, luego de enfrentarse a los retos de obtener esos *datos requeridos* se considera que es posiblemente este el principal obstáculo con el que se encuentran los profesionales o investigadores interesados en reproducir el procedimiento.

Desde el apartado de análisis exploratorio hasta el de análisis de sensibilidad es claro que la consideración de las fuentes de error es muy compleja, en especial cuando se desconoce el proceso específico de control de calidad que llevaron los datos. Más aún cuando en dicho proceso no se han rellenado eventos hidro-meteorológicos importantes que podrían conllevar a resultados disímiles en los análisis de extremos.

Estos errores no sólo podrían llevar a estimar inadecuadamente los parámetros de las distribuciones con las que se realiza inferencia, sino también podrían llevar a escoger metodologías que no necesariamente son válidas para el análisis o bien a escoger familias de distribuciones o cópulas erradas. Es decir, estos errores se transmiten en el proceso de investigación a través de errores estructurales y no estructurales en las estimaciones. Es en parte por esta razón que se decidió no incluir análisis posteriores para el caso de las distribuciones marginales.

También, asociado a la importancia de los datos de entrada, es sorprendente no encontrar literatura de cópulas que indique cuál es la afectación directa de la separación

de los eventos en la dependencia entre caudal pico, volumen y duración. Se presume que algunos de estos análisis o muchos de ellos parten no de la serie de datos original, sino de las series de máximos. Si bien en Nikoletta (2017) se menciona la metodología que se empleó no se es claro cómo se validó el parámetro del filtro para las cuencas investigadas, ni qué procedimientos adicionales se emplearon para la obtención de las crecientes. Nótese que este es un elemento muy relevante en el análisis dado que, al menos con la metodología de separación de flujo y segmentación empleada en este estudio, los resultados del análisis de sensibilidad son claros en identificar que definitivamente  $\alpha$  está inversamente asociada a la dependencia entre el caudal pico y el volumen de las crecientes.

La importancia del proceso de segmentación de la series de tiempo de caudal horario y la incertidumbre asociada al mismo hace evidente la necesidad de que se desarrolle más investigación en esa línea que incluya un análisis de sensibilidad en estimación de parámetros de distribuciones de valores extremos univariadas y multivariadas.

Una idea preliminar que podría aplicarse es emplear información de concentración de trazadores o isótopos en conjunto con datos de precipitación y caudal para generar una base de datos de crecientes segmentadas y con el flujo base separado. Con esta base de datos se podría entonces probar cuáles métodos son mejores estimando la segmentación con información parcial o reducida. Para ello se podrían emplear los algoritmos indicados en el apartado 2.10, el empleado en este documento y otros como la Detección Bayesiana de puntos de Cambio (Barry y Hartigan (1993)). Es necesario también explorar los métodos de segmentación de tormentas de precipitación ya que algoritmos como el de tiempos entre arribos (Chin , *et al* (2016)) implica la elección subjetiva de umbrales. Esto hace que en general la definición del evento como unidad de análisis sea, cuanto mucho, difusa.

Queda pendiente como una línea a investigar, la evaluación de la posibilidad de que existan distintos procesos estocásticos de producción de eventos extremos actuando simultáneamente en la misma cuenca. Si ese fuese el caso, el estudio de mezclas de distribuciones de valores extremos y la relación de estas familias con señales sinópticas podría ser de gran utilidad.

Se considera que el uso de métodos Bayesianos en la inferencia de las propiedades de estos procesos podría ayudar a relajar algunos de los supuestos requeridos por los métodos clásicos. Esto podría no sólo mejorar la robustez en las estimaciones, sino también la cuantificación de la incertidumbre asociada a estas.

Algunas de las extensiones más obvias de este trabajo podrían ser el desarrollo de cópulas trivariadas (incluyendo la duración), emplear precipitación en el análisis y determinar cuál es la estructura de dependencia de valores extremos a lo largo de una cuenca como en el caso de confluencias. Si bien todos estos temas tienen un valor científico considerable es importante conocer si es posible llevar este tipo de análisis a la práctica y si ello tiene sentido. Por ello sería valioso también determinar qué tanto difiere un análisis de cópulas para la creación de tormentas de diseño con un análisis de modelado continuo de la estructura hidráulica que se desea evaluar.

Un ejemplo de la idea anteriormente planteada, sería transitar el registro completo de caudal a través de una presa y realizar un análisis univariado de valor extremo con base en la serie transformada. También, se podría transitar una serie de tormentas sintéticas generadas con base en cópulas para cierta probabilidad de no excedencia. Luego se podría comparar la diferencia entre los resultados obtenidos con ambas metodologías y la complejidad de ejecutarlas (e.g. costo computacional, facilidad para el usuario etc.).

Es fundamental no perder de perspectiva que, si bien los datos de los que se dispuso tienen áreas de oportunidad, es muy rara la ocasión en que del todo dichos registros existen. El esfuerzo requerido para la obtención, procesamiento almacenado y curado de dicha información es titánico. La labor que ha realizado y sigue realizando el ICE en este aspecto es fundamental para el desarrollo de la hidrología del país. Es por ello que se considera que estos datos se encuentren cada vez más disponibles a investigadores y a los practicantes de la hidrología. Esto podría facilitar enormemente el proceso de depuración de la información y permitiría transmitir el valor que ella contiene a la sociedad costarricense que tanto lo necesita.

# Capítulo 5

## Cierre

### 5.1. Conclusiones

- Las series de datos disponibles presentan vacíos y potenciales errores que se propagan a través de todo el análisis.
- El proceso de separación de flujo base y la segmentación de las crecientes puede influir significativamente en las estimaciones de las series de valores de caudal y volumen así como en la estructura de dependencia que estas presentan.
- Existe evidencia de que podrían estar actuando simultáneamente en las cuencas estudiadas distintos procesos de generación de eventos extremos.
- Se encontró una gran similitud en la estructura de dependencia de caudal pico y volumen en las crecientes de las tres estaciones analizadas en la cuenca del río General; sin embargo, en general se observó mayor dependencia y consistencia en el caso de la estación 3103.
- En general la serie de valores de flujo escedente muestra mayor dependencia que la serie de valores de flujo total.
- Se considera que la cópula Gumbel-Hougaard muestra características en general compatibles con la estructura de dependencia de la serie de datos de caudal pico y volumen de las crecientes evaluada.



- El modelo de cópulas para el análisis de la relación de dependencia entre el caudal pico y el volumen de crecientes es una opción viable y rápida siempre y cuando se tengan los datos necesarios para su ajuste.
- Los valores estimados de las probabilidades de excedencia puede diferir significativamente entre el caso univariado y el multivariado.

## 5.2. Recomendaciones

- Para las cuencas de interés construir una base de datos con las características necesarias para realizar análisis univariados y multivariados de eventos extremos.
  - Realizar una revisión exhaustiva de las curvas de gasto para eliminar inconsistencias y cuantificar la incertidumbre asociada a la estimación de caudal para toda la serie de datos disponible.
  - Rellenar los períodos sin registro de eventos meteorológicos extremos mediante el uso de modelos hidrológicos, o marcas de creciente en conjunto con modelos hidráulicos si fuera posible.
  - Medir mediante una metodología apropiada, como trazadores químicos, las características de la reacción de la cuenca ante tormentas.
  - Generar, con la ayuda de especialistas en hidrología y meteorología, una base de datos de tormentas y sus correspondientes crecientes segmentadas apropiadamente en el tiempo y con la separación de flujo base y flujo excedente.
  - Generar, con la ayuda de especialistas en meteorología e hidrología, una base de datos de sequías.
- Calibrar y validar diferentes algoritmos para obtener eventos hidrológicos y sus características de forma automatizada.
  - Desarrollar una definición estándar de evento de precipitación y sequía según las características meteorológicas e hidrológicas de la cuenca analizada.
  - Con los datos de trazadores químicos y base de datos de eventos calibrar distintos métodos de separación de caudal base y segmentación.

- Identificar la calidad de ajuste de de estos métodos dentro y fuera de muestra, para determinar la validez de su uso en cuencas menos estudiadas.
- Realizar un análisis univariado de eventos extremos en cada una de las marginales de interés considerando las particularidades de las series de valores extremos y empleando distintas metodologías.
  - Considerar picos por encima de un umbral como metodología de muestreo de la cola de la distribución.
  - Investigar las diferencias en distribución de crecientes asociadas a diferentes mecanismos de generación de eventos de precipitación extremos.
  - Valorar el uso de mezclas de distribuciones de valor extremo extremo para la obtención de marginales sometidas a diferentes mecanismos.
  - Investigar el efecto de cambios antrópicos, variabilidad climática y cambio climático en el comportamiento de las series de valores extremos.
  - Analizar el comportamiento de eventos mínimos de caudal y sequías.
- Ampliar el análisis de las distribuciones conjuntas de probabilidad de eventos extremos.
  - Evaluar el uso de cópulas para el análisis de sequías con marginales como la duración y la severidad.
  - Analizar la distribución conjunta de crecientes en convergencias de cuerpos de agua.
  - Modelar la relación de dependencia de caudal y duración, y volumen y duración en crecientes.
  - Evaluar el uso otra familias de cópulas para el ajuste de la estructura de dependencia.
  - Modelar la relación de dependencia en eventos de precipitación empleando variables como intensidad máxima y duración.
  - Implementar el análisis trivariado de cópulas para la generación de tormentas de diseño.

## **Capítulo 6**

### **Referencias bibliográficas**

## Artículos

- Barry, D. & Hartigan, J. A. (1993). A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88(421), 309-319.
- Blume, T., Zehe, E. & Bronstert, A. (2007). Rainfall—runoff response, event-based runoff coefficients and hydrograph separation. *Hydrological Sciences Journal*, 52(5), 843-862.
- Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 8, 3-62.
- Brunner, M. I., Seibert, J. & Favre, A.-C. (2016). Bivariate return periods and their importance for flood peak and volume estimation. *Wiley Interdisciplinary Reviews: Water*, 3(6), 819-833.
- Chapman, T. (1999). A comparison of algorithms for stream flow recession and baseflow separation. *Hydrological Processes*, 13(5), 701-714.
- Chin, R. J., Lai, S. H., Chang, K. B., Jaafar, W. Z. W. & Othman, F. (2016). Relationship between minimum inter-event time and the number of rainfall events in Peninsular Malaysia. *Weather*, 71(9), 213-218.
- De Michele, C., Salvadori, G., Canossi, M., Petaccia, A. & Rosso, R. (2005). Bivariate statistical approach to check adequacy of dam spillway. *Journal of Hydrologic Engineering*, 10(1), 50-57.
- Eckhardt, K. (2012). Analytical sensitivity analysis of a two parameter recursive digital baseflow separation filter. *Hydrology and Earth System Sciences*, 16(2), 451-455.
- Eckhardt, K. (2005). How to construct recursive digital filters for baseflow separation. *Hydrological Processes: An International Journal*, 19(2), 507-515.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiémondge, N. & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water resources research*, 40(1).

- Foks, S. S., Raffensperger, J. P., Penn, C. A. & Driscoll, J. M. (2019). Estimation of base flow by optimal hydrograph separation for the conterminous United States and implications for national-extent hydrologic models. *Water*, 11(8), 1629.
- Genest, C. & Favre, A.-C. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, 12(4), 347-368. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:4\(347\)](https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347))
- Genest, C. & Nešlehová, J. G. (2014). On tests of radial symmetry for bivariate copulas. *Statistical Papers*, 55(4), 1107-1119.
- Genest, C., Nešlehová, J. & Quessy, J.-F. (2012). Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics*, 64(4), 811-834.
- Genest, C., Remillard, B. & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2), 199-213. <https://EconPapers.repec.org/RePEc:eee:insuma:v:44:y:2009:i:2:p:199-213>
- Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3).
- Grønneberg, S. & Hjort, N. L. (2014). The Copula Information Criteria. *Scandinavian Journal of Statistics*, 41(2), <https://onlinelibrary.wiley.com/doi/pdf/10.1111/sjos.12042>, 436-459. <https://doi.org/10.1111/sjos.12042>
- Kojadinovic, I., Segers, J. & Yan, J. (2011). Large-sample tests of extreme-value dependence for multivariate copulas. *Canadian Journal of Statistics*, 39(4), 703-720.
- Kojadinovic, I. & Yan, J. (2010). Nonparametric rank-based tests of bivariate extreme-value dependence. *Journal of Multivariate Analysis*, 101(9), 2234-2249.
- Kojadinovic, I. & Yan, J. (2012). A Non-parametric Test of Exchangeability for Extreme-Value and Left-Tail Decreasing Bivariate Copulas. *Scandinavian Journal of Statistics*, 39(3), 480-496.
- Ladson, T. (2013). A standard approach to baseflow separation using the Lyne and Hollick filter. *Australian Journal of Water Resources*.
- Mei, Y. & Anagnostou, E. N. (2015). A hydrograph separation method based on information from rainfall and runoff records. *Journal of Hydrology*, 523, 636-649.
- Oppel, H. & Mewes, B. (2020). On the automation of flood event separation from continuous time series.
- Salvadori, G. & De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water resources research*, 40(12).

- Serinaldi, F. (2008). Analysis of inter-gauge dependence by Kendall's  $\tau$  K, upper tail dependence coefficient, and 2-copulas with application to rainfall fields. *Stochastic Environmental Research and Risk Assessment*, 22(6), 671-688.
- Shao, G., Zhang, D., Guan, Y., Sadat, M. A. & Huang, F. (2020). Application of different separation methods to investigate the baseflow characteristics of a Semi-Arid Sandy Area, Northwestern China. *Water*, 12(2), 434.
- Sklar, A. (1959). Fonctions de Répartition à n Dimensions et Leurs Marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8.
- Steinbakk, G. H., Thorarinsdottir, T. L., Reitan, T., Schlichting, L., Hølleland, S. & Engeland, K. (2016). Propagation of rating curve uncertainty and design flood estimation. *Water Resources Research*, 52(9), 6897-6915.
- Tan, S. B., Lo, E. Y.-M., Shuy, E. B., Chua, L. H. & Lim, W. H. (2009). Hydrograph separation and development of empirical relationships using single-parameter digital filters. *Journal of Hydrologic Engineering*, 14(3), 271-279.
- Thiesen, S., Darscheid, P. & Ehret, U. (2019). Identifying rainfall-runoff events in discharge time series: a data-driven method based on information theory. *Hydrology and Earth System Sciences*, 23(2), 1015-1034.
- Wen, Y., Yang, A., Kong, X. & Su, Y. (2022). A Bayesian-Model-Averaging Copula Method for Bivariate Hydrologic Correlation Analysis. *Frontiers in Environmental Science*, 635.

## Libros

- Blanco, L., Arunachalam, V. & Dharmaj, D. (2012). *Introduction to Probability and Stochastic Processes with Applications*. Springer Berlin Heidelberg.
- Chen, L. & Guo, S. (2018). *Copulas and Its Application in Hydrology and Water Resources*. Springer. <https://books.google.co.cr/books?id=gIJiDwAAQBAJ>
- Chow, V. T., Maidment, D. R. & Mays, L. W. (1996). *Hidrología aplicada*. McGraw-Hill.
- Chow, V. T., Maidment, D. & Mays, L. (1988). *Applied Hydrology* (1.<sup>a</sup> ed.). McGraw-Hill.
- Hamed, K. & Rao, A. R. (1999). *Flood frequency analysis*. CRC press.
- Hofert, M. (2018). *Elements of Copula Modeling with R*. Springer.
- Hosking, J. R. M. & Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511529443>
- Joe, H. (2014). *Dependence modeling with copulas*. CRC press.
- Mays, L. (1999). *Hydraulic Design Handbook* (1.<sup>a</sup> ed.). McGraw-Hill Professional.
- McNeil, A. J., Frey, R. & Embrechts, P. (2005). *Quantitative risk management: Concepts, techniques and tools* (Vol. 3). Princeton university press Princeton.
- Murillo, R. (2006). *Estimación de eventos extremos ena hidrología*. Escuela de Ingeniería Civil. Universidad de Costa Rica.
- Nelsen, R. (2006). *An Introduction to Copulas* (2.<sup>a</sup> ed.). Springer.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. & Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. Wiley.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <http://ggplot2.org>
- World Meteorological Organization, W. (2008). *Guide to Hydrological Practices*. WMO-No.168. World Meteorological Organization.

## Tesis

- García, E. (2015). *Análisis de intensidades de precipitación de la cuenca del río Grande de Térraba*. Universidad de Costa Rica, Facultad de Ingeniería, Escuela de Ingeniería Civil.
- Muñoz, E. (2017). *Elaboración del atlas de sequías para Costa Rica, mediante el análisis regional de frecuencias*. Universidad de Costa Rica, Facultad de Ingeniería, Escuela de Ingeniería Civil.
- Murillo, R. (1994). *Estudio de intensidades de lluvia en la cuenca del río Virilla*. Universidad de Costa Rica, Facultad de Ingeniería, Escuela de Ingeniería Civil.
- Nikoletta, S. (2017). *Bivariate frequency analysis of extreme rainfall and floods using copulas*. Université Grenoble Alpes, University of Thessaly.
- Quirós, E. (1979). *Distribución de frecuencia para volúmenes máximos anuales*. Universidad de Costa Rica, Facultad de Ingeniería, Escuela de Ingeniería Civil.
- Requena, A. I. (2015). *Flood frequency analysis by a bivariate model based on copulas*. Universidad Politécnica de Madrid. Escuela técnica superior de ingenieros de caminos canales y puertos.
- Rudín, F. (1983). *Utilización de correlación múltiple para el diseño de crecientes en Costa Rica*. Universidad de Costa Rica, Facultad de Ingeniería, Escuela de Ingeniería Civil.
- Zegpi, M. (2008). *Modelo analítico para el comportamiento hidrológico de cuencas urbanas*. Pontificia Universidad Católica de Chile, Escuela de Ingeniería.
- Zúñiga, E. (1983). *Metodología para el diseño de crecientes en Costa Rica*. Universidad de Costa Rica, Facultad de Ingeniería, Escuela de Ingeniería Civil.



## Informes, Reportes y Reglamentos

- AyA. (2017). *Norma técnica para diseño y construcción de sistemas de abastecimiento de agua potable y saneamiento y pluvial* (inf. téc.). Instituto Costarricense de Acueductos y Alcantarillados. San José. Costa Rica. [http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm\\_texto\\_completo.aspx?param1=NRTC&nValor1=1&nValor2=59933&nValor3=109664&strTipM=TC](http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?param1=NRTC&nValor1=1&nValor2=59933&nValor3=109664&strTipM=TC)
- MOPT. (s.f.). *Términos de referencia técnicos para estudios hidrológicos e hidráulicos para puentes* (inf. téc.). Ministerio de Obras Públicas y Transportes. <https://www.mopt.go.cr/wps/wcm/connect/27e089c0-24bf-47e4-b3d4-90bb757d77a7/tr-hidrologicos-hidraulicos.pdf?MOD=AJPERES>
- Municipalidad de Heredia. (2015). *Reglamento para el otorgamiento del desfogue pluvial en el Cantón de Heredia* (inf. téc.). Municipalidad de Heredia. San José. Costa Rica. [https://www.heredia.go.cr/sites/default/files/reglamento\\_desfogue\\_pluvial\\_canton\\_central\\_de\\_heredia.pdf](https://www.heredia.go.cr/sites/default/files/reglamento_desfogue_pluvial_canton_central_de_heredia.pdf)
- Rojas, N. (2011a). *Curvas de Intensidad Duración Frecuencia de algunas estaciones meteorológicas automáticas* (inf. téc.). Ministerio de Ambiente y Energía (MINAE). [http://www.mag.go.cr/informacion/curso\\_agua\\_pluvial/cidf\\_automaticas.pdf](http://www.mag.go.cr/informacion/curso_agua_pluvial/cidf_automaticas.pdf)
- Rojas, N. (2011b). *Curvas de Intensidad Duración Frecuencia de algunas estaciones meteorológicas mecánicas* (inf. téc.). Ministerio de Ambiente y Energía (MINAE). [http://www.mag.go.cr/informacion/curso\\_agua\\_pluvial/cidf\\_mecanicas.pdf](http://www.mag.go.cr/informacion/curso_agua_pluvial/cidf_mecanicas.pdf)

# Manuales

- DR, F., MT, W., JA, A., TS, S. & ZM, E. (2018). *EcoHydRology: A Community Modeling Foundation for Eco-Hydrology* [R package version 0.4.12.1]. R package version 0.4.12.1. <https://CRAN.R-project.org/package=EcoHydRology>
- Hofert, M., Hornik, K. & McNeil, A. J. (2018). *qrmtools: Tools for Quantitative Risk Management* [R package version 0.0-10]. R package version 0.0-10. <https://CRAN.R-project.org/package=qrmtools>
- Hofert, M., Kojadinovic, I., Maechler, M. & Yan, J. (2018). *copula: Multivariate Dependence with Copulas* [R package version 0.999-19]. R package version 0.999-19. <https://CRAN.R-project.org/package=copula>
- Hosking, J. R. M. (2019). *L-Moments* [R package, version 2.8]. R package, version 2.8. <https://CRAN.R-project.org/package=lmom>
- Pedersen, T. L. (2020). *patchwork: The Composer of Plots* [R package version 1.1.1]. R package version 1.1.1. <https://CRAN.R-project.org/package=patchwork>
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Tierney, L., Rossini, A. J., Li, N. & Sevcikova, H. (2018). *snow: Simple Network of Workstations* [R package version 0.4-3]. R package version 0.4-3. <https://CRAN.R-project.org/package=snow>
- Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'* [R package version 1.2.1]. R package version 1.2.1. <https://CRAN.R-project.org/package=tidyverse>

## Conferencias

- Ávila, L. & Mono, M. (2018). Análise de frequência bivariada de cheias para estimação de risco hidrológico baseado na modelagem de funções cópulas, En *XXII Simpósio Brasileiro de Recursos Hídricos: Florianópolis*.
- Lyne, V. & Hollick, M. (1979). Stochastic time-variable rainfall-runoff modelling, En *Institute of Engineers Australia National Conference*. Institute of Engineers Australia Barton, Australia.
- Stamatatou, N., Vasiliades, L. & Loukas, A. (2018). Bivariate flood frequency analysis using Copulas, En *3rd EWaS International Conference on Insights on the Water-Energy-Food Nexus*.

## **Apéndice A**

### **Crecientes máximas empleadas en los análisis**

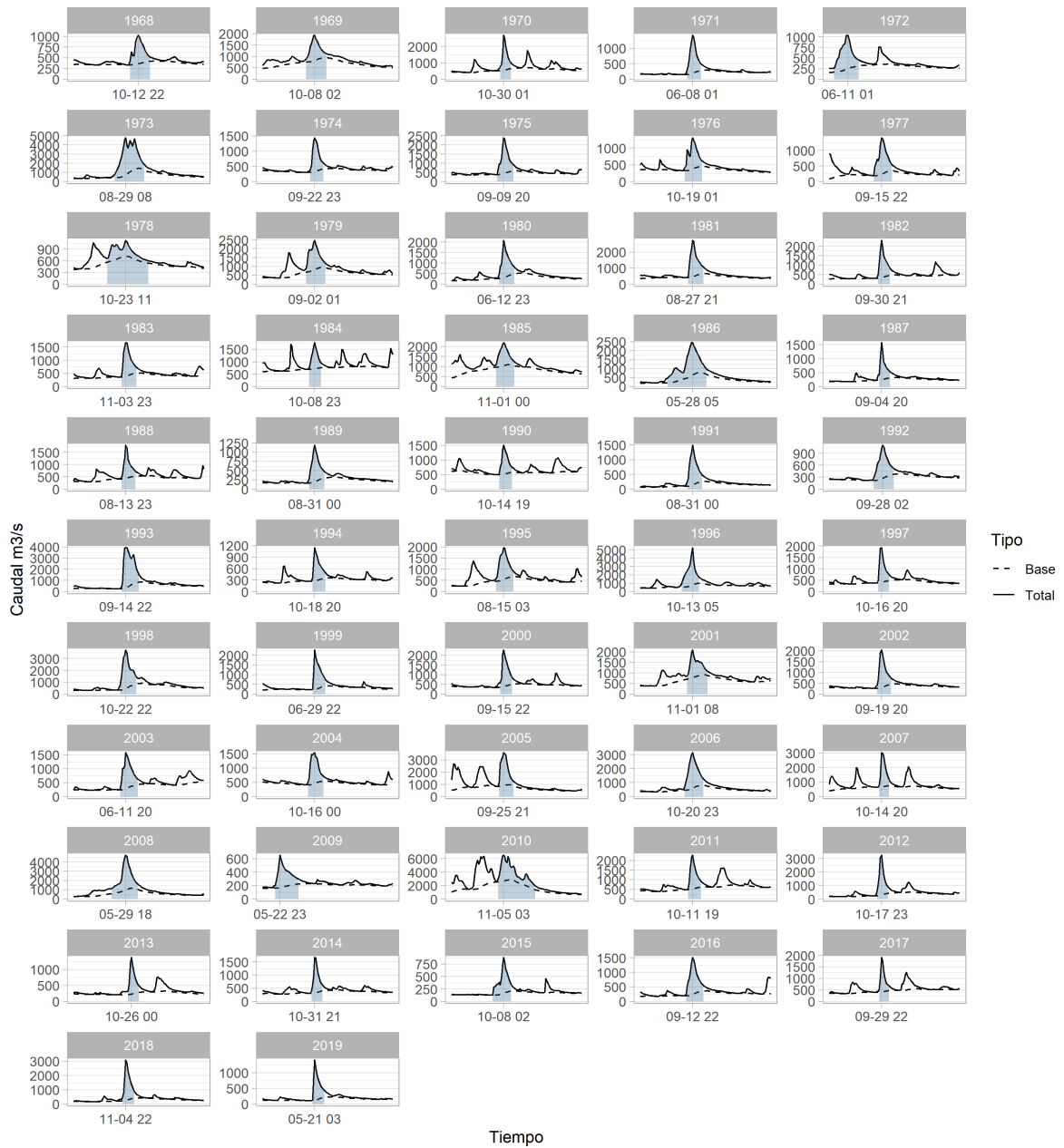


Figura A.1: Crecientes estación 3103  $\{V_b, \max\{Q_b\}\}$ .

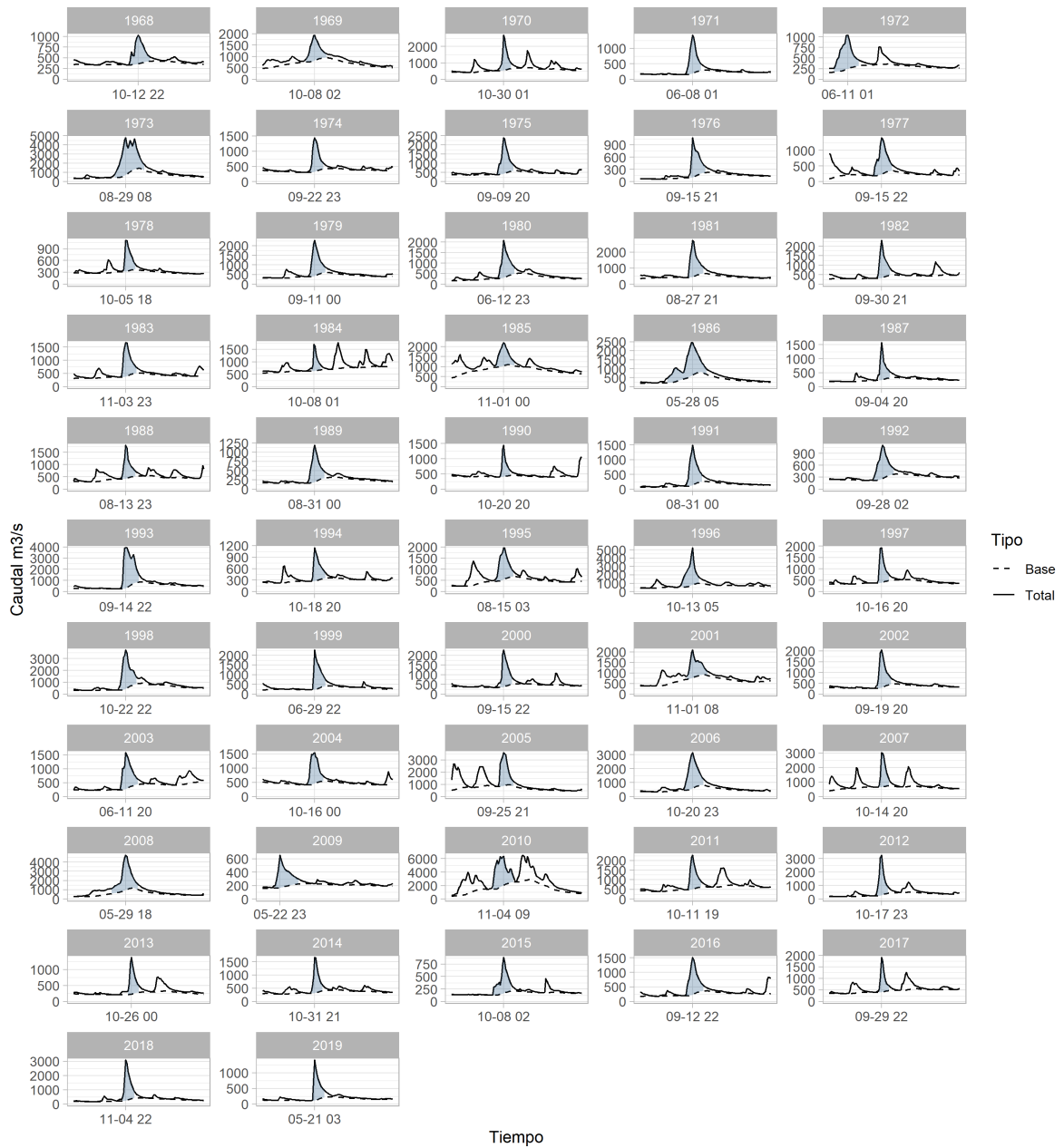


Figura A.2: Crecientes estación 3103  $\{V_e, \max\{Q_e\}\}$ .

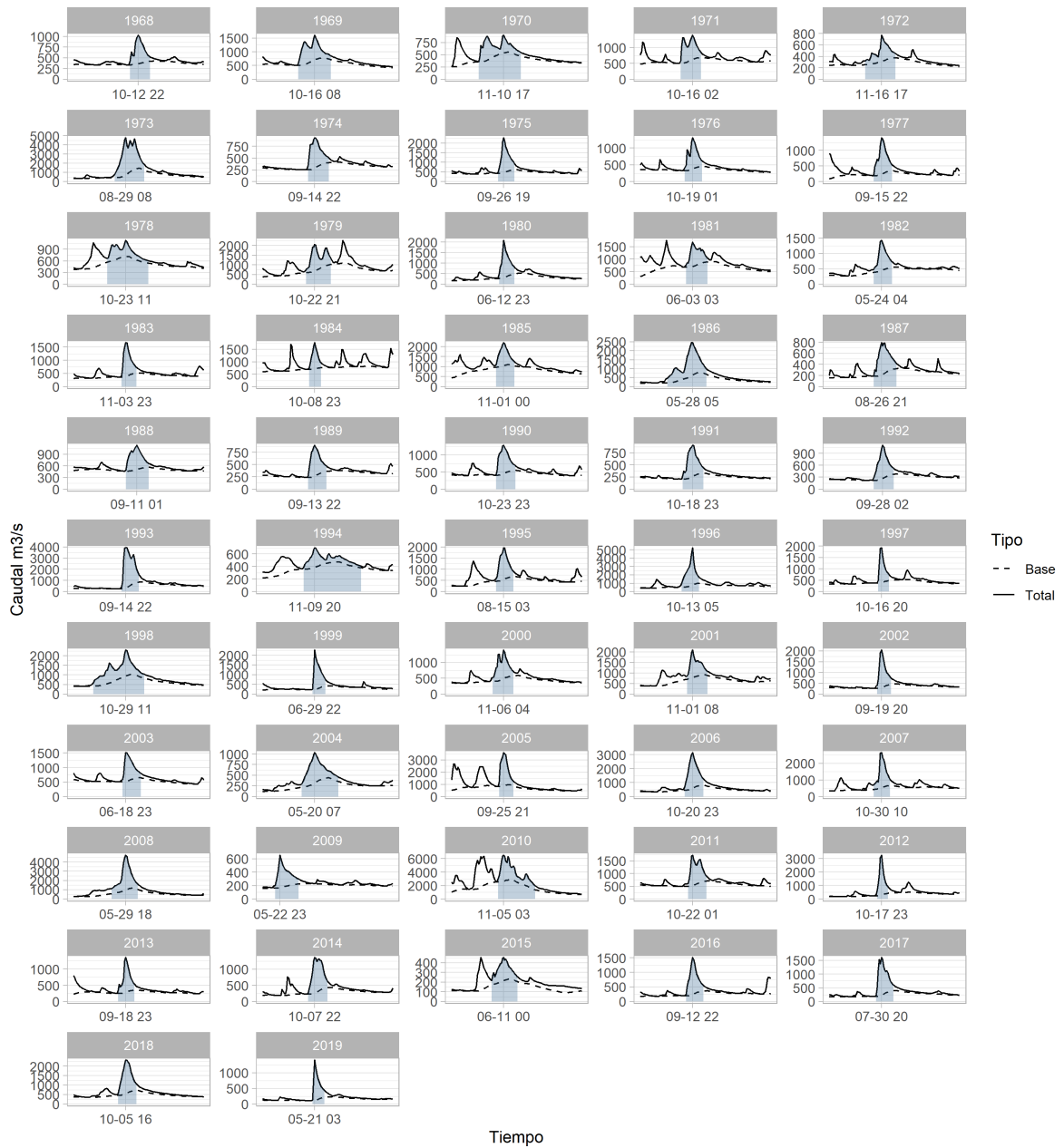


Figura A.3: Crecientes estación 3103  $\{\max\{V_b\}, Q_b\}$ .

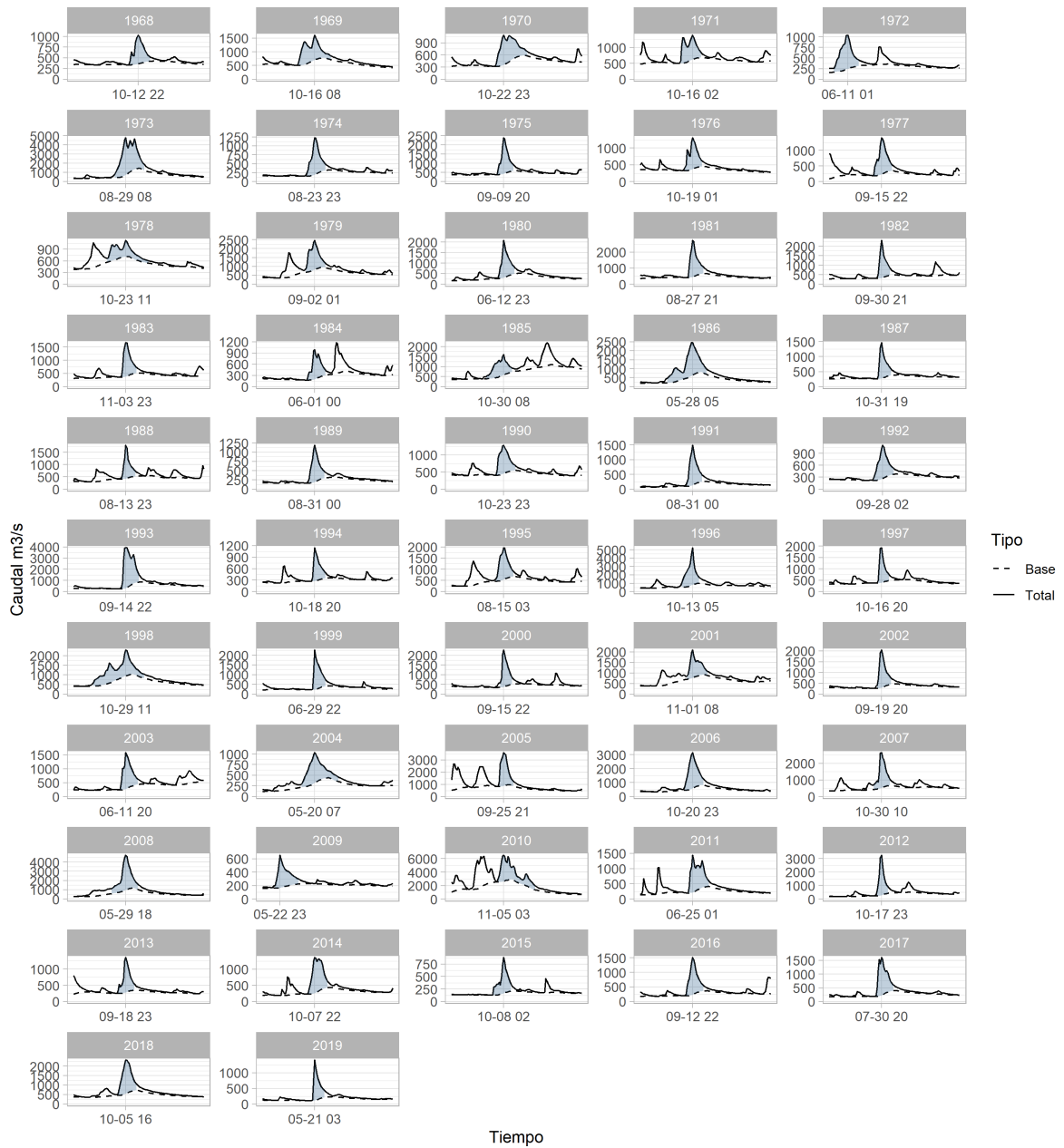


Figura A.4: Crecientes estación 3103  $\{\max\{V_e\}, Q_e\}$ .



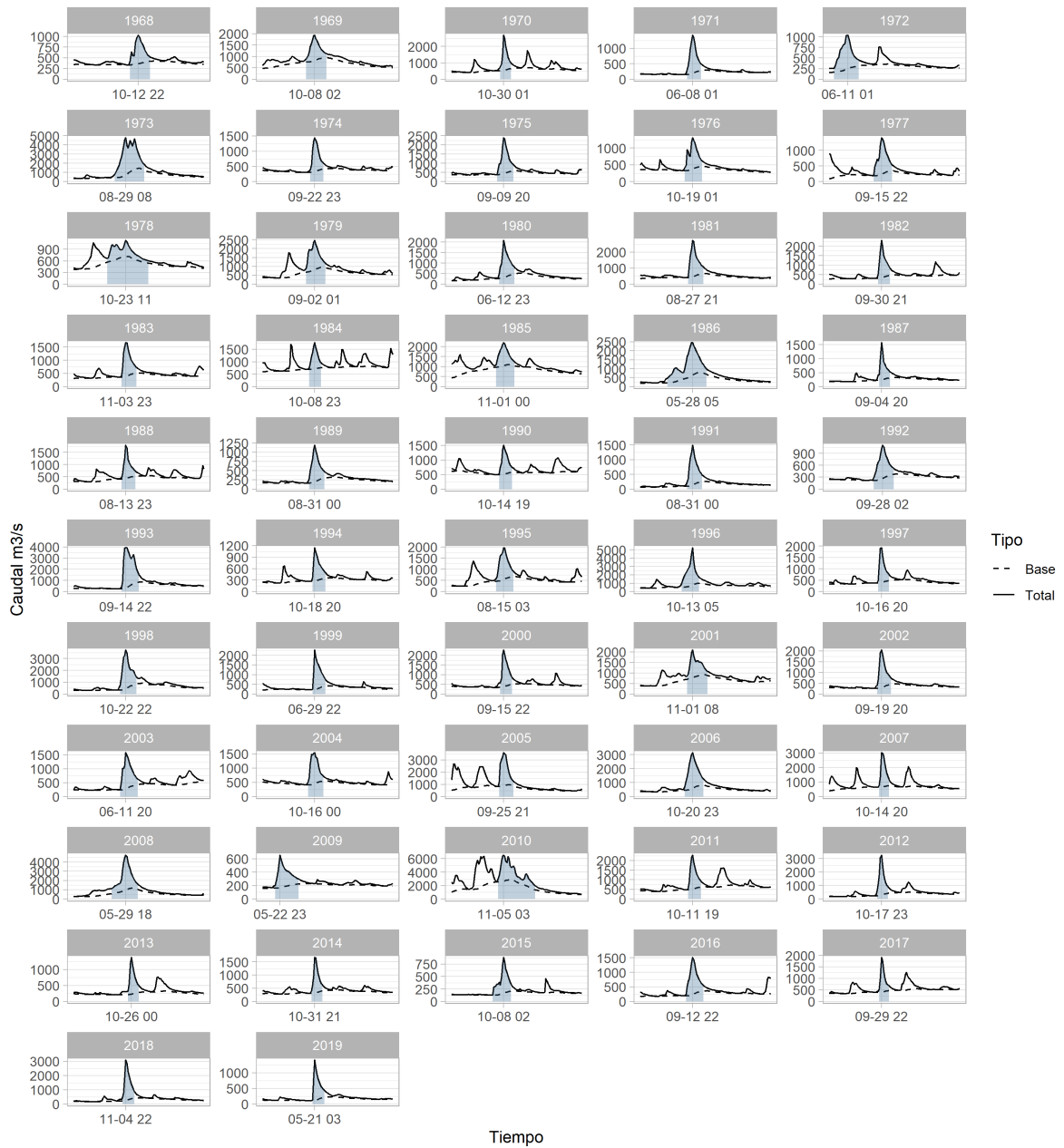


Figura A.5: Crecientes estación 3104  $\{V_b, \max\{Q_b\}\}$ .

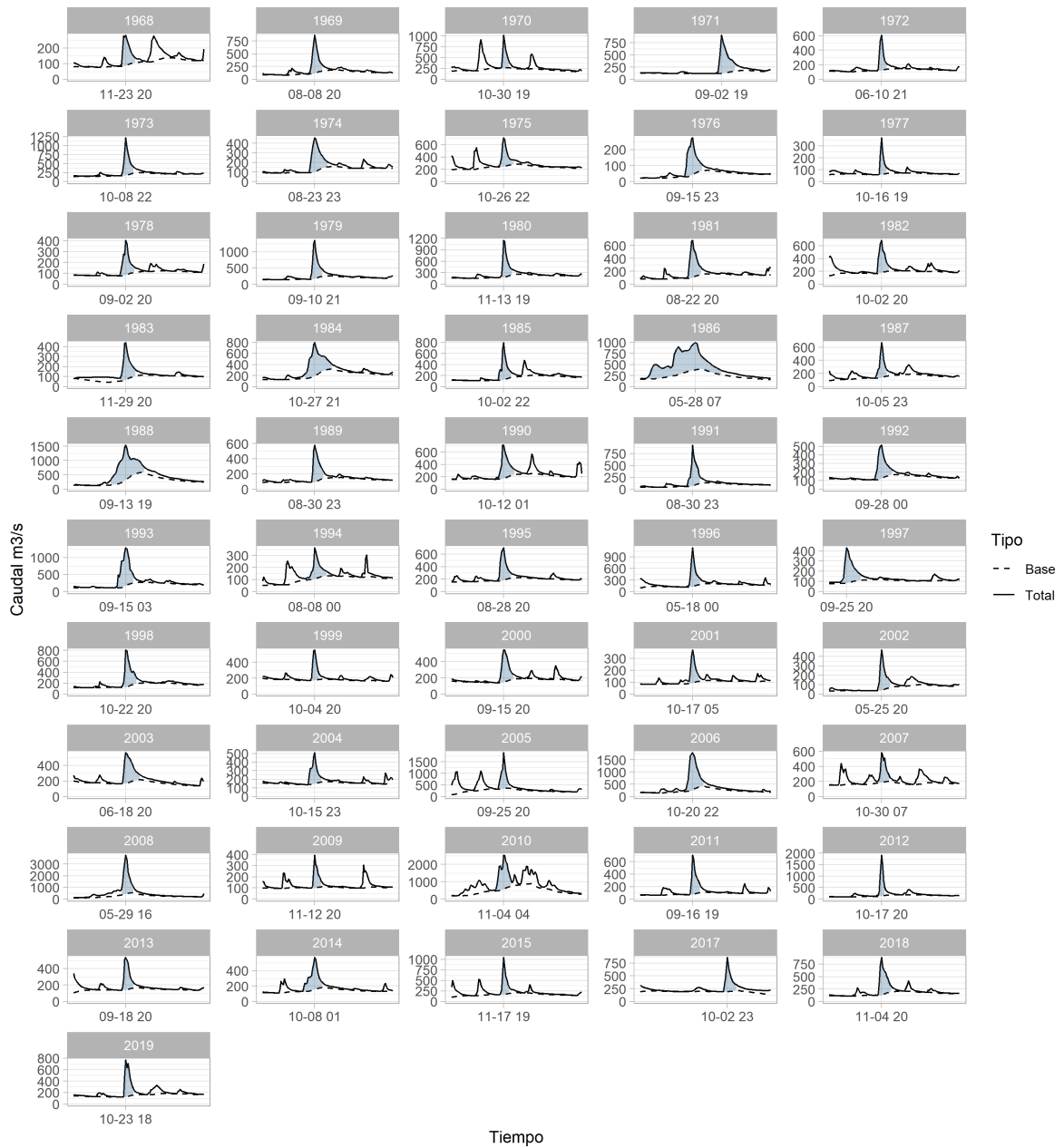


Figura A.6: Crecientes estación 3104  $\{V_e, \max\{Q_e\}\}$ .



Figura A.7: Crecientes estación 3104  $\{\max\{V_b\}, Q_b\}$ .

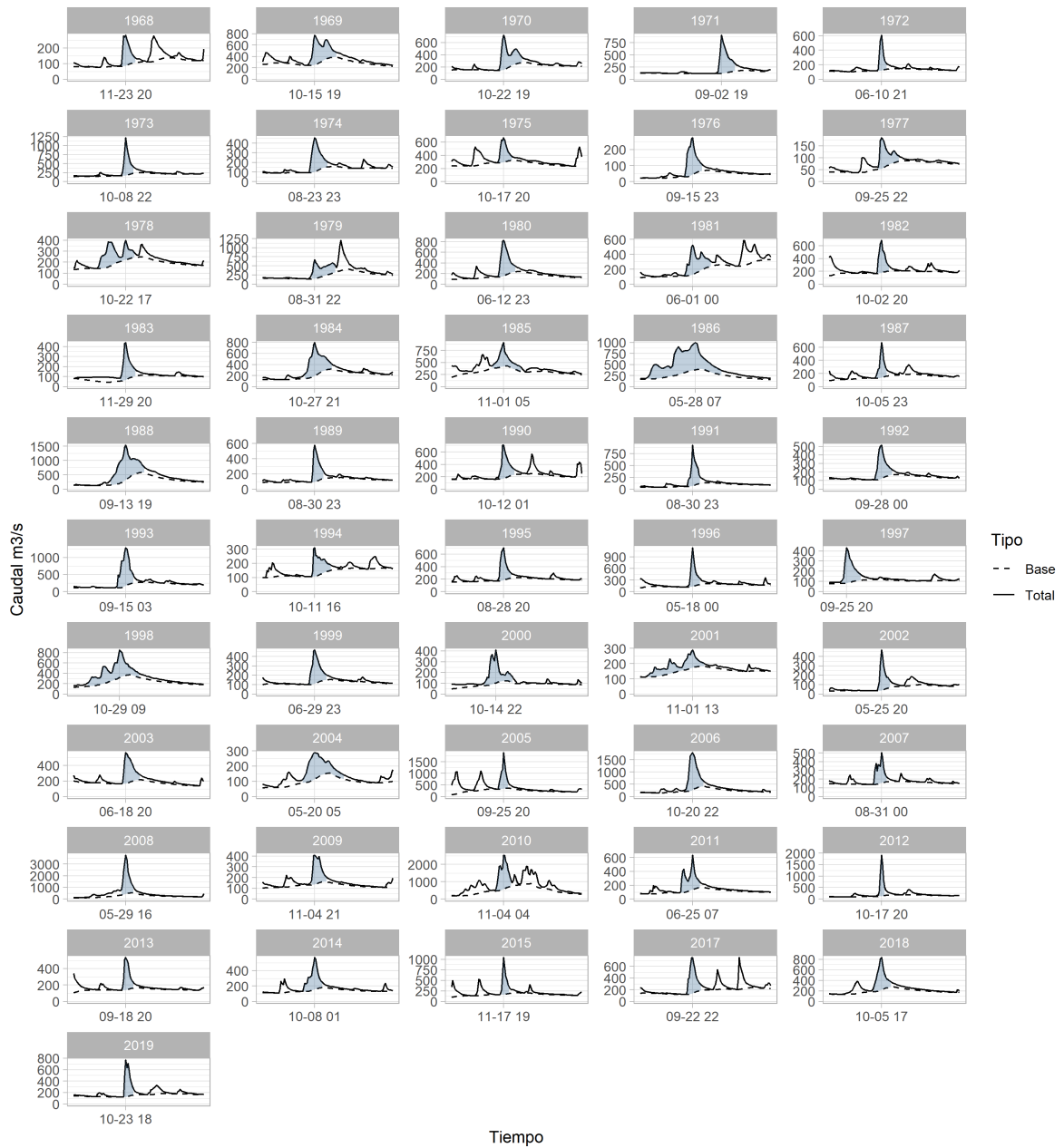


Figura A.8: Crecientes estación 3104  $\{\max\{V_e\}, Q_e\}$ .

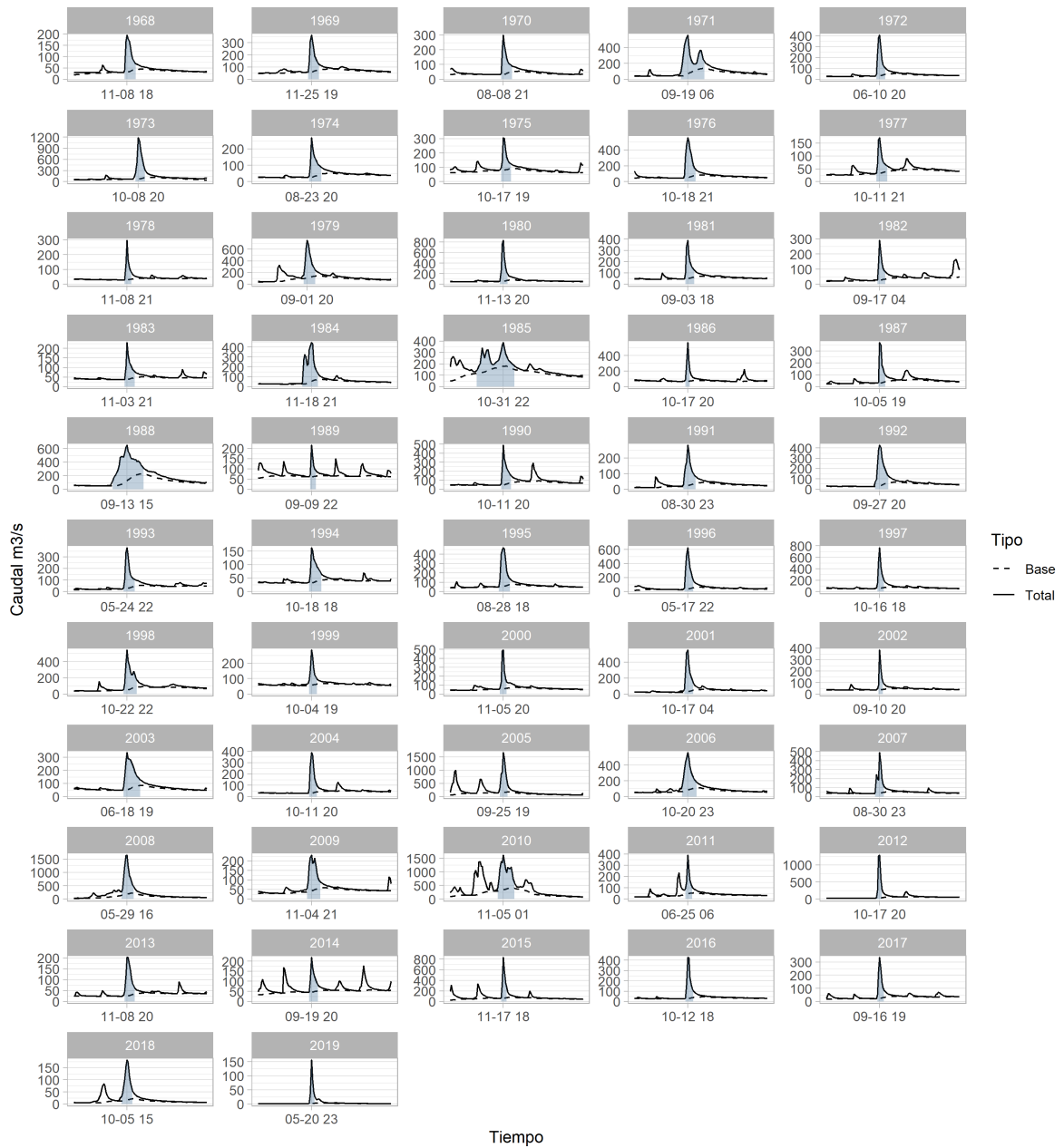


Figura A.9: Crecientes estación 3105  $\{V_b, \max\{Q_b\}\}$ .

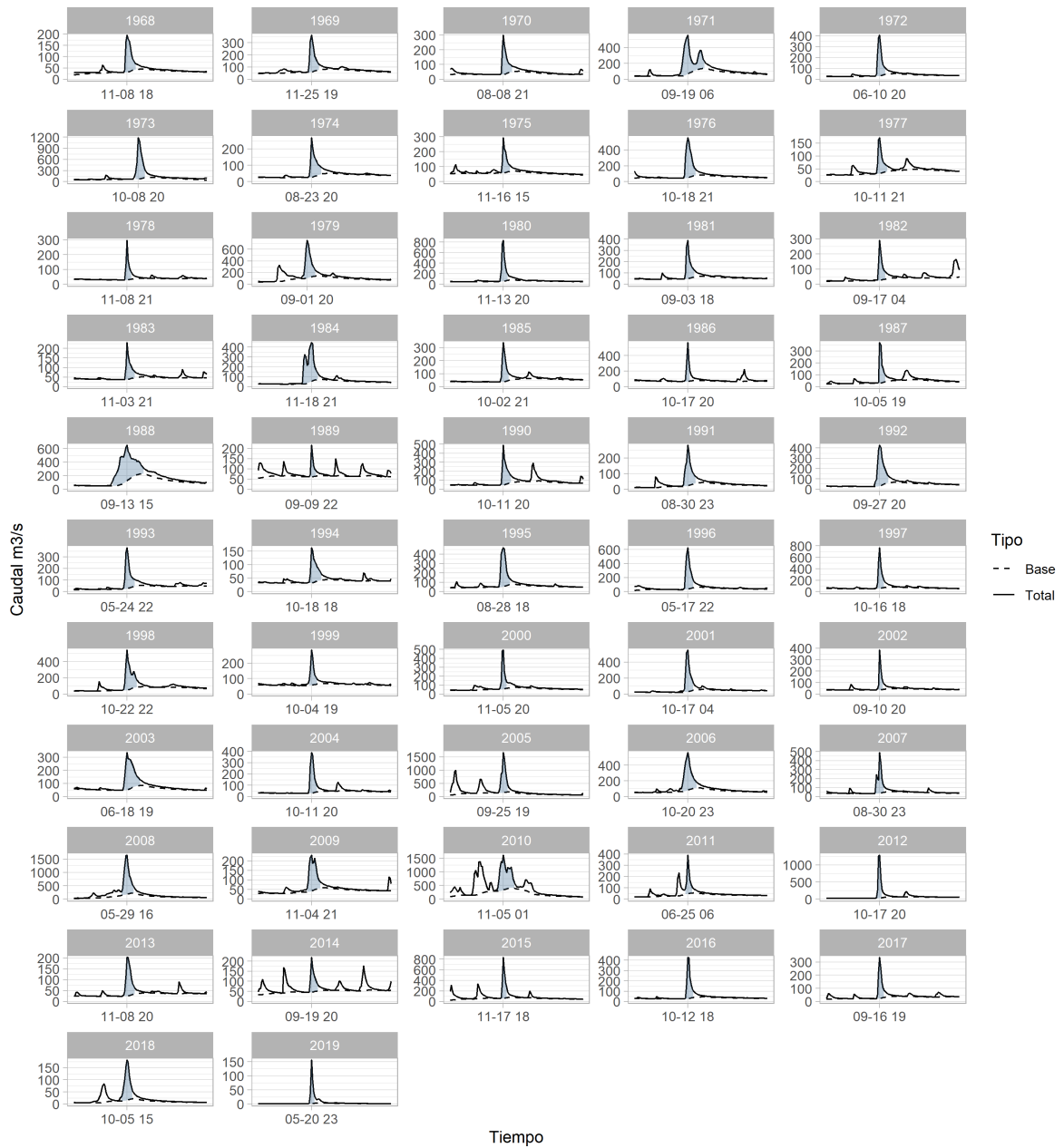


Figura A.10: Crecientes estación 3105  $\{V_e, \max\{Q_e\}\}$ .

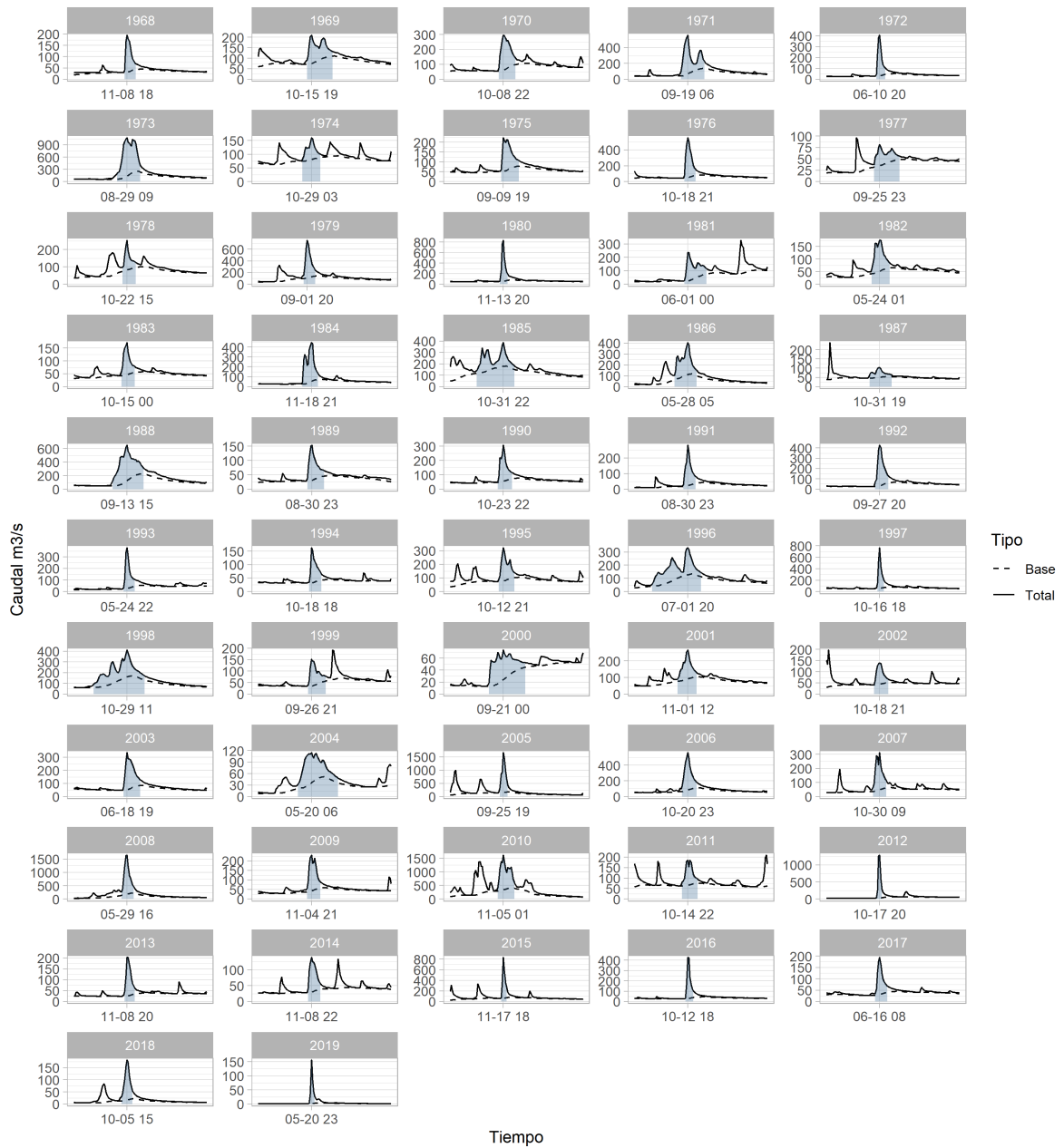


Figura A.11: Crecientes estación 3105  $\{\max\{V_b\}, Q_b\}$ .

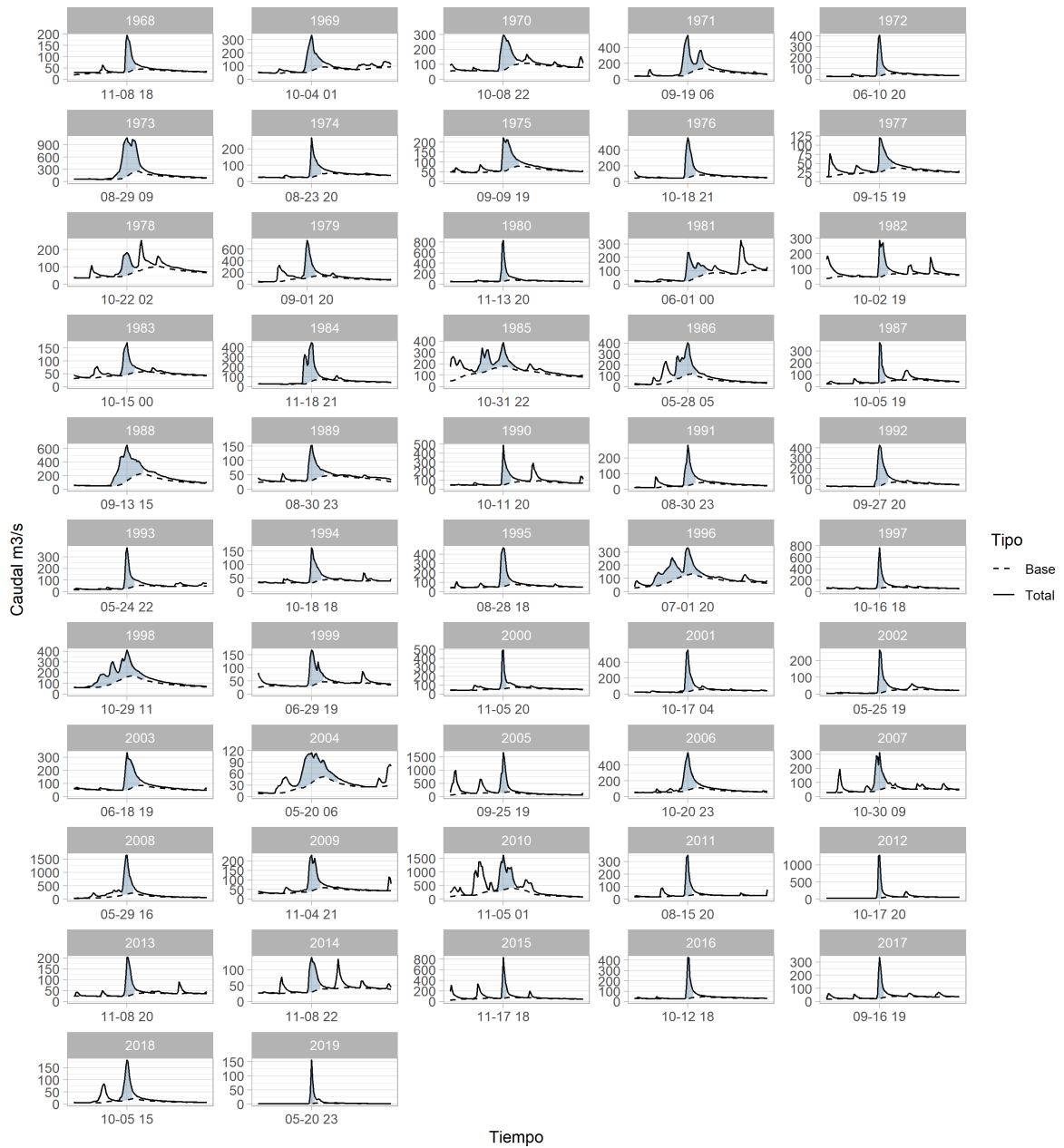


Figura A.12: Crecientes estación 3105  $\{\max\{V_e\}, Q_e\}$ .