

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

SIMULACIÓN DE SERIES DE TIEMPO HIDROLÓGICAS UTILIZANDO
PROCESOS DE POISSON FILTRADOS

Tesis sometida a la consideración de la Comisión del Programa de Estudios
de Posgrado en Estadística para optar al grado de Magíster Scientiae en Estadística

Federico Gómez Delgado

Ciudad Universitaria Rodrigo Facio, Costa Rica

2006

Dedicatoria

A Raquel

A mis padres Manuel y Elizabeth

Agradecimientos

Al Dr. Lars Gottschalk por su permanente y atenta supervisión durante el desarrollo de la investigación. Sus enseñanzas y su apoyo incondicional fueron fundamentales para la culminación de esta tesis. Externo mi gratitud con la Dr. Irina Krasovskaia, quien me ha ayudado de muchas maneras y me ha impulsado al emprendimiento de nuevos proyectos.

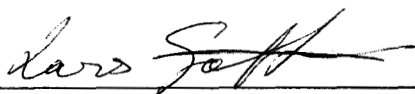
Quiero agradecer al Dr. Juan Muñoz por la revisión detallada del documento y por sus valiosas observaciones y consejos. Al Msc. Juan Bautista Cavaría, por su amable colaboración durante los años que he participado en la Maestría. La confianza y la paciencia del Ing. Javier Orozco Canossa han sido elementos fundamentales para concluir este trabajo.

San José, Noviembre de 2006

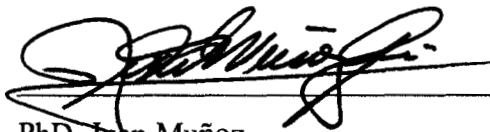
Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Estadística de la Universidad de Costa Rica, como requisito parcial para optar al grado de Magíster Scientiae en Estadística.



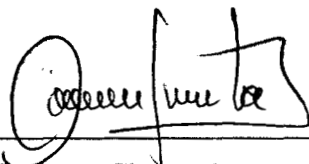
Msc. Sandra Hernández
Representante del Decano del SEP
Presidente del Tribunal de Tesis



PhD. Lars Gottschalk
Director de Tesis



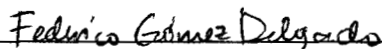
PhD. Juan Muñoz
Lector



MSc. Juan Bautista Chavarría
Lector



MSc. Olman Ramírez
Director de la Maestría Académica en Estadística



Lic. Federico Gómez Delgado
Candidato

Índice general

CAPÍTULO 1	INTRODUCCIÓN	1
1.1	PLANTEAMIENTO DEL PROBLEMA	1
1.2	MODELOS ESTOCÁSTICOS DE CAUDALES Y PROCEDIMIENTOS DE SIMULACIÓN	3
1.3	LA ESCALA DE VARIABILIDAD Y LA MODELACIÓN DE CAUDALES DIARIOS	5
1.4	ANTECEDENTES DE LOS MODELOS ESTOCÁSTICOS DE SIMULACIÓN DE CAUDAL DIARIO	7
1.4.1	<i>Modelos auto-regresivos o markovianos</i>	7
1.4.2	<i>Procesos de Poisson filtrados o procesos de ruido disparado</i>	9
1.4.3	<i>Modelo estocástico-conceptual</i>	9
1.4.4	<i>Modelos de probabilidades de transición</i>	13
1.4.5	<i>Modelo de proceso puntual estocástico con base conceptual</i>	16
1.5	OBJETIVOS DE LA INVESTIGACIÓN	17
CAPÍTULO 2	TEORÍA Y METODOLOGÍA	19
2.1	DEFINICIÓN DE UNA SERIE DE TIEMPO HIDROLÓGICA	19
2.2	EL PROCESO POISSON	19
2.3	EL PROCESO DE POISSON FILTRADO	21
2.3.1	<i>Identificación de la lluvia efectiva en el FPP y estimación de la tasa de eventos v</i>	24
2.3.2	<i>Identificación de la magnitud de los impulsos y estimación del salto promedio θ</i>	30
2.4	PROCESO DE POISSON FILTRADO DE SEGUNDO ORDEN	31
2.5	EL EFECTO DEL MUESTREO PROMEDIADO EN INTERVALOS DISCRETOS PARA MODELOS FPP DE PRIMER Y SEGUNDO ORDEN	34
2.5.1	<i>Modelos FPP de primer orden</i>	35
2.5.2	<i>Modelos FPP de segundo orden</i>	36
2.6	SEPARACIÓN DEL PROCESO DE RUIDO DISPARADO EN EL TIEMPO PARA MODELOS FPP DE PRIMER Y SEGUNDO ORDEN	38
2.6.1	<i>Modelos FPP de primer orden</i>	39
2.6.2	<i>Modelos FPP de segundo orden</i>	39
2.7	AJUSTE DE LOS PARÁMETROS DEL MODELO FPP DE PRIMER Y SEGUNDO ORDEN	41
2.7.1	<i>Modelos FPP de primer orden</i>	41
2.7.2	<i>Modelos FPP de segundo orden</i>	42
2.7.3	<i>Valoración de la estacionalidad en la estimación de los parámetros</i>	42
2.8	PROCEDIMIENTO DE GENERACIÓN DE CAUDALES SINTÉTICOS PARA MODELOS FPP DE PRIMER Y SEGUNDO ORDEN	43
2.9	PROPIEDADES DE CRUCE DE UNA SERIE CRONOLÓGICA	46

CAPÍTULO 3	DATOS Y ANÁLISIS DESCRIPTIVO	49
3.1	CARACTERIZACIÓN DE LAS ZONAS DE ESTUDIO	49
3.1.1	<i>Fisiografía.....</i>	50
3.1.2	<i>Clima.....</i>	53
3.1.3	<i>Perfil fisio-climático.....</i>	55
3.2	DESCRIPCIÓN GENERAL DE LAS SERIES DE TIEMPO.....	56
3.2.1	<i>Estadísticos principales de las series de caudal diario.....</i>	60
3.2.2	<i>Distribución de los caudales promedio diarios.....</i>	62
3.2.3	<i>Evaluación de la condición estacionaria o de tendencia.....</i>	67
3.2.4	<i>Independencia de las observaciones.....</i>	68
3.2.5	<i>Periodicidad de las series de caudal diario.....</i>	70
3.2.6	<i>Correlaciones de las series en el espacio.....</i>	72
3.2.7	<i>Efecto de fenómenos globales climáticos en las series de caudal.....</i>	74
CAPÍTULO 4	RESULTADOS Y DISCUSIÓN	79
4.1	IDENTIFICACIÓN DE LA LLUVIA EFECTIVA EN LAS SERIES DE CAUDAL PROMEDIO DIARIO Y ESTIMACIÓN DE LA TASA DE EVENTOS N	79
4.2	EL ENFOQUE FPOT: PRUEBAS DE INDEPENDENCIA Y DISTRIBUCIÓN POISSON DE LOS EVENTOS DE LLUVIA EFECTIVA.....	81
4.3	IDENTIFICACIÓN DE LA MAGNITUD DE LOS IMPULSOS Y ESTIMACIÓN DEL SALTO PROMEDIO O DE LA DISTRIBUCIÓN EXPONENCIAL ASOCIADA	92
4.4	EVALUACIÓN DE LAS PROPIEDADES DE CRUCE EN LAS SERIES DE CAUDAL PROMEDIO DIARIO	94
4.5	MODELOS DE POISSON FILTRADOS MENSUALES DE PRIMER ORDEN.....	98
4.5.1	<i>Estimación de los parámetros mediante el método de preservación de tres momentos.....</i>	98
4.5.2	<i>Generación y validación de las series sintéticas de caudal diario para modelos FPP mensuales de primer orden con preservación de tres momentos.....</i>	101
4.5.3	<i>Estimación de los parámetros mediante el método de preservación de momentos y considerando la tasa de ocurrencia de eventos observada v.....</i>	106
4.5.4	<i>Estimación mediante el método de preservación de momentos considerando la tasa de ocurrencia de eventos v corregida con el método FPOT.....</i>	111
4.6	MODELOS DE POISSON FILTRADOS MENSUALES DE SEGUNDO ORDEN.....	112
4.6.1	<i>Estimación de los parámetros por preservación mediante el método de preservación de cinco momentos.....</i>	112
4.6.2	<i>Generación y validación de las series sintéticas de caudal diario para modelos FPP mensuales de segundo orden.....</i>	114
4.7	RESUMEN DE LOS RESULTADOS OBTENIDOS.....	115
CAPÍTULO 5	CONCLUSIONES Y RECOMENDACIONES	117

Índice de tablas

TABLA 3.1: CARACTERÍSTICAS FISIGRÁFICAS GENERALES DE LAS CUENCAS ESTUDIADAS	50
TABLA 3.2: DISTRIBUCIÓN PORCENTUAL DE ELEVACIONES Y PENDIENTES EN LAS CUENCAS ESTUDIADAS	51
TABLA 3.3: PRINCIPALES CARACTERÍSTICAS DE LAS ESTACIONES ANALIZADAS Y SUS SUB-CUENCAS TRIBUTARIAS	52
TABLA 3.4: CARACTERÍSTICAS DE LA PRECIPITACIÓN ANUAL EN LAS CUENCAS ESTUDIADAS.....	55
TABLA 3.5: CARACTERÍSTICAS DE LAS SERIES DE CAUDAL PROMEDIO DIARIO PARA LAS ESTACIONES HIDROLÓGICAS	58
TABLA 3.6: PRINCIPALES ESTADÍSTICOS DE LAS SERIES DIARIAS DE CAUDAL	61
TABLA 3.7: AJUSTE DE FUNCIONES DE DISTRIBUCIÓN TEÓRICAS A LAS DISTRIBUCIONES OBSERVADAS DE LAS SERIES ESCOGIDAS	64
TABLA 3.8: PARÁMETROS DEL MODELO LINEAL PROPUESTO PARA ANALIZAR LA TENDENCIA DE LAS SERIES. INTERVALO DE CONFIANZA DEL PARÁMETRO DE PENDIENTE Y VALORACIÓN DE LA TENDENCIA DE LA SERIE	68
TABLA 3.9: PRUEBA DE CORRIDAS (“RUNTEST”) PARA EVALUAR LA INDEPENDENCIA DE LAS OBSERVACIONES EN LAS SERIES ESTUDIADAS.....	68
TABLA 3.10: CORRELACIONES ENTRE EL CAUDAL PROMEDIO DIARIO MEDIDO EN LAS ESTACIONES ESCOGIDAS	73
TABLA 4.1 SELECCIÓN DEL UMBRAL REQUERIDO PARA LA APLICACIÓN DEL MÉTODO FPOT	91
TABLA 4.2 PARÁMETROS MENSUALES DE UN MODELO SNP AJUSTADOS PRESERVANDO TRES MOMENTOS (P3M), A PARTIR DEL REGISTRO HISTÓRICO EN LA ESTACIÓN HIDROLÓGICA 3001 ASUNCIÓN.....	100
TABLA 4.3 PARÁMETROS MENSUALES DE UN MODELO SNP AJUSTADOS CON EL MÉTODO TPD, A PARTIR DEL REGISTRO HISTÓRICO EN LA ESTACIÓN HIDROLÓGICA 3001 ASUNCIÓN.....	107
TABLA 4.4 PARÁMETROS MENSUALES DE UN MODELO SNP DE SEGUNDO ORDEN Y TIPO SOSN, AJUSTADOS CON EL MÉTODO P5M A PARTIR DEL REGISTRO HISTÓRICO EN LA ESTACIÓN HIDROLÓGICA 3001 ASUNCIÓN.....	113
TABLA 4.5: COMPARACIÓN DE LOS RESULTADOS OBTENIDOS CON LOS PRINCIPALES MODELOS FPP	116

Índice de figuras

FIGURA 2.1 EL PROCESO DE RUIDO DISPARADO (SNP) COMO UNA SECUENCIA DE IMPULSOS Y RECESIONES	24
FIGURA 2.2 IDENTIFICACIÓN DE LA LLUVIA EFECTIVA	25
FIGURA 2.3 APLICACIÓN DE FILTRO UMBRAL A LA LLUVIA EFECTIVA, PARA GENERAR LA SERIE FPOT.....	29
FIGURA 3.1 ESTACIONES HIDROLÓGICAS ESCOGIDAS PARA LA SIMULACIÓN.....	49
FIGURA 3.2: UBICACIÓN DE LAS ESTACIONES HIDROLÓGICAS Y SU ELEVACIÓN.....	51
FIGURA 3.3: LLUVIA MEDIA ANUAL EN LA REGIÓN ANALIZADA.....	54
FIGURA 3.4: LÍNEA DE EXTRACCIÓN DEL PERFIL. ELEVACIÓN, PENDIENTE Y PRECIPITACIÓN ANUAL EN LA DIRECCIÓN PACÍFICO-CARIBE.....	56
FIGURA 3.5: PARTE DE LA SERIE DE CAUDAL PROMEDIO DIARIO EN LA ESTACIÓN HIDROLÓGICA 8001 PACUARE ...	57
FIGURA 3.6: REGRESIÓN LINEAL APLICADA A LOS CAUDALES DIARIOS OBSERVADOS.....	59
FIGURA 3.7: REEMPLAZO DE VALORES EN BLANCO DE LA ESTACIÓN 8001 PACUARE, TOMANDO COMO BASE UN REGISTRO OBSERVADO EN LA 8003 DOS MONTAÑAS	60
FIGURA 3.8: RELACIÓN ENTRE LOS COEFICIENTES DE VARIACIÓN PARA LA LLUVIA (C.V. ESPACIAL) Y PARA EL CAUDAL (C.V. TEMPORAL) CON EL TAMAÑO DE LA SUB-CUENCA DE DRENAJE.....	62
FIGURA 3.9: DISTRIBUCIONES DE FRECUENCIA DE LOS CAUDALES DIARIOS EN LAS ESTACIONES HIDROLÓGICAS SELECCIONADAS	63
FIGURA 3.10: PRUEBAS GRÁFICAS DE AJUSTE DE DISTRIBUCIONES TEÓRICAS PARA ESTACIONES EN LOS REGÍMENES CARIBE Y PACÍFICO	65
FIGURA 3.11: CURVAS DE DURACIÓN DE CAUDAL DIARIO EN LAS ESTACIONES ESCOGIDAS	66
FIGURA 3.12: FUNCIONES DE AUTO-CORRELACIÓN PARA TRES SERIES DE CAUDAL.....	69
FIGURA 3.13: SERIES DE CAUDAL PROMEDIO MENSUAL EN ESTACIONES DE LAS VERTIENTES CARIBE Y PACÍFICO .	70
FIGURA 3.14: ANÁLISIS ESPECTRAL DEL CAUDAL PROMEDIO DIARIO.....	72
FIGURA 3.15: PROMEDIOS MENSUALES DE LOS CAUDALES EN LAS ESTACIONES ANALIZADAS	72
FIGURA 3.16: CORRELACIÓN ENTRE EL CAUDAL PROMEDIO MENSUAL Y EL NAO	76
FIGURA 3.17: CORRELACIÓN ENTRE EL CAUDAL PROMEDIO MENSUAL Y EL NIÑO3.4.....	77
FIGURA 3.18: CORRELACIÓN ENTRE EL CAUDAL PROMEDIO MENSUAL Y EL SSN.....	78
FIGURA 4.1 TASA DIARIA DE EVENTOS POR MES PARA LAS ESTACIONES ANALIZADAS	79
FIGURA 4.2 SERIE DE CAUDAL OBSERVADA EN LA ESTACIÓN 29004 SAVEGRE EN EL AÑO 2000, MESES DE MARZO Y ABRIL, AMBOS CON BAJOS PROMEDIOS DE CAUDAL Y ALTAS TASAS DE EVENTOS	80
FIGURA 4.3 TASA ANUAL DE EVENTOS COMO FUNCIÓN DEL CAUDAL UMBRAL NORMALIZADO, UTILIZADO PARA FILTRAR LOS PICOS	82
FIGURA 4.4 TASA ANUAL DE EVENTOS EN FUNCIÓN DE LA PROBABILIDAD DE EXCEDENCIA (PE) DEL CAUDAL UMBRAL DE FILTRO	84
FIGURA 4.5 TASA ANUAL DE EVENTOS COMO FUNCIÓN DEL COEFICIENTE DE VARIABILIDAD (C.V.) DE LOS CAUDALES DIARIOS, PARA TRES PROBABILIDADES DE EXCEDENCIA DEL UMBRAL FILTRO	86

FIGURA 4.6 LA VARIACIÓN DEL UMBRAL Y SU INFLUENCIA EN LA FORMA DE LA FUNCIÓN DE TASA DE EVENTOS...	87
FIGURA 4.7 PRUEBA DE INDEPENDENCIA DE KENDALL EN FUNCIÓN DEL UMBRAL s DE FILTRO DE LOS PICOS.....	88
FIGURA 4.8 PRUEBA DE CUNNANE PARA LA DISTRIBUCIÓN POISSON DE LOS PICOS, EN FUNCIÓN DEL UMBRAL s DE FILTRO	90
FIGURA 4.9 PRUEBAS DE DISTRIBUCIÓN GDP DE TIPO W^2 DE CRAMER-VON MISES Y A^2 DE ANDERSON-DARLING.	93
FIGURA 4.10 PROPIEDADES DE CRUCE PARA UNA SERIE DE CAUDAL DIARIO	95
FIGURA 4.11 CRUCES HACIA ARRIBA Y LONGITUD PROMEDIO DE CORRIDAS COMO FUNCIÓN DE LA PROBABILIDAD DE EXCEDENCIA DEL UMBRAL	96
FIGURA 4.12 ESTADÍSTICOS DE LAS SERIES DIARIAS DE CAUDAL	99
FIGURA 4.13 COEFICIENTE DE VARIACIÓN Y PRIMERA AUTO-CORRELACIÓN DE LAS SERIES ANALIZADAS DE CAUDAL DIARIO.....	100
FIGURA 4.14 UN AÑO DE CAUDALES MEDIOS DIARIOS EN LA ESTACIÓN 3001 ASUNCIÓN PARA LA SERIE OBSERVADA Y PARA LA SERIE SEMC SIMULADA CON EL MÉTODO P3M.....	103
FIGURA 4.15 ESTADÍSTICOS MENSUALES DE LA SERIE OBSERVADA Y DE LAS 100 SERIES MC TIPO P3M.....	105
FIGURA 4.16 COMPARACIÓN DE LOS PARÁMETROS MENSUALES ESTIMADOS PARA UN MODELO SNP DE PRIMER ORDEN, ENTRE EL MÉTODO P3M Y EL MÉTODO TPD	108
FIGURA 4.17 UN AÑO DE CAUDALES MEDIOS DIARIOS EN LA ESTACIÓN 3001 ASUNCIÓN PARA LA SERIE SEMC SIMULADA CON EL MÉTODO TPD	109
FIGURA 4.18 ESTADÍSTICOS MENSUALES DE LA SERIE OBSERVADA Y DE LAS 100 SERIES MC TIPO TPD	109
FIGURA 4.19 PROPIEDADES DE CRUCE COMO FUNCIÓN DE UN UMBRAL s PARA LA SERIE SINTÉTICA DE CAUDAL DIARIO SEMC-TPD Y PARA LA SERIE HISTÓRICA.....	110
FIGURA 4.20 UN AÑO DE CAUDALES MEDIOS DIARIOS EN LA ESTACIÓN 3001 ASUNCIÓN PARA LA SERIE SEMC SIMULADA CON EL MÉTODO P5M	114
FIGURA 4.21 ESTADÍSTICOS MENSUALES DE LA SERIE OBSERVADA Y DE LAS 100 SERIES MC TIPO P5M.....	115

Capítulo 1 Introducción

1.1 Planteamiento del problema

La generación de series sintéticas del caudal en los ríos tiene una gran importancia en el campo de la hidrología y en la gestión integrada de los recursos hídricos. En estas disciplinas el problema de la carencia de información sobre el comportamiento de las corrientes de agua superficial es un tema fundamental. Esta carencia se presenta por la inexistencia de mediciones de caudal en sitios de interés, o por la pobreza del registro en el caso de que sí se hayan efectuado mediciones. Aún en presencia de una buena cantidad de observaciones, las estadísticas del caudal podrían ser poco confiables por el incumplimiento de la homogeneidad de la serie, situación que podría darse, por ejemplo, por la construcción de una represa aguas arriba del punto de medición, o por el cambio en el uso del suelo en la cuenca tributaria asociada (Yue y otros, 1999).

El registro sistemático y sucesivo del caudal en un río genera una serie de tiempo típica que, a su vez, se puede definir como la observación de un fenómeno que se mueve en el tiempo y que produce un conjunto ordenado de valores igualmente espaciados que, en el caso general, denotan persistencia. Un caso especial se da cuando los datos son independientes y pueden ser reorganizados sin pérdida de información (Gottschalk, 2004).

El conocimiento del régimen de caudales resulta indispensable en etapas de planificación, diseño, gestión y operación de sistemas de recursos hidráulicos. La resolución temporal (escala de tiempo) requerida está en función del análisis por efectuar. De ese modo, los estudios que poseen menor sensibilidad a las variaciones de caudal de corto plazo usualmente se basan en datos promediados mensual o semanalmente. En caso de que las variaciones rápidas sean importantes se requerirá contar con valores diarios.

Algunas aplicaciones que requieren trabajar con información diaria son (i) la regulación óptima de crecientes y (ii) el enrutamiento del flujo para el control de la industria pesquera y la calidad del agua en períodos de bajo caudal (Payne y otros, 1969), (iii) la previsión precisa a

corto plazo de la evolución futura de los caudales diarios (Bernier, 1970), (iv) la determinación de parámetros de diseño donde se demande un análisis de sensibilidad basado en simulaciones. Esto se aplica tanto en las etapas de planificación como en las de definición de reglas de operación en redes de suministro de agua para usos como consumo humano, industrial, para riego u otros (Weiss, 1973, 1977), (v) el simulacro de los procesos de erosión-sedimentación de capas en la cama de los ríos utilizando un modelo de transporte de sedimentos (Kron y otros, 1990), (vi) la simulación de esquemas de almacenamiento y bombeo de agua (Aksoy y Bayazit, 2000a,b), (vii) la simulación de la respuesta de los ríos a la contaminación (Xu y otros, 2002), y (viii) la emulación del comportamiento de embalses, así como la evaluación de políticas de operación-regulación alternativas en obras hidráulicas, como parte de un análisis de riesgos (de inundación, de perturbaciones ecológicas, de afectación a actividades socioeconómicas, etc.).

En Costa Rica, un ejemplo de la importancia de conocer y poder modelar el caudal diario en los ríos, es el desarrollo de métodos hidrológicos que, tomando en consideración aspectos ambientales y socio-económicos, permitan evaluar los impactos ocasionados por una eventual regulación de las corrientes de agua sobre las personas y los componentes ecológicos, en una región determinada. Así, el planteamiento de estas regulaciones en la forma de diques o embalses, podría ser analizado más integralmente si se contara con la capacidad de proyectar múltiples escenarios naturales, para así contrastarlos con nuevas propuestas de gestión del recurso hídrico.

Entonces, el problema de esta investigación consiste en evaluar y validar un método de generación de series de tiempo sintéticas, que en la forma de una simulación de Monte Carlo, permita subsanar las carencias de información hidrológica en escala diaria. Si se comprueba que un modelo como los “procesos de Poisson filtrados” es confiable para representar los caudales diarios en Costa Rica, cualquiera de las aplicaciones antes mencionadas podría ser desarrollada como parte de un proceso de planificación, para diversos plazos y escalas. Esto permitiría lograr una gestión del agua más acorde con las necesidades reales de los diversos actores involucrados.

1.2 Modelos estocásticos de caudales y procedimientos de simulación

Todas las aplicaciones anteriores requieren fundamentarse en un modelo de la realidad, pues la abstracción resulta indispensable dada la gran complejidad de la mayoría de los sistemas hidrológicos. Es muy usual clasificar estos modelos en determinísticos o estocásticos, teniendo presente que deben pertenecer inicialmente a las familias de modelos simbólicos (antes que materiales) y matemáticos (en lugar de no matemáticos), según la clasificación de Singh citada por Xu (2005). Otros niveles intermedios de clasificación se aplicarán en función de las necesidades particulares del análisis y, por tanto, del modelo escogido.

Un modelo estocástico es el adecuado para efectos de definir criterios de diseño, o bien una combinación de modelos estocásticos y determinísticos. Lo anterior pues cualquier sistema debe diseñarse para operar en el futuro y, en el caso de los modelos determinísticos, ninguno de sus componentes no paramétricos (la variable de entrada, la variable de salida observada o el término residual entre las variables de salida observada y ajustada) puede presentar variaciones aleatorias. Esta carencia impide asociar distribuciones de probabilidad a los mencionados componentes y por tanto, no permite una generación de posibles valores de entrada en el futuro. Cuando menos no permite obtener dichos valores en forma coherente y simultánea, requerimiento que en este caso podría aplicarse a las entradas a una cuenca en forma de precipitación, radiación solar, humedad del suelo, etc.

Dado lo anterior, con el uso de un modelo estocástico es posible interpretar el registro de caudales con el que se cuenta como una muestra o una realización particular de una población de valores prácticamente infinita, que conforma el proceso aleatorio, y entonces suponer que realizaciones futuras se parecerán al registro histórico sólo en un sentido estadístico, aún cuando el proceso sea estacionario (Haan, 1977). Ese parecido entre datos observados y generados (en términos de parámetros estadísticos de las series) se puede obtener de manera muy precisa por medio ya sea de modelos aleatorios o de series simuladas, pero debe estar claro que esto no es una reproducción exacta de eventos pasados, ni la predicción del futuro (Gottschalk, 2004). Por tanto, los caudales generados de manera estocástica deben valorarse como representativos de flujos futuros sólo en un sentido estadístico (Haan, 1977).

Diffícilmente la secuencia observada de caudal se repetiría en la vida útil de un proyecto o de un escenario de planificación. Estos lapsos pueden ser grandes, y con frecuencia son mayores que el periodo de observación. No obstante, no serán tan grandes como para que se dé una réplica de la realización registrada, la cual puede constar de varias decenas de años. Por tanto, la peor avenida de caudal o la peor sequía observadas en el registro histórico, no son las peores avenidas o sequías posibles.

La distribución de características esenciales para establecer los criterios de diseño deberían entonces basarse en cientos o miles de realizaciones simuladas (por medio de operaciones analíticas o simulaciones de Monte Carlo), en lugar de basarse sólo en una, y la precisión obtenida podría ser controlada por el analista, hasta tanto el modelo estocástico sea correcto. Por consiguiente, la obtención de una modelación adecuada de la realidad es indispensable, y requiere un proceso objetivo que parte de la identificación satisfactoria de los fenómenos importantes que intervienen en el proceso evaluado, hasta llegar a la validación correspondiente con los registros disponibles.

La obtención de un modelo adecuado y la generación de series sintéticas permiten hacer afirmaciones probabilísticas relacionadas con posibles valores futuros. Debe establecerse el conjunto de características del fenómeno estudiado que se desean preservar en la modelación, y que resultan fundamentales para dar validez a la evaluación de posibles escenarios futuros. Algunas de las características que se escogen con más frecuencia para su preservación en la modelación de caudales son la media (anual, estacional, mensual, etc.), la desviación estándar, el sesgo y la primera auto-correlación; pero otras más específicas pueden valorarse según los requerimientos específicos de cada estudio.

Una alternativa adicional ante la adopción de un modelo estocástico o determinístico puro es un modelo estocástico con base conceptual, que no es tan estricto como el modelo teórico-determinístico (que es un modelo de base física o modelo “caja-blanca”), pero tampoco es simplemente empírico (modelo “caja negra” o entrada-salida). El modelo conceptual considera las leyes físicas, pero en una forma altamente simplificada, de modo que las variables físicas actúan sobre las variables de entrada para producir las variables de salida (Xu, 2005).

1.3 La escala de variabilidad y la modelación de caudales diarios

Dado un problema de simulación hidrológica, el estudio de la escala de variabilidad de la serie de tiempo de caudales proporciona la información necesaria para determinar el modelo requerido para describir el escurrimiento superficial observado. Dicha escala determina las propiedades estadísticas y las distribuciones de probabilidad de las características del fenómeno por analizar. La teoría de la probabilidad, y específicamente la de los procesos aleatorios, constituye la base para caracterizar la variabilidad de las series de tiempo hidrológicas, así como la de procesos en el espacio o procesos espacio-tiempo (Gottschalk, 2005).

Un análisis como el anterior aplicado para resolver el problema de la modelación de caudales diarios revela algunas dificultades adicionales, en comparación con series agregadas a mayor escala, que pueden complicar la obtención de un modelo que se desempeñe adecuadamente. Los cambios en las series, asociados con una agregación de la información en mayores escalas, se manifiestan tanto en la función de respuesta, donde varios efectos que ésta determina desaparecen y otros cobran mayor significado; como en la función de entrada al sistema, pues su estructura pasa de un proceso puntual aproximadamente independiente y periódico en la escala diaria a un proceso continuo casi gaussiano en la anual (Murrone y otros, 1997).

En general, cuanto menor sea el intervalo del tiempo de análisis más complicado debería ser el modelo por desarrollar, pues en una serie de intervalo pequeño (por ejemplo diaria) además de las características del flujo presentes en series de mayor plazo (las anuales, por ejemplo), se tornan importantes las fluctuaciones de corto plazo y, por tanto, deben considerarse sus efectos (Aksoy, 2003, 2004).

Claps y Rossi (1995) presentaron un resumen de diversos modelos que fueron identificados, partiendo de su propuesta de marcos conceptuales, para relacionar la lluvia efectiva con la escorrentía. Esta recopilación se encuentra detallada según las escalas de tiempo utilizadas con más frecuencia y permite verificar la gran diversidad de modelos que resultan óptimos en cada

nivel de agregación, partiendo de un mismo procedimiento de identificación. La eficiencia en la estimación de los parámetros de ciertos tipos de modelos para diferentes escalas también es un factor importante que debe considerarse. Claps y Murrone (1993, 1994) estudiaron este efecto, que se presenta especialmente en aquellos modelos que no tienen naturaleza física o, al menos, conceptual. En estos casos la información de los parámetros es transferible dado que su significado conceptual no depende de una escala de tiempo particular.

Puesto que una serie diaria de caudal no es estacionaria, que posee propiedades estadísticas variantes estacionalmente y que existe una compleja estructura de dependencia interna entre sus observaciones (Bernier, 1970), cualquier modelo que se proponga para reproducir su comportamiento deberá considerar estas características hasta algún grado, no obstante el costo analítico que esto implica. Las complicaciones señaladas en la simulación del caudal diario demandarán dar más importancia a los objetivos de menor escala, como serían los procesos físicos involucrados en la formación del caudal. Esto se refleja en el hidrograma, cuyos tramos crecientes se deben a factores externos a la cuenca y pueden ser asociados a la precipitación, y sus tramos decrecientes regidos por características de la cuenca, tales como su capacidad de retención y las fuentes del caudal (Kelman, 1980).

Algunos enfoques iniciales proponían generar precipitación diaria por medio de un modelo estadístico, para luego transformarla en caudal utilizando un modelo determinístico; no obstante, otros componentes del ciclo hidrológico para alimentar el modelo determinístico son más difíciles de simular. La naturaleza de la respuesta de la cuenca, la distribución espacial no uniforme de la precipitación y la carencia de registros suficientemente largos de lluvia son dificultades adicionales en la aplicación de estos métodos (Xu y otros, 2002). Nuevamente se postula que el enfoque determinístico estricto no es el indicado para estos fines. La solución suele estar en los modelos estocásticos.

Las dificultades en la modelación del caudal diario deben tomarse en cuenta sin perder de vista la parsimonia del modelo resultante que, de incumplirse, podría hacer muy engorrosas la etapa de estimación, la transferencia del prototipo obtenido a otros sitios y el cálculo de las grandes cantidades de series simuladas que se desea obtener.

1.4 Antecedentes de los modelos estocásticos de simulación de caudal diario

Una característica esencial que se considera indispensable para aceptar un modelo de simulación de caudal diario es que reproduzca la forma de “diente de sierra” que se observa en las series reales. Este patrón, como ya se mencionó, se constituye por hidrogramas con periodos cortos de caudales crecientes (picos), seguidos por períodos grandes de flujos decrecientes (las recesiones), cuyo comportamiento asimétrico conduce a la propiedad estadística de la irreversibilidad en el tiempo (Lawrence y Kottegoda, 1977).

Weiss (1975) indica que muchos procesos que describen registros geofísicos no son reversibles en el tiempo, como sucede en este caso, de modo que un proceso simplemente gaussiano no sería adecuado para modelar este tipo de series. Se han propuesto y empleado diversos métodos para analizar y simular el caudal diario. Varios de ellos se presentan a continuación.

1.4.1 Modelos auto-regresivos o markovianos

En general el patrón de picos y recesiones es ignorado en los modelos auto-regresivos (AR), dada su naturaleza fundamentalmente empírica. Es posible observar esta omisión en los modelos propuestos por Beard (1967) y por Quimpo (1968). En el primero de estos casos se generaron primero caudales mensuales por medio de una cadena de Markov simple (cada caudal se constituía por un componente correlacionado con el caudal del mes anterior y por un componente aleatorio) y luego se desagregaron las cantidades totales mensuales asignando las correspondientes a cada día (por medio de una cadena de Markov de segundo orden, utilizando variables estandarizadas).

En el segundo caso, se utilizó un modelo libre de tendencia, al que se asocia un componente oscilatorio (detectado y aislado usando análisis espectral de Fourier) y un proceso auto-regresivo (un esquema de Markov ajustado a la serie de residuos estandarizados). Este procedimiento, que en términos más simples es un modelo AR aplicado a los valores de caudal diario estandarizados (se les elimina el componente determinístico) es básicamente un modelo

AR de segundo orden. Payne y otros (1969) utilizaron una cadena de Markov de primer orden utilizando un enfoque similar al de Beard (1967) pero incluyendo un coeficiente para reflejar las características de la cuenca. Este trabajo incluye un aporte de un reajuste de los hidrogramas para aclarar el efecto de los fenómenos en una escala diaria, y un factor de amortiguamiento para producir simulaciones cuyas distribuciones estadísticas anuales fuesen similares a las históricas.

Ninguno de los análisis anteriores resultó satisfactorio para modelar el caudal diario, dado que los modelos AR (o en términos generales los modelos de la familia ARIMA) no son capaces de reproducir las características de los hidrogramas implícitamente (Lawrence y Kottegoda, 1977) ni los valores extremos o las propiedades de cruce (Kottegoda y otros, 1995), y en general sólo pueden preservar características de mayor escala como la media, la desviación estándar y varios órdenes de auto-correlación (Aksoy y Bayazit, 2000a).

Kottegoda y Natale (2000) tratan de corregir este problema utilizando modelos auto-regresivos gamma, con procesos ARIMA no-normales para simular una recesión retrospectivamente hasta alcanzar el nivel del impulso que la originó. Las características de mayor escala temporal y las periódicas son modeladas a través de análisis espectral, mientras la sección creciente del hidrograma se simula utilizando un generador de números aleatorios uniformes. Con este modelo compuesto se logró mejorar el desempeño de procesos AR o ARIMA simples y mantener las características del hidrograma y de las distribuciones del promedio, los caudales bajos y los picos.

También se han empleado modelos PARMA (ARMA periódicos) pero al igual que los anteriores son esencialmente empíricos y su objetivo se limita a obtener el mejor ajuste de la serie analizada, que es lo que determina su tipo y orden (Claps y Rossi, 1995). Por esto, el proceso estocástico cambia y se ajusta a las características de cada serie cronológica, lo cual imposibilita tanto el uso de consideraciones adicionales al estimar los parámetros, como su respectiva validación siguiendo criterios hidrológicos.

1.4.2 Procesos de Poisson filtrados o procesos de ruido disparado

Como respuesta al problema de la forma de “diente de sierra” de un hidrograma típico de caudal diario, que produce un comportamiento altamente asimétrico más que gaussiano, surgen los modelos de ruido disparados (conocidos como shot noise process, SNP). Un modelo SNP se basa en un proceso tipo conteo e intenta reproducir las características de corto plazo de los caudales diarios, conservando a la vez las propiedades de largo plazo.

Estos modelos también se conocen como procesos de Poisson filtrados (FPP) y su origen suele atribuirse a Parzen (1962). Bernier (1970) analizó el uso de los SNP en la representación de caudales diarios, después de hacer una revisión de los modelos teóricos y empíricos existentes en ese momento.

Su conclusión más relevante es la utilidad de los modelos que se basan en la estructura específica de los procesos físicos, que en este caso estaría dada por el carácter Poisson de los procesos de lluvia y por una función de transferencia determinística entre lluvia y caudal. En este trabajo, además, se adelanta la dificultad en el procedimiento de estimación para los modelos conceptuales discretos.

Weiss (1973) aplica los SNP (en una variación denominada modelo doble de ruido disparado) a series de caudal diario y logra comprobar su mayor realismo físico en el efecto de las recesiones simuladas. La estimación de los parámetros se hace por medio del método de momentos. En un trabajo posterior Weiss (1977) le incluye elementos de análisis al modelo clásico SNP para tomar en cuenta la estacionalidad en los caudales diarios y generaliza el modelo doble de ruido disparado en una familia que denomina modelos de ruido disparado de segundo orden.

1.4.3 Modelo estocástico-conceptual

Un tercer tipo de modelo es el estocástico-conceptual (CS). Éste no pretende analizar todos los procesos específicos que actúan en la transformación lluvia-escorrentía como lo hace un

modelo determinístico, pero tampoco pierde de vista consideraciones físicas que son ignoradas por los modelos empíricos (por ejemplo las versiones tradicionales del AR, ARIMA o PARMA).

Los modelos CS se basan en hipótesis conceptuales que restringen las formas que estos pueden adoptar, así como su orden. Pretenden una explicación causal entre las variables de entrada y las de salida. De este modo, son inherentemente parsimoniosos dado que su complejidad está en función de cada característica que se identifique en las series observadas.

Además, ante una estructura poco compleja, las mencionadas hipótesis pueden tomarse como base para la estimación de los parámetros. En caso de carencia de datos sólo estos modelos permiten transferir información de cuencas similares en el contexto del análisis regional.

Por tanto, en los modelos CS se toma como sustento la serie observada de caudal y se separan las secciones crecientes y decrecientes del hidrograma para simularlas por aparte, dado que estas secciones representan procesos físicos diferentes. En las secciones crecientes se considera la naturaleza estadística de los eventos de precipitación que las propician (frecuencia y magnitud de los picos en la serie).

La representación conceptual de la respuesta de la cuenca se vale del vínculo entre modelos lineales de la hidrología estocástica y de la paramétrica, que está dado por la función de almacenamiento del sistema (Quimpo, 1971), y eventualmente por su función de recarga, en modelos más complejos. La función de respuesta o transferencia del sistema con frecuencia combina el efecto de embalses (usualmente) lineales y canales lineales y actúa sobre una entrada estocástica que es lo que define propiamente el modelo estocástico (Claps y Rossi, 1995).

Las ventajas de los modelos CS, que son típicas del enfoque conceptual, son la posibilidad de utilizar información a priori sobre la estructura estocástica de la entrada (en términos de precipitación total o efectiva), tanto para identificar del tipo y orden del modelo, como para limitar los problemas numéricos asociados a la estimación simultánea de los parámetros.

También es útil la posibilidad de establecer vínculos entre parámetros estimados y conceptuales para validar los primeros, hecho que es más valioso que las pruebas estadísticas, en caso de limitaciones en la cantidad o calidad de los datos (Claps y Murrone, 1993; Claps y Rossi, 1995).

Haciendo uso de una representación conceptual de la cuenca como dos embalses lineales Kelman (1980) elaboró un modelo que generó (con algunos problemas en la caracterización estacional) secuencias de caudales recesivos como una salida estocástica de estos embalses.

Claps y Rossi (1993, 1995) profundizaron más en la esquematización conceptual-estocástica al aplicarla en el análisis de caudales en múltiples escalas para procesos markovianos y de conteo. Pero tal vez lo más relevante es la propuesta de un marco de referencia para la construcción de modelos estocásticos-conceptuales que incluye pasos de identificación, estimación y validación adaptados para aprovechar la información a priori contenida en los procesos físicos representados.

Murrone y otros (1997) continúan trabajando con el marco de referencia antes mencionado para reproducir caudales en periodos cortos (de diario a semanal) basándose en una estructura físicamente consistente y requiriendo que el modelo pueda coexistir con aquellos que son adecuados para escalas más agregadas. Lo anterior fue abordado incluyendo en la función de respuesta de la cuenca un tercer componente de tipo embalse lineal.

Según el esquema de identificación se propuso obtener cada conjunto de parámetros recarga-almacenamiento de forma progresiva para cada componente según la secuencia: agua subterránea interanual, agua subterránea intermensual y agua sub-superficial. Para esto, se utilizaron respectivamente los modelos ARMA(1,1), PIR-ARMA(2,2) (modelo ARMA con residuos independientes periódicos) y SN, para una escala variable de T días. Dicha escala se obtuvo maximizando la varianza relativa del caudal superficial. En el caso intermensual se observó una mayor eficiencia del modelo SN, de forma que no se utilizaron los resultados del PIR-ARMA(2,2).

Los resultados de este procedimiento denotan la preservación de los efectos del componente subterráneo interanual en la serie simulada, lo que permite una mejor discriminación entre la influencia que se debe al agua superficial y la debida al escurrimiento directo. La simulación se basó en un modelo Poisson ruido blanco con pulsos exponenciales para reproducir la lluvia efectiva, y mostró un desempeño satisfactorio con respecto a la serie observada, con una excelente replicación de los máximos para distintas duraciones (Murrone y otros, 1997).

Un FPP de naturaleza conceptual más compleja fue planteado por Yue y Hashino (1999), pues combina una secuencia de tormentas en la forma de un proceso puntual marcado (con número de eventos binomial, Poisson ó binomial negativo, dependiendo de su relación media-variancia), con una función de respuesta conceptual tipo lluvia-esorrentía en la que tres tanques en serie representan los flujos sub-superficiales rápidos, los sub-superficiales retardados y los subterráneos, y otro tanque en paralelo describe el escurrimiento superficial. Los parámetros de la función de respuesta son los tamaños de “abertura” de los cuatro tanques, las dos capacidades de infiltración entre los tres tanques en paralelo, y la máxima altura de almacenamiento del tanque superior en el esquema en serie.

Con un enfoque análogo al expuesto en Yue y otros (1999) se planteó la derivación de las estadísticas básicas para caudales promediados J días (media, variancia y coeficiente de correlación) como función de los parámetros de las propiedades estocásticas de la lluvia, la función de respuesta de la cuenca (en forma de los parámetros de los tanques del modelo) y el período sobre el que se promedia el caudal. La aplicación de este modelo fue adecuada en términos la reproducción de las medias y variancias mensuales, y de las promediadas en cinco días, pero no fue así para las correlaciones. No obstante, los autores comentan la utilidad del método para obtener las características del caudal a partir de registros cortos basándose en información de lluvia más completa.

Claps y Laio (2003) propusieron un procedimiento objetivo de eficiencia para la identificación de picos en un registro continuo de caudal, denominado picos filtrados sobre el umbral (FPOT). El FPOT se basa en una prueba de independencia de los picos y en otra de bondad de ajuste con respecto al modelo Poisson-Pareto, y fue utilizado por Claps y otros (2005) para

generar los eventos de lluvia efectiva de forma tal que cumplieran los supuestos del modelo Poisson requeridos por el FPP clásico. También se proponen y discuten herramientas simples para evaluar y comparar modelos por medio de índices objetivos de desempeño tomando en cuenta su parsimonia.

1.4.4 Modelos de probabilidades de transición

Otro grupo de modelos relacionados con los FPP son los que Aksoy y Bayazit (2000a,b) identifican como modelos de probabilidades de transición (TP). Toman su nombre del análisis de las probabilidades de transición entre días “húmedos” y secos (días “húmedos” son aquellos en los que se da un incremento de caudal, y secos los que no presentan esta condición). Es común encontrarlos en la literatura como modelos de Treiber-Plate, pues su origen está en el trabajo de Treiber y Plate (1977). En éste se puso énfasis en varios factores como los picos (definiendo sus días de ocurrencia y sus alturas), la forma de las recesiones, el volumen de las ondas, y el rango.

La base de los modelos TP es la convolución (también conocida como estimación inversa) de pulsos de caudal correlacionados. Estos pulsos se dan en tiempos determinados por una cadena de Markov en vez de un proceso Poisson, su magnitud proviene de procesos auto-regresivos de primer orden, e incluye un componente aleatorio a partir de una distribución exponencial transformada. La convolución se define por medio de una función de transferencia dependiente de la altura del pulso, con parámetros obtenidos por mes para tomar en cuenta la estacionalidad.

Kottegoda y Horder (1980) estudiaron la lluvia diaria con un modelo probabilístico y estimaron los caudales creando otro modelo de tipo lluvia-escorrentía a partir de los pulsos y de una función de transferencia variable estacionalmente. Con estos principios Schneider y Schultz (1982) generaron secuencias simultáneas de escorrentía diaria en múltiples sitios.

Kron y otros (1990) también utilizaron el modelo Treiber-Plate con un enfoque multi-sitio, de forma que correlación entre las series artificiales se determina por medio de probabilidades

condicionales al producir los pulsos. La convolución de los pulsos se efectúa con una función del sistema específica para el sitio correspondiente.

Kottegoda y otros (1995) continúan con la aplicación del modelo TP siguiendo el enfoque anterior de Kottegoda y Horder (1980), pero utilizando datos de caudal para estimar los días secos o “húmedos” dada la carencia de información de lluvia. Su justificación radica en la sensibilidad de la respuesta de la cuenca estudiada. A grandes rasgos su procedimiento es generar secuencias secas y “húmedas” consecutivas, producir los caudales bajos, simular las curvas de recesión y luego la sección creciente de los hidrogramas. Pese a las limitaciones encontradas, esta técnica resulta adecuada para modelar eventos extremos de bajo caudal.

Rajagopalan y otros (1996) introduce una variante en los modelos TP al proponer un modelo no homogéneo de Markov de un paso y dos estados para simular precipitación, estimando las matrices diarias de transición de forma no paramétrica, a través de un promedio ponderado de los conteos de transición en un lapso simétrico alrededor del día evaluado. Con este modelo se lograron reproducir las funciones de probabilidad-masa de las longitudes de los periodos secos y “húmedos”, así como los estadísticos de los totales de lluvia.

Aksoy y Bayazit (2000a,b) elaboraron un modelo TP para caudales intermitentes en el que determinaron el estado de flujo (inexistente, incrementándose, disminuyéndose) por medio de una matriz de transición de probabilidades de una cadena de Markov de tres estados. Luego estimaron la magnitud de los incrementos (suponen distribución gamma) y finalmente las curvas de recesión (como función de decrecimiento determinística de forma exponencial, con parámetros mensuales diferentes en las partes altas y bajas).

Este modelo logró reproducir exitosamente el porcentaje de tiempo sin flujo, las medias y desviaciones estándar diarias y mensuales, las características de las secciones ascendentes del hidrograma, el número, magnitud y altura de los picos y los valores extremos máximos. No obstante, generó caudales mínimos menores que los observados.

Xu y otros (2001) desarrollaron el modelo de pulsos con correlación dependiente de una cadena Markov (CDMCP) con el cual intentaron describir el caudal diario en múltiples sitios. Aquí los pulsos pueden estar auto-correlacionados y también correlacionados de forma cruzada en las escalas temporal y espacial, respectivamente. Los pulsos se constituyen por una cadena de Markov de primer orden y dos estados, que se complementa con una función de respuesta variable. Con el CDMCP se lograron preservar la estructura de corto plazo del proceso de eventos, así como parte de las características del caudal estacional. No obstante los autores consideraron que posiblemente la cadena de primer orden era insuficiente y que los coeficientes de correlación cruzada entre los diferentes sitios debían ser mejorados sustancialmente.

En un intento más de establecer enlaces entre propiedades físicas y modelación estocástica Xu (2003) propuso un modelo llamado pulso de correlación cruzada de Markov. Con él simuló series en múltiples sitios, integrando ocurrencias aleatorias de periodos “húmedos” y secos y evaluando los pulsos según el enfoque de Treiber. El modelo resultó satisfactorio para reproducir la distribución de frecuencia del caudal diario, y la media, desviación estándar y primera auto-correlación de los valores diarios, evaluadas mes a mes, pero no fue del todo satisfactorio para emular las distribuciones de los caudales medios mensuales o de los máximos anuales.

El trabajo de Aksoy (2003, 2004) continúa enfocado en caudales intermitentes, y profundiza en los modelos generadores de pulsos basados en cadenas de Markov, estudiando el comportamiento de tres de ellos. El primero se basa en dos cadenas de dos estados (impulso, no impulso y aumento, disminución), mientras en el segundo son de tres estados (no caudal, aumento o disminución). El tercero utiliza análisis armónico y ajusta series de Fourier a la cadena de Markov de tres estados.

A las diferentes opciones de generación de pulsos se agregaron técnicas como la definición de la curva ascendente con distribución gamma de dos parámetros, y la de la curva de recesión por medio de una función de decrecimiento exponencial. El número de pulsos, así como el de picos (definidos por el autor cuando los dos caudales adyacentes tienen valores inferiores) fue

reproducido satisfactoriamente por los tres modelos generadores de pulsos, aún cuando el que utiliza series de Fourier erró su estimación en un par de meses.

1.4.5 Modelo de proceso puntual estocástico con base conceptual

Xu y otros (2002) pretende resumir los modelos estocástico-conceptual (CS), de ruido disparado (SN) y de transición de probabilidades (TP) en uno sólo, que denominó proceso puntual estocástico con base conceptual (CBSPP). Éste abarca tanto los aspectos estocásticos (tormentas definidas como un proceso puntual de Poisson no homogéneo compuesto), como los relativos a los fenómenos hidrológicos asociados (escurrimiento directo proveniente de un almacenamiento en cascada de embalses lineales superficiales, así como caudal base derivado del almacenamiento sub-superficial y subterráneo).

Antes de aplicar el modelo se comprobó que en cada mes las ocurrencias de las tormentas tuviesen la distribución supuesta teóricamente, lo cual en términos generales resultó satisfactorio. Los parámetros de la distribución de las intensidades asociadas a los pulsos se estimaron suponiendo un comportamiento tipo Gamma. Luego se ejecutó la simulación y se evaluaron estadísticos mensuales. Las medias y las variancias se reprodujeron adecuadamente, no así el sesgo ni la curtosis, en tanto que los máximos y mínimos son coherentes.

También se analizaron la bondad de ajuste entre la distribución de las intensidades de los pulsos observados y la distribución marginal simulada, que resultó satisfactoria. Al estudiar la correlación entre las dos estaciones empleadas, los correlogramas mensuales observados y simulados resultaron similares. Las características observables del flujo reflejan que la simulación reproduce la forma aserrada.

No obstante lo anterior, se concluyó que el modelo debería mejorarse en sus tres aspectos esenciales: (i) que los componentes del proceso puntual sean agrupados, con técnicas como las expuestas por Xu y otros (1998), (ii) que su función de respuesta sea variable, como proponen Kottegoda y Horder (1980) para funciones dependientes del tiempo, y Xu y otros (2001) para

variaciones dependiendo del caudal, y (iii) que se mejore notablemente modelo de correlación cruzada para múltiples sitios.

1.5 Objetivos de la investigación

La gestión de los recursos hídricos es un tema relativamente novedoso y poco analizado en Costa Rica, en comparación con el desarrollo que éste ha tenido en países más desarrollados y/o con mayor escasez de agua dulce, dadas sus condiciones naturales. De ahí que actualmente se conoce una gran cantidad de métodos para evaluar el caudal diario en los ríos, pero estos no han sido desarrollados ni validados para las condiciones tropicales de la región centroamericana.

Por su nivel de detalle, la caracterización del caudal diario que en épocas anteriores no resultaba crítica, ha empezado a serlo en tiempos más recientes, dado que hay mayor demanda de recursos hídricos por parte de la población, requerimientos más especializados en el desarrollo industrial, mayores exigencias en la planificación y operación de proyectos y políticas del agua más severas en términos ambientales y sociales.

Costa Rica cuenta con una red de estaciones hidrológicas que abarca la mayoría de las cuencas del país. No obstante, para muchas de las exigencias antes mencionadas se requieren registros de caudal muy grandes, con los que generalmente no se cuenta. Esta particularidad no sólo afecta a Costa Rica sino que es un problema común en el campo de la hidrología. Los métodos mencionados para analizar el caudal diario existen gracias a esa necesidad, pero no necesariamente estos modelos son directamente generalizables, como sí lo es el problema que les da origen.

Es en este punto donde debe plantearse la importancia de proponer y corroborar un modelo para la simulación del caudal diario en los ríos de Costa Rica, adaptado a características particulares de la región, como las condiciones climáticas tropicales y la fisiografía y uso del suelo de estas cuencas.

El objetivo principal de la presente investigación es aplicar y comprobar la validez de los procesos de Poisson filtrados, con sus posibles variantes, para la generación de series sintéticas de caudal diario. Este método podría preservar no sólo los estadísticos más importantes de las series estudiadas, sino también sus patrones más representativos y fundamentales en términos de los procesos físicos que están detrás, o sea, los picos y las recesiones.

Otros objetivos que se pretende alcanzar son:

Verificar que las series generadas conserven las propiedades de cruce de una serie histórica observada en una estación hidrológica.

Comprobar que el método logra reproducir adecuadamente el patrón de recesiones característico en los caudales diarios y otras propiedades importantes.

Estimar escenarios críticos para periodos de recesión, que puedan ser utilizados para diversas aplicaciones tal como la evaluación de caudales de compensación en proyectos hidroeléctricos.

Capítulo 2 Teoría y metodología

2.1 Definición de una serie de tiempo hidrológica

Se define el caudal como el volumen de agua que transcurre a través de una sección del cauce de un río en un determinado intervalo de tiempo. Posee dimensiones de volumen entre tiempo y según el sistema internacional de unidades se especifica en metros cúbicos por segundo (m^3/s). Así, para diversos momentos t (usualmente espaciados igualmente) se considera que el caudal medido $x(t)$ es una muestra de una población infinita $X(t)$.

La esorrentía natural es un proceso continuo en el tiempo, pero su registro usualmente se hace de manera discreta y se denotará por x_t . La muestra de tamaño n se compone por los valores x_1, x_2, \dots, x_n ; donde todos los valores son contiguos según la unidad de tiempo y denominados en conjunto como una realización de X_t . No obstante, es evidente que este tipo de muestra no contiene observaciones independientes y tanto la descripción de la misma como la inferencia a partir de su análisis y modelación, deben introducir las estructuras autocorrelación necesarias.

Por lo anterior, el análisis de las series de tiempo suele demandar un correcto aislamiento de tendencias, factores estacionales, factores cíclicos, saltos en la media y/o variancia y el estudio de problemas de heterocedasticidad. Estos componentes tienen efectos relativos variantes sobre el proceso observado dependiendo de la escala de análisis. Su adecuada identificación permite separar el fenómeno aleatorio implicado, que es lo que posibilita plantear un modelo estocástico adecuado para dicho proceso.

2.2 El proceso Poisson

Es un proceso de conteo de valores enteros $\{N(t), t \geq 0\}$, que son eventos aleatorios distribuidos por algún mecanismo estocástico. Usualmente se trabaja con el proceso Poisson caracterizado como el proceso de valores enteros con incrementos independientes estacionarios y saltos unitarios (también conocido como el modelo de incrementos independientes). Si los eventos

ocurren en intervalo de tiempo $(0, \infty)$, para $t > 0$ se tiene $N(t)$ como valor entero positivo igual al número de eventos ocurridos en el intervalo $(0, t]$, abierto en 0 y cerrado en t . El proceso estocástico asociado debe satisfacer el siguiente sistema de axiomas para considerarse de tipo Poisson:

1. Puesto que se empiezan a contar los eventos en $t=0$ se define que $N(0)=0$
2. $\{N(t), t \geq 0\}$ posee incrementos independientes
3. Para cualquier $t > 0$, $0 < P[N(t) > 0] < 1$.
4. Para cualquier $t \geq 0$, $\lim_{h \rightarrow 0} \frac{P[N(t+h) - N(t) \geq 2]}{P[N(t+h) - N(t) = 1]} = 0$
5. El proceso de conteo $N(t)$ tiene incrementos estacionarios, o sea, para dos puntos cualquiera $t > s \geq 0$ y cualquier $h > 0$, las variables aleatorias $N(t) - N(s)$ y $N(t+h) - N(s+h)$ están idénticamente distribuidas.

La derivación axiomática de la distribución del proceso Poisson puede hacerse definiendo $P_n(t) = P[N(t) = n]$. Es posible demostrar que para instantes t y s cualesquiera que definen lapsos $(0, s]$ y $(s, t]$ con probabilidades respectivas $P(s)$ y $P(t)$ se cumple que:

$$P_n(t+s) = \sum_{i=0}^n [P_{n-i}(s) \cdot P_i(t)] \quad (2.1)$$

Suponiendo que el lapso analizado se divide en pequeños sub-intervalos de tamaño h de manera que en cada uno puede darse cuando mucho un evento, y que existe un entero positivo ν tal que:

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1 - P[N(h) = 0]}{h} &= \nu \\ \lim_{h \rightarrow 0} \frac{P[N(h) = 1]}{h} &= \nu \\ \lim_{h \rightarrow 0} \frac{P[N(h) \geq 2]}{h} &= 0 \end{aligned} \quad (2.2)$$

Entonces restando $P_n(t)$ de la ecuación (2.1), aplicándole el límite cuando $h \rightarrow 0$, utilizando las relaciones en (2.2) y valiéndose de la definición de derivada se puede llegar a:

$$\frac{dP_n(t)}{dt} = -\nu \cdot P_n(t) + \nu \cdot P_{n-1}(t)$$

Es posible verificar que una solución a la ecuación diferencial anterior es:

$$P[N(t)=n] = e^{-\nu t} \frac{(\nu t)^n}{n!}$$

Que es la típica distribución de probabilidad de un proceso Poisson para un número de eventos $n=0,1,2,\dots$; en un tiempo $t>0$ y con una tasa de ocurrencia $\nu>0$.

Esto es válido para un proceso de Poisson homogéneo, de modo que su intensidad promedio ν es constante. Esto es $E[N(t)]=\nu t$. Se debe agregar que $V[N(t)]=\nu t$.

2.3 El proceso de Poisson filtrado

Un planteamiento general de los procesos de Poisson filtrados (FPP) fue expuesto por Parzen (1962), en el que define un proceso estocástico como FPP si éste se origina a partir de la aplicación de operaciones lineales sobre un proceso Poisson $N(t)$. Si se especifican:

- $\tau_m, m=1,2,\dots$: tiempos en que se producen eventos de $N(t)$
- Y : variable aleatoria con una distribución particular que toma valores y_m para los tiempos τ_m , que son independientes e idénticamente distribuidos entre ellos, así como independientes con respecto a $N(t)$
- $w(t,y)$: función continua dependiente del tiempo y sujeta a los incrementos y .

Entonces un FPP se define por:

$$X(t) = \sum_{m=1}^{N(t)} w(t - \tau_m, y_m)$$

El proceso se parte un punto de referencia $t=0$, con un valor asociado de inicio $X(0)$.

La interpretación de Weiss (1977) sobre el FPP se basa en la asociación directa de la función $w(t,y)$ con la función de transferencia lluvia-escorrentía, conocida en hidrología como el hidrograma unitario. Cada evento de lluvia presente en un momento τ_m genera un volumen de lluvia y_m que es disipado conforme corre el tiempo t a partir del origen del evento. Éste modelo supone, entonces, que las precipitaciones son instantáneas y secuencialmente independientes.

Los FPP pertenecen a una clase más general de modelos llamados procesos puntuales filtrados (FPPN). Como alternativa a la suposición de que el número de eventos se producen con

distribución Poisson, se han estudiado FPNP con distribuciones más complejas como la binomial, la binomial negativa, o una combinación de éstas cuando se aplican FPNP no homogéneos en el tiempo (Yue y otros, 1999). Sin embargo, en estos últimos casos se debe trabajar con un parámetro adicional, de forma que tanto la estimación estacional de los parámetros como el cálculo de los estadísticos promediados (se expondrán en la sección 2.5) se complicarían notablemente. Por esto, si el FPP es adecuado para representar los caudales registrados, es probable que sea la mejor opción para aplicarlo repetitivamente o en múltiples sitios, dado su mayor nivel de parsimonia.

Hasta ahora sólo se ha hablado de procesos de Poisson filtrados estacionarios. El caudal diario no presenta esta característica de modo que el proceso $N(t)$ debería ser de tipo no homogéneo, y tanto la distribución de las intensidades de impulso y_m como la función de transferencia w dependientes de τ_m . Aunque existen análisis para tomar en cuenta de forma detallada la estacionalidad en los FPP, estos pueden tornarse muy complejos al incrementar lo que se conoce como el orden del modelo (este concepto se definirá más adelante). La manera usual de lidiar con este problema es, entonces, estimar parámetros mensuales para $N(t)$, para la distribución de y_m , y eventualmente para w .

La escogencia de la distribución de Y y la forma de w definen diversos tipos de modelos FPP. Uno de los más conocidos es el proceso de ruido disparado (SNP), que consiste en un FPP homogéneo (con tasa de eventos ν) en el que Y posee distribución exponencial con media θ y función de densidad de probabilidad:

$$f(y) = \frac{1}{\theta} \cdot e^{-\frac{y}{\theta}}, \quad y > 0 \quad (2.3)$$

Otras distribuciones que se han sugerido para Y son la pareto generalizada y la gamma (Xu y otros, 2002; Claps y Laio, 2003), pero el incremento en un parámetro introduce una complicación adicional en el análisis que no siempre se justifica (ver la sección 2.3.2).

Si en el análisis del escurrimiento superficial se supone que el decaimiento (el término correspondiente en hidrología es recesión) es exponencial, con una tasa b , la función de

respuesta (que es la amortiguación de un impulso) sería $h(t)=e^{-bt}$. Si la función de transferencia adopta la forma:

$$w(t, y) = y \cdot h(t), \quad t > 0$$

entonces $w(t, y) = y \cdot e^{-bt}$. Todo lo anterior permite formular un SNP como:

$$X(t) = \sum_{m=1}^{N(t)} [y_m \cdot h(t)] \quad (2.4)$$

Para la recesión exponencial: $X(t) = \sum_{m=1}^{N(t)} [y_m \cdot e^{-bt}]$, con punto de partida $t=0$ y valor de inicio $X(0)$.

Los parámetros del SNP son la tasa de eventos v , el impulso promedio θ y la tasa de decaimiento b . Se ha demostrado que la distribución $f(x)$ de $X(t)$ es gamma con parámetros $(1/\theta, v/b)$. Se trata de una distribución no-negativa y sesgada positivamente, con densidad de probabilidad:

$$f(x) = \frac{(1/\theta)^{v/b} x^{v/b-1}}{\Gamma(v/b)} e^{-x/\theta}$$

y con momentos:

$$\begin{aligned} E[X(t)] &= v \cdot \theta / b \\ V[X(t)] &= v \cdot \theta^2 / b \\ P[X(t), X(t+s)] &= e^{-bs} \quad (s > 0) \end{aligned} \quad (2.5)$$

El proceso SNP como una suma de impulsos y recesiones se ejemplifica en la Figura 2.1. La parte a) presenta un hidrograma de caudal diario típico, en tanto la b) resalta los eventos como impulsos indicados según $N(t)$ en la secuencia definida por m . En este caso particular la ocurrencia de los eventos en el tiempo t podría darse como $t_m = \{0, 0, 1, 0, 0, 0, 1, 0, 0, 1, \dots\}$, o simplemente como $\tau_m = \{3, 7, 10, \dots\}$. No obstante, la Figura 2.1b) plantea un problema en la definición de los impulsos, pues la manera en que se presentan sus magnitudes y_m no reflejaría adecuadamente la operación lineal de sumatoria sobre los eventos del proceso Poisson, lo cual es precisamente lo que define a cualquier FPP.

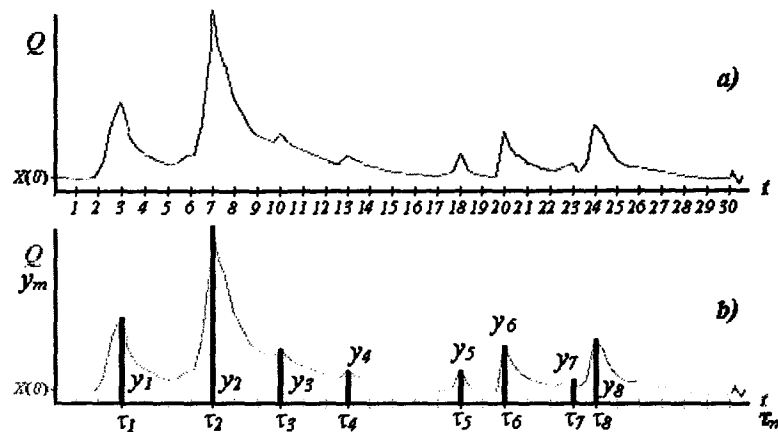


Figura 2.1 El proceso de ruido disparado (SNP) como una secuencia de impulsos y recesiones

De esta manera, tal definición demanda la formulación de un procedimiento de identificación de la lluvia efectiva. Éste término hidrológico se refiere a la fracción de la precipitación que llega al cauce como escorrentía directa, y es la que produce los aumentos repentinos de caudal por encima de un nivel base.

2.3.1 Identificación de la lluvia efectiva en el FPP y estimación de la tasa de eventos ν

Un enfoque utilizado con frecuencia para identificar los eventos de lluvia efectiva consiste en asociar su ocurrencia con cualquier incremento en la serie de caudal. No obstante, Claps y Laio (2003) mencionan que este procedimiento tiende a producir secuencias de pulsos con una tasa de eventos ν muy grande. Lo anterior se debe a que errores de medición pueden producir pequeños incrementos en el caudal, que no deben confundirse con eventos de lluvia efectiva. Además usualmente los resultados del método no respetan las hipótesis de independencia de los eventos y de que están distribuidos de forma Poisson.

Como alternativa, estos autores proponen el método denominado picos filtrados sobre el umbral (FPOT), que garantiza el cumplimiento de las hipótesis anteriores. La determinación de los FPOT se basa en los siguientes pasos. (i) Se identifican todos los máximos locales en la serie, definiendo una secuencia denominada picos reales (AP). (ii) Al restar a los AP los respectivos mínimos locales anteriores, se obtiene una serie de picos filtrados (FP), la cual

podría interpretarse como una aproximación del componente de lluvia efectiva. (iii) Se aplica un filtro umbral a la secuencia FP para mantener sólo los picos suficientemente grandes. El criterio para la escogencia del umbral debe remover el componente de ruido que genera picos falsos en la serie FP, de forma que se genere una muestra Poisson independiente. Ésta constituye la serie FPOT.

Claps y otros (2002) exponen la utilidad del segundo paso del procedimiento anterior para aislar y eventualmente modelar la lluvia efectiva, tomando en cuenta el efecto confundidor de eventos sucesivos dependientes y el fenómeno del caudal base (la parte del caudal que entra al cauce procedente principalmente del agua subterránea, y que es independiente de la precipitación). En este caso, se definiría la lluvia efectiva como la fracción de la lluvia que se transforma en escorrentía a lo largo de un día.

La aplicación de los primeros dos pasos en el procedimiento proporciona una secuencia de impulsos como la que se presenta en la Figura 2.2. Los círculos indican un máximo relativo y los cuadrados un mínimo. Las barras negras corresponden con los picos de lluvia efectiva que componen la serie FP.

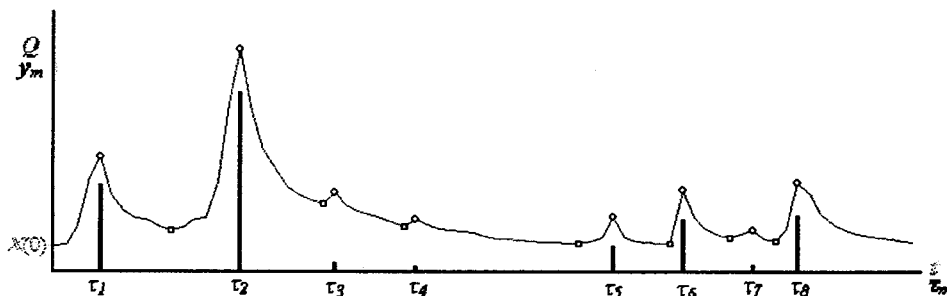


Figura 2.2 Identificación de la lluvia efectiva

Comparando la Figura 2.1 con la Figura 2.2 es posible plantear el supuesto de que los eventos identificados en los tiempos τ_3 , τ_4 , τ_7 de este ejemplo son producto del error de medición del caudal, y que por consiguiente, pueden provocar distorsiones en los análisis estadísticos sobre las frecuencias y las magnitudes de y_m que se requieren para la modelación. El tercer paso del procedimiento FPOT pretende corregir este problema valiéndose de un filtro umbral. Como se indicó antes, cualquier criterio para la escogencia del umbral debería procurar el cumplimiento de las hipótesis de independencia y distribución Poisson de los eventos. Lo anterior es posible

utilizando la prueba τ de Kendall y la prueba de Cunnane respectivamente, de modo que se llegue a un umbral suficientemente alto como para satisfacer ambas hipótesis conjuntamente. El efecto del incremento en el umbral es la exclusión de los eventos menores en la serie FP, lo que aumenta las distancias entre los picos restantes y definen los eventos de la serie FPOT. Por esto, los picos tienden a ser independientes al escogerse umbrales más elevados. Otro efecto que se produce es que el número de cruces del umbral tiende a convertirse en una variable aleatoria independiente, lo cual se manifiesta en su convergencia asintótica al comportamiento Poisson (Claps y otros, 2005).

Una prueba más se puede incorporar al criterio de selección del umbral. Ésta se relaciona con la magnitud de los eventos, que de tener una distribución exponencial según la ecuación (2.3) cumpliría con una distribución pareto generalizada con parámetro de forma igual a cero. Para esto, Choulakian y Stephens (2001) propusieron y probaron el uso del estadístico W^2 de Cramér–von Mises y su versión modificada, el A^2 de Anderson–Darling.

En algunas pruebas, al estimar los parámetros ν y θ asociados con los impulsos del modelo en una forma variable en el tiempo (para incluir los efectos estacionales que presentan las precipitaciones), se produce una considerable reducción en los tamaños de muestra. Por ejemplo, si se estiman parámetros mensuales, la prueba debe aplicarse sobre una muestra de cada mes, que equivaldría a un doceavo de la muestra de un modelo no estacional.

Lo anterior no incide en la significancia de la prueba de Cunnane ni en la τ de Kendall, pues no es necesario dividir la muestra. Así, aunque se calculen ν_ε estacionales, el proceso Poisson subyacente se mantendría, al igual que la independencia entre sus eventos, dado que ésta permanece sujeta a la escogencia de umbrales adecuados.

Sin embargo, al suponer θ_ε variable en intervalos de tiempo estacionales sí se debe efectuar una segregación en la muestra al evaluar la distribución de la intensidad de los pulsos. Esto implica una pérdida excesiva de precisión en las pruebas W^2 o A^2 , pues se tendrían que aplicar en cada intervalo de tiempo en el que se suponen valores de y_m homogéneos. En consecuencia, cuando se aplican modelos con variación estacional, Claps y otros (2005) recomiendan excluir

estas últimas pruebas del procedimiento para la escogencia del umbral FPOT, que inicialmente fue planteado por Claps y Laio (2003).

Evaluación de la independencia de los impulsos

Según Ferguson y otros (2000) la prueba de independencia τ de Kendall es más potente y funciona mejor en muestras pequeñas que otros procedimientos para comprobar la independencia serial (como el rho de Spearman). Su aplicación para evaluar los eventos $i_1, i_2, \dots, i_N, N > 3$, de la serie FPOT consiste en comparar las magnitudes o rangos R_i de los N pulsos a evaluar, de forma que sea posible estimar la dependencia serial de primer orden. Lo anterior se logra calculando el número de discordancias N_d en la variación del tamaño de dos parejas de rangos adyacentes. Por ejemplo, si se tienen dos parejas de rangos (R_i, R_{i+1}) y (R_j, R_{j+1}) que satisfacen el criterio $R_i < R_j$ y $R_{i+1} > R_{j+1}$ (o de manera equivalente $R_i > R_j$ y $R_{i+1} < R_{j+1}$) el número de discordancias se puede estimar por medio de:

$$N_d = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} I(R_i < R_j; R_{i+1} > R_{j+1}) \quad (2.6)$$

Donde $I(\cdot)$ es una función indicadora que vale uno si la expresión lógica de su argumento se cumple y cero si no es así (en este caso el argumento consiste en dos inecuaciones que deben cumplirse de manera obligatoria). Ejemplificando lo anterior, el argumento de $I(\cdot)$ para $i=1$ en la ecuación (2.6) sería la secuencia de comparaciones $R_1 < R_1$ y $R_2 > R_2$, $R_1 < R_2$ y $R_2 > R_3$, $R_1 < R_3$ y $R_2 > R_4$, ..., $R_1 < R_{N-1}$ y $R_2 > R_{N-1}$.

Basándose en las discordancias se puede definir el estadístico de Kendall como:

$$\tau = 1 - \frac{4 \cdot N_d}{(N-1)(N-2)} \quad (2.7)$$

Las ecuaciones expuestas para N_d y τ corresponden al modelo conocido como no circular, que consiste en excluir la secuencia (R_N, R_1) de las comparaciones, y por tanto, variar la cantidad total de eventos comparados en las ecuaciones (2.6) y (2.7).

Según la hipótesis nula de picos subsecuentes independientes para $N \geq 4$, τ se aproxima a una variable aleatoria normal con media y variancia dadas por:

$$E(\tau) = -\frac{2}{3(N-1)}$$

$$V(\tau) = \frac{20 \cdot N^3 - 74 \cdot N^2 + 54 \cdot N + 148}{45 \cdot (N-1)^2 \cdot (N-2)^2}$$

Estos estadísticos permiten la construcción de una prueba de hipótesis de dos colas con la hipótesis nula de independencia, ante la hipótesis alternativa de que haya una dependencia positiva (o negativa), basándose en el estadístico de prueba $T_N = [\tau - E(\tau)] / \sqrt{V(\tau)}$. Si este valor es mayor que una t de student $t_{N,1-\alpha/2}$ (o menor que $t_{N,\alpha/2}$) se rechazaría la hipótesis nula con un nivel predeterminado α tal que si rige H_0 , entonces $P(T_N < t_{N,\alpha/2}) = P(T_N > t_{N,1-\alpha/2}) = \alpha/2$. Para series largas se comprobaría que $\lim_{n \rightarrow \infty} P(T_N > z_{1-\alpha}) = \lim_{n \rightarrow \infty} P(T_N > z_{1-\alpha}) = \alpha/2$, donde z es una normal estándar. Expresado como un intervalo de confianza para τ se obtendría:

$$\tau_{\alpha/2, 1-\alpha/2} = E(\tau) \pm z_{\alpha/2} \cdot \sqrt{V(\tau)} \quad (2.8)$$

Así, la ecuación (2.8) tendría un $z_{\alpha/2}$ igual a 1.96 para evaluar los límites de τ . Éstos delimitarían al τ de Kendall de la ecuación (2.7) con un 95% de confianza de que no haya dependencia entre los pulsos seleccionados.

Evaluación del comportamiento Poisson de los impulsos

La prueba de Cunnane para valorar si una serie de impulsos sigue la distribución Poisson consiste en definir límites de aceptación para el índice de dispersión I_d , que es la razón entre la varianza y la media. Lang y otros (1999) presentan un procedimiento sencillo para aplicar esta prueba, que se basa en la suposición de que el número de eventos anual $N_{e|a}$ tiene una distribución aproximadamente normal cuando el parámetro de Poisson ν_a (con tiempo t en escala anual) es suficientemente grande ($\nu_a > 5$). Tomando en cuenta que la media y la variancia de la distribución Poisson son iguales, se propone que el estadístico de prueba

$I_d \cdot (n_a - 1) = \sum_{i=1}^{n_a} [N_{e|a}(i) - \bar{N}_{e|a}]^2 / \bar{N}_{e|a}$ tiene una distribución χ^2 con $(n_a - 1)$ grados de libertad,

donde $\bar{N}_{e|a} = \sum_{i=1}^{n_a} [N_{e|a}(i)] / n_a$ es el promedio del número de eventos $N_{e|a}(i)$ correspondientes

con los años de registro $i=1,2,\dots,n_a$. En el caso de que $v_a \leq 5$, la prueba funciona siempre y cuando $n_a > 5$. Puede notarse que un valor aproximado de la tasa de eventos anual v_a está dado por $\hat{v}_a = \bar{N}_{ela}$. Si rige la hipótesis H_0 de que se tiene un proceso Poisson, con un nivel de significancia α que cumple $P[\chi^2_{na-1,\alpha} > I_d(n_a-1) > \chi^2_{na-1,1-\alpha}] = \alpha$, entonces un intervalo de confianza para:

$$I_d = \frac{\sum_{i=1}^{n_a} [N_{ela}(i) - \bar{N}_{ela}]^2}{\bar{N}_{ela} \cdot (n_a - 1)} \quad (2.9)$$

estaría dado por:

$$\frac{\chi^2_{na-1,\alpha}}{n_a - 1} < I_d < \frac{\chi^2_{na-1,1-\alpha}}{n_a - 1} \quad (2.10)$$

Si la ecuación (2.10) se cumple se debe aceptar la hipótesis nula, que indica que el número de ocurrencias de la serie de impulsos sigue un proceso Poisson.

Estimación de la tasa de eventos v , para t en escalas diaria y anual

El mínimo umbral que satisfaga las hipótesis de independencia y distribución Poisson de los eventos garantizará la mayor recopilación de eventos de lluvia efectiva posible, que respeten los preceptos de un FPP. Este umbral (línea paralela a la abcisa en la Figura 2.3a) se aplica a la serie simple de picos filtrados FP, y los valores de lluvia efectiva que lo superen se preservan para conformar la serie FPOT (parte b de la Figura 2.3).

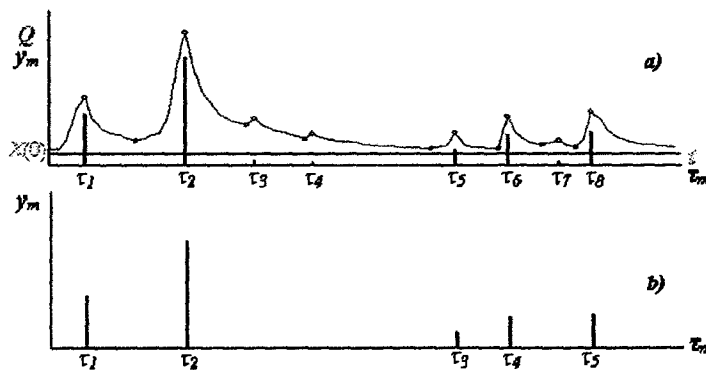


Figura 2.3 (a) Aplicación de filtro umbral a la lluvia efectiva, para generar (b) la serie FPOT

El valor promedio de eventos por unidad de tiempo t en la serie FPOT define la tasa de eventos ν_t del proceso Poisson, con respecto a esa unidad de tiempo. Por ejemplo, al aplicar la prueba de Cunnane se requiere evaluar el patrón de Poisson en un lapso anual. En ese caso se sugirió $\hat{\nu}_a = \bar{N}_{e/a}$ para estimar ν_a , que es la tasa de eventos en una escala de un año. Entonces, para $t=1.5$ (un año y medio) se esperaría observar $E[N(1.5)] = \nu_a \cdot 1.5$ eventos.

Una estimación directa para ν en escala de tiempo diaria sería de utilidad para ajustar el proceso Poisson de un modelo SNP como (2.4). Lo anterior en el caso de que no se utilice el método de preservación de momentos (que se discutirá en la sección 2.7). Entonces, si en el periodo de observación de T días la serie FPOT presentó $N[T]$ eventos, una estimación de la tasa de eventos diaria es $\hat{\nu} = N[T]/T$.

2.3.2 Identificación de la magnitud de los impulsos y estimación del salto promedio θ

La evaluación de magnitud de los impulsos de lluvia efectiva puede ser inconveniente como parte del proceso de elección del umbral de los FPOT para modelos estacionales, según lo que se discutió en la sección 2.3.1. No obstante, una vez determinado ese umbral, es posible estimar si los excesos de lluvia poseen una distribución exponencial, según lo indicado en la ecuación (2.3). Para este fin, suelen hacerse pruebas de bondad de ajuste basadas en los estadísticos de Cramér-von Mises W^2 y de Anderson-Darling A^2 antes mencionados. Se prueba si los saltos FPOT siguen la distribución Pareto generalizada (GPD), que se define por la función acumulada $F(y) = 1 - (1 - k \cdot y/a)^{1/k}$, en la que a es un parámetro de escala positivo y k es el parámetro de forma. La función de densidad correspondiente es $f(y) = (1/a)(1 - k \cdot y/a)^{(1-k)/k}$. Cuando $k=0$ la GPD se convierte en la exponencial, que es lo que en este caso se quiere comprobar. Usualmente se supone que ambos parámetros son desconocidos, de forma que es posible estimarlos utilizando el método de máxima verosimilitud.

Ha sido común valerse de la relación entre los parámetros $a = k/\psi$ para lograr reducir la función de verosimilitud de la GPD a un solo parámetro. Lo anterior lleva a la maximización de la

siguiente función de log-verosimilitud de perfil (la obtención de esta función se presenta en el Anexo A):

$$L(\psi) = -N \cdot \ln \left[\frac{-1}{N \cdot \psi} \sum_{i=1}^N \ln(1 - y_i \cdot \psi) \right] + -N - \sum_{i=1}^N \ln(1 - y_i \cdot \psi) \quad (2.11)$$

de forma que su máximo local en ψ permite obtener:

$$\hat{k} = \frac{-1}{N} \sum_{i=1}^N \ln(1 - y_i \cdot \psi) \quad y \quad \hat{a} = \frac{\hat{k}}{\psi} \quad (2.12)$$

Una vez estimados los parámetros es posible aplicar la prueba de bondad de ajuste, que consiste en evaluar la hipótesis nula H_0 : la muestra aleatoria y_1, y_2, \dots, y_N procede de una distribución GPD, esto es $F(y) = 1 - (1 - k \cdot y/a)^{1/k}$. El procedimiento sugerido por Choulakian y Stephens (2001) consiste en aplicar la transformación $z_{(i)} = F(y_{(i)})$ para $i=1, 2, \dots, N$; donde $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(N)}$ son los estadísticos de orden. Luego se calculan los estadísticos W^2 y A^2 como:

$$W^2 = \sum_{i=1}^N \left[z_{(i)} - \frac{2 \cdot i - 1}{2 \cdot N} \right]^2 + \frac{1}{12 \cdot N} \quad (2.13)$$

$$A^2 = -N - \frac{1}{N} \sum_{i=1}^N \{ [2 \cdot i - 1] \cdot [\ln(z_{(i)}) + \ln(1 - z_{(n+1-i)})] \} \quad (2.14)$$

Choulakian y Stephens (2001) calcularon los puntos porcentuales asintóticos superiores para los W^2 y A^2 , utilizando la teoría asintótica y simulaciones de Monte Carlo con 10000 muestras para cada combinación de N y k . Estos valores se presentan en la Tabla B.1 y en la Tabla B.2 del Anexo B, y se requiere interpolar el valor estimado de \hat{k} para encontrar los límites bajo los cuales no se rechaza H_0 (o sea, la muestra aleatoria evaluada procede de una distribución GPD). Si el estimador de k es relativamente cercano a cero y no se rechaza H_0 , se podría pensar que la distribución de y_m es aproximadamente exponencial. Entonces se consideraría adecuado el uso de esta distribución, en términos globales, para generar los saltos de lluvia efectiva como parte de un procedimiento de simulación de caudal diario. El parámetro θ que representa el impulso promedio en la ecuación (2.3) podría aproximarse entonces como $\theta \approx \hat{a}$.

2.4 Proceso de Poisson filtrado de segundo orden

El orden de un FPP se refiere a la cantidad de escalas de tiempo sucesivas en las que el modelo preserva las propiedades estadísticas de la serie histórica. En el análisis de caudales

diarios, una simulación basada en el modelo SNP como el presentado en la ecuación (2.4) puede preservar la media, la desviación estándar y la primera auto-correlación de la serie histórica a nivel diario, pero al agrupar la información en meses se suelen subestimar los últimos dos estadísticos mencionados. También se generan demasiados picos en la información sintética y las recesiones suelen decaer muy abruptamente, ignorando el componente de caudal base propio de la escorrentía superficial.

Antes se discutió la importancia de la escala de variabilidad en la escorrentía diaria. Ahora se agrega la relación entre la variabilidad en dos escalas sucesivas (la diaria y la mensual) con los diversos procesos que intervienen en la formación de caudales. Autores como Weiss (1977), Claps y Murrone (1994), Murrone y otros (1997), Xu y otros (2002) y Claps y otros (2005), entre otros, han planteado la necesidad de diferenciar los aportes de dichos procesos: el flujo base (procedente de fuentes subterráneas), el flujo subsuperficial (flujo interno que puede provenir de un movimiento rápido de agua a través de conducciones en el suelo, macro-poros y/o zonas de infiltración), y el flujo saturado por tierra (flujo superficial). Cada uno de ellos aporta su variabilidad en la respuesta final de la cuenca.

Ignorando que la distribución relativa de estos componentes no es lineal, se suele hacer una separación conceptual lineal de los flujos. De esta manera, en su modelación se pueden diferenciar las características de auto-correlación de las diversas escalas temporales, que en forma general, pertenecen a un proceso hidrológico determinado. Por ejemplo, en el caso del caudal diario, es posible identificar las fracciones superficial, subsuperficial y, en menor medida, la subterránea. No obstante, este último componente es de gran importancia en la escala de variación mensual que, como se dijo, puede ser un objetivo por cumplir en un procedimiento de simulación a nivel diario, como es caso del SNP.

Se plantea, entonces, que la función de respuesta del sistema $h(t)$ antes definida como una simple exponencial negativa, ahora está constituida por la combinación lineal de las respuestas de diversos elementos conceptuales, que actúan en paralelo. Usualmente se utilizan entre dos y cuatro componentes. Se presentará el caso de dos embalses lineales en paralelo, de forma que uno modele el componente la escorrentía superficial, y el otro el flujo base (agua subterránea y

sub-superficial). Esto define un FPP de segundo orden, pues se espera poder explicar tanto la variabilidad en el comportamiento del caudal diario (relacionado con el escurrimiento superficial y parte del sub-superficial), como las propiedades generales del caudal mensual (que tiene relación con el caudal sub-superficial y subterráneo). Cada componente se supone de tipo SNP como en la ecuación (2.4) y posee tasas de decaimiento exponenciales b_1 y b_2 respectivamente. En este caso, evidentemente $b_1 > b_2$, pues el abatimiento del caudal debe ser mucho más abrupto en el SNP superficial $X_1(t)$ (llamado rápido, con parámetros ν_1, θ_1, b_1), que en el SNP base $X_2(t)$ (llamado lento con parámetros ν_2, θ_2, b_2). El caudal observado se representa entonces como $X(t) = X_1(t) + X_2(t)$. Existen diversas posibilidades de relacionar los saltos y recesiones de los procesos rápido y lento. Weiss (1977) propuso tres formulaciones que producen los siguientes modelos:

- Modelo de ruido disparado de segundo orden (SOSN): considera que las tasas de ocurrencia del proceso Poisson son iguales en ambos SNP, $\nu = \nu_1 = \nu_2$, por lo cual los eventos se producen simultáneamente. Sus saltos aleatorios se relacionan según $y_1/y_2 = \theta_1/\theta_2$. Entonces se tiene $X(t)$ como un FPP con ocurrencias $\{y_m\}$ (impulsos aleatorios distribuidos exponencialmente con media $\theta = \theta_1 + \theta_2$) en los tiempos $\{\tau_m\}$ (tiempos de los eventos Poisson con tasa $\nu = \nu_1 = \nu_2$). Tomando en cuenta que $b_1 \neq b_2$. La función de respuesta del sistema es:

$$h(t) = \frac{\theta_1}{\theta} e^{-b_1 t} + \frac{\theta_2}{\theta} e^{-b_2 t} \quad (2.15)$$

- Modelo doble de ruido disparado (DSN): los procesos rápido y lento se asumen totalmente independientes, cada uno con su propia serie de eventos y saltos. Aquí $b_1 \neq b_2$, $\nu_1 \neq \nu_2$ y $\theta_1 \neq \theta_2$.

- Modelo modificado de ruido disparado de segundo orden (MSOSN): surge de la combinación de los dos anteriores, de forma que $\nu_1 > \nu_2$, y los eventos del proceso lento (cuyos saltos cumplen $\theta = \theta_1 + \theta_2$) ocurren simultáneamente con algunos de los eventos del proceso rápido.

Suponiendo que todos los eventos (modelo SOSN), o algunos de ellos (MSOSN) ocurren simultáneamente en los procesos rápido $X_1(t)$ y lento $X_2(t)$, y por tanto, producen saltos correlacionados y_1 y y_2 , la correlación entre $X_1(t)$ y $X_2(t)$ se puede estimar por medio de:

$$\rho_{12} = \frac{2(b_1 \cdot b_2)^{1/2}}{b_1 + b_2} \cdot \frac{\nu_{12}}{(\nu_1 \cdot \nu_2)^{1/2}} \cdot \frac{c_{12} + 1}{2} \quad (2.16)$$

donde para los procesos rápido (subíndice 1) y lento (subíndice 2) se tiene que:

ν_{12} : tasa a la que ocurren los eventos que se producen simultáneamente en ambos procesos.

Debe notarse que $\nu_{12} \leq \nu_1$ y $\nu_{12} \leq \nu_2$.

c_{12} : coeficiente de correlación de las magnitudes de los saltos y_1 y y_2 asociados con un evento que se da simultáneamente en ambos procesos.

El primer componente de ρ_{12} toma en cuenta el efecto de las diferentes tasas de decaimiento en la correlación (es decir, el efecto que se produce por ser diferentes las dos formas de recesión), mientras los otros dos se deben a la correlación entre las dos series de eventos y saltos.

La correlación rezagada entre $X_1(t)$ y $X_2(t)$ es:

$$\rho[X_1(t), X_2(t+s)] = \rho_{12} \cdot e^{-b_2 \cdot s} \quad (2.17)$$

Entonces, pese a que los saltos de $X_1(t)$ y $X_2(t)$ para los modelos SOSN y MSOSN están relacionados debido a sus tiempos de ocurrencia, la correlación entre ellos según la ecuación (2.16) es muy pequeña (especialmente por las diferencias entre las tasas de decaimiento y entre la magnitud promedio de los impulsos), de modo que no existe mucha diferencia en este aspecto entre tales modelos y el DNS.

2.5 El efecto del muestreo promediado en intervalos discretos para modelos FPP de primer y segundo orden

El caudal es un proceso natural continuo en el tiempo, como también es el modelo FPP, en su forma de un SNP. No obstante, los registros históricos almacenados en las bases de datos suelen ser más bien de naturaleza discreta. En el caso del caudal diario, la información que se conserva es discreta, usualmente en la forma de promedios diarios para el lapso entre los tiempos de observación de las muestras tomadas. También los valores sintéticos por generar suelen demandarse en este formato.

2.5.1 Modelos FPP de primer orden

Si el caudal sigue un SNP de primer orden en tiempo continuo $X(t)$, y las observaciones (y los valores generados) son promedios de este SNP en un periodo J días, entonces para $j \geq J$ se puede plantear que el valor promediado del caudal está dado por:

$$X_j = \frac{1}{J} \int_{j-J}^j X(t) dt \quad (2.18)$$

Si se sustituye la ecuación (2.4) en la (2.18) se llega a:

$$X_j = \frac{1}{J} \int_{j-J}^j \sum_{m=1}^{N(t)} [y_m \cdot h(t - \tau_m)] dt$$

$$X_j = \sum_{m=1}^{N(t)} \left[y_m \cdot \frac{1}{J} \int_{j-J}^j h(t - \tau_m) dt \right]$$

Yue y otros (1999) presentan el desarrollo de esta última ecuación, de modo que ahora el valor promediado del caudal en un lapso de J días, para una recesión exponencial de tipo $h(t) = e^{-bt}$ es:

$$X_j = \sum_{m=1}^{N(t)} [y_m \cdot h_j(t - \tau_m)] \quad (2.19)$$

con:

$$h_j(t) = \frac{1 - e^{-bt}}{J \cdot b} \quad (0 < t \leq J) \quad (2.20)$$

$$h_j(t) = \frac{1 - e^{-bJ}}{J \cdot b} \cdot e^{-b(t-J)} \quad (t > J) \quad (2.21)$$

Se pueden obtener la media, la varianza y primera auto-correlación para los valores X_j dados por la ecuación (2.19), valiéndose de la función general característica $\varphi_{X(t)}(z)$ para el proceso continuo de caudal $X(t)$. Suponiendo que los tiempos de ocurrencia de eventos τ_m son independientes e idénticamente distribuidos al igual que las intensidades y_m , es posible determinar la función característica específica para $N(t)$ siguiendo un proceso Poisson (que define $X(t)$ como un FPP), planteada por Parzen (1962) como:

$$\varphi_{X(t)}(y) = \exp \left\langle \nu \int_{t_0}^{t_1} E \{ \exp [i \cdot y \cdot h(t - \tau)] - 1 \} d\tau \right\rangle \quad (2.22)$$

A partir de la ecuación (2.22), para un tiempo t tal que $t \in [t_0 \leq t \leq t_1]$ se pueden estimar los primeros dos acumulados y la función de covarianza del caudal $X(t)$:

$$\gamma_1[X(t)] = v \cdot E(y) \cdot \int_{t_0}^{t_1} h_J(t - \tau_m) d\tau \quad (2.23)$$

$$\gamma_2[X(t)] = v \cdot E(y^2) \cdot \int_{t_0}^{t_1} h_J(t - \tau_m)^2 d\tau \quad (2.24)$$

$$C[X(t), X(t+s)] = v \cdot E(y^2) \cdot \int_{t_0}^{t_1} h_J(t - \tau_m) \cdot h_J(t+s - \tau_m) d\tau \quad (2.25)$$

Nótese que en las anteriores ecuaciones se reemplazó $h(t-\tau)$ por $h_J(t-\tau_m)$ al tratarse de los valores promediados para J días. La media $E[X(t)]$, la variancia $V[X(t)]$ y la primera autocorrelación $\rho[X(t), X(t+s)]$ que se buscan, pueden obtenerse a partir de (2.23)-(2.25) valiéndose de las relaciones:

$$\begin{aligned} E[X(t)] &= \gamma_1[X(t)] = \frac{v \cdot \theta}{b} \\ V[X(t)] &= \gamma_2[X(t)] = \frac{2 \cdot v \cdot \theta^2}{J^2 \cdot b^3} \cdot (J \cdot b + e^{-J \cdot b} - 1) \\ \rho[X(t), X(t+s)] &= \frac{C[X(t), X(t+s)]}{\sqrt{\gamma_2[X(t)] \cdot \gamma_2[X(t+s)]}} = \frac{(1 - e^{-J \cdot b})^2}{2 \cdot (J \cdot b + e^{-J \cdot b} - 1)} \end{aligned} \quad (2.26)$$

En el caso de la evaluación del caudal diario, tanto las ecuaciones (2.20) y (2.21) que definen la función de respuesta del proceso promediado (2.19), como las ecuaciones (2.26), que representan sus estadísticos básicos, toman el valor $J=1$.

2.5.2 Modelos FPP de segundo orden

Si ahora se considera que el caudal sigue un SNP de segundo orden en tiempo continuo $X(t)$, que considera dos componentes (el rápido denotado con un 1 y el lento con un 2) con las observaciones promediadas en J días, entonces para $j \geq J$ se puede plantear que el valor promediado del caudal está dado por la ecuación (2.19). En este caso, dada la mayor complejidad que introduce el segundo orden del modelo, solo se presentarán ecuaciones para el análisis de series promediadas diariamente, o sea, con $J=1$. De este modo, partiendo de la ecuación (2.15) para $h(t)$ de un modelo SOSN se desarrollaron siguientes funciones de

respuesta (el método propuesto para su determinación y el cálculo correspondiente se presenta en el Anexo C):

$$h_{<1}(t) = \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1 - e^{-b_1 t}) + \frac{\theta_2}{b_2} \cdot (1 - e^{-b_2 t}) \right] \quad (0 < t \leq 1) \quad (2.27)$$

$$h_{>1}(t) = \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1 - e^{-b_1}) + \frac{\theta_2}{b_2} \cdot (1 - e^{-b_2}) \right] \cdot \left\{ 1 + \frac{1}{\theta} \cdot \left[\theta_1 \cdot (e^{-b_1(t-1)} - 1) + \theta_2 \cdot (e^{-b_2(t-1)} - 1) \right] \right\} \quad (t > 1) \quad (2.28)$$

Weiss (1977) presenta una correlación análoga a la correlación ρ_{12} entre $X_1(t)$ y $X_2(t)$ que se expuso en la ecuación (2.16), pero ahora establecida entre los procesos promediados $X_{1,t}$ y $X_{2,t}$:

$$\bar{\rho}_{12} = \rho_{12} \cdot \frac{d_1 + d_2}{2 \cdot (d_1 \cdot d_2)^{1/2}} \quad (2.29)$$

con $d_p = \{2 \cdot [b_p \cdot (1 - e^{-b_p})]\} / b_p^2$ y $p=1,2$. Entonces, a partir de las ecuaciones (2.26) y (2.29) para el proceso promediado, que definen respectivamente los estadísticos básicos, y la correlación entre los componentes rápido y lento; y definiendo para $i,j=1,2$:

$$d_{ij} = \frac{2 \cdot [b_i \cdot T_j - (1 - e^{-b_i T_j})]}{(b_i \cdot T_j)^2} \quad \text{y} \quad e_{ij} = \frac{(1 - e^{-b_i T_j})^2}{(b_i \cdot T_j)^2} \quad (2.30)$$

es posible plantear:

$$\begin{aligned} E[X_{1,t}] &= \frac{v_1 \cdot \theta_1}{b_1} + \frac{v_2 \cdot \theta_2}{b_2} \\ V[X_{1,t}] &= \sigma_{T_1}^2 \quad \text{y} \quad V[X_{2,t}] = \sigma_{T_2}^2 \\ \rho[X_{1,t}, X_{1,t+s}] &= \rho_{T_1} \quad \text{y} \quad \rho[X_{2,t}, X_{2,t+s}] = \rho_{T_2} \end{aligned} \quad (2.31)$$

que se encuentran relacionados por medio de las ecuaciones:

$$\begin{aligned} ss_1 &= \frac{v_1 \cdot \theta_1^2}{b_1} + \frac{2 \cdot v_{12} \cdot \theta_1 \cdot \theta_2}{b_1 + b_2} \quad \text{y} \quad ss_2 = \frac{v_2 \cdot \theta_2^2}{b_2} + \frac{2 \cdot v_{12} \cdot \theta_1 \cdot \theta_2}{b_1 + b_2} \\ ss_1 &= \frac{\sigma_{T_1}^2 \cdot d_{22} - \sigma_{T_2}^2 \cdot d_{21}}{d_{22} \cdot d_{11} - d_{12} \cdot d_{21}} \quad \text{y} \quad ss_2 = \frac{\sigma_{T_2}^2 \cdot d_{11} - \sigma_{T_1}^2 \cdot d_{12}}{d_{22} \cdot d_{11} - d_{12} \cdot d_{21}} \\ \rho_{T_1} \cdot \sigma_{T_1}^2 &= ss_1 \cdot e_{11} + ss_2 \cdot e_{21} \quad \text{y} \quad \rho_{T_2} \cdot \sigma_{T_2}^2 = ss_1 \cdot e_{12} + ss_2 \cdot e_{22} \end{aligned} \quad (2.32)$$

En las ecuaciones (2.30) T_j con $j=1,2$ se refiere a los lapsos evaluados según su respectiva escala, que en este análisis se refiere a $T_1=1$ día y $T_2=1$ mes ≈ 30.44 días. En las ecuaciones

(2.31) $X_{p,t}$ con $p=1,2$ se refiere a la serie promediada en términos de la escala 1 ó 2, o sea que en este análisis $X_{1,t}$ se refiere a la serie de caudal promedio diario, y $X_{2,t}$ a la de caudal promedio mensual. Finalmente, en las ecuaciones (2.32) el término v_{12} tendría que tomar valores coherentes con las propiedades del modelo seleccionado, según lo discutido en la sección 2.4. De este modo, la ecuación para cada modelo de segundo orden es:

$$\begin{array}{ll} \text{SOSN:} & v_{12}=v_1=v_2 \\ \text{DSN:} & v_{12}=0 \\ \text{MSOSN:} & v_{12}=v_2 \text{ (con } v_2 < v_1) \end{array} \quad (2.33)$$

2.6 Separación del proceso de ruido disparado en el tiempo para modelos FPP de primer y segundo orden

Para desarrollar un procedimiento de simulación es necesario contar con una articulación en el tiempo, de modo que sea posible construir las series sintéticas en pasos sucesivos discretos. El modelo SNP tiene una naturaleza continua con eventos discretos, y es posible agregarle a su forma general planteada en la ecuación (2.4), un componente que tome en consideración el valor del proceso después de un lapso s a partir del tiempo t . Para el proceso continuo se plantea que:

$$X(t+s) = h(s) \cdot X(t) + \sum_{k=N(t)+1}^{N(t+s)} [y_k \cdot h(t+s-\tau_k)] \quad (2.34)$$

mientras para el muestreo promediado, el SNP separado en el tiempo resulta ser:

$$X_{t+s} = h_{\leq 1}(s) \cdot X(t) + \sum_{k=N(t)+1}^{N(t+s)} [y_k \cdot h_{\leq 1}(t+s-\tau_k)] \quad (2.35)$$

El primer término de $X(t+s)$ y X_{t+s} en (2.34) y (2.35) representa el efecto de los eventos en el lapso $[0,t]$, mientras el segundo abarca el lapso $(t,t+s]$, y es conocido como el término de innovación. Como señala Weiss (1977) la ecuación (2.34) denota un proceso auto-regresivo de primer orden en el cual el término de innovación, en lugar de ser de tipo gaussiano, es no negativo con distribución sesgada (gamma) y una probabilidad positiva de ser exactamente cero (cuando no se dan eventos entre t y $t+s$). Lo anterior se extiende para la ecuación (2.35).

La separación en el tiempo depende directamente de la función de respuesta de la cuenca evaluada. Se ha visto que estas funciones son distintas si el modelo es de primer o de segundo orden, o si el modelo es de tiempo continuo o promediado. En las sub-secciones siguientes se presentan las posibles combinaciones, tomando en cuenta la definición de $X(t)$ dada en la ecuación (2.4).

2.6.1 Modelos FPP de primer orden

El proceso continuo para un FPP de primer orden tiene la función de respuesta indicada en la sección 2.3, de la forma $h(t)=e^{-bt}$. Entonces, sustituyendo $h(\cdot)$ en la ecuación (2.34) se obtiene:

$$X(t+s) = e^{-bs} \cdot \sum_{m=1}^{N(t)} [y_m \cdot e^{-b(t-\tau_m)}] + \sum_{k=N(t)+1}^{N(t+s)} [y_k \cdot e^{-b(t+s-\tau_k)}] \quad (2.36)$$

El muestreo promediado a nivel diario ($J=1$) sigue la ecuación (2.35) con $h_{\leq 1}$ definido por la ecuación (2.20), de modo que:

$$X_{t+s} = \frac{1-e^{-bs}}{b} \cdot \sum_{m=1}^{N(t)} [y_m \cdot e^{-b(t-\tau_m)}] + \sum_{k=N(t)+1}^{N(t+s)} \left[y_k \cdot \frac{1-e^{-b(t+s-\tau_k)}}{b} \right] \quad (2.37)$$

2.6.2 Modelos FPP de segundo orden

Para el caso específico del modelo SOSN, que es el presenta una expresión explícita para $h(\cdot)$ y $h_{\leq 1}(\cdot)$, se tiene que la separación del proceso continuo se obtiene al combinar (2.34) y (2.15) en la forma:

$$\begin{aligned} X(t+s) = & \left(\frac{\theta_1}{\theta} e^{-b_1 s} + \frac{\theta_2}{\theta} e^{-b_2 s} \right) \cdot \sum_{m=1}^{N(t)} \left\{ y_m \cdot \left[\frac{\theta_1}{\theta} e^{-b_1(t-\tau_m)} + \frac{\theta_2}{\theta} e^{-b_2(t-\tau_m)} \right] \right\} + \\ & + \sum_{k=N(t)+1}^{N(t+s)} \left\{ y_k \cdot \left[\frac{\theta_1}{\theta} e^{-b_1(t+s-\tau_k)} + \frac{\theta_2}{\theta} e^{-b_2(t+s-\tau_k)} \right] \right\} \end{aligned} \quad (2.38)$$

El muestreo promediado tiene una forma separada en el tiempo según la combinación de (2.35) con (2.27):

$$\begin{aligned}
X_{t+s} = & \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1 - e^{-b_1 \cdot s}) + \frac{\theta_2}{b_2} \cdot (1 - e^{-b_2 \cdot s}) \right] \cdot \sum_{m=1}^{N(t)} \left\{ y_m \cdot \left[\frac{\theta_1}{\theta} e^{-b_1 \cdot (t-\tau_m)} + \frac{\theta_2}{\theta} e^{-b_2 \cdot (t-\tau_m)} \right] \right\} + \\
& + \sum_{k=N(t)+1}^{N(t+s)} \left\{ y_k \cdot \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1 - e^{-b_1 \cdot (t+s-\tau_k)}) + \frac{\theta_2}{b_2} \cdot (1 - e^{-b_2 \cdot (t+s-\tau_k)}) \right] \right\}
\end{aligned} \quad (2.39)$$

El modelo DSN separado en el tiempo requiere una pequeña modificación de (2.34) y (2.35) para tomar en cuenta que los componentes rápido y lento suceden independientemente. El proceso continuo toma la forma:

$$\begin{aligned}
X(t+s) = & e^{-b_1 \cdot s} \cdot \sum_{m_1=1}^{N_1(t)} [y_{m_1} \cdot e^{-b_1 \cdot (t-\tau_{m_1})}] + e^{-b_2 \cdot s} \cdot \sum_{m_2=1}^{N_2(t)} [y_{m_2} \cdot e^{-b_2 \cdot (t-\tau_{m_2})}] + \\
& + \sum_{k_1=N_1(t)+1}^{N_1(t+s)} [y_{k_1} \cdot e^{-b_1 \cdot (t+s-\tau_{k_1})}] + \sum_{k_2=N_2(t)+1}^{N_2(t+s)} [y_{k_2} \cdot e^{-b_2 \cdot (t+s-\tau_{k_2})}]
\end{aligned} \quad (2.40)$$

y el promediado:

$$\begin{aligned}
X_{t+s} = & \frac{1 - e^{-b_1 \cdot s}}{b_1} \cdot \sum_{m_1=1}^{N_1(t)} [y_{m_1} \cdot e^{-b_1 \cdot (t-\tau_{m_1})}] + \frac{1 - e^{-b_2 \cdot s}}{b_2} \cdot \sum_{m_2=1}^{N_2(t)} [y_{m_2} \cdot e^{-b_2 \cdot (t-\tau_{m_2})}] + \\
& + \sum_{k_1=N_1(t)+1}^{N_1(t+s)} \left[y_{k_1} \cdot \frac{1 - e^{-b_1 \cdot (t+s-\tau_{k_1})}}{b_1} \right] + \sum_{k_2=N_2(t)+1}^{N_2(t+s)} \left[y_{k_2} \cdot \frac{1 - e^{-b_2 \cdot (t+s-\tau_{k_2})}}{b_2} \right]
\end{aligned} \quad (2.41)$$

El modelo MSOSN también requiere variar (2.34) y (2.35) de manera que se tome en cuenta que algunos eventos presentan únicamente el componente rápido (estos eventos se denotarán con el subíndice 1), mientras los restantes eventos se constituyen por ambos componentes, el lento y el rápido (y se identifican con 1+2). Estos dos tipos de eventos son excluyentes entre sí, por lo que en un mismo tiempo de ocurrencia τ debe evaluarse un impulso en sólo uno de los procesos continuos $X_1(t)$ o $X_{1+2}(t)$:

$$\begin{aligned}
X(t+s) = & e^{-b_1 \cdot s} \cdot \sum_{m_1=1}^{N_1(t)} [y_{m_1} \cdot e^{-b_1 \cdot (t-\tau_{m_1})}] + \left(\frac{\theta_1}{\theta} e^{-b_1 \cdot s} + \frac{\theta_2}{\theta} e^{-b_2 \cdot s} \right) \cdot \\
& \cdot \sum_{m_{1+2}=1}^{N_{1+2}(t)} \left\{ y_{m_{1+2}} \cdot \left[\frac{\theta_1}{\theta} e^{-b_1 \cdot (t-\tau_{m_{1+2}})} + \frac{\theta_2}{\theta} e^{-b_2 \cdot (t-\tau_{m_{1+2}})} \right] \right\} + \sum_{k=N_1(t)+1}^{N_1(t+s)} [y_{k_1} \cdot e^{-b_1 \cdot (t+s-\tau_{k_1})}] + \\
& + \sum_{k_{1+2}=N_{1+2}(t)+1}^{N_{1+2}(t+s)} \left\{ y_{k_{1+2}} \cdot \left[\frac{\theta_1}{\theta} e^{-b_1 \cdot (t+s-\tau_{k_{1+2}})} + \frac{\theta_2}{\theta} e^{-b_2 \cdot (t+s-\tau_{k_{1+2}})} \right] \right\}
\end{aligned} \quad (2.42)$$

o bien, X_{t_1} o $X_{t_{1+2}}$ para la generación promediada:

$$\begin{aligned}
X_{t+s} = & \frac{1-e^{-b_1 s}}{b_1} \cdot \sum_{m_1=1}^{N_1(t)} \left[y_{m_1} \cdot e^{-b_1(t-\tau_{m_1})} \right] + \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1-e^{-b_1 s}) + \frac{\theta_2}{b_2} \cdot (1-e^{-b_2 s}) \right] \cdot \\
& \cdot \sum_{m_{1+2}=1}^{N_{1+2}(t)} \left\{ y_{m_{1+2}} \cdot \left[\frac{\theta_1}{\theta} e^{-b_1(t-\tau_{m_{1+2}})} + \frac{\theta_2}{\theta} e^{-b_2(t-\tau_{m_{1+2}})} \right] \right\} + \sum_{k=N_1(t)+1}^{N_1(t+s)} \left[y_{k_1} \cdot \frac{1-e^{-b_1(t+s-\tau_{k_1})}}{b_1} \right] + \\
& + \sum_{k=N_{1+2}(t)+1}^{N_{1+2}(t+s)} \left\{ y_{k_{1+2}} \cdot \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1-e^{-b_1(t+s-\tau_{k_{1+2}})}) + \frac{\theta_2}{b_2} \cdot (1-e^{-b_2(t+s-\tau_{k_{1+2}})}) \right] \right\}
\end{aligned} \quad (2.43)$$

2.7 Ajuste de los parámetros del modelo FPP de primer y segundo orden

2.7.1 Modelos FPP de primer orden

Los modelos FPP tipo SNP de primer orden se definen al conocer (o en este caso, estimar) los valores de los parámetros ν , θ y b . El método de ajuste propuesto por Weiss (1977) consiste en estimarlos de forma que se preserven la media μ , la variancia σ^2 y la primera auto-correlación $\rho(1)$ de los datos. Así, en el caso del proceso continuo definido por la ecuación (2.4), se calculan los estimadores $\hat{\mu}$, $\hat{\sigma}^2$ y $\hat{\rho}(1)$ para $E[X(t)]$, $V[X(t)]$ y $\rho[X(t), X(t+s)]$ respectivamente en las ecuaciones (2.5) a partir de los caudales históricos. De la tercera expresión se obtiene \hat{b} con facilidad, que es el valor estimado de la tasa de decaimiento. Las otras dos ecuaciones pueden resolverse simultáneamente para encontrar $\hat{\nu}$ y $\hat{\theta}$, que estiman la tasa de ocurrencias del proceso Poisson y la magnitud promedio de los eventos de lluvia efectiva.

Resulta un poco más complicado evaluar el modelo para el muestreo promediado de la ecuación (2.19). Tomando como base los momentos expresados en la ecuación (2.26) y valiéndose de $\hat{\mu}$, $\hat{\sigma}^2$ y $\hat{\rho}(1)$ calculados de la serie histórica, \hat{b} se extrae de la tercera de estas expresiones utilizando métodos numéricos, y $\hat{\nu}$ y $\hat{\theta}$ de las primeras dos al resolverlas simultáneamente.

También es posible obtener \hat{b} , $\hat{\nu}$ y $\hat{\theta}$ a partir, respectivamente, de las recesiones observadas en el registro histórico y de los métodos propuestos al final de las secciones 2.3.1 y 2.3.2. No

obstante, este segundo método no asegura la preservación de los momentos estadísticos antes expuestos, correspondientes con las series observadas de caudal diario.

2.7.2 Modelos FPP de segundo orden

Las resolución simultánea de las expresiones dadas en (2.30)-(2.33) para un SNP de segundo orden y proceso promediado, permiten estimar los parámetros $\nu_1, \theta_1, b_1, \nu_2, \theta_2$ y b_2 del modelo propuesto en la ecuación (2.19), con funciones de respuesta según las ecuaciones (2.27) y (2.28), de forma que se preserven los momentos básicos de las series de caudales promedio diario y promedio mensual, o sea $X_{1,t}$ y $X_{2,t}$ respectivamente. No se abordará el ajuste de los SNP basados en un proceso continuo debido a su poca aplicación en la práctica. Los momentos por mantener son la media general μ de toda la serie, así como las variancias $\sigma_{T_1}^2$ y $\sigma_{T_2}^2$ y las primeras auto-correlaciones ρ_{T_1} y ρ_{T_2} para las dos escalas evaluadas, la denotada con 1 representando al componente rápido-diario del flujo y la 2 al componente lento-mensual. Cuando se estima el modelo SOSN el sistema de ecuaciones recién planteado es suficiente. Cuando se trata de los modelos DSN o MSOSN hace falta una ecuación, de modo que Weiss (1977) recomienda introducir al modelo un valor de la tasa de ocurrencia de los eventos ν_1 observado en forma directa de la serie histórica de datos diarios. Otra alternativa es utilizar el estimador propuesto al final de la sección 2.3.1, que corresponde a los eventos de lluvia efectiva seleccionados con el procedimiento FPOT.

2.7.3 Valoración de la estacionalidad en la estimación de los parámetros

En modelos complejos, como podría ser un SNP de segundo orden para el proceso promediado, la incorporación de la estructura periódica a las diversas escalas de interés puede generar un modelo muy difícil formular, y aún más, de estimar. Una idea de este problema se puede obtener al imaginar el conjunto de parámetros en la función característica específica para $N(t)$ dada por (2.22) y las funciones de respuesta (2.27) y (2.28) como funciones del tiempo t , o sea: $\nu_1(t), \theta_1(t), b_1(t), \nu_2(t), \theta_2(t)$ y $b_2(t)$. De hecho, en (2.22) el término ν tendría que estar dentro de la integral, con la forma $\nu(t)$. También los momentos a preservar en el

procedimiento de estimación basado en (2.31)-(2.33) requerirían valores estimados a partir de la serie histórica $\mu(t), \sigma_{T_1}^2(t), \sigma_{T_2}^2(t), \rho_{T_1}(t)$ y $\rho_{T_2}(t)$, variantes en el tiempo.

Lo anterior justifica que el método para analizar la estacionalidad sea aproximado, encontrando valores para el conjunto de parámetros en forma separada para cada mes o estación: $v_{i,\varepsilon}(t), \theta_{i,\varepsilon}(t), b_{i,\varepsilon}(t)$; donde el subíndice ε indica la estación o mes de base para la estimación estacional, y el subíndice $i=1,2$ el componente del caudal. Así, los momentos a preservar $\mu_\varepsilon(t), \sigma_{T_1,\varepsilon}^2(t), \sigma_{T_2,\varepsilon}^2(t), \rho_{T_1,\varepsilon}(t)$ y $\rho_{T_2,\varepsilon}(t)$, se calcularían a partir de la serie histórica para cada mes o estación ε .

2.8 Procedimiento de generación de caudales sintéticos para modelos FPP de primer y segundo orden

Una vez que se conocen los parámetros correspondientes con los modelos FPP de tipo SNP, estimados con base en los datos históricos y en los procedimientos descritos en la sección 2.7, la generación de datos sintéticos sugerida por Weiss (1977) se basa en pasos de tiempo $s=1$ día. Entonces $X_t, t=1,2,\dots,T$ son los valores del proceso promediado que serán producidos en días consecutivos, y $X(t), t=1,2,\dots,T$ los valores del proceso continuo. Aquí es donde resulta valiosa la separación de los SNP en el tiempo, que fue expuesta en la sección 2.6. Las ecuaciones (2.36) y (2.37) para los modelos de primer orden y (2.38)-(2.43) para los de segundo orden, con el valor de $s=1$, son la base del procedimiento, de modo que el primer término (o sub-grupo de términos) de los procesos X_t y $X(t)$ ahí definidos representa la contribución de los eventos que preceden a t , y el segundo término (o sub-grupo) es el aporte de los eventos en el lapso $(t,t+1)$. El procedimiento parte de un valor inicial arbitrario $X(0)$ (se sugiere la media del mes en que se inicia la simulación), y con éste se deben calcular X_1 y $X(1)$. Hecho esto se puede descartar $X(0)$ y utilizar $X(1)$ para obtener X_2 y $X(2)$. Aplicando este método sucesivamente se puede producir una serie de T caudales promedio diarios, descritos por los valores X_1, X_2, \dots, X_T .

El componente estocástico del modelo, relacionado con la presencia o ausencia de eventos de lluvia y con las magnitudes de estos eventos, se puede incluir en el SNP partiendo de un valor cualquiera y produciendo los X_{t+1} y $X(t+1)$ según la siguiente secuencia de pasos:

(i) Para el modelo de orden uno se calculan los primeros términos de las ecuaciones (2.36) y (2.37), que a continuación se presentan explícitamente en función de $X(t)$ y para un paso de tiempo diario $s=1$:

$$X(t+1) = e^{-b} \cdot X(t) + \sum_{m=N(t)+1}^{N(t+1)} [y_m \cdot e^{-b \cdot (t+1-\tau_m)}] \quad (2.36')$$

$$X_{t+1} = \frac{1-e^{-b}}{b} \cdot X(t) + \sum_{m=N(t)+1}^{N(t+1)} \left[y_m \cdot \frac{1-e^{-b \cdot (t+1-\tau_m)}}{b} \right] \quad (2.37')$$

Si más bien se escoge un modelo de segundo orden de tipo SOSN deben aplicarse (2.38) y (2.39) en la forma:

$$X(t+1) = \left(\frac{\theta_1}{\theta} e^{-b_1} + \frac{\theta_2}{\theta} e^{-b_2} \right) \cdot X(t) + \sum_{m=N(t)+1}^{N(t+1)} \left\{ y_m \cdot \left[\frac{\theta_1}{\theta} e^{-b_1 \cdot (t+1-\tau_m)} + \frac{\theta_2}{\theta} e^{-b_2 \cdot (t+1-\tau_m)} \right] \right\} \quad (2.38')$$

$$X_{t+1} = \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1-e^{-b_1}) + \frac{\theta_2}{b_2} \cdot (1-e^{-b_2}) \right] \cdot X(t) + \sum_{m=N(t)+1}^{N(t+1)} \left\{ y_m \cdot \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1-e^{-b_1 \cdot (t+1-\tau_m)}) + \frac{\theta_2}{b_2} \cdot (1-e^{-b_2 \cdot (t+1-\tau_m)}) \right] \right\} \quad (2.39')$$

Así, se obtienen valores iniciales para $X(t+1)$ y X_{t+1} . Para los modelos de segundo orden tipo DNS se tienen las siguientes fórmulas, a partir de (2.40) y (2.41):

$$X(t+1) = e^{-b_1} \cdot X_1(t) + e^{-b_2} \cdot X_2(t) + \sum_{m_1=N_1(t)+1}^{N_1(t+1)} [y_{m_1} \cdot e^{-b_1 \cdot (t+1-\tau_{m_1})}] + \sum_{m_2=N_2(t)+1}^{N_2(t+1)} [y_{m_2} \cdot e^{-b_2 \cdot (t+1-\tau_{m_2})}] \quad (2.40')$$

$$X_{t+1} = \frac{1-e^{-b_1}}{b_1} \cdot X_1(t) + \frac{1-e^{-b_2}}{b_2} \cdot X_2(t) + \sum_{m_1=N_1(t)+1}^{N_1(t+1)} \left[y_{m_1} \cdot \frac{1-e^{-b_1 \cdot (t+1-\tau_{m_1})}}{b_1} \right] + \sum_{m_2=N_2(t)+1}^{N_2(t+1)} \left[y_{m_2} \cdot \frac{1-e^{-b_2 \cdot (t+1-\tau_{m_2})}}{b_2} \right] \quad (2.41')$$

Los MSOSN se basarían en las ecuaciones (2.42) y (2.43) de la manera que se presentan a continuación:

$$X(t+1) = e^{-b_1} \cdot X_1(t) + \left(\frac{\theta_1}{\theta} e^{-b_1} + \frac{\theta_2}{\theta} e^{-b_2} \right) \cdot X_{1+2}(t) + \sum_{m=N_1(t)+1}^{N_1(t+1)} \left[y_{m_1} \cdot e^{-b_1(t+1-\tau_{m_1})} \right] + \\ + \sum_{m_{1+2}=N_{1+2}(t)+1}^{N_{1+2}(t+1)} \left\{ y_{m_{1+2}} \cdot \left[\frac{\theta_1}{\theta} e^{-b_1(t+1-\tau_{m_{1+2}})} + \frac{\theta_2}{\theta} e^{-b_2(t+1-\tau_{m_{1+2}})} \right] \right\} \quad (2.42')$$

para el proceso continuo, y

$$X_{t+1} = \frac{1-e^{-b_1}}{b_1} \cdot X_1(t) + \frac{1}{\theta} \left[\frac{\theta_1}{b_1} (1-e^{-b_1}) + \frac{\theta_2}{b_2} (1-e^{-b_2}) \right] \cdot X_{1+2}(t) + \sum_{m=N_1(t)+1}^{N_1(t+1)} \left[y_{m_1} \frac{1-e^{-b_1(t+1-\tau_{m_1})}}{b_1} \right] + \\ + \sum_{m=N_{1+2}(t)+1}^{N_{1+2}(t+1)} \left\{ y_{m_{1+2}} \cdot \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} (1-e^{-b_1(t+1-\tau_{m_{1+2}})}) + \frac{\theta_2}{b_2} (1-e^{-b_2(t+1-\tau_{m_{1+2}})}) \right] \right\} \quad (2.43')$$

para el proceso promediado. Cabe destacar que las ecuaciones (2.36')-(2.43') podrían ser deducidas para un paso de tiempo mayor ($s > 1$), pero tomando en consideración funciones de respuesta con $J > 1$, que no se presentan en este trabajo para modelos de segundo orden.

Como se indicó antes, los primeros términos (o sub-grupos de términos para el DNS y MSOSN) de todas estas ecuaciones pueden calcularse a partir de un dato de inicio $X(0)$, y representan el valor del proceso hasta antes de t . Eventualmente, en lugar de depender de $X(t)$ pueden estar en función de $X_1(t)$ y $X_2(t)$ para los DNS, o de $X_1(t)$ y $X_{1+2}(t)$ para los MSOSN. Una vez calculados estos términos, como valor inicial de los respectivos procesos, se puede desechar $X(t)$.

(ii) En el primer paso no es necesario recordar el tiempo del último evento antes de t , de modo que tanto el contador m como los tiempos en que se producen los eventos τ_m se inicializan en cero. Esto es $m=0$, $\tau_m=0$.

(iii) El siguiente tiempo en que se produce un evento τ_m , se debe calcular con base en la expresión $\tau_{m+1} = \tau_m + E$, donde E es generado de manera aleatoria. Para el modelo de primer orden se utiliza una distribución exponencial con media $1/\nu$ para generar el E . Para los modelos de segundo orden de tipo SOSN la media es $1/\nu_1 = 1/\nu_2$, de modo que se producen

tiempos $\tau_{1,m+1}$ y $\tau_{2,m+1}$. En el DNS deben utilizarse dos distribuciones exponenciales con medias $1/\nu_1$ y $1/\nu_2$, para generar eventos independientes en $\tau_{1,m+1} = \tau_{1,m} + E_1$ y $\tau_{2,m+1} = \tau_{2,m} + E_2$. Finalmente, para el MSOSN también se utilizan dos distribuciones con medias $1/\nu_1$ y $1/\nu_2$, pero con el cuidado de que todos los eventos del componente 2 deben ser generados en tiempos $\tau_{1+2,m}$ que coincidan con algunos de los tiempos $\tau_{1,m}$ para el componente 1. Debe recordarse que la tasa de eventos ν_2 es inferior que ν_1 .

De aquí en adelante, para efectos de simplificar la explicación de este procedimiento, también se denominará τ_{m+1} a los tiempos de ocurrencia de los modelos DNS ($\tau_{1,m+1}$ y $\tau_{2,m+1}$) y MSOSN ($\tau_{1,m+1}$ y $\tau_{1+2,m+1}$).

(iv) Si $\tau_{m+1} > 1$, no se registrarán eventos en $(t, t+1)$, de modo que la generación de X_{t+1} y $X(t+1)$ está completa.

(v) Si por el contrario $\tau_{m+1} < 1$, una magnitud de evento y_m (ó y_{m_1}, y_{m_2} ó $y_{m_{1+2}}$) se genera de forma aleatoria a partir de una exponencial con media θ (ó θ_1 y θ_2 según sea el caso).

(vi) Las contribuciones de y_{m+1} (ó y_{m_1+1}, y_{m_2+1} ó $y_{m_{1+2}+1}$) a las ecuaciones para $X(t)$ y X_t escogidas en el paso (i) se calculan utilizando los segundos términos (o sub-grupos de términos) de dichas ecuaciones. Este aporte se debe sumar a los valores iniciales estimados en el paso (i) para X_{t+1} y $X(t+1)$.

(vii) Ahora m toma el valor $m+1$ y se repiten los pasos del (iii) al (vii) para obtener el siguiente caudal promedio diario. Este procedimiento se repite hasta que se considere necesario.

2.9 Propiedades de cruce de una serie cronológica

Las propiedades de cruce (CP) de una serie de tiempo se definen a partir del trazado de un umbral predeterminado. Gottschalk (2004) presenta las principales CP, que son (i) los cruces hacia abajo DC que es una secuencia de datos que precisa cuándo una serie cruza el nivel

umbral desde valores superiores hacia valores inferiores, (ii) los cruces hacia arriba UC que es una secuencia de datos que señala los cruces del umbral de valores menores a valores mayores, (iii) la cantidad de corridas n_{RN} en la serie, tomando en cuenta que las corridas RN son una secuencia ininterrumpida de datos por debajo (o por arriba) del nivel de umbral, acotados por un DC (o UC) a la izquierda y por un UC (o DC) a la derecha, y finalmente (iv) la longitud media de una corrida μ_{RN} , o sea, el promedio de los lapsos en que valores consecutivos de la serie están por debajo (o por arriba) del umbral.

Gottschalk (2004) también plantea un método para estimar las CP de un proceso X_t . Así, define $p=P(X \leq x_p) = F_{x_t}(x_p)$ como la probabilidad de observar datos debajo de un nivel cuantil x_p . Transformando la secuencia original en una serie de ceros y unos (de forma que el valor puntual del proceso x se reemplaza por 0 si $x \leq x_p$ y por 1 si $x > x_p$) se puede obtener el número esperado de UC en función de la cantidad de observaciones n , y de las probabilidades p y $(1-p)$ asociadas a un umbral x_p . Este autor plantea estimaciones similares para las demás propiedades de cruce (DC, n_{RN} ó μ_{RN}), tanto para secuencias aleatorias de datos independientes como para modelos auto-regresivos de primer orden, AR(1). En este último caso se vale de una distribución bi-normal $F_{x_t, x_{t+1}}(x_p, x_p)$ para calcular las probabilidades p en dos observaciones sucesivas, utilizando siempre para un nivel fijo x_p .

Los valores teóricos para las CP de modelos aleatorios sólo dependen del número de observaciones n y de la probabilidad p . Para los AR(1) debe utilizarse, además, la primera auto-correlación $\rho(1)$ de la serie. Por lo anterior, las pruebas basadas en las propiedades teóricas no consideran ninguna distribución subyacente, y se les conoce como libres de distribución, o libres de parámetros. No obstante, la suposición de un simple modelo AR(1) puede ser muy simplista para evaluar caudales diarios, en los que los picos rompen y cambian de nivel las recesiones. En este caso los valores esperados teóricos podrían no ser de utilidad para evaluar un modelo como el FPP.

Aún en el caso de que no se cuente con un desarrollo teórico para obtener las propiedades de cruce analíticamente, aún pueden ser de utilidad si se define un modelo para la serie observada, como el FPP que se propone en este trabajo. Como las CP no son parámetros

directos en un FPP, constituyen una buena base para probarlo valiéndose de la técnica de Monte Carlo. Gottschalk (2004) indica que con este enfoque las derivaciones teóricas pueden reemplazarse por un análisis estadístico directo de las propiedades de interés (aquí serían las de cruce) observadas en series de gran longitud y/o generadas en gran cantidad.

Entonces, para un grupo de series simuladas (o simplemente para una de ellas que se juzgue representativa del comportamiento medio de ese grupo) puede aplicarse una secuencia de umbrales, y para cada uno de ellos calcular los DC, UC, n_{RN} y μ_{RN} . Estos valores podrían compararse con las propiedades de cruce extraídas de la serie histórica en estudio, y así juzgar en qué medida, el modelo logra reproducir parámetros que no fueron incluidos explícitamente en las etapas de identificación, estimación y aplicación.

Capítulo 3 Datos y análisis descriptivo

3.1 Caracterización de las zonas de estudio

Se decidió trabajar con estaciones hidrológicas ubicadas en cuencas cercanas entre sí, dispuestas dentro de una franja relativamente central, y transversal al eje principal del territorio de Costa Rica (ver la Figura 3.1).

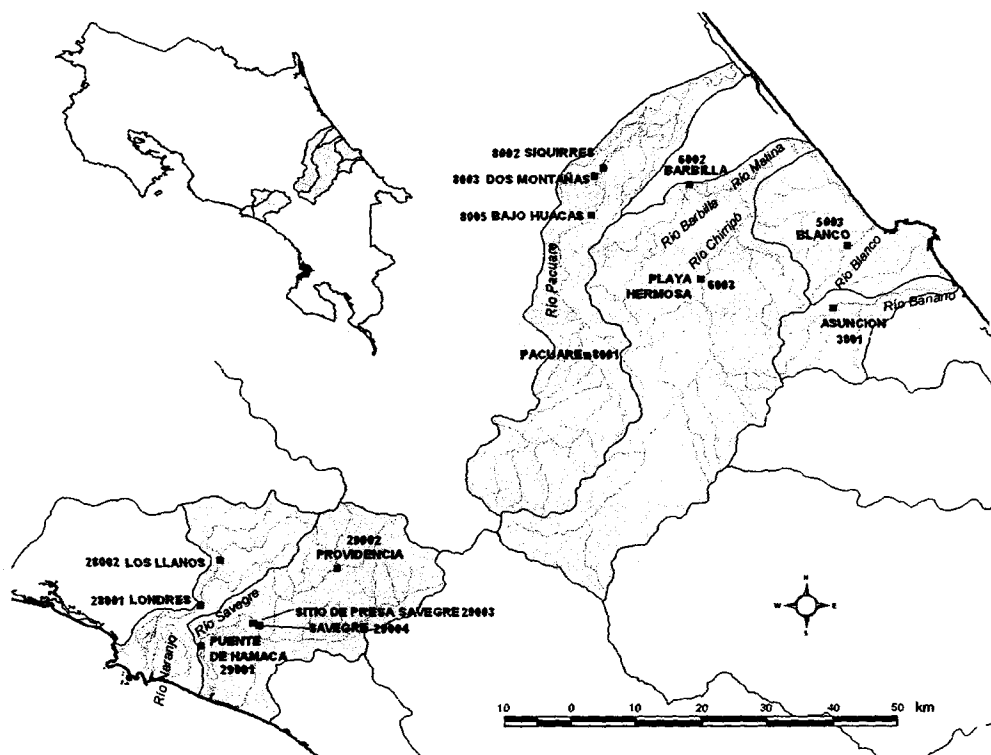


Figura 3.1 Estaciones hidrológicas escogidas para la simulación

La escogencia de las cuencas tiene que ver con el contraste entre la relativa cercanía de las estaciones en esta franja (las dos más alejadas se encuentran a una distancia proyectada en el plano horizontal de 116 km) y las grandes diferencias que se pueden observar en sus registros de caudales en cuanto a la distribución temporal del caudal a lo largo del año. Las estaciones ubicadas al noreste tienen influencia directa del clima del Caribe, en tanto las que están al suroeste tienen una influencia del Pacífico.

Otro factor tomado en cuenta en la selección de la región de trabajo y las estaciones de medición de caudal fue que los respectivos registros no hubiesen sufrido alteración alguna, en términos de afectaciones antropogénicas. Esto ocurre en estaciones ubicadas aguas abajo de un proyecto hidroeléctrico, por ejemplo. Asegurado esto, se contará con información del comportamiento natural de las cuencas y más concisamente de los ríos, en su condición de medios para encauzar los excesos de agua de lluvia hacia el mar.

Se estudiarán estaciones ubicadas en seis cuencas de Costa Rica, dos que desembocan en el Océano Pacífico, y las otras cuatro que desaguan hacia el mar Caribe.

3.1.1 Fisiografía

Las cuencas estudiadas poseen las características que se describen en la Tabla 3.1. De estas cuencas la más grande es la del río Matina, y la más pequeña la del río Banano, que es la de menor tamaño de las treinta y cuatro cuencas definidas en Costa Rica.

El índice de compacidad, que da una idea de la forma de la cuenca (relacionando su perímetro con el de un círculo de la misma superficie), presenta valores muy alejados de uno, lo cual indica que todas estas cuencas poseen una forma alargada, en especial las de los ríos Naranjo y Pacuare. La razón entre las elevaciones promedio y máxima refleja que la cuenca del río Savegre tiene elevaciones relativamente altas, en promedio, con respecto a su punto más alto, en tanto la del río Moín es bastante plana, y su mayor altitud (que no llega a los 500 *m.s.n.m.*) es casi siete veces mayor a la elevación promedio.

Tabla 3.1: Características fisiográficas generales de las cuencas estudiadas

Nº cuenca nacional	Nombre	Posición relativa según el tamaño	Área (km^2)	I_c	$\frac{h_m}{h_{m\acute{a}x}}$ (<i>m.s.n.m.</i>)	$h_{m\acute{a}x}/h_m$	Nº de estac.	
3	Río Banano	34	199.3	1.59	574	2103	3.7	1
5	Río Moín y otros	30	363.6	1.54	73	486	6.7	1
6	Río Matina	16	1410.8	1.72	1095	3704	3.4	2
8	Río Pacuare	20	874.4	1.95	1001	3216	3.2	4
28	Río Naranjo	31	339.9	2.11	869	2923	3.4	2
29	Río Savegre	23	638.4	1.56	1285	3423	2.7	4

I_c : índice de compacidad, h_m : elevación promedio, $h_{m\acute{a}x}$: elevación máxima

Como se mencionó, las cuencas están en una franja que representa aproximadamente el 7.5% del territorio del país, y está orientada de noreste a suroeste.

Las distribuciones de elevaciones y pendientes de terreno en las cuencas evaluadas se presentan en la Tabla 3.2. La cuenca del río Moín se destaca porque todo su territorio presenta elevaciones inferiores a los 1000 *m.s.n.m.* También porque casi el 80% posee pendientes menores o iguales al uno por ciento. La cuenca que difiere más de esta condición es Naranjo, en la cual casi una cuarta parte del área presenta elevaciones entre 2000 y 3000 *m.s.n.m.* y un 4% con 3000 *m.s.n.m.* o más. Además, la mitad de su espacio posee pendientes entre 5 y 15%.

Tabla 3.2: Distribución porcentual de elevaciones y pendientes en las cuencas estudiadas

Nombre	Elevación (<i>m.s.n.m.</i>)				Pendiente (%)			
	≤1000	1000-2000	2000-3000	≥3000	≤1	1-5	5-15	≥15
Río Banano	82	18	0	0	18	34	38	10
Río Moín y otros	100	0	0	0	79	15	6	0
Río Matina	50	36	11	3	12	30	46	12
Río Pacuare	55	35	10	0	20	28	42	10
Río Naranjo	47	26	24	4	9	26	50	15
Río Savegre	57	33	10	0	32	18	36	14

Se trabajará con catorce estaciones hidrológicas, la mayoría de las cuencas Pacuare y Savegre. Trece de ellas se encuentran en elevaciones inferiores a los 1000 *m.s.n.m.* (ver la Figura 3.2), de modo que en la mayoría de los casos se integra la información hidrológica de las partes altas de las cuencas.

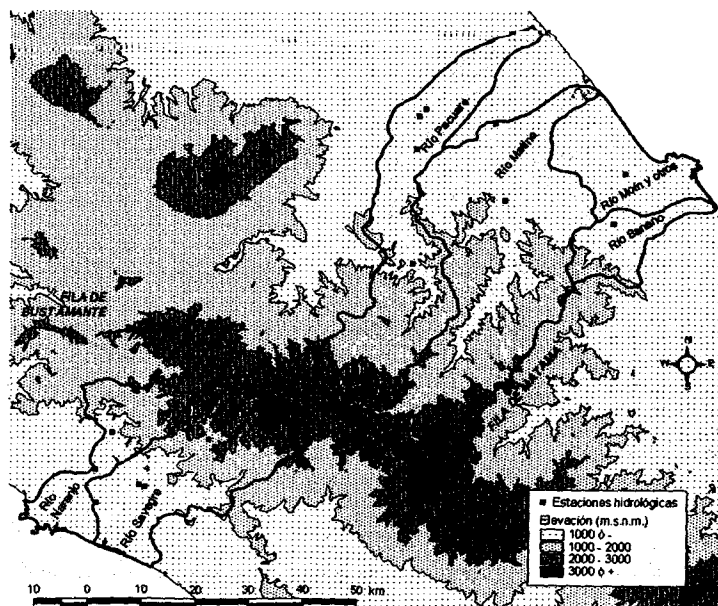


Figura 3.2: Ubicación de las estaciones hidrológicas y su elevación

La Tabla 3.3 resume las características principales de las estaciones, así como de sus cuencas tributarias. Se cuenta con registros para sitios de medición desde elevaciones de 27 *m.s.n.m.* en Puente de Hamaca (cuenca Savegre), hasta 1440 *m.s.n.m.* en Providencia (ubicada en la misma cuenca). Las elevaciones promedio de las cuencas tributarias se encuentran en un rango entre 199 *m.s.n.m.* para la estación Blanco hasta 2627 *m.s.n.m.* en Providencia. Las áreas de las sub-cuencas tributarias van desde los 43 *km²* para la estación 5003 Blanco (cuenca Moín), hasta los 821 *km²* para 6003 Playa Hermosa (Matina).

Tabla 3.3: Principales características de las estaciones analizadas y sus sub-cuencas tributarias

Nombre cuenca	Código estación	Nombre estación	Fecha instalac.	Elevación (m.s.n.m.)	A_d (km ²)	h_{min}	h_m	$h_{máx}$	$h_{máx}/h_m$
						(m.s.n.m.)			
Banano	3001	Asunción	01/07/57	157	91.4	162	841	1895	2.3
Moín	5003	Blanco	26/08/76	40	42.6	67	199	483	2.4
Matina	6002	Barbilla	29/11/76	30	212.0	39	623	1573	2.5
	6003	Playa Hermosa	21/10/81	140	820.6	138	1548	3715	2.4
Pacuare	8001	Pacuare	11/06/54	582	367.4	611	1671	3125	1.9
	8002	Siquirres	09/06/63	53	657.3	90	1281	3216	2.5
	8003	Dos Montañas	27/05/70	69	651.8	90	1280	3135	2.4
	8005	Bajo Huacas	23/02/02	191	615.9	181	1332	3135	2.4
Naranja	28001	Londres	30/07/70	170	210.2	192	1311	2923	2.2
	28002	Los Llanos	28/03/92	480	147.0	528	1510	2923	1.9
Savegre	29001	Pte de Hamaca	31/07/70	27	535.0	45	1511	3423	2.3
	29002	Providencia	24/10/79	1440	127.5	1500	2627	3401	1.3
	29003	S. P. Savegre	25/11/82	170	473.0	168	1672	3423	2.0
	29004	Savegre	01/04/97	240	468.1	186	1682	3423	2.0

A_d : área de drenaje, h_m : elevación promedio, $h_{máx}$: elevación máxima, h_{min} : elevación mínima

La relación entre la elevación máxima y la promedio es relativamente homogénea en las sub-cuencas definidas por las estaciones, con la mayoría de los valores entre 2.2 y 2.5. Un caso excepcional es nuevamente Providencia (Savegre), en la que ésta relación es cercana a uno. Lo anterior se debe a que su cuenca tributaria es relativamente pequeña, y a que casi toda su área se encuentra en la parte alta de la cuenca del río Savegre, por lo que la elevación promedio tiene un valor alto. Un contraste con el caso anterior y dentro de la misma cuenca es Puente Hamaca ($h_{máx}/h_m=2.3$), cuya área incluye la de Providencia, pero también abarca toda la parte media y una proporción de la parte baja de la cuenca general del río Savegre. En estos casos el promedio general tiende a disminuir, y el máximo gana valor relativo.

3.1.2 *Clima*

Como se mencionó antes, la variación climática en el área escogida se puede considerar muy elevada, pues en esta franja de unos 130 *km* de costa a costa, está presente la influencia de los vientos y corrientes marinas del océano Pacífico y del mar Caribe. Debido a que estos factores actúan en una franja de territorio muy angosta, no se da la atenuación de sus efectos al adentrarse en el territorio continental. Su interacción directa y con el sistema de cordilleras (en este caso particular con la parte noroeste de la cordillera de Talamanca y con la fila de Matama) produce un comportamiento muy complejo en las variables climáticas. Lo anterior se comprueba al comparar el clima prevaleciente en las llanuras costeras de las cuencas analizadas de la vertiente atlántica (Banano, Moín, Matina y Pacuare) llamado clima de la región Caribe, con el que prevalece en la zona litoral Pacífico del resto de las cuencas (Naranjo y Savegre) denominado clima lluvioso del Pacífico.

En términos generales, el clima de la región Caribe se caracteriza por una temperatura promedio de 27 °C, precipitaciones anuales entre 3500 y 4000 *mm* en las llanuras (que aumentan de oeste a este) y por la inexistencia de una estación seca definida. Los vientos alisios del noreste y las brisas marinas del Caribe inciden directamente originando fuertes lluvias en julio, diciembre y enero.

El clima lluvioso del Pacífico presenta una temperatura promedio de 27 °C y lluvias anuales en el rango 2000-3500 *mm*, que son aportadas de abril a noviembre por los vientos alisios del suroeste. Esto define una estación seca de entre tres y cinco meses. La estación lluviosa es muy intensa y torrencial. (CADETI-MINAE, 2004).

Pero también el clima de montaña presenta variantes entre las dos vertientes, dependiendo del tipo de exposición al viento en un momento determinado. Entonces según el viento dominante se da una vertiente húmeda, cuando se está a favor del viento (barlovento) o una de abrigo en caso contrario (sotavento). Naturalmente la altitud y la pendiente también influyen en la precipitación, así como en otras variables como la temperatura (disminución de unos 6 °C por cada 1000 de ascenso), y la presión atmosférica, creando un sistema climático complejo para

las partes medias y altas de las cuencas. Así, la precipitación puede llegar hasta 6000 mm en las partes altas de las cuencas del Caribe aquí analizadas, en tanto en el caso del Pacífico este valor es 5500 mm, pero se presenta en la zona media de las cuencas correspondientes.

La precipitación promedio anual en la región de estudio se presenta en la Figura 3.3. Los valores distribuidos para todo el país se obtuvieron a partir de una interpolación estocástica kriging de la precipitación anual en 178 estaciones meteorológicas de Costa Rica, para el periodo 1979-2000 (Gómez y otros, 2006).

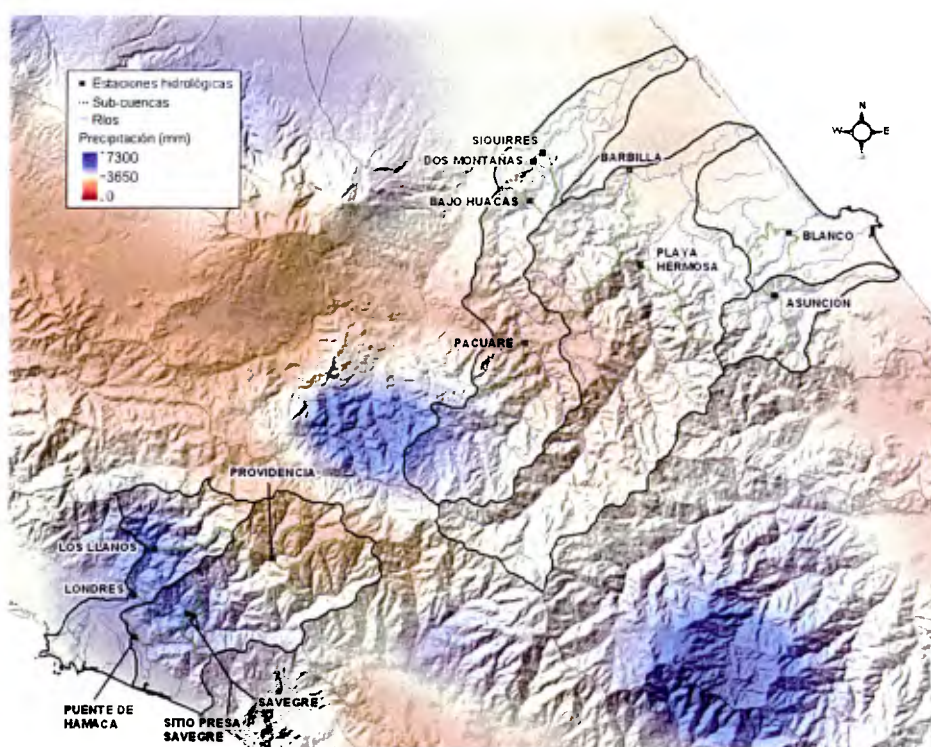


Figura 3.3: Lluvia media anual en la región analizada

Los principales estadísticos de la lluvia anual para cada una de las sub-cuencas tributarias a las estaciones analizadas se detallan en la Tabla 3.4. La precipitación mínima en el Caribe se observa en las áreas tributarias de las estaciones en las cuencas Matina y Pacuare (poco menos de 2600 mm), mientras en el Pacífico corresponde a la cuenca Savegre (unos 2500 mm). Los valores máximos en el Caribe están en la cuenca Pacuare (aproximadamente 5860 mm) en tanto que en el Pacífico esto se da en Savegre (5310 mm).

Tabla 3.4: Características de la precipitación anual en las cuencas estudiadas

Código estación	Nombre estación	P_{min} (mm)	$P_{máx}$ (mm)	P_m (mm)	C.V. (%)
3001	Asunción	3505	3752	3636	1.5
5003	Blanco	3595	3728	3663	1.0
6002	Barbilla	2777	3889	3418	7.6
6003	Playa Hermosa	2568	3948	3258	7.8
8001	Pacuare	2572	5857	3793	21.5
8002	Siquirres	2572	5857	3515	20.9
8003	Dos Montañas	2572	5857	3512	21.0
8005	Bajo Huacas	2572	5857	3497	21.6
28001	Londres	2649	5438	4401	17.2
28002	Los Llanos	2649	5438	4149	17.7
29001	Puente de Hamaca	2507	5310	3651	24.0
29002	Providencia	2507	3194	2655	5.5
29003	S. P. Savegre	2507	5273	3457	21.7
29004	Savegre	2507	5149	3445	21.5

P_{min} , $P_{máx}$, P_m : precipitaciones mínima, máxima y promedio, C.V.: Coeficiente de variación

La comparación de los coeficientes de variación (C.V.) por sub-cuenca con la Figura 3.3 ilustra la variabilidad espacial presente en el área analizada. Para Los Llanos, por ejemplo, con un área que no llega a los 150 km^2 , se tiene un C.V. de casi un 18%. En esta pequeña sub-cuenca se observan sitios con lluvias anuales promedio entre 2650 y 5440 mm (casi 2800 mm de diferencia en una distancia de sólo 18 km). Por el contrario Playa Hermosa, aún siendo la más grande de las sub-cuencas analizadas con un área de 820 km^2 (es 5.6 veces más grande) tiene un coeficiente de variación de sólo el 8%, de modo que su distribución espacial promedio presenta lluvia anual mínima de 2570 y máxima de 3950 (poco menos de 1400 mm de diferencia, es la mitad de lo observado en Playa Hermosa).

3.1.3 Perfil fisio-climático

En la Figura 3.4 se muestra un perfil fisiográfico-climático del tramo comprendido entre las dos estaciones más alejadas, de todas las elegidas para el análisis. Este tramo se aprecia en la parte a) de esa figura, y parte de la estación 29001 Puente de Hamaca (en el Pacífico), hasta llegar a la 5003 Blanco (en el Caribe). Como se había mencionado, la separación es de 116 km, y en esta distancia se encuentran importantes diferencias en las variables discutidas en las dos secciones anteriores, así como en las pendientes (con valores entre 0.3 y 31%). Los máximos de lluvia anual no se encuentran en los puntos de mayor elevación sino en ciertas

zonas medias-altas o medias-bajas de las cuencas. También queda claro el ascenso más abrupto a partir de las regiones costeras del Pacífico que a partir de las del Caribe. Esto se comprueba no sólo con el perfil de territorio, sino con la mayor frecuencia de altas pendientes en el sub-tramo que corresponde con la vertiente Pacífica (los primeros 37 km). En éste la pendiente cruza siete veces el umbral del 18%, que es el mismo número de cruces encontrado en los restantes 79 km de la vertiente Caribe.

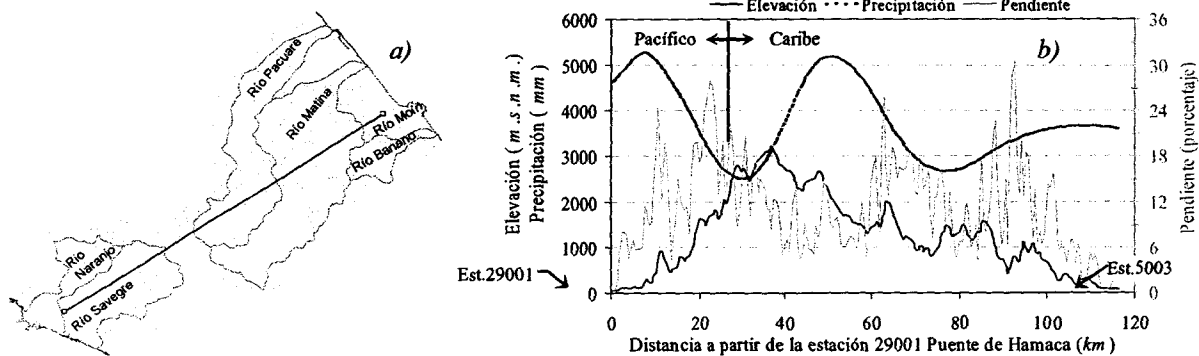


Figura 3.4: a) Línea de extracción del perfil igual al tramo recto entre las estaciones en estudio más alejadas. b) Elevación, pendiente y precipitación anual en la dirección Pacífico-Caribe (de la estación 29001 a la 5003)

3.2 Descripción general de las series de tiempo

Las series seleccionadas de caudal promedio diario corresponden, como se dijo antes, a mediciones del flujo en los cauces de los ríos, que reflejan la tasa a la que el agua sale de la respectiva sub-cuenca tributaria. El comportamiento de las series es importante en la etapa de estimación de un modelo como el proceso de Poisson filtrado, pues de las series observadas se extraerán los momentos estadísticos que se quieren preservar al ejecutar una simulación, en la forma expuesta en la sección 2.7.

Por circunstancias especiales, eventualmente una estación medidora puede dejar de operar, y esto ocasiona una pérdida de información que trunca la serie de tiempos, según se definió en la sección 2.1. Esto ocasiona huecos en el registro, que en casos especiales pueden ser completados utilizando diversas técnicas. La manera más común de hacerlo es correlacionando el caudal de la estación que tiene el registro incompleto, con el de una estación cercana que sea afectada de manera similar por los fenómenos climáticos, y cuya área de drenaje tenga condiciones hidrológicas semejantes. Cuando no existe una estación cercana,

o existe pero no registró caudales en el lapso de interés, aún podría utilizarse un modelo de lluvia-esorrentía calibrado de forma adecuada, para estimar los datos faltantes con base en las precipitaciones (y otras variables climáticas e hidrológicas). No obstante, esta última opción más bien podría considerarse excepcional por la alta demanda de información para hacer tal calibración, y lo engorroso de este proceso.

Una serie típica de caudal diario se observa en la Figura 3.5. Las líneas gruesas junto a la abcisa corresponden con lapsos sin registro de caudal. Es posible reconocer el patrón de diente de sierra que producen los saltos y recesiones, y la mayor frecuencia de eventos en un momento determinado del año, en el que además parece haber un nivel medio mayor en la serie.

En la Figura D.1 del 0 se presentan las series de caudal completas para las catorce estaciones seleccionadas. La cobertura temporal de cada una de las estaciones se puede ver en la Figura D.2. Resulta evidente que las series tienen diversas longitudes y fechas de medición. Otras comparaciones visuales entre las series de información diaria son más difíciles de efectuar, dada su complejidad y variabilidad. Por esto se recurrirá a propiedades de probabilidad y estadística para describirlas.

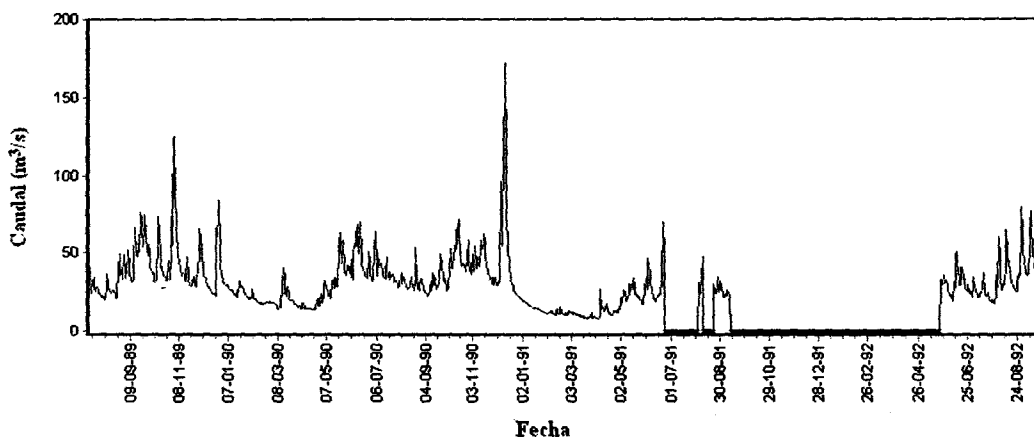


Figura 3.5: Una parte de la serie de caudal promedio diario en la estación hidrológica 8001 Pacuare

Las características generales de las series de las catorce estaciones hidrológicas analizadas se presentan en la Tabla 3.5. Los periodos de registro de estas series pueden ser tan extensos

como 48 años (estación Pacuare), o tan escasos como 4 años (estación Bajo Huacas). Esto permitirá evaluar la sensibilidad de la simulación de series como función de la cantidad de observaciones que se tienen en un sitio determinado. La situación de los huecos en los registros (lapsos de datos faltantes) afecta a cinco de las catorce series analizadas, de modo que en una misma secuencia pueden darse hasta cuatro interrupciones. La estación más perjudicada es Asunción, en la que los blancos representan un 3.5% de los días en el lapso de observación. En segundo lugar está Pacuare, que en la que faltan el 1.8% de los registros.

Tabla 3.5: Características de las series de caudal promedio diario para las estaciones hidrológicas

Código estación	Nombre de estación	Lapso de registro	Días (Años) de registro	Lapsos con faltantes	Cantidad y % de días faltantes	Estación(es) relacionadas
3001	Asunción	19/07/57-30/09/88	11415 (31)	4	399 (3.5%)	8001, 6002
5003	Blanco	01/02/77-30/04/95	6663 (18)	0	-	-
6002	Barbilla	01/12/76-11/05/98	7832 (21)	0	-	-
6003	Playa Hermosa	26/07/82-30/04/91	3201 (9)	0	-	-
8001	Pacuare	01/08/58-30/04/06	17440 (48)	4	316 (1.8%)	8003
8002	Siquirres	09/06/63-30/04/70	2518 (7)	1	17 (0.7%)	8001
8003	Dos Montañas	01/05/70-30/04/00	10958 (30)	0	-	-
8005	Bajo Huacas	23/02/02-30/04/06	1528 (4)	0	-	-
28001	Londres	01/08/70-26/04/06	13053 (36)	2	10 (0.1%)	29004
28002	Los Llanos	28/03/92-14/01/06	5041 (14)	2	35 (0.7%)	29004, 29002
29001	Pte. de Hamaca	01/08/70-30/04/89	13053 (36)	0	-	-
29002	Providencia	24/10/79-26/04/06	9682 (27)	0	-	-
29003	S. P. Savegre	24/11/82-27/07/96	4995 (14)	0	-	-
29004	Savegre	01/05/98-28/02/06	2891 (8)	0	-	-

En todos los casos fue posible completar los valores en blanco de las series, basándose en información de estaciones cercanas que registran valores significativamente correlacionados. Para lo anterior se utilizó el método de regresión lineal antes discutido, pero se introdujo una variante para mejorar su eficiencia. Si se utiliza el conocimiento del alto nivel de sesgo que presentan las distribuciones de caudal diario, es posible obtener una relación lineal más clara después de hacerles una transformación logarítmica. Esto se debe a que al aplicar tal transformación se disminuye la variabilidad de la serie, preservando las tendencias generales subyacentes que corresponden con los impulsos externos que actúan sobre el sistema hidrológico. La comprobación se ejemplifica en la Figura 3.6, donde el coeficiente de determinación de una regresión lineal (R^2) entre las estaciones relacionadas, mejora de 0.69 en la parte *a*) de la figura para los datos sin transformación, hasta 0.81 en la parte *b*) que muestra

el resultado con los datos transformados. Aunque siempre se mejora el R^2 al aplicar esta transformación, no en todos los casos el cambio es tan significativo.

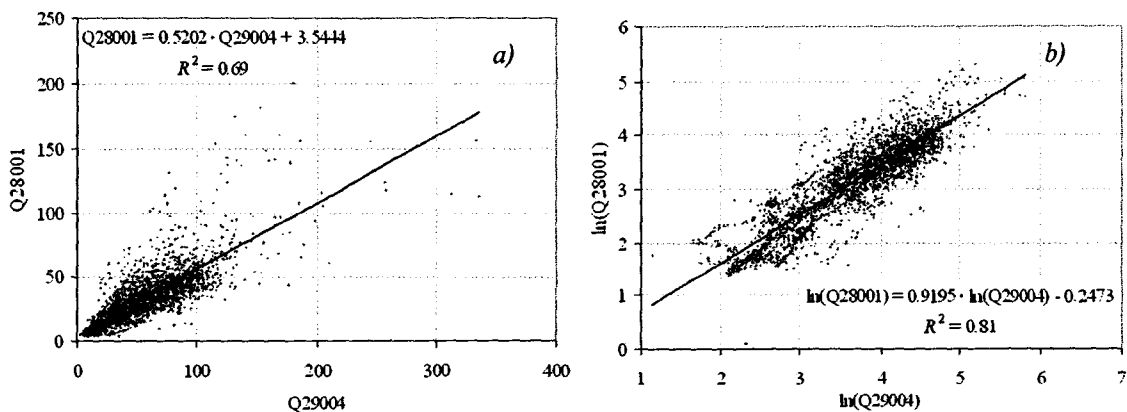


Figura 3.6: Regresión lineal aplicada a los caudales diarios observados en las estaciones 28001 Londres y 29004 Savegre a) sin transformación, y b) con una transformación logarítmica

La regresión presentada en la Figura 3.6 b) permite estimar valores no registrados en la estación 28001 Londres, a partir de la 29004 Savegre, utilizando la expresión:

$$Q^*_{28001} = e^{0.9195 \cdot \ln(Q_{29004}) - 0.2473}$$

Donde Q^*_{28001} es el caudal estimado en la estación Londres y Q_{29004} el caudal observado en la estación 29004 Savegre.

Una relación como la anterior se estableció para rellenar los lapsos faltantes de las estaciones que lo requirieron (ver las columnas quinta y sexta de la Tabla 3.5), tomando como base una estación relacionada (séptima columna de la misma tabla). En algunos casos se utilizaron distintas estaciones relacionadas para reponer diversos huecos en una misma serie, pues no siempre la estación más correlacionada estaba operando en todos los lapsos en que se requería. Un ejemplo de reconstrucción de un registro de caudal diario se observa en la Figura 3.7, de modo que los valores observados para la estación Dos Montañas (serie color gris) se utilizan para completar el registro en Pacuare (la línea negra continua representa los valores observados y la discontinua los generados con la regresión lineal).

Todos los huecos en las series escogidas fueron llenados utilizando la técnica de regresión lineal. Cada caso particular se ilustra en el Anexo E. Una vez que se cuenta con las series

completas, es posible describirlas por medio de técnicas estadísticas básicas y métodos conocidos en el análisis de series de tiempos. Para cada serie se intentará encontrar funciones de distribución teóricas para los valores diarios, evaluar si es estacionaria en la media o si presenta tendencia, el nivel de independencia de sus observaciones y si presenta componentes periódicos o casi-periódicos.

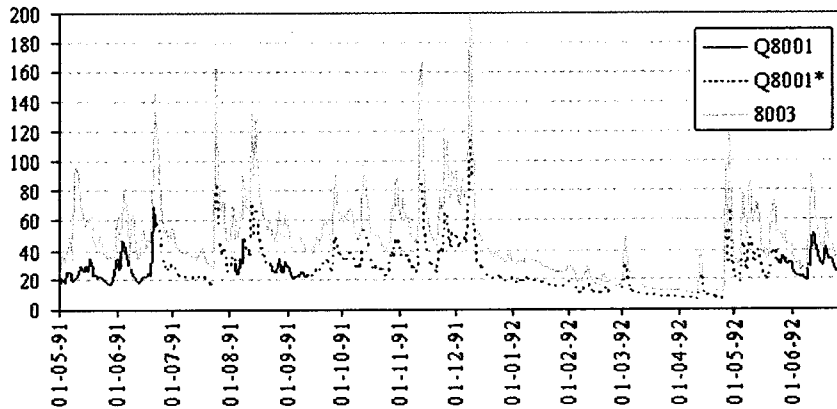


Figura 3.7: Reemplazo de valores en blanco de la estación 8001 Pacuare, tomando como base un registro observado en la 8003 Dos Montañas. Q8001* es la serie generada con regresión lineal

3.2.1 Estadísticos principales de las series de caudal diario

Las principales propiedades estadísticas de la series de tiempos en estudio se resumen en la Tabla 3.6. De este grupo de estaciones, por la que discurre mayor caudal promedio diario es la 8002 Siquirres, con $67.1 \text{ m}^3/\text{s}$ y la menos caudalosa es la 5003 Blanco, con $3.6 \text{ m}^3/\text{s}$. Siquirres también tiene la mayor desviación estándar ($61.6 \text{ m}^3/\text{s}$), pero la menor más bien se observa en la estación 29002 Providencia ($6.7 \text{ m}^3/\text{s}$). Blanco presenta una desviación muy similar a Providencia, pero su promedio es cercano a la mitad del de dicha estación.

Es interesante que los caudales promedio diarios en Siquirres y en Dos Montañas deberían ser casi iguales, pues están ubicadas casi en el mismo sitio. No obstante hay una diferencia de casi $12 \text{ m}^3/\text{s}$ que parece indicar que ha habido un cambio en la media. Debe considerarse que los periodos de observación no son simultáneos y su longitud es considerablemente distinta. En la serie de la estación 8002 Siquirres parece existir cierta distorsión (ver la Figura D.1) provocada por la llegada de un año extraordinariamente lluvioso, 1970, en el que hubo una

gran afectado por el fenómeno meteorológico La Niña. Incluso, los eventos de abril de ese año provocaron la salida de operación del equipo medidor en el río. Dado que el registro de Siquirres es tan corto (unos siete años) la afectación es considerable, lo cual se comprueba al eliminar el último medio año de registro y verificar que la media desciende a $63.2 \text{ m}^3/\text{s}$. Como la serie medida en Dos Montañas consta de 30 años de medición (incluyendo gran parte de 1970) se considera que la media de $55.5 \text{ m}^3/\text{s}$ es más confiable para este sitio de observación común.

Tabla 3.6: Principales estadísticos de las series diarias de caudal

Código estación	Nombre de estación	<i>n</i>	Media (m^3/s)	Desv. Est. (m^3/s)	C.V.	Sesgo	Curtosis	Min.	1°Q	Med.	3°Q	Max.
3001	Asunción	11415	15.6	19.7	1.3	8.0	115	2.2	7.3	11.0	16.6	516
5003	Blanco	6663	3.6	7.2	2.0	6.3	60	0.1	0.8	1.5	3.3	121
6002	Barbilla	7832	25.1	25.5	1.0	4.6	36	3.9	11.8	18.3	28.1	363
6003	Playa Hermosa	3201	62.5	48.2	0.8	7.5	118	2.5	36.5	54.3	76.8	1094
8001	Pacuare	17440	32.4	24.7	0.8	7.2	180	4.3	16.9	27.7	41.6	1006
8002	Siquirres	2518	67.1	61.6	0.9	9.5	188	10.9	35.7	57.8	81.4	1620
8003	Dos Montañas	10958	55.5	44.2	0.8	5.7	73	4.7	29.2	47.1	69.5	940
8005	Bajo Huacas	1528	62.7	50.1	0.8	4.0	32	10.7	33.5	51.6	75.5	630
28001	Londres	13053	28.5	22.5	0.8	2.1	15	3.0	10.1	24.4	40.4	401
28002	Los Llanos	5041	16.1	12.1	0.8	2.1	14	1.2	6.4	14.0	22.0	155
29001	Pte. de Hamaca	6848	61.8	52.5	0.9	2.1	12	6.4	19.2	50.4	87.7	566
29002	Providencia	9682	6.7	6.2	0.9	2.6	15	0.6	2.5	4.5	8.6	81
29003	S. P. Savegre	4995	40.4	32.3	0.8	2.4	17	5.5	14.6	33.6	56.5	432
29004	Savegre	2861	50.6	35.5	0.7	1.4	8	3.2	20.9	43.3	73.7	335

Un caso muy similar al de Siquirres y Dos Montañas se da para las estaciones 29003 Sitio de Presa Savegre y 29004 Savegre. Aquí la diferencia es cercana a los $10 \text{ m}^3/\text{s}$, y los periodos de medición son casi consecutivos y de longitudes no tan distintas (catorce años de la primera contra ocho de la segunda). Sin embargo, tampoco puede concluirse con propiedad que la media ha variado hasta tanto no se examinen posibles señales cíclicas en mayores escalas de tiempo.

Es posible comprobar con facilidad que la media y la desviación estándar de las series de tiempo aumentan de forma casi lineal con el área de la subcuenca tributaria. La variabilidad temporal relativa de las observaciones, expresada por medio del coeficiente de variación (C.V.) es bastante uniforme en las estaciones del Pacífico y del Caribe. Sin embargo, sobresale la estación Blanco por presentar un C.V. aproximado a 200%. Esto podría ser explicado al

tomar en cuenta que Blanco mide la información de la cuenca más pequeña entre las catorce aquí evaluadas. De hecho, es posible relacionar el tamaño de la cuenca con el C.V. espacial en la variable de impulso (la precipitación) y con el C.V. temporal en la variable respuesta (el caudal). Esta relación se expone en la Figura 3.8, y denota que en el primer caso si una cuenca abarcar más área, tendrá mayores contrastes climáticos y un mayor C.V. Lo anterior rige hasta que ya no se continúan dando tales contrastes y, por tanto, la variabilidad tiende a disminuir al abarcar más área que tendrá un comportamiento más homogéneo. Este efecto es el que se produce en el C.V. temporal para el caudal, pues aunque presenta magnitudes de variación relativa mayor en un orden de magnitud que en el caso espacial, parece disminuir y estabilizarse más rápido en función del aumento en el área tributaria. En este caso siempre se espera que al tener mayor capacidad colectora, una cuenca se comporte de forma más estable.

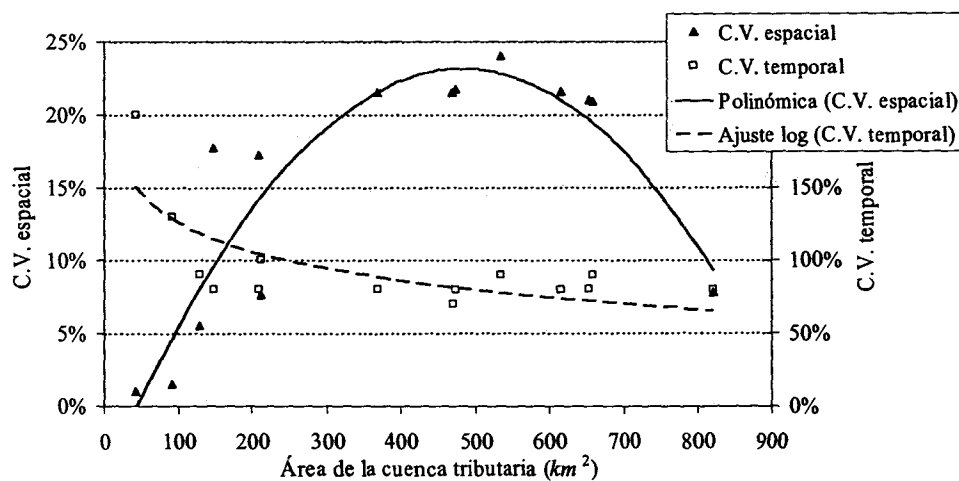


Figura 3.8: Relación entre los coeficientes de variación para la lluvia (C.V. espacial) y para el caudal (C.V. temporal) con el tamaño de la sub-cuenca de drenaje

Ni el sesgo ni la curtosis del caudal dependen del área de la cuenca tributaria a la estación medidora, pero entre ellas están altamente correlacionadas con coeficiente de correlación de Pearson igual a 0.92. Esto implica que poseen un alto nivel de asociación lineal. La mediana y otros valores percentiles se presentan en la última columna de la Tabla 3.6.

3.2.2 Distribución de los caudales promedio diarios

Las distribuciones de los caudales diarios se muestran en la Figura 3.9. De aquí se obtiene información gráfica adicional sobre el sesgo y la curtosis de las series, que fueron presentados

en la Tabla 3.6. Niveles de curtosis (que puede verse como concentración alrededor de algún valor) y sesgo positivo muy grandes se reconocer en las distribuciones de las estaciones 8002 Siquirres, 3001 Asunción, 6003 Playa Hermosa, 8001 Pacuare y 5003 Blanco. Nótese que todas pertenecen a la región Caribe, que posee un patrón climático más estable que la Pacífica, como se comentó antes. En todas las estaciones de esta última región se observan niveles de sesgo muy inferiores (cuencas número 28 y 29).

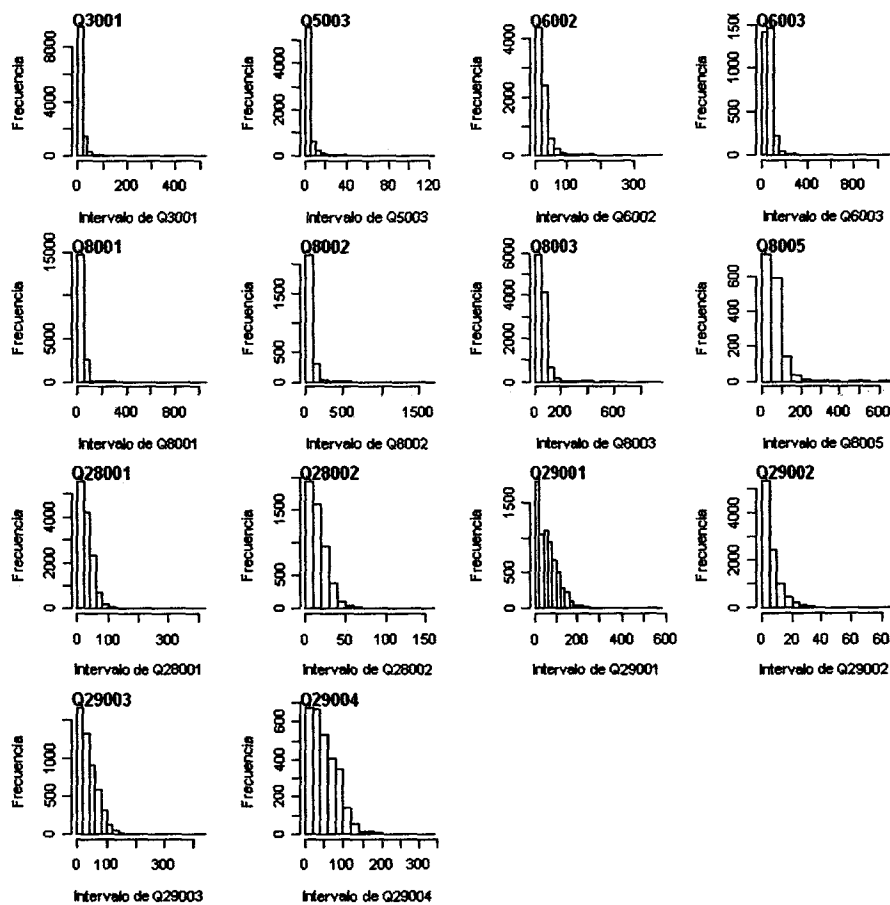


Figura 3.9: Distribuciones de frecuencia de los caudales diarios en las estaciones hidrológicas seleccionadas

Se intentó ajustar funciones de distribución teóricas a los valores observados. Evidentemente el comportamiento de las distribuciones empíricas (Figura 3.9) es muy diferente al normal, por lo que en todos los casos una prueba de Kolmogorov-Smirnov de dos colas revela diferencias significativas al 95% de confianza (o en términos operativos, que los valores del estadístico p

resultaron muy inferiores que 0.05). Otras de las distribuciones más comunes para el análisis de caudales se evaluaron, y los resultados se presentan en la Tabla 3.7.

Tabla 3.7: Ajuste de funciones de distribución teóricas a las distribuciones observadas de las series escogidas

Código estación	Nombre de estación	normal	Log-normal	gamma	gev
3001	Asunción	x	x	-	x
5003	Blanco	x	x	-	√
6002	Barbilla	x	x	-	x
6003	Playa Hermosa	x	x	x	x
8001	Pacuare	x	x	x	x
8002	Siquirres	x	x	x	x
8003	Dos Montañas	x	x	x	x
8005	Bajo Huacas	x	√	x	√
28001	Londres	x	x	x	x
28002	Los Llanos	x	x	x	x
29001	Pte. de Hamaca	x	x	x	x
29002	Providencia	x	x	x	x
29003	S. P. Savegre	x	x	x	x
29004	Savegre	x	x	x	x

*El símbolo √ quiere decir que la distribución empírica se acerca a la teórica, y el símbolo x indica lo contrario. Un valor ausente indica que la rutina para estimar los parámetros de la distribución teórica no convergió

La prueba de Kolmogorov-Smirnov es una prueba de adecuación a una distribución, y se utiliza para decidir si una muestra proviene de una población con determinada distribución. Se aplica sobre la función de distribución acumulada (cdf) empírica, y se basa en su distancia máxima con respecto a la cdf teórica, evaluando las funciones en los puntos muestreados.

En general, prácticamente ninguna de las distribuciones teóricas evaluadas logra explicar adecuadamente el comportamiento del caudal diario. Un ejemplo de los mejores ajustes obtenidos para los grupos de estaciones en el Caribe y en el Pacífico se presenta en la Figura 3.10. El procedimiento se ilustra utilizando gráficos de probabilidad (llamados P-P), gráficos de cuantiles (conocidos como Q-Q), gráficos de recurrencia de eventos, y la superposición del histograma de valores observados contra la función teórica de densidad de probabilidad.

En la parte a) de la figura anterior se presenta el ajuste de la distribución general de valores extremos (gev) para el registro de la estación Bajo Huacas. En ese caso el estadístico D de la prueba de Kolmogorov-Smirnov tiene un valor de 0.029, lo cual resulta en un valor p de 0.16.

Esto quiere decir que la máxima diferencia encontrada entre las distribuciones no es estadísticamente significativa. La parte *b)* de la figura presenta un ajuste *gev* para la estación Savegre, que sin embargo no resultó adecuado, pues se considera que la desviación entre las distribuciones es altamente significativa ($D=0.063$ y $p=2 \times 10^{-10}$).

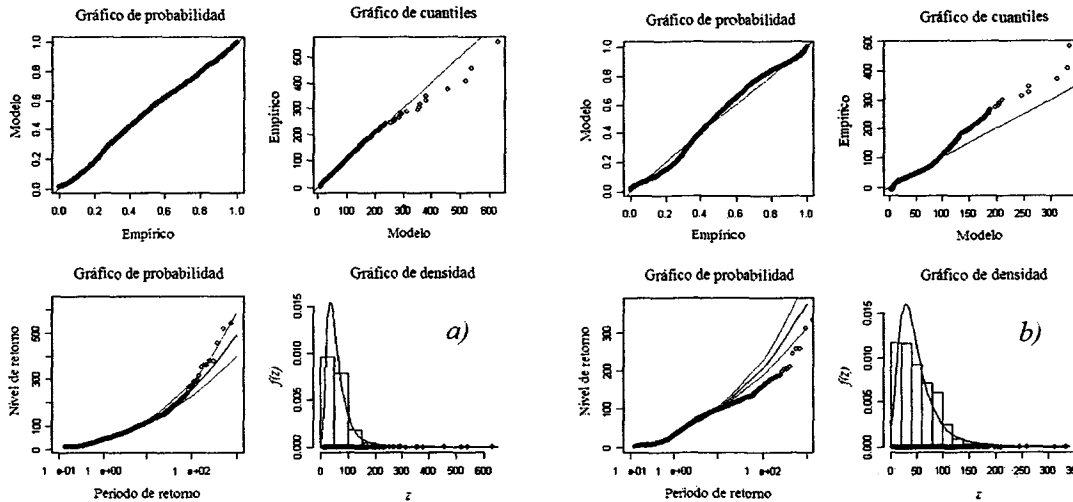


Figura 3.10: pruebas gráficas de ajuste de distribuciones teóricas para estaciones en los regímenes Caribe y Pacífico *a)* *gev* para la estación 8005 Bajo Huacas y *b)* *gev* para la estación 29004 Savegre

Analizando los resultados de todas las pruebas y los diversos gráficos P-P, Q-Q y de distribuciones de frecuencia superpuestas, se comprueba que las distribuciones teóricas aquí evaluadas son aún menos adecuadas para describir las distribuciones de las series de la vertiente Pacífico que las del Caribe. Una vez más, el comportamiento del Pacífico es más complejo en términos del efecto de las dos estaciones climáticas claramente diferenciadas, que aportan sus características particulares a la distribución general obtenida con todo el registro de caudal.

La caracterización del proceso $X(t)$, el caudal promedio diario, por medio de su curva de duración (una herramienta de uso común en hidrología), se presenta en la Figura 3.11. Ésta es una caracterización parcial en la forma de una función unidimensional de distribución de los datos observados, y se especifica como la densidad de probabilidad de primer orden. Dado que la caracterización completa sólo se alcanza con una distribución multi-variada, el emplear una uni-variada implica el uso de una distribución marginal de los datos. La curva de duración

indica en el eje horizontal cuál es el porcentaje del tiempo en que un determinado caudal (en el eje vertical) fue igualado o excedido. Por tanto, haciendo algunas consideraciones, se pueden suponer como curvas de probabilidad, o como un gráfico de la función de cuantiles empíricos X_p . Esto es, el p -ésimo cuantil o percentil del caudal para cierta duración, contra la probabilidad de excedencia p , donde $p = 1 - P(X \leq x) = 1 - F_X(x)$, y $F_X(x)$ es la función acumulada de $X(t)$ (Gottschalk, 2005). No obstante, se reconoce el hecho de que las observaciones poseen un patrón de auto-correlación, y son afectadas por la estacionalidad, por lo que no se trata de curvas estrictamente probabilísticas.

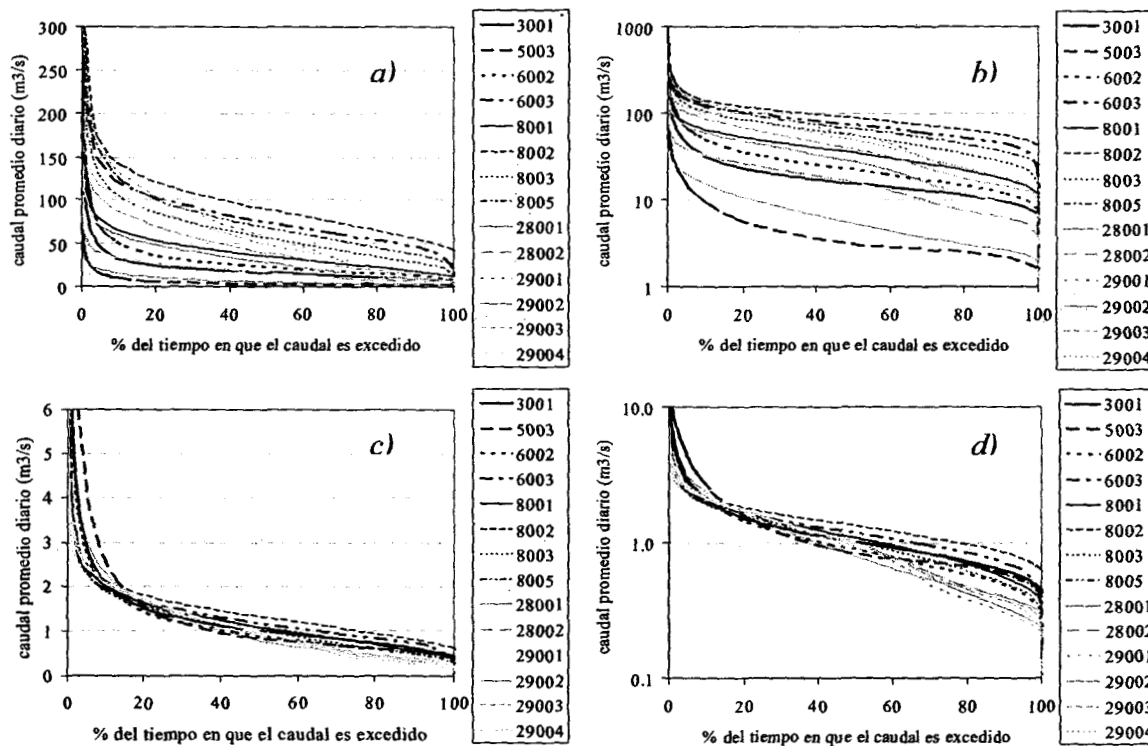


Figura 3.11: Curvas de duración de caudal diario en las estaciones escogidas. Colores negro: Pacífico, gris: Caribe. *a)* Curvas tradicionales para cada serie, *b)* curvas tradicionales y escala logarítmica en la vertical, *c)* curvas adimensionales (caudales normalizados) y *d)* curvas adimensionales con eje vertical logarítmico

La parte *a)* de la figura anterior reseña las curvas de duración en escala original para las catorce estaciones en estudio (se recortó la escala vertical para diferenciar mejor las curvas). Las líneas color gris representan las curvas de las estaciones en la vertiente del Caribe, mientras las de color negro están asociadas a las estaciones del Pacífico. La sección *b)* presenta las mismas curvas, pero con el eje vertical en escala logarítmica, lo que permite observar con más claridad su comportamiento para caudales bajos, que en análisis diarios

constituyen la gran mayoría de los valores. Aquí es posible notar la variación en los niveles de las curvas dentro de rangos distintos, y definidos por las respectivas magnitudes del caudal en cada estación. También que existe una diferencia en la forma de las curvas del Caribe (color gris) con respecto a las del Pacífico (color negro), pues las primeras tienen una pendiente menor, y caen más abruptamente al acercarse al porcentaje de excedencia del 100%. Al normalizar las curvas (se utilizó la media), la parte *c)* de la Figura 3.11 revela que la diferencia en la forma de las curvas en escala natural es casi imperceptible. Sólo después de dibujarlas con escala logarítmica en el eje del caudal es posible diferenciar la mayor rapidez con que caen las curvas de las estaciones en la vertiente Pacífica, en comparación con la vertiente Caribe. Llama la atención en las secciones *c)* y *d)* de esa figura el comportamiento de la estación 5003 Blanco, que en concordancia con las características de variabilidad relativa analizadas en la Tabla 3.6, presenta un comportamiento más volátil en sus máximos, en comparación con las demás estaciones de cualquiera de las dos vertientes.

3.2.3 Evaluación de la condición estacionaria o de tendencia

Se propone un modelo lineal simple para evaluar posibles tendencias en el nivel medio de las series. Los parámetros de tal modelo se muestran en la Tabla 3.8, así como el intervalo de 95% de confianza para la pendiente. En el caso de que este intervalo excluya el cero se manifiesta una tendencia significativa en la serie. Esto se observa en todas las estaciones en estudio, con excepción de Los Llanos y Savegre. La pendiente representa la tasa de cambio sostenido del caudal promedio diario por día, o sea $[m^3/s]/[día]$. Entonces, para la máxima tendencia absoluta encontrada ($3.68 \times 10^{-4} m^3/s/día$, observada en Bajo Huacas), en un año se espera que el caudal medio diario aumente $0.13 m^3/s$, llegando a $1.3 m^3/s$ en una década.

La prueba no paramétrica de Mann-Kendall, para buscar tendencias en los datos, refuerza el resultado anterior. En todas las estaciones el estadístico de prueba z distribuido normal estándar se encuentra fuera del intervalo de confianza al 95%, que se define por los límites -1.96 y 1.96 . La excepción son Londres ($z = -1.4$) y nuevamente Savegre ($z = -0.7$). La conclusión de estos dos enfoques para el análisis de tendencias es que para la mayoría de las series estudiadas no se puede garantizar la condición de media estacionaria.

Tabla 3.8: Parámetros del modelo lineal propuesto para analizar la tendencia de las series. Intervalo de confianza del parámetro de pendiente y valoración de la tendencia de la serie

Código Estac.	Nombre de estación	Años de registro	Intercep	Error Est.	Pendiente	Error Est.	L.I. Pend	L.S Pend	Tend?
3001	Asunción	1957-1988	18.9	0.37	-2.08×10^{-5}	2.03×10^{-6}	-2.48×10^{-5}	-1.69×10^{-5}	sí
5003	Blanco	1977-1995	3.0	0.18	7.03×10^{-6}	1.67×10^{-6}	3.77×10^{-6}	1.03×10^{-5}	sí
6002	Barbilla	1976-1998	20.6	0.57	4.24×10^{-5}	4.62×10^{-6}	3.33×10^{-5}	5.14×10^{-5}	sí
6003	Playa Hermosa	1982-1991	66.0	1.70	-7.85×10^{-5}	3.35×10^{-5}	-1.44×10^{-4}	-1.28×10^{-5}	sí
8001	Pacuare	1958-1906	34.9	0.37	-1.02×10^{-5}	1.35×10^{-6}	-1.28×10^{-5}	-7.54×10^{-6}	sí
8002	Siquirres	1963-1970	54.6	2.44	3.63×10^{-4}	6.11×10^{-5}	2.43×10^{-4}	4.83×10^{-4}	sí
8003	Dos Montañas	1970-2000	60.4	0.84	-3.27×10^{-5}	4.85×10^{-6}	-4.22×10^{-5}	-2.31×10^{-5}	sí
8005	Bajo Huacas	2002-2006	55.0	2.55	3.68×10^{-4}	1.05×10^{-4}	1.61×10^{-4}	5.74×10^{-4}	sí
28001	Londres	1970-2006	29.5	0.39	-5.40×10^{-6}	1.91×10^{-6}	-9.14×10^{-6}	-1.66×10^{-6}	sí
28002	Los Llanos	1992-2006	15.5	0.34	8.27×10^{-6}	4.26×10^{-6}	-7.89×10^{-8}	1.66×10^{-5}	no
29001	Pte. de Hamaca	1970-1989	70.4	1.26	-9.16×10^{-5}	1.16×10^{-5}	-1.14×10^{-4}	-6.88×10^{-5}	sí
29002	Providencia	1979-2006	6.2	0.13	3.85×10^{-6}	8.21×10^{-7}	2.24×10^{-6}	5.46×10^{-6}	sí
29003	S. P. Savegre	1982-1996	37.7	0.91	3.86×10^{-5}	1.15×10^{-5}	1.60×10^{-5}	6.12×10^{-5}	sí
29004	Savegre	1998-2006	49.6	1.33	2.65×10^{-5}	2.93×10^{-5}	-3.10×10^{-5}	8.39×10^{-5}	no

3.2.4 Independencia de las observaciones

La prueba de corridas (conocida como “runtest”) es útil para evaluar la independencia serial de las observaciones de caudal diario. El estadístico de prueba z evalúa si las magnitudes de las observaciones se obtienen en un orden aleatorio, contando las corridas que se dan por arriba y por debajo de un umbral (usualmente la mediana). Como se espera para este tipo de datos el z es sumamente pequeño, lo cual refleja una correlación serial positiva grande. El estadístico se encuentra muy lejos de los límites de confianza $[a,b]$ al 95%, que acotan el z asociado con una secuencia independiente (ver la Tabla 3.9).

Tabla 3.9: Prueba de corridas (“runtest”) para evaluar la independencia de las observaciones en las series estudiadas

Código Estac.	Nombre de estación	a	z	b
3001	Asunción	5604	1688	5813
5003	Blanco	3253	875	3412
6002	Barbilla	3830	927	4004
6003	Playa Hermosa	1546	290	1657
8001	Pacuare	8592	1475	8850
8002	Siquirres	1211	341	1309
8003	Dos Montañas	5377	1159	5583
8005	Bajo Huacas	727	183	803
28001	Londres	6416	754	6639
28002	Los Llanos	2452	281	2591
29001	Pte. de Hamaca	3344	386	3506
29002	Providencia	4746	362	4938
29003	S. P. Savegre	2429	246	2568
29004	Savegre	1379	165	1484

Lo anterior implica que ninguna estación tiene una distribución idénticamente independiente a escala diaria y, en principio, ninguna de estas series debería tratarse como una muestra x de su respectiva variable aleatoria X . Dada esta condición, la distribución de probabilidad unidimensional no proporciona una caracterización completa de la variabilidad del X (Gottschalk, 2005). Otras propiedades, como las de cruce discutidas en el Capítulo 2, deben ser tomadas en cuenta para valorar los rasgos característicos de los caudales diarios, y eventualmente evaluar su simulación a partir de un modelo propuesto.

La función de auto-correlación permite ver posibles estructuras de dependencia entre las observaciones de la series. La Figura 3.12 presenta tres de estas funciones. En *a)* y en *b)* se tienen estaciones de la vertiente del Caribe (Blanco y Pacuare, respectivamente), que tienen estructuras de correlación altamente variantes. En la primera, la auto-correlación es significativa hasta un rezago de aproximadamente un mes (las líneas discontinuas alrededor del eje horizontal limitan la región donde este efecto se puede considerar no significativo), en tanto que en la segunda existen fuertes auto-correlaciones aún después de un año, pero de signo variante dependiendo si se está en la misma estación climática (lluviosa o seca). La parte *c)* de la figura (estación Londres) representa el comportamiento generalizado en las estaciones de la región del Pacífico, que consiste en un patrón de muy altas auto-correlaciones, que varían de manera más gradual y estable, y en las que es aún más marcada la estructura estacional. En esta estructura se observan valores negativos con un máximo cercano a los 183 días, o los seis meses.

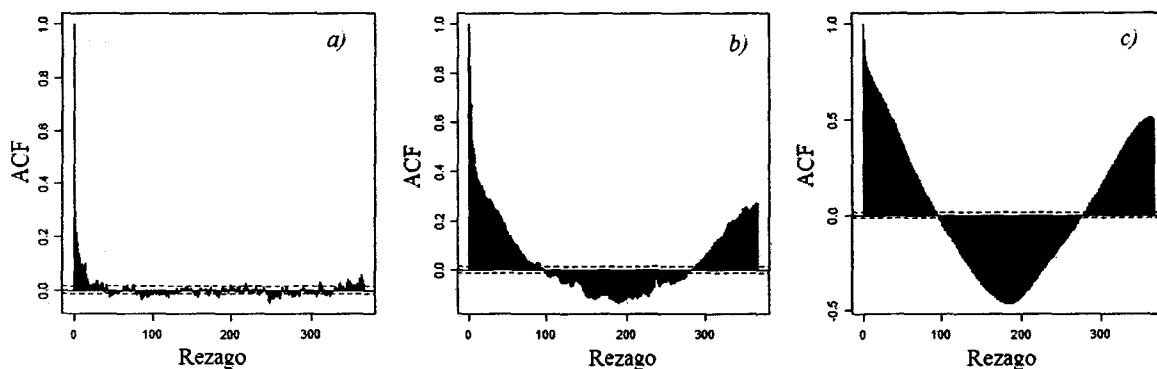


Figura 3.12: Funciones de auto-correlación para tres series de caudal *a)* 5003 Blanco, *b)* 8001 Pacuare y *c)* 28001 Londres

Un análisis de todos los auto-correlogramas en la Figura F.1 del Anexo F permite validar lo anterior, así como profundizar en los contrastes de las características de persistencia en los caudales diarios en el Caribe. Así, un primer grupo conformado por las estaciones Asunción, Blanco y Barbilla tienen la misma estructura, que se puede deber a que son cuencas pequeñas, ubicadas en la parte baja de la vertiente y cercanas a la zona costera. El segundo grupo lo constituyen Siquirres y Bajo Huacas, las cuales tienen mayores áreas de drenaje, pero aún se ubican en la parte baja de la cuenca Pacuare, muy cercanas entre sí. En este grupo debería estar la estación Dos Montañas que es la sustituta de Siquirres, pero más bien se comporta según el patrón de un tercer grupo. Las estaciones de este tercer grupo también poseen áreas tributarias grandes, pero en el caso de Playa Hermosa y Pacuare, están ubicadas más cerca de la zona media de sus respectivas cuencas, al pie del sistema montañoso (ver la Figura 3.2). La caracterización basada en la persistencia de los caudales en el tiempo es más complicada para la zona Caribe, caso contrario al estudio de las series según su función de distribución.

3.2.5 Periodicidad de las series de caudal diario

La Figura 3.13 presenta dos series de caudal mensual obtenidas al promediar los caudales diarios en cada mes de registro. La serie *a)* corresponde a la estación Dos Montañas en la vertiente del Caribe, mientras la serie *b)* pertenece a la estación Providencia, en el Pacífico.

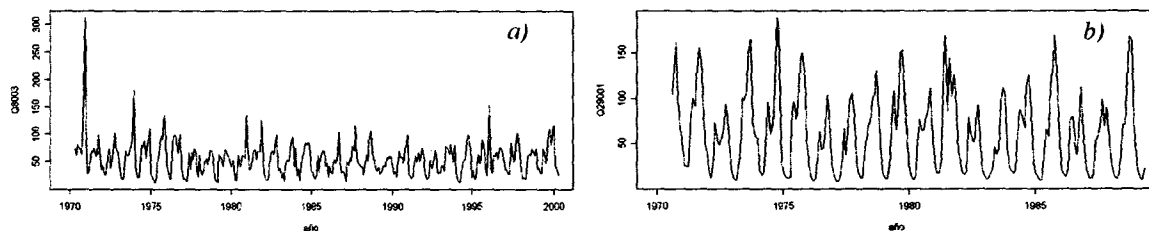


Figura 3.13: Series de caudal promedio mensual en las estaciones *a)* Dos Montañas en la vertiente del Caribe y *b)* Providencia en la vertiente del Pacífico

Parece evidente que ambas series presenta un comportamiento cíclico, que de forma lógica responde al comportamiento estacional del clima. Así, aunque ya no es posible distinguir los picos y recesiones propios de los caudales diarios, sí se reconoce con claridad un aumento en el nivel promedio mensual una vez por año.

En la Figura G.1 del Anexo G se presentan las series mensuales de todas las estaciones analizadas. Al igual que en los ejemplos de la figura anterior, resulta evidente que la periodicidad está aún más marcada en la vertiente Pacífica, mientras en la del Caribe hay un patrón más confuso. No obstante, las funciones de auto-correlación de largo plazo para caudales diario y mensual, mostradas de manera respectiva en la Figura H.1 y en la Figura H.2 del Anexo H, terminan de confirmar la periodicidad anual de las series de caudal (365 días o 12 meses, en cada caso). Estos auto-correlogramas se trazaron para un rezago máximo de cinco años, por lo que es posible apreciar cinco ciclos en cualquiera de las dos escalas de las series.

La Figura H.3 y la Figura H.4 muestran las funciones de densidad espectral de las series diarias y mensuales, respectivamente. La densidad espectral indica cómo se distribuye la variabilidad del proceso completo $X(t)$ (o sea, el caudal promedio diario) con respecto a la frecuencia. Ésta se dibuja en el eje horizontal y en este caso es el recíproco de los ciclos por unidad de tiempo, por lo que corresponde de forma exacta con los días (o meses) de fase. Para las series de caudal en ambas escalas es evidente el efecto de los ciclos de 365 días y 12 meses, con picos notables en el espectro, y en muchos de los casos es notable el ajuste de fases de 183 días y seis meses. Nuevamente el grupo de estaciones 3001 Asunción, 5003 Blanco, 6002 Barbilla, 8002 Siquirres y 8005 Bajo Huacas, todas de la vertiente Caribe, registran un comportamiento de gran complejidad cíclica. Para este grupo la formación de picos múltiples en la densidad espectral, asociados a frecuencias altas (periodos menores a un año) revela una gran cantidad de señales periódicas dentro del ciclo principal, que se pueden detectar en la respuesta del sistema hidrológico de las respectivas sub-cuencas. Una estación representativa de este comportamiento es Asunción, cuya función espectral se presenta en la parte *a*) de la Figura 3.14.

Las demás estaciones del Caribe tienen un comportamiento estacional más claro (un ejemplo es Pacuare, en la parte *b*) de la misma figura), sin llegar a ser tan evidente como el patrón periódico de las estaciones del Pacífico (como en el caso de Londres, mostrada en la parte *c*) de la figura Figura 3.14).

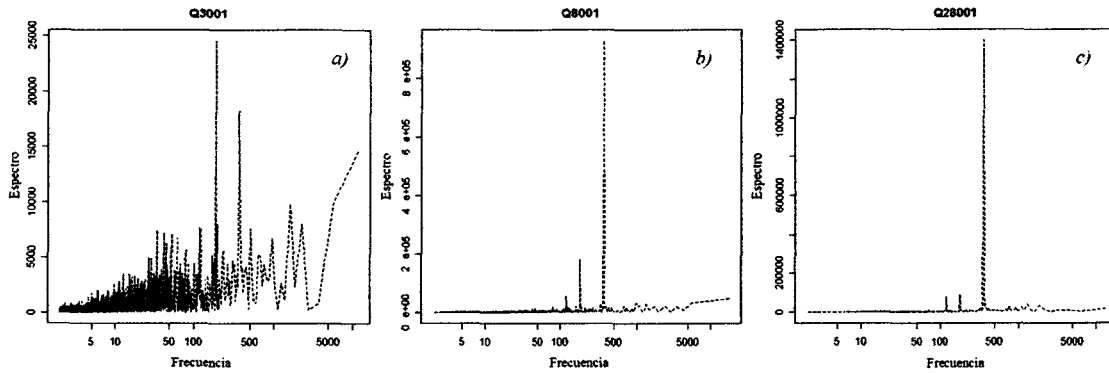


Figura 3.14: Análisis espectral del caudal promedio diario en dos estaciones del Caribe a) 3001 Asunción b) 8001 Pacuare, y en una del Pacífico c) 28001 Londres

Poniendo énfasis en la periodicidad anual claramente establecida en los análisis anteriores, se presenta, en la Figura 3.15, el patrón de escorrentía mensual para las catorce estaciones estudiadas. Las primeras ocho, pertenecientes a la vertiente del Caribe, presentan rangos de variación mensual del caudal entre los meses más seco y más lluvioso de tres veces, como máximo. En el Pacífico puede ser hasta siete veces. Además es evidente que en esta última región las épocas secas y lluviosas son coincidentes en todas las estaciones, con máximos en el mes de octubre, un máximo local en junio y mínimos en abril. Para muchas de las estaciones en el Caribe este comportamiento es más caótico, con picos dobles en junio y octubre (o noviembre) para algunas estaciones, o en julio (agosto) y diciembre para otras.

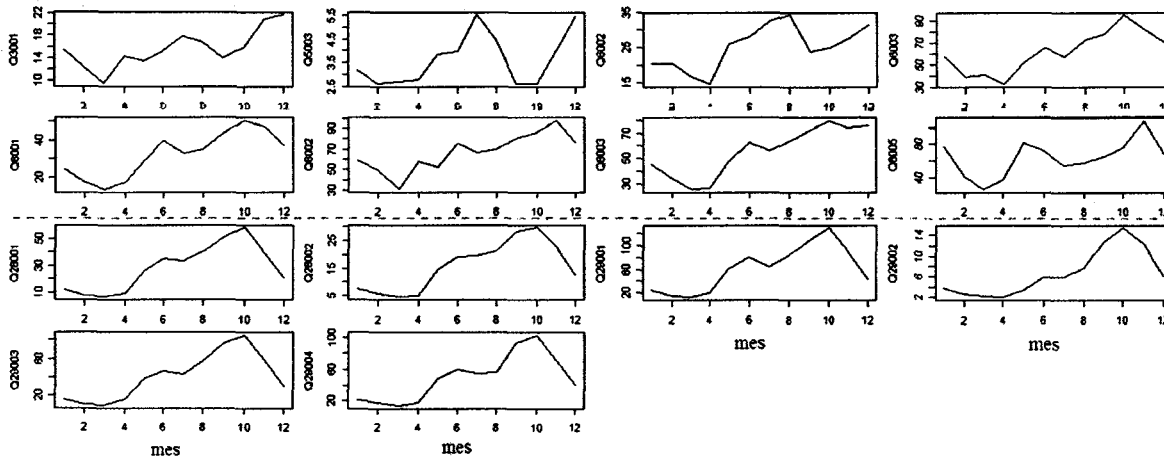


Figura 3.15: Promedios mensuales de los caudales en las estaciones analizadas

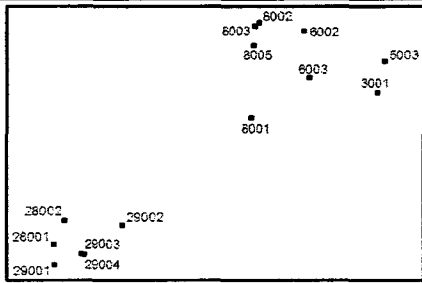
3.2.6 Correlaciones de las series en el espacio

Se puede establecer una correlación de las series de caudal diario en aquellas estaciones que cuentan mediciones hechas en periodos comunes. La Tabla 3.10 expone tales correlaciones, y

en ella se encuentra inserto un esquema básico de la ubicación espacial de las estaciones. La tabla presenta en negrita las correlaciones sobre un umbral (fijado en 0.70) escogido arbitrariamente para resaltar asociaciones fuertes entre los registros de caudal promedio diario. Las correlaciones negativas se muestran en color gris.

Tabla 3.10: Correlaciones entre el caudal promedio diario medido en las estaciones escogidas

	3001	5003	6002	6003	8001	8002	8003	8005	28001	28002	29001	29002	29003	29004
3001	1													
5003	0.67	1												
6002	0.78	0.72	1											
6003	0.73	0.56	0.75	1										
8001	0.53	0.27	0.46	0.69	1									
8002	0.68	-	-	-	0.93	1								
8003	0.55	0.45	0.61	0.77	0.87	-	1							
8005	-	-	-	-	0.89	-	-	1						
28001	0.14	-0.03	0.10	0.33	0.49	-	0.35	0.24	1					
28002	-	-0.08	0.09	-	0.44	-	0.30	0.27	0.85	1				
29001	0.13	-0.03	0.09	0.34	0.51	-	0.35	-	0.90	-	1			
29002	0.17	-0.03	0.08	0.32	0.56	-	0.43	0.42	0.76	0.76	0.79	1		
29003	0.16	-0.04	0.11	0.35	0.62	-	0.44	-	0.93	0.83	0.98	0.84	1	
29004	-	-	0.33	-	0.45	-	0.37	0.25	0.83	0.81	-	0.80	-	1



Los grupos de correlaciones cercanas a la diagonal que están delimitados por polígonos cerrados, corresponden con mediciones hechas en una misma cuenca. En todos los casos, las correlaciones entre estaciones de la misma cuenca resultaron mayores que 0.7. En el caso de las cuencas Pacuare (número 8), Naranjo (número 28) y Savegre (número 29), todas las estaciones analizadas se encuentran en una estructura anidada, lo cual explica las altas correlaciones, de 0.79 en adelante.

Cuando no hay un valor en la tabla es porque las estaciones no realizaron mediciones en un periodo común. Esto es una condición necesaria cuando una estación es el reemplazo de otra, por ejemplo la 8002 Siquirres que fue reemplazada por la 8003 Dos Montañas, y la 29003 Sitio de Presa Savegre se reemplazó con la 29004 Savegre (Estos dos grupos se identifican con una línea más gruesa en el borde de la tabla).

Una vez más se evidencia la poca relación existente entre los regímenes de caudal de la vertiente del Caribe y la del Pacífico, para los que no se detectó ninguna correlación mayor que 0.7. La estación 8001 Pacuare, de la región caribeña, es la que posee mayor similitud con las del Pacífico, llegando a un coeficiente de 0.62 al compararla con la estación 29003 Sitio de Presa Savegre. Por el contrario, la estación caribeña 5003 Blanco presenta correlaciones prácticamente nulas (incluso levemente negativas) con todas las estaciones de la región del Pacífico.

Con esta tabla se refuerza la similitud en las mediciones de caudal efectuadas en la vertiente del Pacífico, lo que denota un régimen estable y común para todas las estaciones. En tanto, en el Caribe parecen definirse dos patrones de variación del caudal, que se manifiestan espacialmente (ver el esquema de ubicación agregado en la tabla) en cuanto a la cercanía y similitud fisiográfica entre las cuencas Banano, Moín y Matina (cuencas 3, 5 y 6), en contraposición de la cuenca del río Pacuare (cuenca 8). En las primeras el régimen de caudales es muy particular, categóricamente distinto al del Caribe, y está medianamente relacionado con el de la cuenca Pacuare (correlaciones entre 0.27 y 0.77). Ésta, en cambio, sí exhibe cierta relación con el régimen del Pacífico, presentando coeficientes entre 0.24 y 0.62.

En el Anexo I se pueden apreciar los gráficos de matriz de correlaciones que ilustran las relaciones arriba establecidas entre los caudales promedio diarios de las estaciones hidrológicas seleccionadas. En la Figura I.1 se evidencia la escasa linealidad entre las estaciones 5003 y 6003, en tanto que la Figura I.2 muestra la enorme asociación lineal entre las estaciones de la cuenca 8 Pacuare, y cierta relación con la estación 6003 Playa Hermosa. La Figura I.3 también hace evidente las fuertes relaciones lineales en los caudales promedio diarios registrados en las estaciones en las cuencas de los ríos Naranjo y Savegre.

3.2.7 Efecto de fenómenos globales climáticos en las series de caudal

Como se ha mencionado, un registro de caudal promedio diario es una medición de la respuesta de la cuenca a un estímulo externo, de tipo climático. Por esto, siguiendo con el

supuesto que el sistema hidrológico tiene una función lineal de almacenamiento del sistema, se esperaría que la respuesta tenga alguna relación con el estímulo, y a la vez, con los factores que están relacionados con él. Se buscó, entonces, establecer una relación entre los caudales promedio mensuales de las diversas estaciones, y fenómenos globales climáticos como el NAO (Oscilación del Atlántico Norte), el NIÑO y el SSN (el número suavizado de manchas solares).

El NAO es uno de los patrones climáticos más antiguos que se conocen, y es el modo dominante de la variabilidad climática de invierno en la región Atlántica Norte. Consiste en una diferencia de presión entre el aire en Islandia y el aire en los Azores (ubicados en la mitad del océano Atlántico, en la misma latitud que Lisboa, Portugal). Usualmente se mide por medio de un índice de teleconexión estandarizado.

En el caso del NIÑO, se trata del calentamiento cíclico de la temperatura del agua del Pacífico Oriental, en la costa oeste de Suramérica. El agua tibia ecuatorial desplaza al agua fría de la corriente Humboldt, interrumpiendo el proceso de ascensión de aguas profundas, de forma que puede alterar el clima en diversas regiones del planeta. El NIÑO se acostumbra medir como temperatura superficial del mar en grados Celsius. Finalmente, el SSN un índice climático espacial, definido como el promedio del número de manchas solares durante los seis meses precedentes al mes en evaluación, ese mes, y los 6 meses posteriores. Se usa como un indicador del nivel global de actividad solar.

Dado que las relaciones de estos índices con las variables climáticas que inciden en los procesos hidrológicos no necesariamente son instantáneas, se analizaron los coeficientes de correlación de Pearson valorando ciertos desfases entre las series climáticas y las de caudal. Entonces en la Figura 3.16 se presentan dichas correlaciones, en forma separada para la vertiente del Caribe y la del Pacífico, de modo que el eje horizontal indica el desfase hecho a la serie climática, mientras el eje vertical presenta la correlación obtenida entre el caudal y el índice de Oscilación del Atlántico Norte.

Por ejemplo, un desfase de -6 implica que la correlación se estableció entre cada par de observaciones del mes m en la serie de caudal promedio mensual y el mes $m-6$ en la variable climática, mientras un desfase de 0 representa la comparación de las series con una correspondencia exacta en los meses de sus respectivas observaciones. Un desfase positivo de k unidades implica que el caudal mensual en m se comparó con el mes $m+k$ de la serie climática. Debe aclararse que cualquier correlación relevante encontrada al aplicar el desfase positivo no quiere decir que el caudal del mes m haya incidido en forma alguna en el índice climático del mes $m+k$. Más bien, debe interpretarse como una posible coincidencia en el patrón de comportamiento estacional, que completa la información proporcionada por los desfases negativos, en los que sí podría establecerse una relación clima-escorrentía.

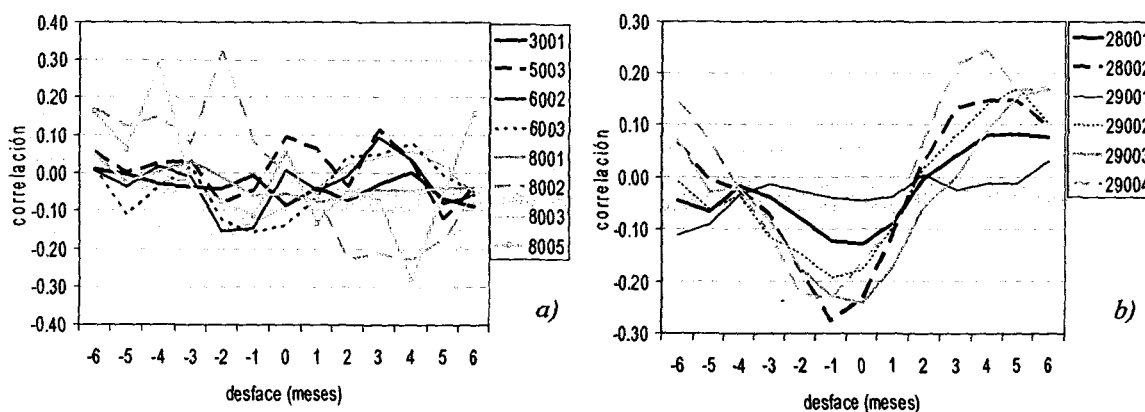


Figura 3.16: Correlación entre el caudal promedio mensual y el NAO para a) las estaciones de la vertiente del Caribe y b) las de la vertiente del Pacífico

En la parte a) de la figura anterior, para las estaciones en el Caribe, sobresalen sólo tres correlaciones cercanas a ± 0.3 . En la estación 8005 Bajo Huacas se produce una correlación positiva de 0.28 para un desfase de -4, y de -0.28 para un desfase 4. Dado que la correlación sin desfase es despreciable, parece establecerse un ciclo común a las series del NAO y del caudal mensual de Bajo Huacas, desfasado en múltiplos de cuatro meses entre los estímulos de la primera y los efectos en la segunda. Un caso similar, pero definido con menos claridad, se da para la estación 8002 Siquirres, pero aquí resulta importante el NAO con dos meses de anticipación al caudal (correlación igual a 0.32).

En la parte b) de la Figura 3.16 se hace la misma asociación para las estaciones del Pacífico, que presentan relaciones bastante más claras con el índice de Oscilación del Atlántico Norte.

Cuatro de las seis estaciones estudiadas aumentan su correlación a valores entre -0.2 y -0.3 para un desfase de -1 mes (28002 Los Llanos, 29002 Providencia, 29003 Sitio de Presa Savegre y 29004 Savegre). El ciclo parece completarse en el desfase positivo 4, de forma que se define un ciclo total de seis meses, que correlaciona con un mes de diferencia y de forma leve al NAO con los caudales en las estaciones mencionadas.

En el caso del fenómeno el NIÑO en las zonas tres y cuatro (NIÑO3.4), la Figura 3.17a) indica una mejor coincidencia en el la respuesta de las sub-cuencas del Caribe con un patrón climático global. Un grupo de estaciones conformado por las estaciones 6003 Playa Hermosa, 8001 Pacuare y 8003 Dos Montañas muestra sus correlaciones más relevantes en rezagos -5 y 2 que sugiere un ciclo común de ocho meses. La estación 8005 Bajo Huacas manifiesta correlaciones entre -0.3 y 0.3 para rezagos de -1, 2 y 5, definiendo un ciclo de cuatro meses.

Con correlaciones máximas cercanas a ± 0.2 se observan 5003 Blanco y 6002 Barbilla, para rezagos negativos de magnitud 1 y 2, y para rezagos positivos de 5 y 6, lo cual insinúa la existencia de ciclos entre 7 y 9 meses. En la estación 8002 Siquirres también se encuentran correlaciones máximas de magnitudes ± 0.2 con rezagos -5 y 3, que definen un ciclo común con el NIÑO de unos nueve meses. Finalmente la estación 3001 Asunción no parece estar influida en absoluto por éste fenómeno climático.

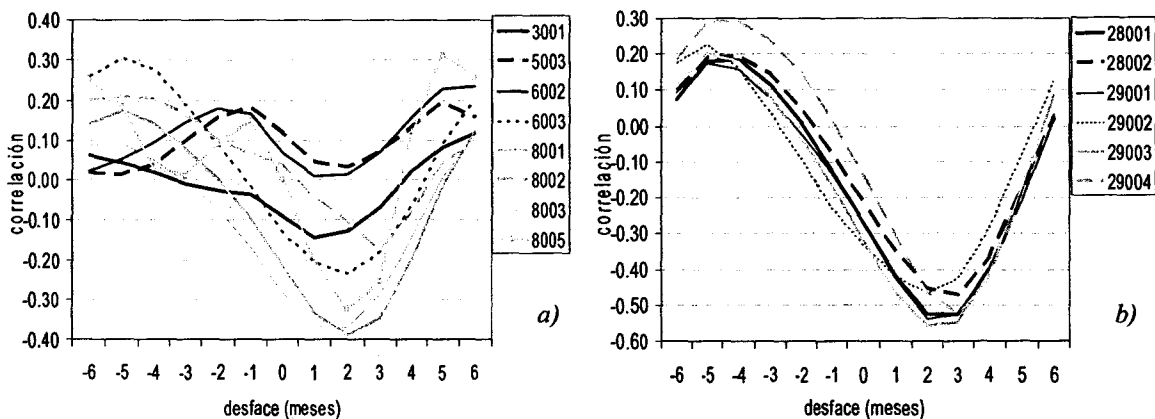


Figura 3.17: Correlación entre el caudal promedio mensual y el NIÑO3.4 para a) las estaciones de la vertiente del Caribe y b) las de la vertiente del Pacífico

La parte b) de la Figura 3.17 sugiere una relación clara entre el comportamiento de la escorrentía media mensual en la vertiente del Pacífico y el fenómeno NIÑO3.4, lo cual es

lógico pues éste fenómeno actúa en el océano correspondiente. En todas las estaciones se observan correlaciones positivas (entre 0.2 y 0.3) para un desfase negativo de 5 meses, y fuertes correlaciones negativas (entre -0.43 y -0.56) para rezagos positivos de 2 y 3 meses. Esto define un ciclo mutuo de unos ocho meses, o sea $2/3$ de un año.

Entonces podría decirse que el comportamiento del caudal que emana de la vertiente pacífica está directamente relacionado con el fenómeno NIÑO en las regiones 3 y 4, medido con cinco meses de anticipación, y que si el patrón con ciclo de ocho meses es permanente, habrá una correlación inversa máxima un año antes (o dos meses después).

En ninguna de las dos vertientes evaluadas se logró establecer un patrón para las relaciones entre el número suavizado de manchas solares (SSN) y el caudal promedio mensual. La Figura 3.18 confirma lo anterior.

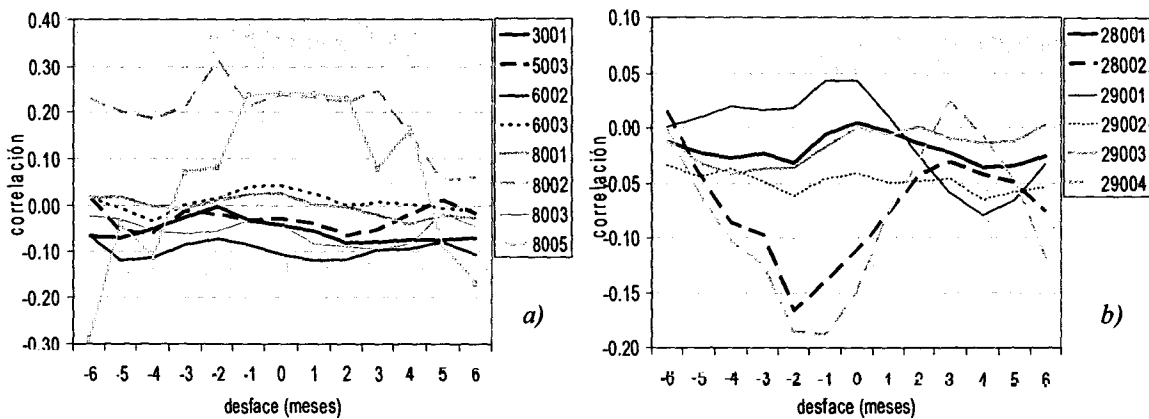


Figura 3.18: Correlación entre el caudal promedio mensual y el SSN para a) las estaciones de la vertiente del Caribe y b) las de la vertiente del Pacífico

Capítulo 4 Resultados y discusión

4.1 Identificación de la lluvia efectiva en las series de caudal promedio diario y estimación de la tasa de eventos ν

Como se discutió en el Capítulo 2, un proceso Poisson filtrado (FPP), o específicamente uno de ruido disparado (SNP), es un modelo que puede aplicarse en hidrología para reproducir el comportamiento de una serie de caudal diario. En la sección 2.3 se describió que la suma de impulsos y recesiones que caracterizan a estas series demandan el conocimiento de la llamada lluvia efectiva.

En una primera aproximación, cualquier aumento por mínimo que sea podría considerarse un evento de lluvia efectiva. Este primer enfoque se aplicó en las series de las catorce estaciones estudiadas, y el resultado se observa en la Figura 4.1. Las líneas de color negro representan a las estaciones de la vertiente del Caribe, mientras las líneas grises punteadas se asocian con las del Pacífico. Con la excepción los meses de marzo y abril, no parece haber una diferencia sistemática en las tasas de ambas vertientes. En estos dos meses las tasas parecen ser sistemáticamente mayores en el Pacífico, lo cual contrasta con los respectivos promedios del caudal (Figura 3.15) que son especialmente bajos en esta vertiente.

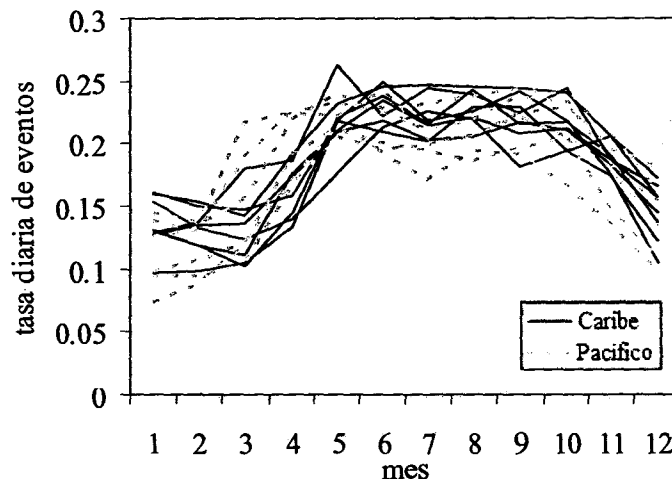


Figura 4.1 Tasa diaria de eventos por mes para las estaciones analizadas

Lo anterior se explica por el hecho de que el registro en estudio está constituido por caudales promedio, que en las condiciones más secas pueden sufrir pequeñas fluctuaciones de un día a otro, en cualquier dirección. Esto da como resultado la detección de eventos espurios que en realidad no están asociados a ninguna señal de lluvia.

El resultado es la estimación de tasas de eventos v_e sobreestimadas. Un ejemplo de esto se da en la Figura 4.2a), en la que se presenta el caudal de la estación 29004 Savegre para un año arbitrario (el 2000). Es evidente que en los meses más secos, marzo y abril, se presenta una gran densidad de líneas verticales grises, que en este caso indican máximos relativos en la serie. La parte b) de la figura muestra con más detalle el comportamiento del caudal en tales meses, de modo que parece que las fluctuaciones efectivamente no corresponden a eventos de lluvia, sino a la variabilidad natural de la media de los caudales de estiaje. Las líneas negras que se superponen a las grises indican la magnitud efectiva de los incrementos, que en estos meses demuestra ser muy pequeña en relación con el nivel en el que se estabiliza la serie (asociado directamente con el caudal base).

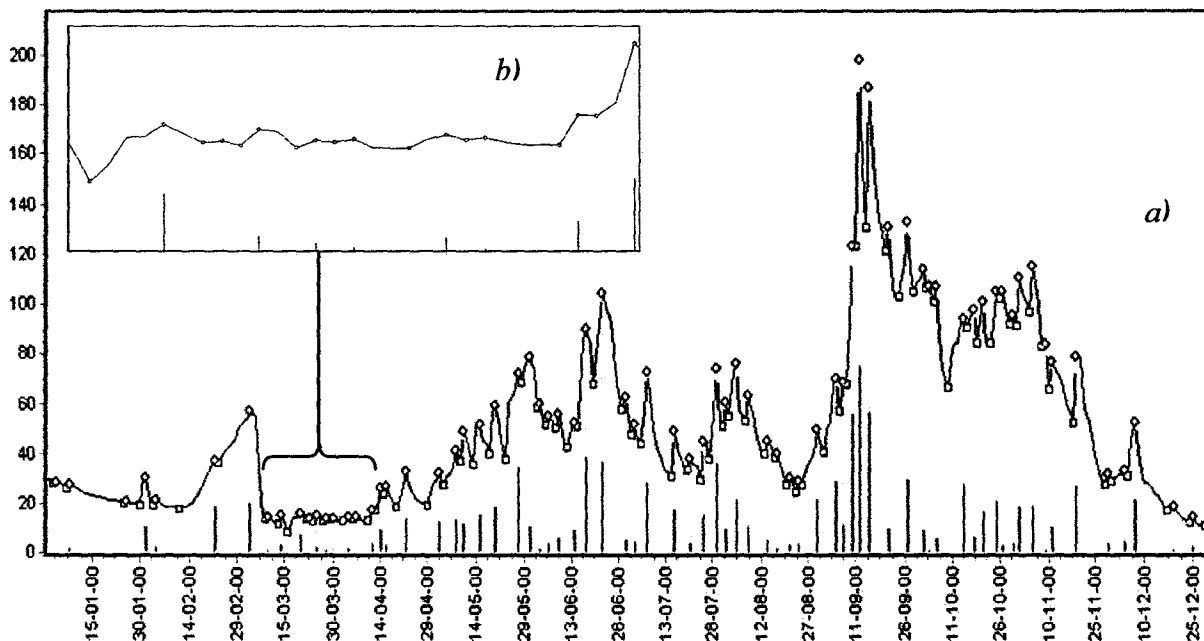


Figura 4.2 A partir de a) la serie de caudal observada en la estación 29004 Savegre en el año 2000, se hace un acercamiento en el recuadro b) que corresponde con el mes de marzo y parte de abril, ambos con bajos promedios de caudal y altas tasas de eventos

Regresando al análisis original, en el que se incorporan todos los incrementos observados al calcular la tasa media de eventos, la Figura 4.1 indica que las tasas más bajas se encontraron en enero, con valores entre 0.07 y 0.16 *eventos/día*, mientras las más grandes corresponden a mayo, y se encuentran entre 0.18 y 0.26 *eventos/día*. En términos mensuales, estos rangos son aproximadamente 2.1 - 4.9 y 5.5 - 7.9 *eventos/mes*, respectivamente.

4.2 El enfoque FPOT: Pruebas de independencia y distribución Poisson de los eventos de lluvia efectiva

Una forma alternativa a la anterior para definir los eventos de lluvia efectiva se expuso en la sección 2.3.1 y se conoce como el método FPOT (Claps y Laio, 2003). Se basa en una prueba de independencia de los eventos y la comprobación de que siguen una distribución Poisson. Ambas consideraciones se evalúan para una secuencia de umbrales de caudal que permiten definir diversos conjuntos de eventos de lluvia efectiva, a los que se aplican las pruebas estadísticas. Como se mencionó en la sección respectiva, la escogencia de un umbral en el que se cumplan esas dos premisas asegurará que la lluvia efectiva tenga una naturaleza coherente con la supuesta en un modelo SNP.

Se estudiaron entonces los efectos del filtrado de los picos debajo de un umbral dado, para las catorce estaciones de interés. La primera característica que se presenta es la cantidad de eventos detectados en función del caudal umbral normalizado (umbral dividido entre el caudal medio de la estación). En la Figura 4.3 se aprecia que ésta es una función decreciente, pues al incrementarse el umbral se van excluyendo los eventos menores. Partiendo de un valor de cero, pequeños incrementos en el caudal umbral producen una gran cantidad de exclusiones de los picos menores, por lo que la tasa de eventos por año decae sustancialmente en un primer tramo. Gradualmente la curva empieza a ser menos abrupta en un tramo de transición, hasta adoptar un comportamiento asintótico en el tramo final. En cualquiera de los casos estudiados la tasa de eventos es inferior de 6 *eventos/año* cuando el umbral supera los 100 m^3/s .

La comparación entre las funciones de tasa de eventos para las estaciones en las vertientes del Pacífico (series de color gris en la Figura 4.3) y del Caribe (series de color negro) permite hacer una distinción categórica entre ellas (especialmente a partir de la observación de la parte

b) de la misma figura). Las curvas de las series pertenecientes al Caribe poseen un decaimiento más gradual, lo cual es lógico dado que sus caudales muestran un comportamiento más irregular. Esto implica la existencia de picos con magnitudes muy variables y, por tanto, con menor susceptibilidad ante el proceso de filtrado como función de la variación del umbral en rangos particulares. Lo anterior se comprueba especialmente en las estaciones 5003 Blanco, 3001 Asunción y 6002 Barbilla, que respectivamente presentan las curvas más atenuadas y al mismo tiempo poseen los mayores coeficientes de variación según la Tabla 3.6

También es muy evidente que presentan las menores auto-correlaciones tanto de corto plazo (Anexo F) como de largo plazo (Anexo H). Estas dos propiedades contrastan en las estaciones del Pacífico, en las que las series son más estables y por tanto, poseen mayores auto-correlaciones y baja variabilidad relativa. El resultado son curvas rápidamente decrecientes como las encontradas para las estaciones 29002 Providencia, 29004 Savegre y 28002 Los Llanos.

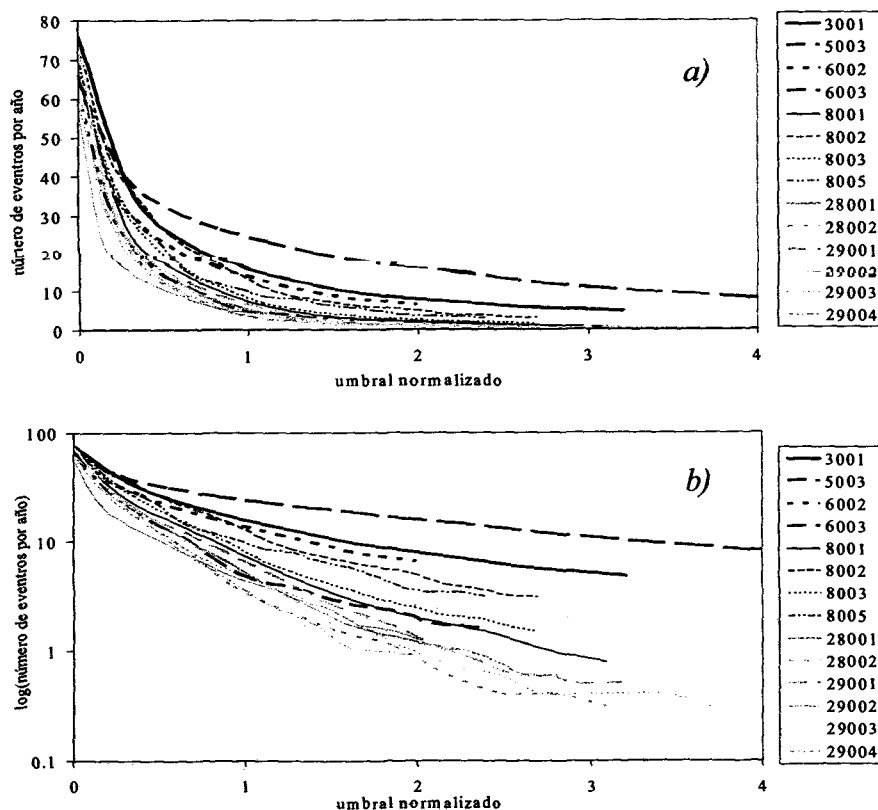


Figura 4.3 Tasa anual de eventos en escalas a) natural y b) logarítmica como función del caudal umbral normalizado, utilizado para filtrar los picos

Resulta interesante la comparación de las estaciones que registran caudales en periodos distintos, pero en el mismo sitio. Como se había mencionado antes, la estación 8002 Siquirres fue reemplazada por la 8003 Dos Montañas, por lo que sus funciones de tasa de eventos deberían ser similares, ante la suposición de que el comportamiento de los eventos es una característica estacionaria en el tiempo. No obstante son notables las diferencias entre estas dos estaciones, que podrían deberse a variaciones del fenómeno que genera la lluvia efectiva en el tiempo.

Otro factor que no es excluyente del anterior es el periodo de registro de la series, ya que la estación 8002 posee un registro relativamente corto (siete años) en comparación con la mayoría de las otras estaciones estudiadas. Lo anterior podría implicar que se cuente con una muestra muy pequeña que no permite inferir características como las tasas de eventos. Las otras estaciones ubicadas en el mismo sitio en periodos sucesivos son la 29003 Sitio de Presa Savegre y la 29004 Savegre. La primera cuenta con observaciones a lo largo de catorce años y la segunda de ocho.

En este caso las funciones de eventos son similares para umbrales menores que la media de la serie, y a partir de ahí divergen. Una vez más se podría desconfiar de la longitud de registro de la 29004, aunque no se puede descartar que las series diarias no sean estacionarias en esta característica. Un último factor puede asociarse directamente con la posibilidad de errores sistemáticos en la medición de los caudales, pues en muchos casos los equipos utilizados para obtener estos registros se cambian por completo al construir una nueva estación. De esta forma, la utilización de equipos de distinta calidad o en un estado de utilidad diferente, o la nueva definición de la sección transversal de medición y su respectivo proceso de aforo, pueden influir en los registros, de modo tal que los estadísticos de la serie y/u otras de sus características se vean sistemáticamente afectadas.

Otra relación se intentó establecer entre la tasa de eventos y la probabilidad de excedencia (PE) del caudal de filtro (o sea, la PE con base en la distribución de los caudales diarios de la respectiva estación). Esto quiere decir que para un umbral s utilizado para filtrar los picos efectivos en una serie hidrológica se busca la probabilidad empírica p_s de observar un caudal

mayor que él (tomando como base todas las mediciones de esa serie) y se asocia con la tasa anual de eventos después del filtrado. Esto se representa en la Figura 4.4. Aquí es más clara la separación en el comportamiento de la tasa de eventos del Caribe (que además exhibe patrones altamente variables).

Las funciones de las estaciones del Pacífico parecen ser más homogéneas entre ellas. Aunque en casi todas las curvas se reconocen tramos aproximadamente lineales, un análisis detallado de cada una evidencia que una asociación lineal sólo se cumple en ciertos tramos. Los altos valores del coeficiente de determinación (magnitudes de R^2 entre 0.80 y 0.99) de regresiones lineales elaboradas para cada curva pueden ser engañosos, dadas las diferencias observadas al comparar gráficamente la regresión con la curva empírica.

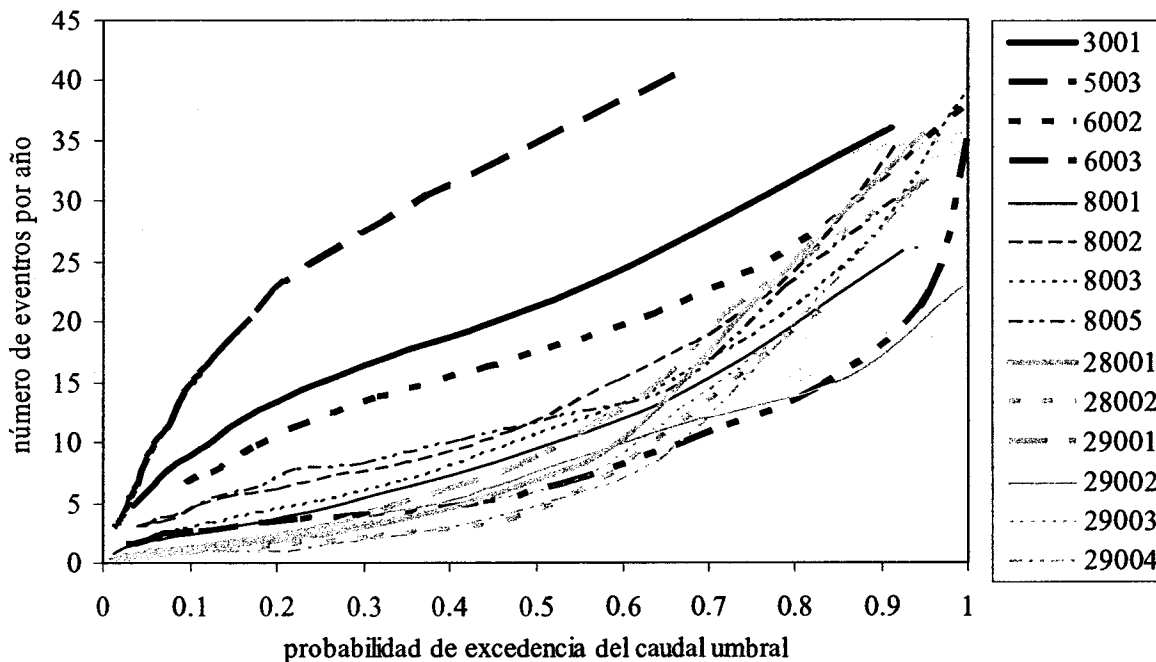


Figura 4.4 Tasa anual de eventos en función de la probabilidad de excedencia (PE) del caudal umbral de filtro

Debe considerarse entonces el comportamiento cóncavo o convexo como un posible indicador de cuán abruptos son los saltos y recesiones en la serie. La fuerte concavidad en curva de la estación 5003 para probabilidades de excedencia bajas indicaría que la tasa puede crecer aceleradamente cuando los filtros aún se encuentran entre el 10 y 20% de los mayores caudales registrados. Por tanto, muchos caudales diarios de gran magnitud son eventos

relativamente cortos y con recesiones agudas y de corta duración. Claramente, esto se produce en las cuencas pequeñas, como son precisamente las que tributan a las estaciones 5003, 3001 y 6002. No es así para la 6003 Playa Hermosa, que aunque está ubicada muy cerca de las anteriores, posee el área tributaria más grande de las catorce estudiadas. Sin embargo este efecto no parece influir en la región del Pacífico, pues las sub-cuencas más pequeñas (28002 y 29002) no reportan grandes concavidades en filtros con bajas probabilidades de excedencia.

Lo anterior se explica porque en el Pacífico la magnitud relativa de los eventos es baja con respecto al nivel medio de la serie, de modo que hay más homogeneidad entre los caudales comunes y los caudales pico. Además las cuencas tienen una topografía que se espera más regular que las cuencas pequeñas del Caribe, pues su relación de elevaciones máxima/promedio es baja (ver la Tabla 3.3).

Se intentó establecer, además, una relación entre las tasas de eventos para diversos umbrales de filtro (por ejemplo los de PE igual que 25, 50 y 75%) y alguna de las características de las sub-cuencas tributarias a las estaciones, debidamente resumidas en la Tabla 3.3 (características fisiográficas), la Tabla 3.4 (se evaluó la variabilidad espacial de la precipitación) y la Tabla 3.6 (estadísticos de la serie de tiempo).

De todas estas características sólo se encontró una relación clara con el coeficiente de variabilidad (C.V.) de la serie diaria. O sea, si se conoce la variabilidad relativa a la media en una estación hidrológica, entonces se puede definir su respectiva función de tasa de eventos con respecto a la PE de diversos umbrales. Con esta información es posible elaborar una curva como las presentadas en la Figura 4.4

La Figura 4.5 presenta las regresiones lineales aplicadas a tres cuartiles de PE, de modo que en cada uno se tiene un grupo de puntos que representa las catorce estaciones en estudio. Estas curvas podrían deducirse para más valores de PE, como por ejemplo los deciles. Al evaluar una nueva estación bastaría evaluar su C.V. en estas curvas, y así se obtendrían los puntos que definen su función de tasa de eventos contra la PE del filtro.

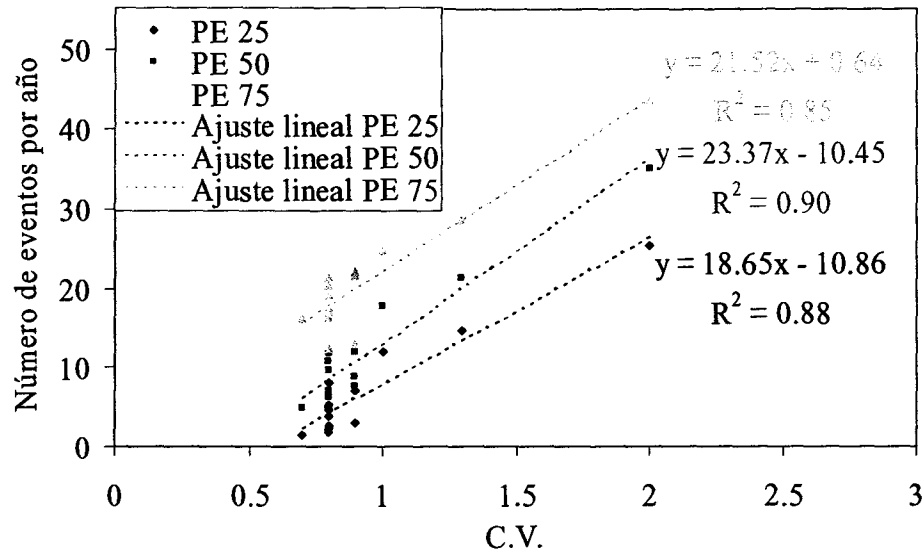


Figura 4.5 Tasa anual de eventos como función del coeficiente de variabilidad (C.V.) de los caudales diarios en una estación hidrológica, para tres probabilidades de excedencia del umbral filtro

Un aspecto adicional que puede ser altamente influyente en las curvas de tasa de eventos, es el fenómeno de los picos espurios (sección 4.1 y Figura 4.2). Podría ser que este fenómeno incida en las series del Pacífico (que tienen estacionalidad ampliamente marcada y largos estiajes), en cuanto al severo decrecimiento de las respectivas curvas en el intervalo de umbrales normalizados entre 0 y 0.5 (Figura 4.3). El efecto de los picos espurios sobre dos diversos niveles de estacionalidad se puede ver en la Figura 4.6. En la parte *a*) se presenta un año de caudal medio diario observado en la estación caribeña 5003 Blanco.

La baja estacionalidad de muchas estaciones en el Caribe se asocia con la dificultad de encontrar lapsos grandes de estiaje en la serie (niveles de agua mínimos y sostenidos en el cauce), que son propicios para la formación de picos espurios. Por tanto, al ir incrementando el umbral de filtrado se obtendrá una variación gradual en el número de eventos remanentes. Lo contrario sucede en las series del Pacífico, ejemplificadas en la parte *b*) de la misma figura (con un año de registro en la estación 29002 Providencia) dado que aquí sí se reconoce un gran periodo de estiaje con muchos picos espurios.

Rápidamente en el proceso FPOT se llega a un umbral que los excluye. Éste podría ser el caso del umbral proporcional común que se presenta como ejemplo en ambas partes de la Figura 4.6 (en este caso se utilizó la media de cada serie completa). La mayoría de las líneas verticales trazadas en cada máximo local, y asociadas con la lluvia efectiva son superadas por este umbral en la estación 29002 Providencia, y por tanto descartadas. En el caso de la estación 5003 Blanco es grande aún el número de eventos excedentes con un umbral igual a la media de la serie completa.

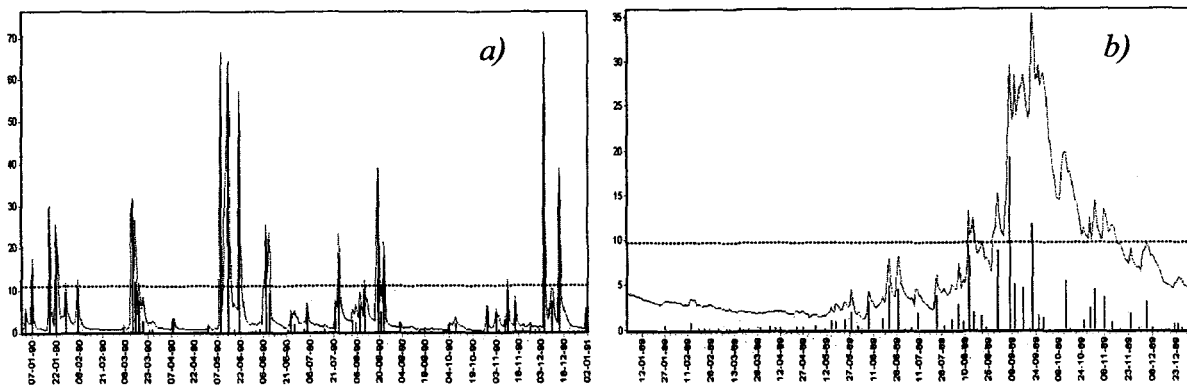


Figura 4.6 La variación del umbral y su influencia en la forma de la función de tasa de eventos en series de las vertientes a) Caribe, estación 5003 Blanco y b) Pacífico, estación 29002 Providencia

Así, la heterogeneidad de los registros con respecto a la media y la presencia de lapsos grandes de estiaje hacen que en la serie del Pacífico solo se observen cuatro eventos sobre este umbral. El valor esperado de ésta cantidad de eventos, basándose en la serie completa, se obtiene encontrando la tasa anual en la ordenada de la Figura 4.3a) asociada con valor igual a uno con en la abcisa (o sea, un umbral igual a una vez la media), de modo que en este caso corresponde se esperaría observar 4.2 *eventos/año*. Por el contrario, la estación del Caribe muestra una cantidad muchas veces mayor de eventos sobre un umbral análogo, de modo que se contaron 32 picos no filtrados por el procedimiento.

El valor esperado basándose en el registro completo de la estación difiere del observado en el año graficado, pues indica 24.4 eventos anuales. No obstante, lo que interesa en esta evaluación es la forma en la que el procedimiento FPOT puede incidir en la cantidad de eventos escogidos en cada paso, dependiendo de que el patrón de los caudales sea Pacífico o Caribe. Una observación adicional es que la etapa de cambio más severa en la definición de la

cantidad de eventos se presenta para cualquiera de las series estudiadas (del Pacífico o del Caribe) en el rango de umbrales comprendidos entre cero y la media de la serie de caudal.

Una vez aclarado el procedimiento FPOT que considera los incrementos sucesivos del caudal umbral, se puede proceder a la aplicación de las pruebas estadísticas estudiadas en la sección 2.3.1. Así, para evaluar la independencia entre los impulsos se implementó la prueba del τ de Kendall, que se aplica a cada nuevo conjunto de picos de lluvia efectiva resultantes después de variar el umbral de filtro. Lo anterior se logra obteniendo el número de discordancias según la ecuación (2.6) y el estadístico de Kendall por medio de la ecuación (2.7).

El intervalo de confianza al 95% se estima basándose en la ecuación (2.8). Como resultado se obtienen las pruebas de independencia como función del umbral (aquí llamado s), que pueden tener distinta forma, según la naturaleza del fenómeno subyacente de lluvia efectiva.

En la Figura 4.7 se aprecian dos ejemplos de la prueba de Kendall como parte de la aplicación del procedimiento FPOT. En la parte *a*) de esa figura (estación 3001 Asunción) se aprecia cómo la secuencia de eventos puede requerir un filtrado severo para que estos sean independientes (esto es que la línea continua se encuentre dentro del rango señalado por las dos líneas punteadas). En ese caso se requeriría un umbral de aproximadamente $32 \text{ m}^3/\text{s}$ (el doble del caudal medio en la estación). Por el contrario, la parte *b*) de la Figura 4.7 (estación 29004 Savegre) sugiere que con sólo $7 \text{ m}^3/\text{s}$ (el 14% del caudal promedio anual) ya la serie de picos de lluvia efectiva es independiente.

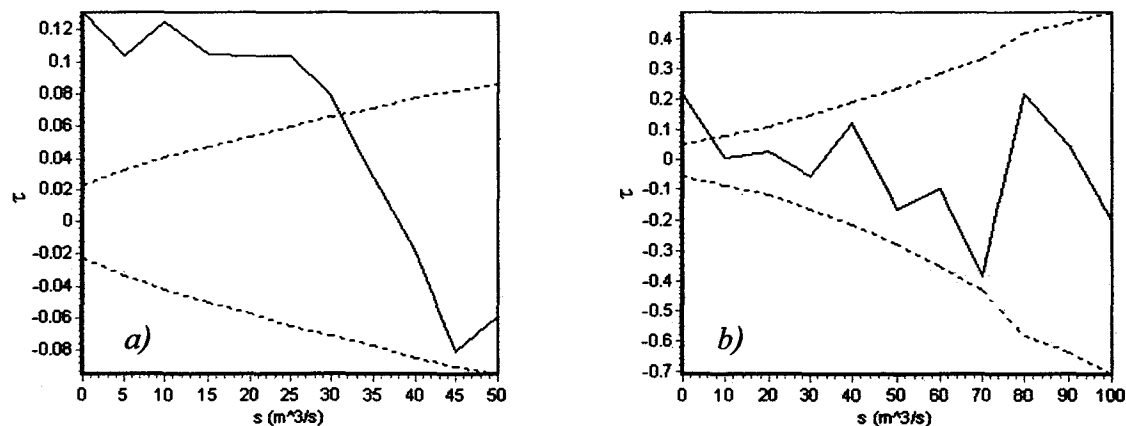


Figura 4.7 Prueba de independencia de Kendall en función del umbral s de filtro de los picos, para las estaciones *a*) 3001 Asunción (Caribe) y *b*) 29004 Savegre (Pacífico). El estadístico τ está representado por la línea continua y los límites de aceptación al 95% de confianza son las líneas punteadas

Las pruebas de independencia para las catorce estaciones estudiadas se encuentran en el Anexo J. Aquí es posible reconocer que existen casos en que la serie de eventos es independiente casi sin necesidad de filtrarla (como el caso de las estaciones 6003 Playa Hermosa y 8002 Siquirres). La estación que requiere un mayor umbral para cumplir los requerimientos de independencia sigue siendo la 3001 Asunción.

Un segundo requerimiento del método FPOT se relaciona con la distribución que poseen los picos de caudal. La prueba de Cunnane es de utilidad para evaluar si los impulsos de lluvia efectiva ocurren según la distribución Poisson, conforme lo que estipula la teoría de los procesos de Poisson Filtrados (FPP) propuesta aquí.

Para lo anterior se empleó el método de cálculo de Lang y otros (1999) para obtener el índice de dispersión I_d (la razón entre la varianza y la media en una escala de tiempo anual) con sus respectivos límites de aceptación para la prueba estadística con confianza del 95%. Se aplicaron las ecuaciones (2.9) y (2.10) en cada conjunto de eventos definidos por el umbral correspondiente, y los resultados se ejemplifican en la Figura 4.8. En la parte *a*) se observa que en su condición original la serie de picos de la estación 3001 Asunción sigue una distribución Poisson (pues la línea continua está incluida en la región de aceptación de la hipótesis nula definida por las líneas punteadas).

Sin embargo, esta condición se pierde al utilizar filtros mayores que $2.3 \text{ m}^3/\text{s}$. El caso contrario se observa en la parte *b*) de esa figura que corresponde a la estación 29004 Savegre, pues más bien el uso de umbrales a partir de los $18.1 \text{ m}^3/\text{s}$ produce picos filtrados con la distribución requerida. En el Anexo J se presentan los gráficos de las pruebas de Cunnane para las catorce estaciones evaluadas en este estudio. Otras estaciones cumplen la distribución Poisson en su condición natural sin filtrar (además de la 3001 Asunción), como es el caso de la 8003 Dos Montañas y la 28002 Los Llanos. Algunas otras no siguen esta distribución sin importar cuál umbral se aplique, dentro del rango de filtros propuesto para cada estación. Esta condición se presenta en las estaciones 6002 Barbilla y 28001 Londres.

Volviendo a la sección 2.3.1, se propone obtener para cada serie un umbral suficientemente alto como para satisfacer las hipótesis de independencia y distribución Poisson conjuntamente. Este criterio es obviamente imposible de aplicar en la serie de la estación 3001 Asunción, pues al contrastar la Figura 4.7 con la Figura 4.8 se verifica el hecho de que no existe un rango de umbrales s común en el que ambas pruebas sean satisfactorias.

Esto quiere decir que la prueba de Kendall se cumple en el intervalo de umbrales $[32,50+)$ mientras que la de Cunnane en $[0,2.3]$. Es evidente que los intervalos son mutuamente excluyentes y que nunca se obtendría un cumplimiento simultáneo de ambas pruebas en el rango de análisis propuesto (umbrales mayores de $50 \text{ m}^3/\text{s}$ se consideran poco razonables, dado que el caudal promedio en la serie diaria es unas tres veces inferior, con un valor de $15.6 \text{ m}^3/\text{s}$).

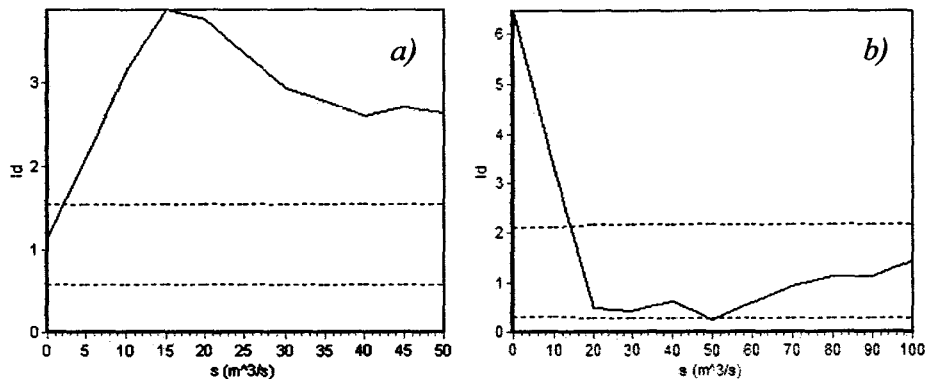


Figura 4.8 Prueba de Cunnane para la distribución Poisson de los picos, en función del umbral s de filtro, para las estaciones *a)* 3001 Asunción (Caribe) y *b)* 29004 Savegre (Pacífico). El estadístico I_d está representado por la línea continua y los límites de aceptación al 95% de confianza son las líneas punteadas

Una aplicación exitosa del método FPOT puede hacerse, más bien, estudiando la otra estación que se ha presentado como ejemplo, o sea, la 29004 Savegre. Aquí las regiones de aceptación de las pruebas son $R_1=[7,100+)$ para la prueba de independencia, y $R_2=[18.1,100+)$ para la prueba Poisson. La región de interés es entonces $R_3=R_1 \cap R_2=[18.1,100+)$, y el umbral buscado es el mínimo que pertenezca a la región R_3 , esto es $\text{Mín}(R_3)=18.1$. Se quiere el mínimo pues aún la intención es alterar lo menos posible el conjunto inicial de picos encontrados en la serie, y eso se logra con el mínimo caudal en la región conjunta R_3 que contiene los valores que cumplen simultáneamente las dos pruebas propuestas. Suponiendo que ambas pruebas son totalmente independientes, este resultado cuenta con una confianza del 95%.

La Tabla 4.1 indica, para cada estación en el análisis, si las pruebas (como función del umbral) se cumplen para algún(os) valor(es) de los filtros evaluados. La tercera columna se refiere a la prueba independencia de Kendall y la cuarta a la Poisson de Cunnane. La quinta columna evalúa, en el caso de que las dos condiciones anteriores se cumplan, si se producen simultáneamente para algún rango de umbrales. En este caso, es posible calcular el umbral mínimo como $Mín(R_3)$, según la definición expuesta arriba. La prueba que definió dicho umbral (ya sea la τ o la I_d) se identifica con un sombreado en la columna correspondiente, para cada estación. La última columna presenta el umbral FPOT normalizado utilizando la media de la serie diaria.

Tabla 4.1 Selección del umbral requerido para la aplicación del método FPOT

Código estación	Nombre de estación	Cumple τ ?	Cumple I_d ?	Cumple simult.?	Umbral escogido	Umbral normalizado
3001	Asunción	√	√	x	-	-
5003	Blanco	√	√	√	8.6	2.4
6002	Barbilla	√	x	-	-	-
6003	Playa Hermosa	√	√	√	53.6	0.9
8001	Pacuare	√	√	√	40*	1.2
8002	Siquirres	√	√	√	38	0.6
8003	Dos Montañas	√	√	√	15.4	0.3
8005	Bajo Huacas	√	√	√	38	0.6
28001	Londres	√	x	-	-	-
28002	Los Llanos	√	√	√	3.6	0.2
29001	Pte. de Hamaca	√	√	√	62.2	1.0
29002	Providencia	√	√	√	4.1	0.6
29003	S. P. Savegre	√	√	√	10.8	0.3
29004	Savegre	√	√	√	14.1	0.3

*Se acepta un I_d que no cumple pero es muy cercano al límite superior, para evitar utilizar el primer umbral que sí cumple estrictamente la prueba, que sería 58.5 m^3/s , pero que se considera muy elevado

**El símbolo √ indica que la prueba se cumple para algunos (o todos) los umbrales evaluados. El símbolo x indica que nunca se cumplen.

La Tabla 4.1 sugiere que los umbrales normalizados obtenidos a partir del método FPOT suelen ser relativamente más altos para las hidrologías del Caribe, en comparación con las del Pacífico. Exceptuando la estación 8003 Dos Montañas, las demás de su misma vertiente ameritaron la aplicación de un umbral relativamente alto, de 0.6 veces la media o más. Por el contrario, en la vertiente Pacífica tres de cinco estaciones cumplieron los requerimientos de independencia y distribución Poisson con el uso de filtros bajos (menores o iguales que 0.3). Una vez más este resultado coincide con el de la Figura 4.3, en cuanto a que las series del

Caribe demandan mayores caudales relativos de filtro para llegar a obtener valores semejantes a las series del Pacífico en diversas propiedades estadísticas. Nuevamente la Figura 4.6 sugiere que la alta frecuencia en la ocurrencia de picos de carácter abrupto en el Caribe influye en esta respuesta.

4.3 Identificación de la magnitud de los impulsos y estimación del salto promedio θ de la distribución exponencial asociada

En adición a los estadísticos expuestos en la sección anterior, se calcularon las pruebas de bondad de ajuste de Cramér-von Mises (W^2) y de Anderson-Darling (A^2) mencionados en la sección 2.3.2 para cada uno de los umbrales de filtrado. Esto con la intención de evaluar si la magnitud de los impulsos de lluvia efectiva siguen la distribución pareto generalizada (GPD), que incluye a la exponencial como caso particular. El supuesto de los modelos de Poisson filtrado es que la distribución de la magnitud de sus picos es exponencial.

Entonces, para cada nuevo conjunto de eventos de lluvia efectiva generado en un paso del procedimiento FPOT, se maximizó la función de log-verosimilitud mostrada en la ecuación (2.11) utilizando análisis numérico. Luego se obtuvieron los estimadores \hat{a} y \hat{k} de los parámetros de la GPD tomando como base las ecuaciones (2.12), para finalizar calculando los estadísticos W^2 y A^2 con las ecuaciones (2.13) y (2.14). Para evaluar la hipótesis nula de que la muestra aleatoria de eventos procede de una distribución GPD se calcularon los puntos asintóticos superiores sugeridos por Choulakian y Stephens (2001) utilizando la Tabla B.1 y la Tabla B.2 del Anexo B.

El resultado obtenido al evaluar las catorce estaciones en estudio con este procedimiento confirma la proposición de Claps y otros (2005), que indica que las pruebas en cuestión no son adecuadas en el proceso de escogencia de un umbral para filtrar los picos de lluvia efectiva. Aún sin hacer una desagregación estacional (ver la sección 2.3.1) se encontró que en muy pocos casos las distribuciones siguen una GPD (según las pruebas de W^2 y A^2). En la Figura 4.9 se presentan dos casos opuestos. Las partes *a*) y *b*) indican que para la estación hidrológica 3001 Asunción (en el Caribe) la hipótesis de una distribución GPD nunca se cumple, dentro

del rango de filtros evaluado (la línea continua nunca está en el rango de aceptación de la prueba, que tiene a la línea punteada como límite superior). Las partes *c)* y *d)* corresponden con la estación 28002 Los Llanos (en el Pacífico), que por lo menos en términos prácticos se aproxima a cumplir la hipótesis de una distribución GPD, para un umbral de cerca de $35 \text{ m}^3/\text{s}$ (poco más de dos veces la media de la serie diaria).

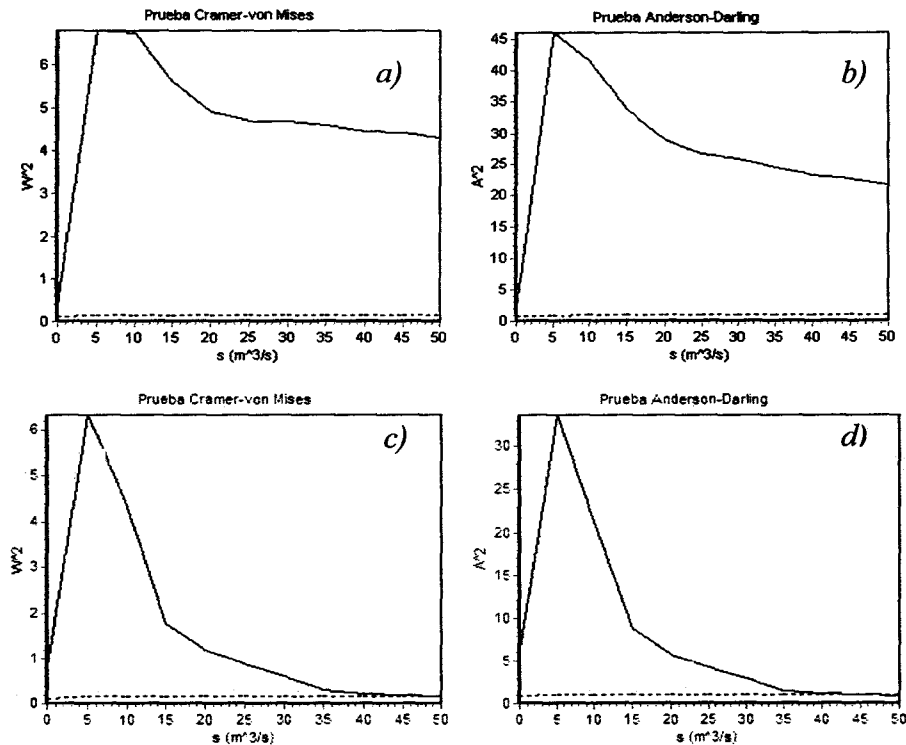


Figura 4.9 Pruebas de distribución GPD de tipo *a)* W^2 de Cramer-von Mises y *b)* A^2 de Anderson-Darling para la estación 3001 Asunción. En ese mismo orden *c)* y *d)* presentan las pruebas respectivas para la estación 28002 Los Llanos. La línea continua indica el valor del estadístico calculado para la serie de eventos filtrados utilizando el umbral s . La línea punteada indica el límite superior de la prueba respectiva al 95% de confianza.

Como se comentó en la sección 2.3.2, la densidad de una GPD cuando el parámetro k es cero es la función de densidad de una distribución exponencial, lo cual se comprueba a través del límite de la GPD cuando k tiende a cero:

$$\lim_{k \rightarrow 0} \left[\frac{1}{a} \cdot \left(1 - \frac{k \cdot y}{a} \right)^{\frac{1-k}{k}} \right] = \frac{1}{a} \cdot e^{-\frac{y}{a}} \quad (4.1)$$

Las pruebas de hipótesis basadas en W^2 y A^2 se cumplieron sólo en tres de las catorce estaciones estudiadas: la 28002 Los Llanos y la 29003 S.P.Savegre; ambas en el Pacífico. En

el caso de que \hat{k} (el estimador del parámetro k) tome un valor relativamente cercano a cero para el mínimo umbral de cumplimiento, entonces el parámetro θ que representa el impulso promedio en la ecuación (2.3) podría aproximarse como $\hat{\theta} \approx \hat{a}$. Esto se aplica en la estación 28002, pues en el umbral de cumplimiento $s=35$ el valor de \hat{k} es relativamente cercano a cero (-6.6×10^{-9}). Así, el salto promedio tomaría el valor $\hat{\theta} \approx \hat{a} = 60.8 \text{ m}^3/\text{s}$. Al igual que en el caso anterior, el cumplimiento de las hipótesis es aproximado para la estación 29003, que requiere un umbral de $100 \text{ m}^3/\text{s}$ (2.5 veces el promedio de la serie) para que sus eventos FPOT sigan una GPD. Como \hat{k} se estimó en -1.2×10^{-8} (valor relativamente cercano a cero) se aproxima el salto medio como $\hat{\theta} \approx 205$.

4.4 Evaluación de las propiedades de cruce en las series de caudal promedio diario

Una sucesión de umbrales s análoga a la utilizada en el procedimiento anterior se puede emplear para evaluar las propiedades de cruce (CP) descritas en la sección 2.9. En este caso, si se traza un nivel umbral como los que se presentaron en la Figura 4.6, cada una de las CP podrían entonces ser directamente observadas en la serie, siguiendo su definición respectiva. Luego se seguiría el mismo procedimiento para un nuevo umbral y así sucesivamente, de modo que se cuente con una función específica para cada propiedad, como una función de s . Tales funciones permiten caracterizar tanto la serie observada en la estación hidrológica como la(s) serie(s) simulada(s) por un modelo FPP.

Las funciones de cruces hacia arriba (UC), cruces hacia abajo (DC), número de corridas en la serie (n_{RN}) y longitud media de una corrida (μ_{RN}), obtenidas para umbrales desde $5 \text{ m}^3/\text{s}$ hasta $50 \text{ m}^3/\text{s}$, se pueden apreciar en la Figura 4.10. Como se esperaría, la función UC mostrada en la parte *a*) de la figura es prácticamente igual a la función DC en la parte *b*), ya que después de contar un cruce hacia arriba se necesitará contar uno hacia abajo para poder continuar agregando otros UC. O sea, siempre hay una compensación. La cantidad observada de corridas está relacionada con los UC y DC, pues cada vez que se produce alguno de estos se genera una nueva corrida. Entonces, teóricamente $n_{RN} = UC + DC + 1$, lo que es coherente con la función graficada en *c*).

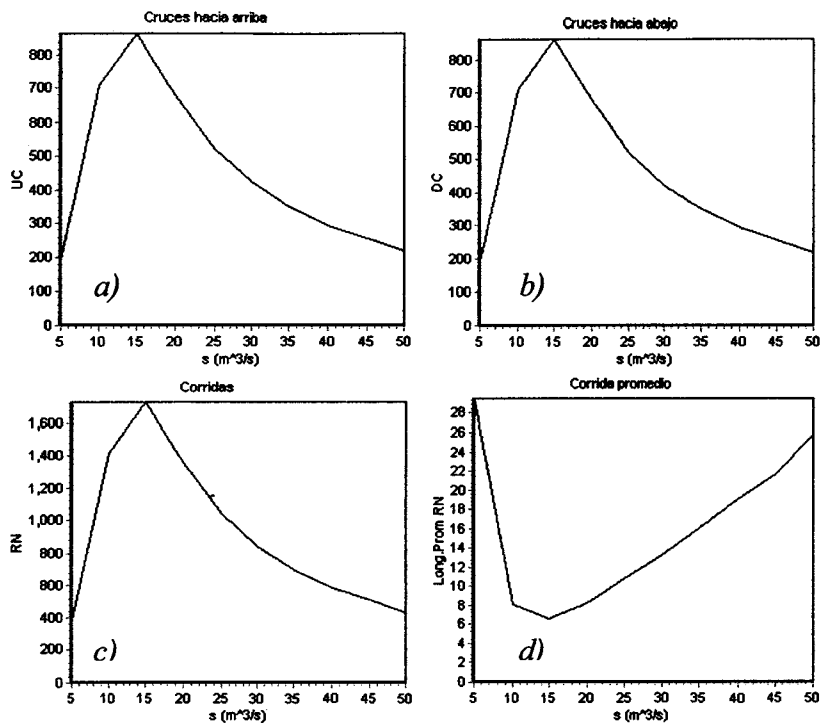


Figura 4.10 Propiedades de cruce para la serie de caudal diario correspondiente con la estación 3001 Asunción, como función de un umbral s : *a*) cruces hacia arriba *b*) cruces hacia abajo *c*) número de corridas *d*) longitud promedio de corridas

La cuarta propiedad de cruce μ_{RN} tiene la función dada en la parte *d*) de la figura anterior. Todas las funciones mostradas presentan un comportamiento lógico, pues para un umbral s igual que cero no existen cruces. Tomando un s muy bajo sólo se atraviesa la serie en los mínimos locales más bajos (o sea, a través de las recesiones más secas). Conforme se utilizan mayores valores de s se “corta” la serie en más puntos, y por eso se las funciones UC, DC y n_{RN} crecen, en tanto que lo lapsos entre cruces se hacen más pequeños y μ_{RN} disminuye. Este proceso tiene un máximo, a partir del cual nuevamente se reducen los cruces, pues sólo se atraviesa la serie por los picos más elevados. Eventualmente las tres primeras funciones tienden a un valor de cero, y la última depende del proceso de cálculo (si se incluyen los extremos para el cálculo de la longitud, la μ_{RN} tiende a la longitud de la serie completa).

En adelante, dado que para las catorce estaciones evaluadas se comprobó la dependencia entre UC, DC y n_{RN} , sólo se presentará la función de UC junto con la de μ_{RN} para caracterizar una serie diaria de caudal. Estas dos funciones de cruce se presentan respectivamente en los gráficos *a*) y *b*) de la Figura 4.11 para todas las estaciones estudiadas, y en función de

probabilidades comunes de excedencia de umbral. Las estaciones en la vertiente del Caribe se representan en color negro y las del Pacífico en gris. Con esto presente se puede observar que tanto la cantidad de cruces hacia arriba por año, como la longitud media de las corridas, son distintas en ambas vertientes. Como es lógico dada la inestabilidad de las series del Caribe, los UC son más aquí para cualquier umbral, en tanto que las longitudes promedio de las corridas son notable y sistemáticamente mayores en las estaciones del Pacífico.

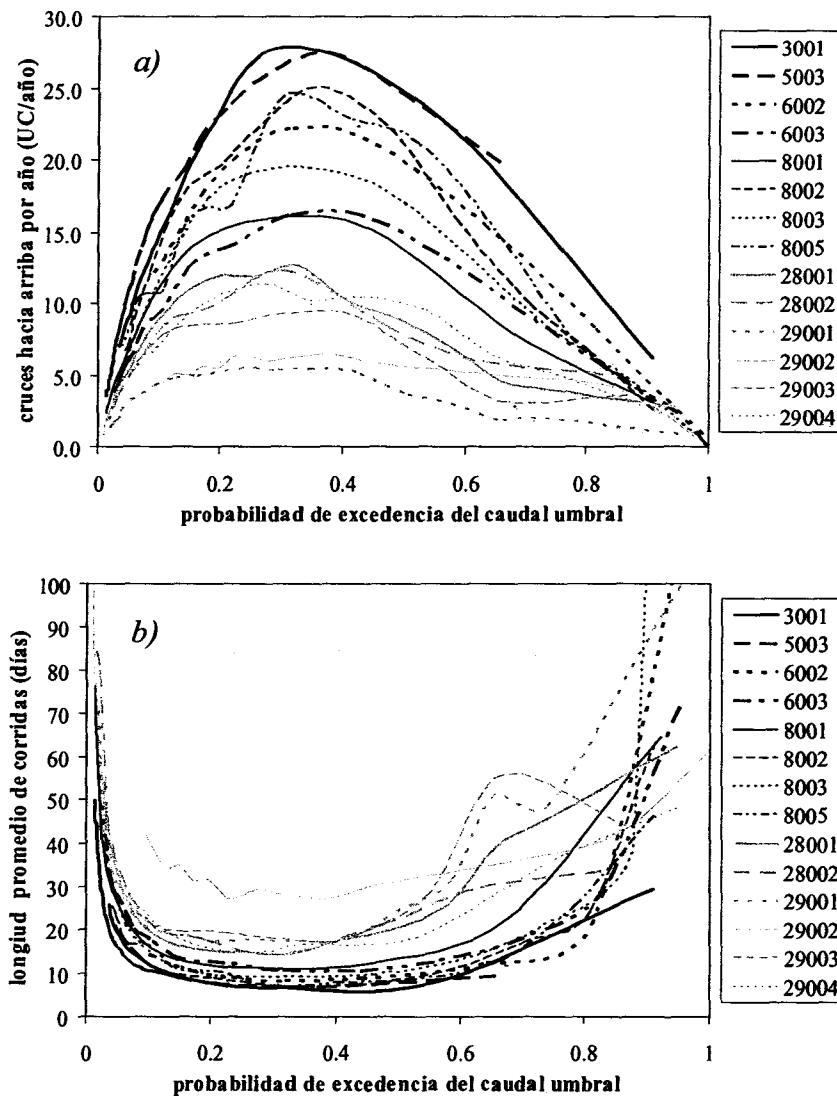


Figura 4.11 a) Cruces hacia arriba y b) longitud promedio de corridas como función de la probabilidad de excedencia del umbral

Los cruces hacia arriba son especialmente altos en algunas series como las de las estaciones 3001 Asunción y 5003 Blanco, que poseen áreas de drenaje, exposición a los vientos marinos

y fisiografías muy similares. Las estaciones con menores funciones de cruce por año son la 29001 Puente de Hamaca y la 29002 Providencia, que representan condiciones muy diferentes de la misma cuenca, pero que aún así tienen una estabilidad similar y relativamente grande.

Si se comparan estos resultados con los obtenidos en el análisis de la Figura 4.4 no se puede establecer una relación directa entre la tasa de eventos y la cantidad de cruces hacia arriba. Lo anterior se sustenta en que, como se expuso antes, las mayores tasas se presentaron en las estaciones 5003 Blanco, 3001 Asunción y 6002 Barbilla y las menores en la 29002 Providencia, 6003 Playa Hermosa (para umbrales bajos, o sea, con probabilidad de excedencia mayor que 70%), 28002 y 29004 (para umbrales altos). Aunque hay algunas coincidencias, en general no se puede hablar de un patrón claramente establecido.

Las funciones de longitud de corridas evidencian el fenómeno estacional que afecta las series de caudal del Pacífico, pues su mayor nivel implica que se registran lapsos mayores con caudales consecutivos por debajo (o arriba) de un umbral relativo cualquiera. Lo anterior refleja una estabilidad en los lapsos con o sin lluvia, que son característicos de estaciones climáticas claramente establecidas. La estación 29002 Providencia tiene longitudes de corrida especialmente altas para umbrales elevados (probabilidades de excedencia menores del 50%), que pueden deberse a la presencia periódica de eventos máximos de magnitud relativamente grande (con respecto a la media) con una distribución muy regular en el tiempo. Así, la curva respectiva sugiere que, en promedio, casi cualquier umbral es cruzado con una frecuencia de entre 30 y 40 días (con excepción de umbrales muy pequeños que tienen probabilidad de excedencia mayor que el 80%). Este alto grado de estabilidad es coherente con las altas auto-correlaciones que se encontraron en los caudales diarios del Pacífico (ver el Anexo F). Las demás estaciones del Pacífico, que también miden una estacionalidad bien definida, tienen patrones de longitud de corridas bastante variables en los umbrales bajos.

Paradójicamente las estaciones del Caribe, con patrones de caudal más complejos e inestables, presentan longitudes de corrida muy similares ante las variaciones del umbral relativo. Para este grupo de estaciones, cuyas cuencas tributarias tienen muy diversas formas de exposición

climática, área y fisiografía; prácticamente podría generalizarse que si se trazan umbrales con excedencia entre el 10 y el 50% se tendrá una longitud de corridas alrededor de los diez días.

4.5 Modelos de Poisson filtrados mensuales de primer orden

Los procesos de Poisson filtrados (FPP) de primer orden se caracterizaron en la sección 2.3. Su proceso de estimación fue descrito en la sección 2.7.1 para el caso específico de los modelos de ruido disparado (SNP).

4.5.1 Estimación de los parámetros mediante el método de preservación de tres momentos

Partiendo de cada una de las series observadas en las catorce estaciones hidrológicas en estudio, se ajustó un grupo de parámetros para cada mes del año. Con esto se espera tomar en cuenta el efecto de variación estacional en cada estación (sección 2.7.3). El ajuste con el método de preservación de momentos requirió el cálculo de los tres principales estadísticos de los respectivos caudales diarios registrados. La forma de obtenerlos es agrupando los caudales diarios en unidades de un mes, y así se forma una serie con unidades mes-año, de estadísticos con variabilidad y auto-correlación en escala diaria.

Luego, todos los valores de esta serie que pertenecen al mes del año $\varepsilon=1, \dots, 12$ se promedian para obtener un estadístico general para cada ε . Éste valor es el que se presenta en el eje horizontal de la Figura 4.12. El término “Total” indica que los estadísticos con variabilidad diaria se obtuvieron promediando todos los valores de la serie mes-año, sin importar su mes del año ε . En este caso se obtienen los estadísticos generales por mes, con una escala de variabilidad diaria.

Regresando a la Figura 4.12, en ella se presentan dos ejemplos del cálculo de momentos estadísticos para una estación en la vertiente caribeña y otra en la vertiente pacífica. Se seleccionaron como ejemplo pues ambas estaciones poseen caudales medios similares. Esto permite observar la clara diferencia en la distribución mensual, que en el caso del estimador del promedio μ identifica las series según la vertiente en que se encuentran (ver el comentario

de la Figura 3.15). El estimador de la desviación estándar σ presenta un comportamiento muy variable en las series del Caribe, mientras que en las series del Pacífico se tienen desviaciones proporcionales a la media muy similares a las del gráfico *b)* de la Figura 4.12.

Los estimadores de las auto-correlaciones con rezago uno $\rho(1)$ no parecen exhibir un patrón estacional, con excepción de los valores mínimos que se presentan en el Caribe en julio y en el Pacífico en setiembre.

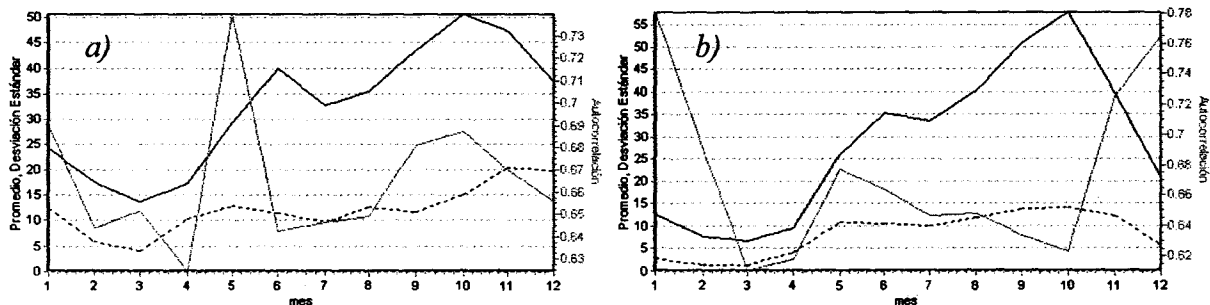


Figura 4.12 Estadísticos de las series diarias de caudal para las estaciones *a)* 8001 Pacuare y *b)* 28001 Londres. La línea continua de color negro representa el promedio, la línea punteada la desviación estándar y la de color gris la primera auto-correlación

Un resumen comparativo de los coeficientes de variabilidad $C.V. = \sigma/\mu$ y de $\rho(1)$ para las catorce estaciones en estudio se presenta respectivamente en los gráficos *a)* y *b)* de la Figura 4.13. En el Pacífico las series tienen un C.V. relativamente bajo (entre 0.2 y 0.5), en tanto que en la mayoría de estaciones del Caribe oscila entre 0.4 y 1 para los primeros cinco meses del año, y entre 0.2 y 0.9 entre junio y diciembre. La estación 5003 Blanco sigue siendo excepcional por su alta variabilidad relativa, pero es lógico pues posee la cuenca tributaria más pequeña de todas las estaciones en estudio.

Los estadísticos de las series observadas en escala diaria, promediados mes a mes: μ_ε , σ_ε y $\rho(1)_\varepsilon$, se utilizan para estimar $E_\varepsilon[X(t)]$, $V_\varepsilon[X(t)]$ y $\rho_\varepsilon[X(t), X(t+s)]$ respectivamente. Siguiendo el procedimiento de la sección 2.7.1, con estos parámetros se puede plantear el sistema de ecuaciones (2.26) con $J=1$ (valor promediado para un día) para obtener \hat{b}_ε , ν_ε y θ_ε , que son la tasa de decaimiento, la tasa de eventos y el salto medio correspondientes con cada mes ε . Esto puede ser resuelto numéricamente como un sistema de ecuaciones no lineales, cuyas

propiedades le dan nombre al método, al cumplirse la preservación de tres momentos estadísticos en la escala diaria. En siglas, este método se identificará como P3M.

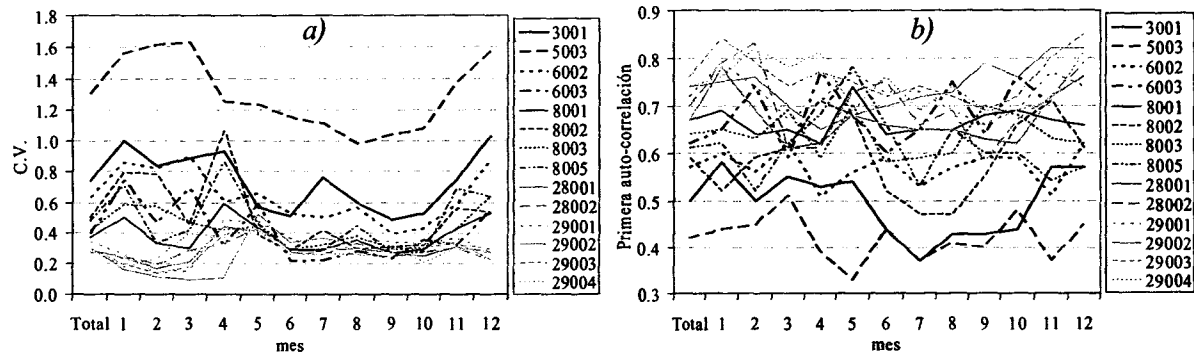


Figura 4.13 Coeficiente de variación y primera auto-correlación de las series analizadas de caudal diario

Así, se aplicó este método para ajustar los parámetros mensuales de las catorce series en estudio. El proceso se ilustra de forma general utilizando la estación 3001 Asunción, pues los resultados de las simulaciones (presentados en la siguiente sección) son muy similares para las demás series históricas registradas, en cuanto a la preservación de sus momentos estacionales. La Tabla 4.2 contiene, para cada mes del año, un grupo de parámetros ajustados para el modelo FPP de tipo SNP.

Tabla 4.2 Parámetros mensuales de un modelo SNP ajustados preservando tres momentos (P3M), a partir del registro histórico en la estación hidrológica 3001 Asunción

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
\hat{b}_ϵ	0.88	1.15	0.98	1.04	1.01	1.40	1.75	1.44	1.44	1.40	0.91	0.91
$\hat{\theta}_\epsilon$	20.2	12.2	10.0	17.0	5.99	6.06	17.2	9.30	5.21	6.88	15.4	30.5
$\hat{\nu}_\epsilon$	0.67	1.16	0.93	0.88	2.28	3.50	1.81	2.60	3.87	3.23	1.24	0.65

Los valores obtenidos para las tasas de eventos $\hat{\nu}_\epsilon$ pueden compararse directamente con las tasas observadas en la Figura 4.1. De aquí se concluye que las tasas diarias de ocurrencia de eventos en cualquiera de las estaciones nunca superaron el valor máximo de 0.26 *eventos/día* observado en la estación 8002 Siquirres para el mes de mayo. Por lo tanto, los valores de $\hat{\nu}_\epsilon$ estimados con el método P3M son excesivos (última fila de la Tabla 4.2), y no se espera que reproduzcan el comportamiento real de los eventos en el momento de realizar la simulación.

En este punto debe recordarse que el supuesto principal del método P3M es que la distribución de los caudales diarios es de tipo gamma (sección 2.3). La Tabla 3.7 indica que ese supuesto no se cumple de manera estricta para las series completas. Otro inconveniente puede ser que el cálculo de los parámetros mensuales se basa en los momentos obtenidos para cada mes del año. Dichos momentos se calcularon a partir de la serie observada, que se supone estrictamente estacionaria para todos ellos. Esta condición podría incumplirse levemente, cuando menos en el caso de la media (ver la sección 3.23). La aplicación del modelo SNP con los parámetros obtenidos y su posterior validación con una simulación de Monte Carlo, son entonces pasos necesarios para juzgar cuánto afecta el incumplimiento de las dos hipótesis anteriores durante la etapa de estimación.

4.5.2 Generación y validación de las series sintéticas de caudal diario para modelos FPP mensuales de primer orden con preservación de tres momentos

Según lo indicado en la sección anterior, e ignorando los valores excesivos que se encontraron para las tasas ν_e , se generó un conjunto de series simuladas de caudal medio diario siguiendo el procedimiento detallado en la sección 2.8.

El modelo de procesos de Poisson filtrados (FPP) de tipo ruido disparado (SNP) se ilustra por medio de las ecuaciones de los procesos continuo y promediado del caudal diario, con paso unitario y para el mes de enero en la estación 3001 Asunción:

$$X(t+1) = e^{-0.88} \cdot X(t) + \sum_{m=N(t)+1}^{N(t+1)} \left[y_m \cdot e^{-0.88(t+1-\tau_m)} \right]$$

$$X_{t+1} = \frac{1 - e^{-0.88}}{0.88} \cdot X(t) + \sum_{m=N(t)+1}^{N(t+1)} \left[y_m \cdot \frac{1 - e^{-0.88(t+1-\tau_m)}}{0.88} \right] \quad (4.1)$$

Tal como se enunció en la sección 2.3, en estas ecuaciones los eventos de lluvia efectiva y_m son independientes e idénticamente distribuidos entre ellos, y también son independientes con respecto al proceso Poisson de $N(t)$ eventos. Su distribución de está dada por:

$$f(y_m) = \frac{1}{20.2} \cdot e^{\frac{-y_m}{20.2}}, \quad y_m > 0 \quad (4.2)$$

Finalmente, los tiempos τ_m en los que se producen los $N(t)$ eventos siguen la distribución de probabilidad del proceso Poisson:

$$P[N(t) = n] = e^{-0.67 \cdot t} \frac{(0.67 \cdot t)^n}{n!} \quad (4.3)$$

La longitud de cada serie sintética se escogió igual a la de la serie observada, lo que implica que para la estación 3001 Asunción se simularon 11415 caudales medios (31 años) en días consecutivos, por cada serie simulada. Así, se generaron cien series sintéticas partiendo de un valor inicial $X(0)$ igual al primer registro de la serie histórica. A cada una de estas series se les calcularon los principales estadísticos mensuales en escala de variación diaria (como aquellos en la Figura 4.12), así como en escala mensual. Para cada mes $\varepsilon=1,2,\dots,12$ se calculó el promedio de cada uno de los estadísticos en escalas de variación diaria ($\mu_{d,\varepsilon}$, $\sigma_{d,\varepsilon}$ y $\rho(1)_{d,\varepsilon}$) y mensual ($\sigma_{m,\varepsilon}$ y $\rho(1)_{m,\varepsilon}$), tomando como base los valores respectivos de cada una de las cien series sintéticas. Lógicamente el promedio en escala diaria para el mes ε reporta los mismos valores que en escala mensual, de modo que de aquí en adelante se ignora este último.

Sobre las series simuladas se aplicó el criterio de cuadrados mínimos para encontrar aquella cuyo conjunto de estadísticos de escala diaria μ_{ε}' , σ_{ε}' y $\rho(1)_{\varepsilon}'$ fuese más similar al promedio de los cien conjuntos de estadísticos antes mencionados. Este criterio se define como:

$$K = \{k | \text{Min}[SSE(k)]\} \quad (4.4)$$

donde K es el índice de la serie con menor suma de errores cuadráticos, y $k=1,2,\dots,N_s$ representa el índice de cada una de las N_s series simuladas (cien, en este caso) La función de suma del error cuadrático es la siguiente:

$$SSE(k) = [\mu_{\varepsilon}'(k) - \mu_{d,\varepsilon}]^2 + [\sigma_{\varepsilon}'(k) - \sigma_{d,\varepsilon}]^2 + [\rho(1)_{\varepsilon}'(k) - \rho(1)_{d,\varepsilon}]^2 \quad (4.5)$$

La aplicación de este criterio indicó que, por ejemplo, la serie sintética que se creó en la simulación número 87 de 100, fue la que registró estadísticos mensuales (con escala de variabilidad diaria) más similares a los estadísticos medios según la simulación de Monte Carlo (MC). A esta serie se le denominará SEMC (similitud de estadísticos mensuales en la serie de Monte Carlo).

La Figura 4.14 permite hacer una comparación de los rasgos de *a)* la serie observada, con *b)* la SEMC (o sea, la serie número 87), para un año de registro arbitrario (1958). Aunque parece haber cierta similitud en la frecuencia y magnitud de los mayores eventos en ambas series (especialmente durante abril y diciembre), resulta evidente la mayor incidencia de picos de cualquier magnitud en el gráfico *b)* asociado con la serie simulada, en la que también se pueden apreciar recesiones muy agudas después de los picos. Esto confirma que la preservación de tres momentos (P3M) expuesta en la sección anterior sobreestimó las tasas de eventos ν_ϵ . Dichas tasas se utilizaron como parámetros para la generación de los tiempos Poisson τ_m en los que se producen eventos, y que se describen por la ecuación (4.3).

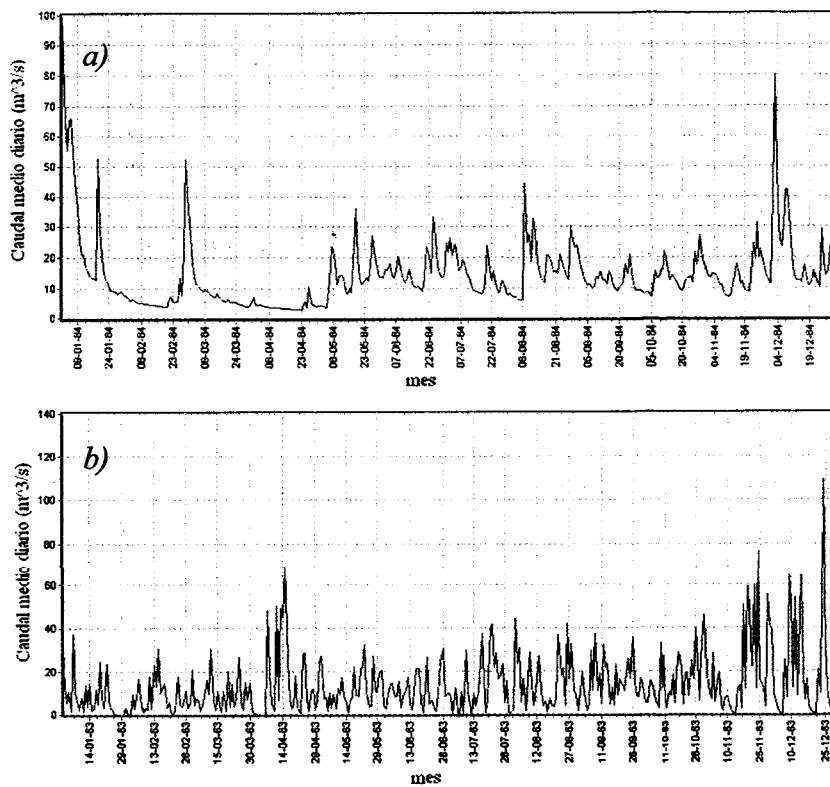


Figura 4.14 Un año de caudales medios diarios en la estación 3001 Asunción para *a)* la serie observada y *b)* la serie SEMC simulada con el método P3M

Aún ante un escenario de incumplimiento de las características hidrológicas de la serie (o sea, del patrón de eventos y la intensidad de las recesiones), es posible evaluar en cuánto el modelo SNP logró preservar los momentos estadísticos. La simulación de N_s series sintéticas y el respectivo cálculo de sus estadísticos en escalas diaria (ρ_ϵ' , σ_ϵ' y $\rho(1)_\epsilon'$) y mensual (σ_ϵ'' y

$\rho(1)_\varepsilon$ "), hacen posible plantear que cada serie es una muestra y que sus N_s conjuntos de estadísticos estimados conforman un conjunto de distribuciones muestrales (DM), cada una asociada a un parámetro.

De la teoría del muestreo se conoce que para cualquier parámetro, si la muestra es suficientemente grande (la serie en este caso) y se han minimizado los errores denominados como "no de muestreo", se tendrá una DM aproximadamente normal (teorema del límite central). De esta manera, es posible conformar las DM de cada estadístico con N_s muestras de un tamaño dependiente de la longitud de la serie histórica, y entonces calcular su promedio y su error estándar por medio de las expresiones:

$$\bar{E}_\varepsilon = \frac{1}{N_s} \sum_{i=1}^{N_s} E_{\varepsilon,i}$$

$$EE(E_\varepsilon) = \frac{1}{(N_s - 1)} \sum_{i=1}^{N_s} (E_{\varepsilon,i} - \bar{E}_\varepsilon)^2$$

donde $E_{\varepsilon,i}$ es el estadístico en estudio para el mes ε y la serie simulada i . Un intervalo con 95% de confianza $[L_I, L_S]$ para cada estadístico se puede obtener con la ecuación $L_{S,I} = \bar{E}_\varepsilon \pm 1.96 \cdot EE(E_\varepsilon)$.

En el caso de la estación 3001 Asunción, utilizando el modelo SNP con estimación paramétrica tipo P3M, se obtuvieron los estadísticos en escalas diaria y mensual que se muestran en la Figura 4.15.

Este resultado indica claramente que el modelo logró reproducir el promedio, en tanto que sobreestimó levemente la desviación estándar y subestimó notablemente la primera auto-correlación de la serie diaria (gráficos *a*), *b*) y *c*) respectivamente en la Figura 4.15). En el caso de la auto-correlación, es razonable pensar que una simulación con excesiva cantidad de eventos (se sobreestimó la tasa $\hat{\nu}_\varepsilon$) y recesiones extraordinariamente agudas (gráficamente se dedujo que se sobreestimaron las tasas de decaimiento b_ε), redundarán en niveles mucho menores de auto-correlaciones diarias.

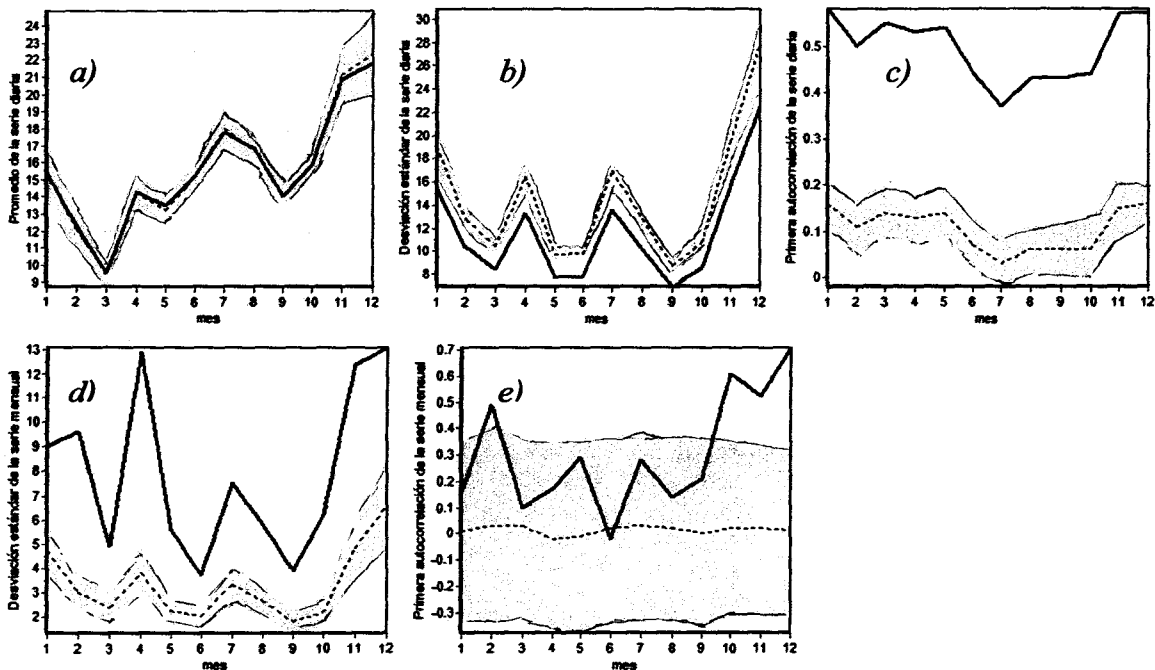


Figura 4.15 Estadísticos mensuales de la serie observada (línea gruesa continua) y de las 100 series MC (línea delgada punteada) tipo P3M, con su respectivo intervalo del 95% de confianza (región gris). En escala de variación diaria se tienen *a)* el promedio, *b)* la desviación estándar y *c)* la primera auto-correlación, y en escala mensual *d)* la desviación estándar y *e)* la primera auto-correlación

Los estadísticos de la serie mensual no se especificaron explícitamente en este modelo de primer orden, y por lo tanto no debe asumirse que el modelo los preservará de manera directa. En este caso la desviación estándar fue fuertemente subestimada, en tanto que la primera auto-correlación simulada se mantiene entre los amplios rangos de variación probable, en ocho de los doce meses del año.

Una vez más se recurre a la serie SEMC, ahora para evaluar sus propiedades de cruce (CP), y compararlas con las de la serie histórica de la estación 3001 Asunción. Aunque no existe una relación directa y claramente establecida entre la tasa de eventos y propiedades como los cruces de un umbral hacia arriba (UC) y la longitud promedio de las corridas (μ_{RN}), es lógico pensar que a mayores tasas de eventos y de decaimiento, se producirán más cruces de ciertos umbrales (especialmente los intermedios, en términos de su probabilidad de excedencia). También se producirían longitudes de corridas menores. El resultado de la estimación de las CP para la SEMC confirmó el principio anterior, pues se alcanzaron hasta 2400 UC y una μ_{RN} mínima de 2.4 días para un umbral cercano a $15 \text{ m}^3/\text{s}$. Examinando la Figura 4.10 se puede

establecer que en la serie histórica se presentan como máximo, unos 860 UC y una μ_{RN} mínima de 6.6 días, también para los umbrales cercanos a los $15 m^3/s$.

Las propiedades del modelo estimado para la estación 3001 Asunción, en cuanto a su capacidad de preservar los estadísticos explícitos (los de la serie diaria) e implícitos (los de la serie mensual y las propiedades de cruce), se reproduce en las demás estaciones hidrológicas tanto del Caribe como del Pacífico. Incluso en esta última vertiente se encontraron tasas de eventos aún mayores y, por tanto, condiciones de auto-correlación y cruce más alejadas de las observadas en las series históricas. Por todo lo anterior, se considera que el ajuste de los parámetros del modelo SNP es inadecuado cuando se basa en el método de P3M, ya que este objetivo podría ser muy ambicioso si se trabaja en las condiciones de incumplimiento parcial de los supuestos ya comentadas en la sección 4.5.1. El uso de este método llevó a parámetros poco realistas (como aquellos mostrados en la Tabla 4.2.) en la estimación de los catorce modelos SNP elaborados en el presente trabajo.

4.5.3 Estimación de los parámetros mediante el método de preservación de momentos y considerando la tasa de ocurrencia de eventos observada v

Se observó que la estimación de ν_ϵ (la tasa de eventos de lluvia efectiva) es poco realista cuando se utiliza el criterio de preservación de los tres momentos estadísticos antes mencionados (también citado como P3M). Ante esto, se propuso un nuevo enfoque en el que dichas tasas se obtienen directamente de la serie histórica para cada mes ϵ , y luego se utilizan dos de las tres ecuaciones (2.26) para obtener los parámetros restantes \hat{b}_ϵ y θ_ϵ . Cada una de esas ecuaciones preserva un momento específico, de modo que existen tres posibles combinaciones, cada una de las cuales permitiría preservar dos de los tres momentos diarios.

Tales combinaciones se exponen en las siguientes sub-secciones. Antes, se aclara que en esta sección siempre se calcularon las tasas de eventos considerando que cualquier aumento del caudal, seguido de inmediato por una disminución, constituye un pico o evento. Ya este

cálculo se había hecho para cada una de las estaciones en la sección 4.1, y los resultados se presentaron en la Figura 4.1.

Preservando el promedio y la desviación estándar de las series diarias

Es éste, entre todos los métodos estudiados en el presente trabajo, el que produjo los mejores resultados. Como pretende conservar la tasa de eventos de la serie, y su promedio y desviación estándar, se le nombrará TPD. A partir de un ν conocido se resuelven simultáneamente las primeras dos de las ecuaciones (2.26), para $E[X(t)]$ y $V[X(t)]$ respectivamente. La Tabla 4.3 muestra los parámetros mensuales ajustados para la estación 3001 Asunción.

Tabla 4.3 Parámetros mensuales de un modelo SNP ajustados con el método TPD, a partir del registro histórico en la estación hidrológica 3001 Asunción

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
\hat{b}_ε	0.17	0.11	0.12	0.18	0.08	0.07	0.15	0.09	0.06	0.07	0.12	0.19
$\hat{\theta}_\varepsilon$	16.3	8.95	7.72	13.1	4.51	4.09	10.9	6.26	3.47	4.65	12.1	24.5
$\hat{\nu}_\varepsilon$	0.16	0.15	0.14	0.19	0.23	0.25	0.25	0.25	0.25	0.24	0.21	0.17

Comparando este modelo con el anterior se tiene claro que las tasas de eventos son muy inferiores en el TPD, ya que corresponden con las tasas ν_ε observadas en la serie histórica. La Figura 4.18 permite ver con mayor claridad el cambio en los tres parámetros al fijar ν_ε en el procedimiento TPD. En cada uno de los gráficos de la figura se tiene en el eje horizontal el parámetro del modelo inicial (preservación de tres momentos) y en el vertical el del TPD. La línea inclinada indica la dimensión lineal en la que el parámetro evaluado tiene el mismo valor para ambos modelos. Así, valores por debajo de la línea implican que los respectivos meses presentaron valores mayores en el modelo P3M que en TPD, y viceversa.

Entonces, la Figura 4.16a) evidencia la reducción sustancial de las tasas de eventos para el modelo TPD, que se produjo en todos los meses y en una forma no proporcional a sus valores inicialmente ajustados con el P3M. El nuevo método hace que el parámetro de salto promedio ajustado $\hat{\theta}_\varepsilon$ sea sistemáticamente inferior, presentando una reducción aproximadamente proporcional con respecto a sus valores de P3M.

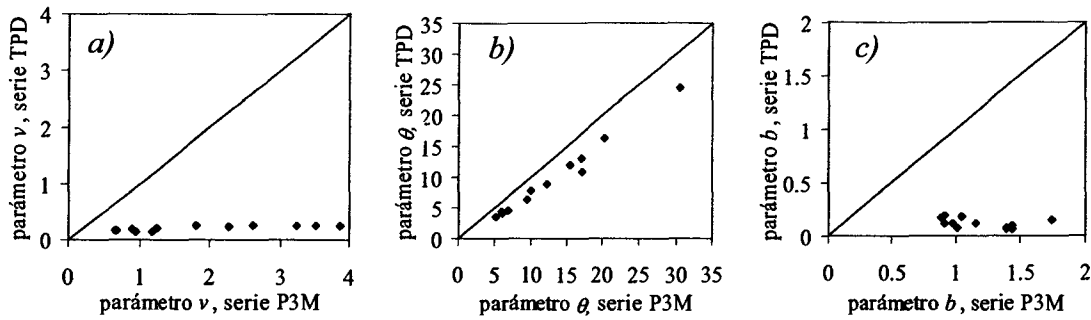


Figura 4.16 Comparación de los parámetros mensuales estimados para un modelo SNP de primer orden, entre el método P3M y el método TPD. En *a)* se comparan las tasas de eventos, en *b)* los saltos promedio y en *c)* las tasas de decaimiento

Finalmente, se evidencia que el cambio sustancial en las tasas de ocurrencia ν_e ahora pre-fijadas, es equilibrado con una disminución notable en las respectivas tasas de decaimiento estimadas \hat{b}_e (ver la parte *c)* de la Figura 4.16). La lógica detrás de esto es que si se producen menos eventos se requerirá que su efecto sea más duradero, para así mantener un nivel en la serie que preserve el promedio correspondiente. Lo anterior es controlado, a su vez, por la preservación de la desviación estándar, característica que impediría una recesión excesivamente gradual.

El modelo TPD también se utilizó para generar una simulación Monte Carlo (MC) de cien series sintéticas. Nuevamente se buscó la serie SEMC tomando como base las ecuaciones (4.4) y (4.5). Anteriormente se presentaron para el año de 1963 los caudales observados y simulados con P3M en la estación 3001 Asunción, (ver la Figura 4.14). Ahora se exponen en la Figura 4.17 los valores simulados con TPD. La comparación de las dos series simuladas no deja duda de las diferencias entre ellas. Con TPD se producen menos eventos con recesiones mucho más graduales. En general, se puede decir que aunque el patrón no es idéntico al de la serie histórica, tiene un mayor parecido en ciertas recesiones y la frecuencia de los picos.

Con éste antecedente, que hace suponer un comportamiento auto-regresivo más razonable para el método de estimación TPD, se calcularon los estadísticos y bandas de confianza de las cien series sintéticas, esperando un cambio radical en el comportamiento de las auto-correlaciones. Esto se comprueba al examinar la Figura 4.18*c)*, pues deja de presentarse un error sistemático en la primera auto-correlación de la escala diaria

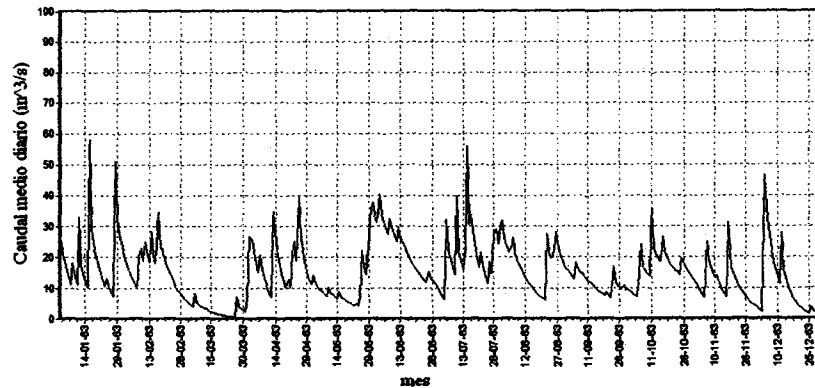


Figura 4.17 Un año de caudales medios diarios en la estación 3001 Asunción para la serie SEMC simulada con el método TPD

Además, para la serie mensual y en casi todos los meses del año, la Figura 4.18e) señala que la región de confianza MC de las auto-correlaciones de orden uno (franja gris) incluye a los estadísticos de la serie histórica (línea gruesa continua). No está de más decir que se sigue preservando adecuadamente el promedio en el gráfico *a)*, que la desviación en escala diaria en *b)* es ahora levemente subestimada por el MC-TPD, y finalmente que la desviación de la serie mensual en *d)* también perdió su error sistemático, incluyendo en la región de confianza a los estadísticos de la mayoría de los meses del año.

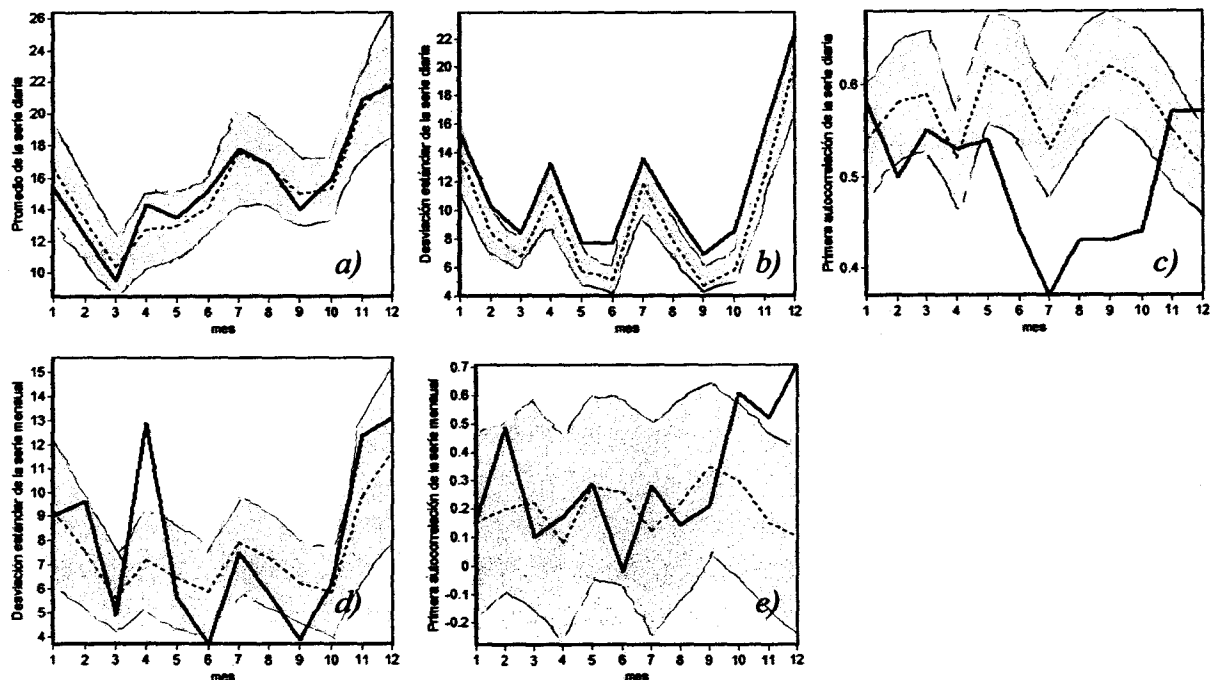


Figura 4.18 Estadísticos mensuales de la serie observada (línea gruesa continua) y de las 100 series MC (línea delgada punteada) tipo TPD, con su respectivo intervalo del 95% de confianza (región gris). En escala de variación diaria se tienen *a)* el promedio, *b)* la desviación estándar y *c)* la primera auto-correlación, y en escala mensual *d)* la desviación estándar y *e)* la primera auto-correlación

Es importante recordar que la primera auto-correlación de la serie diaria no se definió explícitamente como un estadístico a preservar, como tampoco se hizo con las propiedades de la serie mensual. Su aproximación aquí se debe más bien a las bondades del modelo para reproducir un patrón de caudal diario cercano a la realidad.

Las principales propiedades de cruce de la serie SEMC para la técnica TPD también fueron valoradas y se presentan en la Figura 4.19. En la misma figura se suministran las funciones observadas respectivas. Se puede concluir que, a diferencia del modelo P3M, el modelo TPD preserva razonablemente bien los cruces hacia arriba (UC) y las longitudes promedio de corrida (μ_{RN}). El máximo número de UC en esta serie sintética es 660 para un umbral de $15 \text{ m}^3/\text{s}$, contra 860 UC de la serie histórica en ese mismo umbral. En tanto, el mínimo μ_{RN} es 8.6 días en la serie SEMC contra 6.6 de de la serie histórica, ambos para $15 \text{ m}^3/\text{s}$.

Éste método de estimación, como el que proporcionó los resultados más precisos, fue aplicado para obtener los parámetros de los FPP de tipo SNP en las catorce estaciones analizadas en este trabajo. La aplicación en forma de una simulación MC de cien series sintéticas se reiteró para cada estación. Las comparaciones de los estadísticos y las bandas de confianza así obtenidos, con los valores correspondientes observados en las series históricas, se pueden consultar en el Anexo K.

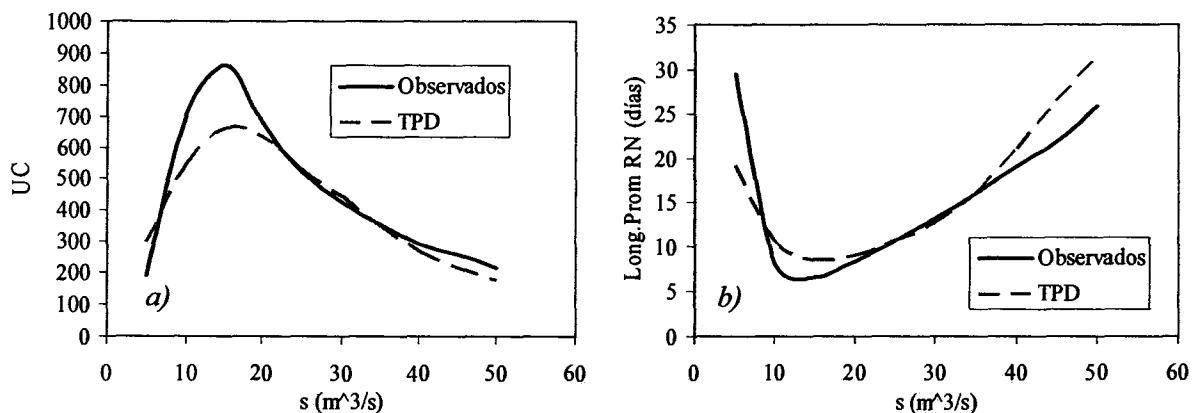


Figura 4.19 Propiedades de cruce como función de un umbral s para la serie sintética de caudal diario SEMC estimada por medio del método TPD, y para la serie histórica. Estación 3001 Asunción. *a)* cruces hacia arriba y *b)* longitud promedio de corridas

Preservando la media y la primera autocorrelación de las series diarias, o bien su desviación estándar y su primera autocorrelación

El hecho de que las primeras auto-correlaciones de las series diarias no fueron preservadas adecuadamente por el modelo P3M, se reflejó otra vez en los resultados obtenidos cuando ese estadístico se quiso preservar junto con el promedio (método aquí llamado TPA), o en conjunto con la desviación estándar (método definido como TDA). En estos dos casos nuevamente se fijan los valores conocidos de ν_ϵ . Entonces, para el modelo TPA se resuelven simultáneamente la primera y la tercera de las ecuaciones (2.26) para $E[X(t)]$ y $\rho[X(t), X(t+s)]$ respectivamente. Para el modelo TDA se resuelven la segunda y tercera ecuaciones (2.26), que pretenden preservar los momentos $V[X(t)]$ y $\rho[X(t), X(t+s)]$.

Como ejemplo se analiza la estación 3001 Asunción. Los estadísticos resultantes de ambos modelos (TPA y TDA) pueden apreciarse en el 0. Es evidente que con el modelo TPA se preservó la media, pero no la desviación estándar ni la primera auto-correlación de las series diarias. Lo anterior pese a que la auto-correlación debía preservarse, según la definición explícita en la etapa de estimación de parámetros. En cambio en el modelo TDA fue la desviación estándar la que se respetó, pero no la desviación ni la auto-correlación. En las demás estaciones se obtuvieron resultados similares a estos, por lo que se concluyó que al utilizar el método de preservación de momentos no es adecuado intentar reproducir la primera auto-correlación de la serie diaria. Así, si se desea mantener la tasa de eventos, es mejor emplear un procedimiento como el TPD expuesto en la sub-sección anterior.

4.5.4 Estimación mediante el método de preservación de momentos considerando la tasa de ocurrencia de eventos ν corregida con el método FPOT

En algunos casos, cuando la modelación basada en el método TPD no es satisfactoria (en cuanto a que no preserva razonablemente bien al menos tres de los cinco momentos evaluados) se recurrió a la aplicación del método FPOT presentado en la sección 2.3.1. Con él se definieron los umbrales de filtrado de eventos dados en la Tabla 4.1. Así, para las nuevas series de picos FPOT se calcularon nuevas tasas de eventos ν_ϵ y con ellas se aplicó el procedimiento TPD. Con este procedimiento se intentó eliminar las distorsiones introducidas por eventos espurios, así como garantizar que se respetan las hipótesis de independencia y

distribución Poisson que los picos de lluvia efectiva deberían tener, para así modelarlos con un FPP.

Según el simple criterio de aceptación recién definido, en este trabajo se consideró que las estaciones 8001 Pacuare, 8003 Dos Montañas y todas las del Pacífico no se pudieron modelar adecuadamente con el método TPD aplicado sobre picos sin filtrar. Entonces, estas fueron las estaciones candidatas a la aplicación del método FPOT, para intentar mejorar los resultados.

Si se examina la última columna de la Tabla 4.1 se verá que la magnitud relativa del umbral de filtro, con respecto a la media de la serie, varía entre 0.3 y 2.4 para las estaciones en que el método resultó aplicable. No obstante, en forma general al trabajar con umbrales mayores que la mitad de la media (o sea umbral normalizado mayor que 0.5) las tasas de eventos obtenidas se redujeron excesivamente, registrando incluso un valor de cero para varios meses. Así, un nuevo criterio se estableció para no exceder el filtrado de los picos, limitando el uso del método FPOT a aquellas series en que el umbral normalizado no fuese mayor que 0.5. De ese modo, se presentan dos ejemplos (la estación 8003 en la vertiente del Caribe y la 28002 en el Pacífico) en los que se combinaron los procedimientos FPOT y TPD.

Nuevamente, a partir de las series simuladas MC se elaboraron intervalos de confianza para los principales estadísticos de las series diaria y mensual. Los resultados se encuentran en el Anexo M. Paradójicamente, el método FPOT empeoró las estimaciones de los modelos TPD en todos los casos (comparar la Figura K. 6 con la Figura M. 1 y la Figura K. 9 con la Figura M. 2). Por lo tanto, de manera empírica se puede concluir que no es adecuado utilizar el método FPOT para filtrar los picos en los eventos de lluvia efectiva, antes de aplicar un modelo de tipo proceso de Poisson filtrado SNP.

4.6 Modelos de Poisson Filtrados mensuales de segundo orden

4.6.1 Estimación de los parámetros por preservación mediante el método de preservación de cinco momentos

Los modelos FPP tipo SNP de segundo orden que se detallaron en la sección 2.4, y específicamente los de ruido disparado de segundo orden (SOSN) se ajustaron para las catorce

series de tiempos evaluadas aquí, siguiendo el método descrito en la sección 2.7.2. La generación de caudales sintéticos se basó en el procedimiento de la sección 2.8 para modelos SOSN. Como aquí el objetivo es preservar cinco momentos estadísticos en dos escalas diferentes (el promedio, desviación estándar y primera auto-correlación de la serie diaria, y la desviación y auto-correlación de la mensual), a este método se le llamará P5M.

En este caso se encontró una complicación adicional en el proceso de ajuste de los parámetros. El sistema de ecuaciones no lineales (2.30)-(2.33) para encontrar los parámetros $\nu_1, \theta_1, b_1, \nu_2, \theta_2$ y b_2 , es complejo y su convergencia fue solo eventual, para cada uno de los meses de las series estudiadas. Como ejemplo, en la estación 3001 Asunción no se obtuvo convergencia para ningún mes (el código para resolver el sistema no lineal con un programa informático se puede consultar en el Anexo N, para el mes de enero). No obstante en la mayoría de las otras series se obtuvo entre tres y cuatro meses con convergencia. Los problemas de convergencia en el análisis numérico se mantuvieron al intentar calibrar también los modelos DSN o MSOSN.

Para los meses en los que el sistema con fue resuelto numéricamente dentro de la tolerancia especificada, se propuso obtener soluciones razonables aceptando una convergencia débil. Esto se hizo minimizando las ecuaciones implícitas para cada uno de los parámetros. También se cuidó que valores resultantes en los parámetros de tasa de eventos ν , impulso promedio θ y tasa de decaimiento b , fueran conceptualmente posibles según los procesos rápido (identificado con un 1) y lento (identificado con 2). La convergencia débil permite, además, intentar acercar la tasa de eventos a los valores observados. Así, en la Tabla 4.4 se muestran los parámetros mensuales obtenidos para el modelo SNP tipo SOSN utilizando el método P5M.

Tabla 4.4 Parámetros mensuales de un modelo SNP de segundo orden y tipo SOSN, ajustados con el método P5M a partir del registro histórico en la estación hidrológica 3001 Asunción.

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
$b_{\varepsilon,1}$	0.23	0.50	0.27	0.50	0.40	0.31	0.62	0.23	0.23	0.21	0.30	0.21
$b_{\varepsilon,2}$	0.06	0.04	0.05	0.03	0.07	0.07	0.07	0.02	0.02	0.04	0.04	0.01
$\theta_{\varepsilon,1}$	15.2	6.00	8.30	9.30	10.7	11.0	11.87	15.0	13.2	14.7	15.0	19.0
$\theta_{\varepsilon,2}$	2.35	6.00	1.42	5.00	2.74	4.00	5.91	1.80	0.80	1.29	6.00	4.00
$\nu_{\varepsilon,2}$	0.16	0.15	0.14	0.19	0.23	0.25	0.33	0.25	0.24	0.24	0.21	0.17
$\nu_{\varepsilon,1}$	0.16	0.15	0.14	0.19	0.23	0.24	0.33	0.25	0.24	0.24	0.20	0.17

4.6.2 Generación y validación de las series sintéticas de caudal diario para modelos FPP mensuales de segundo orden

Los parámetros de la Tabla 4.4 se utilizaron entonces para componer el modelo SOSN:

$$X(t+1) = (0.87 \cdot e^{-0.23} + 0.13 \cdot e^{-0.06}) \cdot X(t) + \sum_{m=N(t)+1}^{N(t+1)} \{y_m \cdot [0.87 \cdot e^{-0.23(t+1-\tau_m)} + 0.13 \cdot e^{-0.06(t+1-\tau_m)}]\}$$

$$X_{t+1} = 0.057 \cdot [66.1 \cdot (1 - e^{-0.23}) + 39.2 \cdot (1 - e^{-0.06})] \cdot X(t) + \sum_{m=N(t)+1}^{N(t+1)} \{y_m \cdot 0.057 \cdot [66.1 \cdot (1 - e^{-0.23(t+1-\tau_m)}) + 39.2 \cdot (1 - e^{-0.06(t+1-\tau_m)})]\}$$

con magnitud de eventos y_m distribuidos según:

$$f(y) = \frac{1}{17.6} \cdot e^{\frac{-y}{17.6}}, \quad y > 0$$

y tiempos de ocurrencia τ_m que siguen un proceso Poisson con distribución:

$$P[N(t) = n] = e^{-0.16t} \frac{(0.16 \cdot t)^n}{n!}$$

De los caudales así generados, se presenta un año en la Figura 4.20. Si se compara con la Figura 4.14a) que presenta los caudales observados en ese mismo año (1963), se aprecia cierta similitud en las recesiones obtenidas para los eventos más grandes, aunque las magnitudes simuladas son parecen excesivas. Además, las recesiones también llevan al caudal a un nivel excesivamente bajo, de modo que no se reproduce bien el componente de flujo lento que se esperaba mejorar al utilizar un modelo de segundo orden.

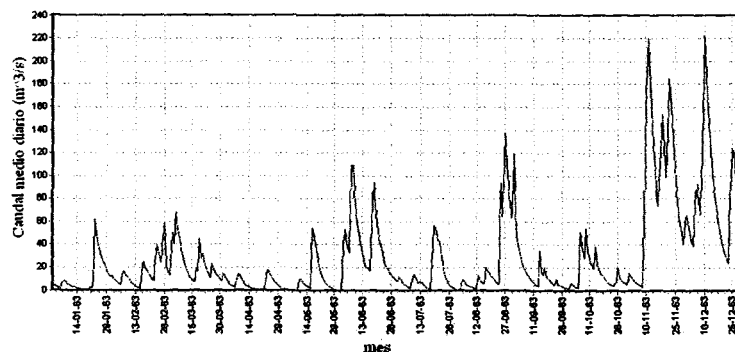


Figura 4.20 Un año de caudales medios diarios en la estación 3001 Asunción para la serie SEMC simulada con el método P5M

Desavenencias como las anteriores, que se detectaron en el patrón de las series simuladas, se pueden apreciar también al verificar la preservación de los momentos estadísticos. La Figura

4.21 hace evidente las fallas en el cumplimiento de prácticamente todos los momentos en escala diaria o mensual. Todas estas evidencias sugieren que el incumplimiento de los supuestos de la distribución gamma para el proceso $X(t)$ de caudales diarios (sección 2.3), la distribución Poisson de los eventos, su independencia y la distribución exponencial de sus magnitudes afectan con mucha más intensidad, el proceso de calibración de un modelo más complejo, como con los SNP de segundo orden.

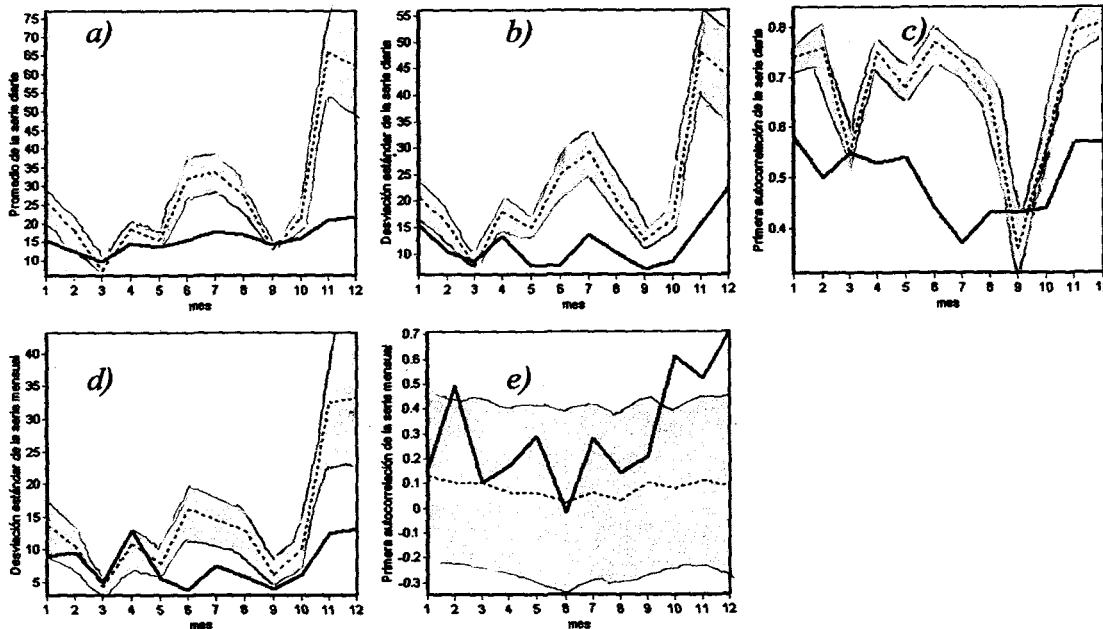


Figura 4.21 Estadísticos mensuales de la serie observada (línea gruesa continua) y de las 100 series MC (línea delgada punteada) tipo P5M, con su respectivo intervalo del 95% de confianza (región gris). En escala de variación diaria se tienen *a)* el promedio, *b)* la desviación estándar y *c)* la primera auto-correlación, y en escala mensual *d)* la desviación estándar y *e)* la primera auto-correlación

Resultados similares se encontraron en las restantes estaciones, por lo que, en general, los modelos SNP de segundo método parecen ser muy exigentes en su etapa de estimación, como para lograr su implementación de manera adecuada, para simular caudales diarios.

4.7 Resumen de los resultados obtenidos

La aplicación de modelos FPP de primer y segundo orden, con diversas variantes, demostró la dificultad de modelar caudal diario. Esto fue discutido en el capítulo de introducción al tema.

En la Tabla 4.5 se hace una valoración general del desempeño de los diversos modelos que fueron empleados para simular las series diarias. Es evidente que de todos ellos, solo el TPD resultó satisfactorio para preservar tanto los momentos de las series diarias y mensuales, como las propiedades de cruce. Debe aclararse que solo el modelo P5M debía preservar los momentos mensuales, pues en los demás casos se trata de modelos de primer orden.

Tabla 4.5: Comparación de los resultados obtenidos con los principales modelos FPP

Código estación	Nombre de estación	P3M	TPD	TPD-FPOT	P5M
3001	Asunción	x	√	-	x
5003	Blanco	x	√	-	x
6002	Barbilla	x	√	-	x
6003	Playa Hermosa	x	√	-	x
8001	Pacuare	x	x	x	x
8002	Siquirres	x	√	-	x
8003	Dos Montañas	x	x	x	x
8005	Bajo Huacas	x	√	-	x
28001	Londres	x	x	x	x
28002	Los Llanos	x	x	x	x
29001	Pte. de Hamaca	x	x	x	x
29002	Providencia	x	x	x	x
29003	S. P. Savegre	x	x	x	x
29004	Savegre	x	x	x	x

Este análisis permite concluir que los modelos TPD son los únicos que logran reproducir adecuadamente el caudal promedio diario. Esto se cumple específicamente en la región caribeña del país. El P3M parece dejar claro que la aspiración de preservar tres momentos es excesiva cuando las series observadas incumplen varios de los supuestos de un SNP. El modelo TPD-FPOT no logró solventar los problemas al modelar caudales con altos niveles de estacionalidad. Finalmente, los modelos P5M (de segundo orden) estimados con un método de cinco momentos, presentaron problemas desde la etapa de estimación. Se considera que para aplicar cualquiera de los P5M (que permitirían una reproducción fiel de las recesiones) deberán cumplirse con mayor claridad los supuestos generales de un SNP.

Capítulo 5 Conclusiones y recomendaciones

El análisis de series de tiempo hidrológicas en Costa Rica debe considerar la compleja estructura fisiográfica del país, y su gran variabilidad climática en diversas escalas. La interacción de vientos y corrientes marinas provenientes tanto del océano Pacífico como del mar Caribe, y la relación con el sistema montañoso del país, hacen que las lluvias presenten una gran variabilidad espacial, y se distribuyan en el tiempo de forma muy distinta en las dos vertientes. Esto se agrava por el hecho de que las mediciones hidrológicas se requieren en sitios que tienen cuencas tributarias de muy diversas formas, áreas, coberturas del suelo y patrones de elevación, factores que repercuten en la respuesta hidrológica de las mismas.

La longitud de las series de tiempo no resultó ser un factor asociado con el desempeño de los respectivos modelos de ruido disparado (SNP). Tampoco se pudo explicar el rendimiento del modelo tomando como base las propiedades de las cuencas tributarias asociadas a las respectivas estaciones (área, elevaciones, pendientes), ni las propiedades estadísticas de las series diarias (media, desviación estándar, coeficiente de variabilidad, sesgo, curtosis, distribución de frecuencias, curva de duración, etc).

La cantidad de eventos que provocan incrementos en los caudales diarios medidos en los ríos no varía entre las vertientes Caribe y Pacífico, pero sí presentan una variación estacional de modo que para el mes con menor cantidad de eventos se observaron tasas de ocurrencia entre 2 y 5 *eventos/mes*. A partir del mes de mayo, que es el primero del año hidrológico en Costa Rica y un mes de transición, se presentan las mayores tasas de ocurrencia de eventos con entre 5.5 y 8 *eventos/mes*.

El procedimiento FPOT para filtrar aquellos picos que no cumplen con las premisas de un proceso de ruido disparado (SNP), refleja la existencia de eventos espurios en las series que tienen largas recesiones, típicamente las del Pacífico. La tasa de eventos como función del umbral de filtrado decrece más rápidamente en las series de esta vertiente, con respecto a las del Caribe, ya que en las primeras la magnitud relativa de los eventos es menor con respecto al

nivel medio de la serie. Esta mayor homogeneidad entre los caudales comunes y los caudales pico hace que estos últimos se descarten con niveles de filtrado más bajos.

Para una probabilidad fija de excedencia del umbral de filtro FPOT, existe una relación lineal entre la tasa de eventos y el coeficiente de variabilidad de los caudales diarios de la serie. Es posible, entonces, ajustar ecuaciones lineales que expresen esta relación para diversos percentiles de excedencia.

El procedimiento de selección de picos filtrados sobre un umbral (FPOT) proporcionó resultados muy variables en las diversas estaciones analizadas, sin importar en cual vertiente se encontraron. Cuando resultó aplicable, se logró obtener un umbral que garantiza los supuestos del modelo SNP sobre independencia y distribución Poisson de los eventos. Estos umbrales registraron valores desde el 20% del promedio de la serie, hasta el 240%.

Las pruebas de bondad de ajuste de Cramér-von Mises (W^2) y de Anderson-Darling (A^2) no son adecuadas para evaluar la distribución de las magnitudes de una secuencia de eventos de lluvia efectiva, en el caso de series de caudal medio diario. Adicionalmente, es erróneo utilizar el modelo de distribución pareto-exponencial para aproximar el parámetro de salto promedio de los picos. Dichos valores son altamente sobreestimados, si se comparan con aquellos resultantes de aplicar los métodos de estimación de momentos.

Al definir las propiedades de cruce de las series de caudal como función de la probabilidad de excedencia del umbral utilizado, se puede comprobar que éstas reflejan la inestabilidad de las series de caudal diario. Las estaciones del Caribe siguen este patrón y es notable la diferencia con las del Pacífico, con estacionalidad definida y altos niveles de auto-correlación. Como máximo, se llegaron a contar hasta 28 cruces hacia arriba UC (o hacia abajo DC), como promedio anual, para todo el registro de la estación 3001 Asunción. En cambio, la estación 29001 Puente de Hamaca presenta una cantidad máxima de UC (ó DC) de 5.5 cruces anuales, para un umbral con probabilidad de 24% de ser excedido. La longitud promedio de las corridas denota que, siguiendo una proporcionalidad inversa con los UC (ó DC), se dan lapsos mucho mayores entre cruces en las estaciones del Pacífico. Así, la estación 29002 Providencia

presenta, para cualquier umbral, al menos periodos de 27 días entre cruces. En contraposición, en la 5003 Blanco pueden verse cruces cada 7 días.

La generación de caudales sintéticos utilizando procesos de Poisson filtrados (FPP) del tipo SNP de primer orden, permite establecer variantes en su etapa de estimación de parámetros. Para esto debe tomarse como base el método de preservación de momentos estadísticos. Un primer enfoque llamado P3M, que pretende conservar la media, la desviación estándar y la primera auto-correlación en escala diaria, resultó poco satisfactorio para la preservación de estos momentos, y los de escala mensual. También erró notablemente al estimar las tasas de eventos, que se sobreestimaron de manera considerable. Esto incidió principalmente en el parámetro de la recesión, que tomó valores relativamente grandes y produjo recesiones muy agudas después de los picos. Lógicamente esto repercute en una sobreestimación los UC (ó DC) de aproximadamente 2.8 veces, misma magnitud en que se subestima la longitud promedio de corridas.

La estimación del SNP de primer orden preservando sólo dos momentos (la media y la desviación de la serie diaria) y fijando la tasa de eventos (método TPD) suministró resultados más realistas. Lo anterior se aplica en la reproducción tanto de las tasas de eventos, como de los momentos en sí. En este último caso, para las series simuladas se evaluaron los momentos explícitos (los tres de las series en escala diaria) y los implícitos (la desviación estándar y la primera auto-correlación en escala mensual). Cuando el modelo SNP-TPD fue satisfactorio (seis de las catorce estaciones analizadas) las propiedades de cruce se lograron reproducir adecuadamente.

También en el caso del modelo SNP tipo TPD, se alcanzó el objetivo de investigación inicialmente planteado, al verificar la adecuada reproducción del patrón de picos y recesiones característico en las series de caudal promedio diario.

Con el método TPD no se logró preservar la media en las series de la cuenca Pacífico (especialmente en los meses secos, diciembre a mayo) ni los demás momentos en la mayoría de los meses del año. Tampoco se tuvo éxito en las estaciones 8001 Pacuare y 8003 Dos

Montañas, que tienen influencia del Pacífico, y que presentan las mayores correlaciones con los patrones hidrológicos de aquella vertiente. Estas dos se son afectadas por el fenómeno el NIÑO3.4 de manera similar a las estaciones del Pacífico, y tienen estructuras de auto-correlación y estacionalidad también similares a estas últimas.

Otras variantes del SNP de segundo orden y tasa de eventos fija no resultaron exitosos (modelos TPA y TDA). Esto se debe a que, ante el incumplimiento de los supuestos que se presenta en todas las series estudiadas, la preservación de la primera auto-correlación de la serie diaria es una condición muy exigente. Así, se introduce una severa distorsión en la estimación de los parámetros de decaimiento y salto medio.

La aplicación del método FPOT para seleccionar los picos de lluvia efectiva que deben modelarse con un SNP no produjo buenos resultados. Su implementación en los modelos TPD más deficientes sólo logró empeorar los resultados de la preservación de momentos observados.

No se recomienda la aplicación de los SNP de segundo orden, ya que la gran cantidad de parámetros que demandan dificulta la aplicación de los métodos de preservación de momentos. En consecuencia, la estimación de estos parámetros por medio de métodos numéricos difícilmente converge. La debilidad en el cumplimiento de los supuestos sugiere modelos más parsimoniosos, que permitan verificar con más precisión los valores de los parámetros desde el punto de vista conceptual.

Una generalización del procedimiento para fijar las tasas de eventos podría hacerse en el modelo SOSN aquí expuesto, tomando ventaja del hecho de que éste considera que los impulsos del componente rápido (escala de variación diaria) y los del componente lento (escala de variación mensual) se producen simultáneamente. Al fijar estas dos tasas de ocurrencia en valores iguales a los observados, se puede prescindir de una de las expresiones en el sistema de ecuaciones no lineales correspondiente, y entonces estimar los demás parámetros restantes.

Una nueva aplicación de los modelos SNP podría hacerse, pero esta vez estimando sus parámetros a partir de las series observadas y sin suponer la preservación de sus momentos. Esto posibilitaría evaluar si un SNP puede explicar empíricamente el comportamiento de diente de sierra específico para los caudales diarios en Costa Rica, y en caso de ser así, qué repercusiones se observaría sobre los grupos de tres y cinco momentos que el método original de estimación pretendía preservar.

A partir del conocimiento de que las distribuciones de caudal diario en Costa Rica no siguen una distribución gamma, y que la magnitud de sus eventos efectivos no siguen una distribución exponencial (entre otros supuestos incumplidos), se sugiere desarrollar un nuevo modelo de tipo FPP (de primer o segundo orden) que, a diferencia del SNP, consideren específicamente las condiciones locales en el componente estocástico.

De manera implícita, al ejecutar las simulaciones de un gran número de series sintéticas se establecieron escenarios críticos de eventos máximos (siempre que el modelo haya sido validado de forma adecuada). También, la permanencia de caudales excepcionalmente bajos se puede estudiar a partir de estas series.

Los FPP pueden aplicarse en entidades que intervienen en la gestión de los recursos hídricos de forma apreciable, como el Instituto Costarricense de Electricidad. Las metodologías de estimación de caudales de compensación requieren simulaciones de las condiciones naturales encontradas en los ríos. Con esto se pueden conocer las condiciones hidrológicas más críticas posibles encontradas en la naturaleza (de escasez de agua, en este caso). Sobre esta base se pueden valorar las afectaciones ambientales y socioeconómicas en que se incurriría con diversos esquemas de regulación del caudal natural.

Para la ejecución del análisis mencionado en el párrafo anterior se recomienda la utilización de un modelo SNP de tipo TPD, siempre y cuando la serie por simular se encuentre en el Caribe de Costa Rica y no presente influencia del Pacífico en su comportamiento. Si estas condiciones no se cumplen, las variantes teóricas y prácticas sugeridas aquí podrían permitir el desarrollo de nuevos y mejores modelos para las condiciones tropicales del país.

Referencias bibliográficas

- Aksoy, H. (2003) Markov chain-based modeling techniques for stochastic generation of daily intermittent streamflows. *Adv. Water Resour.* 26[6], 663–671.
- Aksoy, H. (2004) Pulse generation models for daily intermittent streamflows. *Hydrological Sciences Journal* 49[3], 399–411.
- Aksoy, H., Bayazit, M. (2000a) A model for daily flows of intermittent streams. *Hydrol. Process.* 14, 1725–1744.
- Aksoy, H., Bayazit, M. (2000b) A daily intermittent streamflow simulator. *Turk Journal of Engineering and Environmental Science* 24[4], 265-276.
- Beard, L.R. (1967) Simulation of daily streamflow. *Proceedings, The International Hydrology Symposium, Fort Collins, Colorado*, 624–632.
- Bernier J. (1970) Inventaire des modèles de processus stochastiques applicables a la description des débits journaliers des rivières. *Revue Inst Intern Stat* 38[1], 49–61.
- Choulakian, V., Stephens, M. A. (2001) Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics* 43[4], 478-484.
- Claps, P., Murrone, F. (1993) Univariate conceptual–stochastic models for spring runoff simulation. En *Modelling and Simulation*, Hamza MH (ed.). *International Conference on Modeling and Simulation, IASTED*, 10–12 May, Pittsburgh, 491–494.
- Claps, P., Murrone, F. (1994) Optimal parameter estimation of conceptually-based streamflow models by time series aggregation. En *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, vol. 3, Hipel KW, McLeod AL, Panu US, Singh VP (eds). *Kluwer Acad., Dordrecht, The Netherlands*, 421–434.
- Claps, P., Rossi, F. (1995) Conceptually-based univariate stochastic modelling of river runoff. *Excerpta* 8, 95–133.
- Claps P., Laio, F., Villani, P. (2002) Assessment of extreme flood production mechanisms through POT analysis of daily data, *Proceedings International conference on flood estimation, Berne, Switzerland*. CHR report II-17, 75-83.
- Claps P., Laio, F. (2003) Can continuous streamflow data support flood frequency analysis? An alternative to the Partial Duration Series approach. *Water Resources Research* 39[8].

- Claps, P., Giordano, A., Laio, F. (2005) Advances in shot noise modeling of daily streamflows. *Advances in Water Resources*, 28, 992-1000.
- Comisión Asesora sobre Degradación de Tierras (CADETI)/Ministerio de Ambiente y Energía (MINAE) (2004) Programa de Acción Nacional de Lucha contra la Degradación de Tierras en Costa Rica. San José, Costa Rica.
- Ferguson, T.S., Genest, C., Hallin, M. (2000) Kendall's tau for serial dependence. *Can. J. Stat.* 28, 587-604.
- Gómez, F., Krasovskaia, I., Gottschalk, L., Leblois, E. (2006) Interpolation of water balance components for Costa Rica. Proc. 5th FRIEND World Conference, Water resource variability: Processes, analyses and impacts. Havana, Cuba. IAHS Publ. 308.
- Gottschalk, L. (2004) Ch. 7 Low flow and drought time series modelling. En "Hydrological Drought" by L. Tallaksen and H.A.J. van Lanen (Eds.), Elsevier.
- Gottschalk, L. (2005) Ch. 7 Methods for analysing variability. En Anderson, M.G. & McDonnell, J.J. *Encyclopedia of Hydrological Sciences*.
- Haan, C.T. (1977) *Statistical methods in hydrology*, The Iowa State University Press, Ames.
- Kelman, J. (1980) A stochastic model for daily streamflow. *Journal of Hydrology* 47[3/4], 235-249.
- Kottegoda, N.T., Horder M.A. (1980) Daily flow model based on rainfall occurrences using pulses and a transfer function. *Journal of Hydrology* 47[3/4], 215-234.
- Kottegoda, N.T., Natale, L., Raiteri, E. y Saccardo, I. (1995) A Stochastic Model of Daily Flows for Simulating Low Flows in a Highly Developed Basin. Proc. Int. Conf. Statistical and Bayesian Methods in Hydrological Sciences in Honor of Jacques Bernier, UNESCO, Paris.
- Kottegoda, N.T., Natale, L. (2000) Daily Streamflow Simulation Using Recession Characteristics. *J. Hydrologic Engrg.*, 5[1], 17-24.
- Kron, W., Plate E.J., Ihringer, J. (1990) A model for the generation of simultaneous daily discharges of two rivers at their point of confluence. *Stochastic Hydrology and Hydraulics* 4, 255-276.
- Lang, M., Ouarda, T.B.M.J., Bobée, B. (1999) Towards operational guidelines for over-threshold modeling, *J. Hydrol.* 225, 103-117.

- Lawrance, A.J., Kottegoda N.T. (1977) Stochastic modeling of riverflow time series, *Journal of the Royal Statistical Society*, A140, 1-47.
- Murrone, F., Rossi, F., Claps, P. (1997) Conceptually-based Shot Noise modelling of streamflows at short time interval. *Stoch. Hydrol. and Hydraul.* 11[6], 483-510.
- Parzen, E. (1962) *Stochastic Processes*. Holden Day Inc., San Francisco, CA.
- Payne K., Neumann W.R., Kerri K.O. (1969) Daily stream flow simulation. *Journal of the Hydraulic Division, American Society of Civil Engineers* 95[4], 1163-1179.
- Quimpo, R.G. (1968) Stochastic analysis of daily river flows. *J. Hydraul. Div. Proc. Amer. Soc. Civ. Eng. Paper No. 5719,94, (HYD)*, 43-57.
- Quimpo, R.G. (1971) Structural relation between parametric and stochastic hydrology. *Mathematical models in hydrology, vol.1, Int. Assoc. Hydrol. Sci. Publ.* 100, 151-157.
- Rajagopalan, B., Lall, U., Tarboton, D.G. (1996) A Nonhomogeneous Markov Model for Daily Precipitation Simulation. *ASCE, Journal of Hydrologic Engineering* 1[1], 33-40.
- Schneider K., Schultz G.A. (1982) Multisite data generation model for daily discharges. *Proceedings of the Exeter Symposium: Optimal Allocation of Water Resources, IAHS Publ. No. 135, International Association of Hydrological Sciences, Wallingood*, 69-78.
- Treiber B., Plate E.J. (1977) A stochastic model for the simulation of daily flows. *Hydrological Sciences Bulletin* 22[1], 175-192.
- Weiss, G. (1973) Shot noise models for synthetic generation of multisite daily streamflow data. *Proc. of the Symposium on Design of Water Resources Project with Inadequate Data, Unesco, WMO, IAHS, Madrid, Vol. 2*.
- Weiss, G. (1975) Time-Reversibility of Linear Stochastic Processes. *Journal of Applied Probability* 12[4], 831-836.
- Weiss, G. (1977) Shot noise models for the generation of synthetic streamflow data. *Water Resources Research* 13[1], 101-108.
- Xu, C.Y. (2005) *Hydrological Models*. Lecture book for "Hydrological Modelling" Course. Department of Geosciences, University of Oslo.
- Xu, Z.X., Li, J.Y., Ito, K. (1998) Clustering stochastic point process model for flood risk analysis. *Stochastic Hydrology and Hydraulics* 12, 53-64.
- Xu, Z.X., Schumann, A., Brass, C., Li, J.Y., Ito, K. (2001) Chain-dependent Markov correlation pulse model for daily streamflow generation. *Adv. Water Resour.* 24, 551-564.

- Xu, Z.X., Schultz, G.A., Schumann, A. (2002) A conceptually based stochastic point process model for daily stream-flow generation. *Hydrological Processes* 16[15], 3003-3017.
- Xu, Z.X., Schumann, A., Li, J.Y. (2003) Markov cross-correlation pulse model for daily streamflow generation at multiple sites. *Adv. Water Resour.* 26, 325–335.
- Yue, S., Hashino, M. (1999) A stochastic model for deriving the basic statistics of J-averaged streamflow. *Water Resour Res* 35[10], 3127–3137.
- Yue, S., Hashino, M., Bobée, B., Rasmussen, P.F., Ouarda, T.B.M.J. (1999) Derivation of streamflow statistics based on a filtered point process. *Stochastic Environmental Research and Risk Assessment* 13, 317-326.

Anexo A: Definición de la función de log-verosimilitud de perfil para la distribución Pareto generalizada

Partiendo de la función de densidad $f(y)=(1/a)(1-k \cdot y/a)^{(1-k)/k}$, para una muestra aleatoria y_1, y_2, \dots, y_N se puede establecer la verosimilitud como:

$$l(a, k) = \frac{1}{a} \left(1 - \frac{k \cdot y_1}{a}\right)^{\frac{1-k}{k}} \cdot \frac{1}{a} \left(1 - \frac{k \cdot y_2}{a}\right)^{\frac{1-k}{k}} \cdot \dots \cdot \frac{1}{a} \left(1 - \frac{k \cdot y_N}{a}\right)^{\frac{1-k}{k}}$$

$$l(a, k) = \left(\frac{1}{a}\right)^N \prod_{i=1}^N \left(1 - \frac{k \cdot y_i}{a}\right)^{\frac{1-k}{k}}$$

Aplicando el logaritmo natural a la ecuación anterior se obtiene la log-verosimilitud:

$$\ln[l(a, k)] = L(a, k) = N \cdot \ln\left(\frac{1}{a}\right) + \left(\frac{1-k}{k}\right) \cdot \sum_{i=1}^N \ln\left(1 - \frac{k \cdot y_i}{a}\right)$$

$$L(a, k) = -N \cdot \ln(a) + \left(\frac{1}{k} - 1\right) \cdot \sum_{i=1}^N \ln\left(1 - \frac{k \cdot y_i}{a}\right) \quad (\text{A.1})$$

Si se acepta la relación entre los parámetros $a=k/\psi$, se puede plantear:

$$L(k) = -N \cdot \ln\left(\frac{k}{\psi}\right) + \left(\frac{1}{k} - 1\right) \cdot \sum_{i=1}^N \ln(1 - y_i \cdot \psi)$$

El estimador máximo verosímil de k está definido por:

$$\frac{\partial L(k)}{\partial k} = -N \cdot \frac{\partial[\ln(k/\psi)]}{\partial k} + \sum_{i=1}^N \ln(1 - y_i \cdot \psi) \cdot \frac{\partial(1/k)}{\partial k} - 0 = 0$$

$$\frac{\partial L(k)}{\partial k} = -N \cdot \frac{1}{k} + \sum_{i=1}^N \ln(1 - y_i \cdot \psi) \cdot \frac{-1}{k^2} = 0$$

$$-N = \sum_{i=1}^N \ln(1 - y_i \cdot \psi) \cdot \frac{1}{k}$$

$$\hat{k} = \frac{-1}{N} \sum_{i=1}^N \ln(1 - y_i \cdot \psi) \quad (\text{A.2})$$

y entonces

$$\hat{a} = \frac{\hat{k}}{\psi} \quad (\text{A.3})$$

Al sustituir en (A.1) sucesivamente $a=k/\psi$ y luego (A.2) se obtiene la log-verosimilitud de perfil (así llamada pues es la log-verosimilitud antes maximizada sobre k):

$$L(k) = -N \cdot \ln\left(\frac{k}{\psi}\right) + \left(\frac{1}{k} - 1\right) \cdot \sum_{i=1}^N \ln(1 - y_i \cdot \psi)$$

$$L(\psi) = -N \cdot \ln\left[\frac{-1}{N \cdot \psi} \sum_{i=1}^N \ln(1 - y_i \cdot \psi)\right] + -N - \sum_{i=1}^N \ln(1 - y_i \cdot \psi) \quad (\text{A.4})$$

Anexo B: Límites superiores porcentuales asintóticos para los estadísticos de Cramér-von Mises W^2 y de Anderson-Darling A^2

Tabla B.1: Puntos porcentuales asintóticos para W^2 , donde $P(W^2 \geq z) = \alpha$ y z es el valor leído en la tabla

$k \setminus \alpha$	0.5	0.25	0.1	0.05	0.025	0.01	0.005	0.001
-0.9	0.046	0.067	0.094	0.115	0.136	0.165	0.187	0.239
-0.5	0.049	0.072	0.101	0.124	0.147	0.179	0.204	0.264
-0.2	0.053	0.078	0.111	0.137	0.164	0.2	0.228	0.294
-0.1	0.055	0.081	0.116	0.144	0.172	0.21	0.24	0.31
0	0.057	0.086	0.124	0.153	0.183	0.224	0.255	0.33
0.1	0.059	0.089	0.129	0.16	0.192	0.236	0.27	0.351
0.2	0.062	0.094	0.137	0.171	0.206	0.254	0.291	0.38
0.3	0.065	0.1	0.147	0.184	0.223	0.276	0.317	0.415
0.4	0.069	0.107	0.159	0.201	0.244	0.303	0.349	0.458
0.5	0.074	0.116	0.174	0.222	0.271	0.338	0.39	0.513

Tabla B.2: Puntos porcentuales asintóticos para A^2 , donde $P(A^2 \geq z) = \alpha$ y z es el valor leído en la tabla

$k \setminus \alpha$	0.5	0.25	0.1	0.05	0.025	0.01	0.005	0.001
-0.9	0.339	0.471	0.641	0.771	0.905	1.086	1.226	1.559
-0.5	0.356	0.499	0.685	0.83	0.978	1.18	1.336	1.707
-0.2	0.376	0.534	0.741	0.903	1.069	1.296	1.471	1.893
-0.1	0.386	0.55	0.766	0.935	1.11	1.348	1.532	1.966
0	0.397	0.569	0.796	0.974	1.158	1.409	1.603	2.064
0.1	0.41	0.591	0.831	1.02	1.215	1.481	1.687	2.176
0.2	0.426	0.617	0.873	1.074	1.283	1.567	1.788	2.314
0.3	0.445	0.649	0.924	1.14	1.365	1.672	1.909	2.475
0.4	0.468	0.688	0.985	1.221	1.465	1.799	2.058	2.674
0.5	0.496	0.735	1.061	1.321	1.59	1.958	2.243	2.922

Anexo C: Estimación de las funciones de respuesta para procesos promediados en el tiempo

Partiendo de la función de respuesta para un modelo de segundo orden de tipo SOSN dado en la ecuación (2.15), es posible plantear que, para un paso de tiempo s , el valor del proceso promediado en el intervalo ($0 < t \leq s$) tiene el valor:

$$\begin{aligned} h_j(t) &= \int_0^t h(\tau) d\tau = \int_0^t \left(\frac{\theta_1}{\theta} e^{-b_1 \tau} + \frac{\theta_2}{\theta} e^{-b_2 \tau} \right) d\tau = \frac{1}{\theta} \cdot \left(\theta_1 \int_0^t e^{-b_1 \tau} d\tau + \theta_2 \int_0^t e^{-b_2 \tau} d\tau \right) \\ &= \frac{1}{\theta} \cdot \left[\frac{-\theta_1}{b_1} \cdot (e^{-b_1 \tau} |_0^t) + \frac{-\theta_2}{b_2} (e^{-b_2 \tau} |_0^t) \right] = \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (e^{-b_1 \cdot 0} - e^{-b_1 t}) + \frac{\theta_2}{b_2} \cdot (e^{-b_2 \cdot 0} - e^{-b_2 t}) \right] \\ h_j(t) &= \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1 - e^{-b_1 t}) + \frac{\theta_2}{b_2} \cdot (1 - e^{-b_2 t}) \right] \quad (\text{C.1}) \end{aligned}$$

que coincide con la ecuación (2.27), con $s=1$. Ésta es una función creciente, que a partir del instante $t=s$ debe cambiar su forma para reproducir la recesión correspondiente, que estaría definida por la respuesta del sistema $h(t)$. Entonces, para ($t > s$) la función complementaria $h_j^{(c)}(t)$ tiene la forma $h_j^{(c)}(t) = h_j(s) \cdot H(t-s)$, donde $H(t-s) = h(t-s) + C$ y sujeto a que en $t=s$ $H(0)=1$. Entonces, empleando la función de respuesta $h(t)$ dada en (2.15) se plantea que:

$$H(0) = h(0) + C = \frac{\theta_1}{\theta} e^{-b_1 \cdot 0} + \frac{\theta_2}{\theta} e^{-b_2 \cdot 0} + C = 1$$

$$C = 1 - \frac{\theta_1 + \theta_2}{\theta}$$

Entonces:

$$H(t-s) = \frac{\theta_1}{\theta} e^{-b_1(t-s)} + \frac{\theta_2}{\theta} e^{-b_2(t-s)} + 1 - \frac{\theta_1 + \theta_2}{\theta} = 1 + \frac{1}{\theta} (\theta_1 \cdot e^{-b_1(t-s)} + \theta_2 \cdot e^{-b_2(t-s)} - \theta_1 - \theta_2)$$

$$H(t-s) = 1 + \frac{1}{\theta} \cdot [\theta_1 \cdot (e^{-b_1(t-s)} - 1) + \theta_2 \cdot (e^{-b_2(t-s)} - 1)]$$

Finalmente, para $t > s$:

$$\begin{aligned} h_j^{(c)}(t) &= \frac{1}{\theta} \cdot \left[\frac{\theta_1}{b_1} \cdot (1 - e^{-b_1 s}) + \frac{\theta_2}{b_2} \cdot (1 - e^{-b_2 s}) \right] \\ &\quad \left\{ 1 + \frac{1}{\theta} \cdot [\theta_1 \cdot (e^{-b_1(t-s)} - 1) + \theta_2 \cdot (e^{-b_2(t-s)} - 1)] \right\} \quad (\text{C.2}) \end{aligned}$$

que con $s=1$ coincide con la ecuación (2.28), y rige para $t > 1$.

Anexo D: Series de caudal diario analizadas

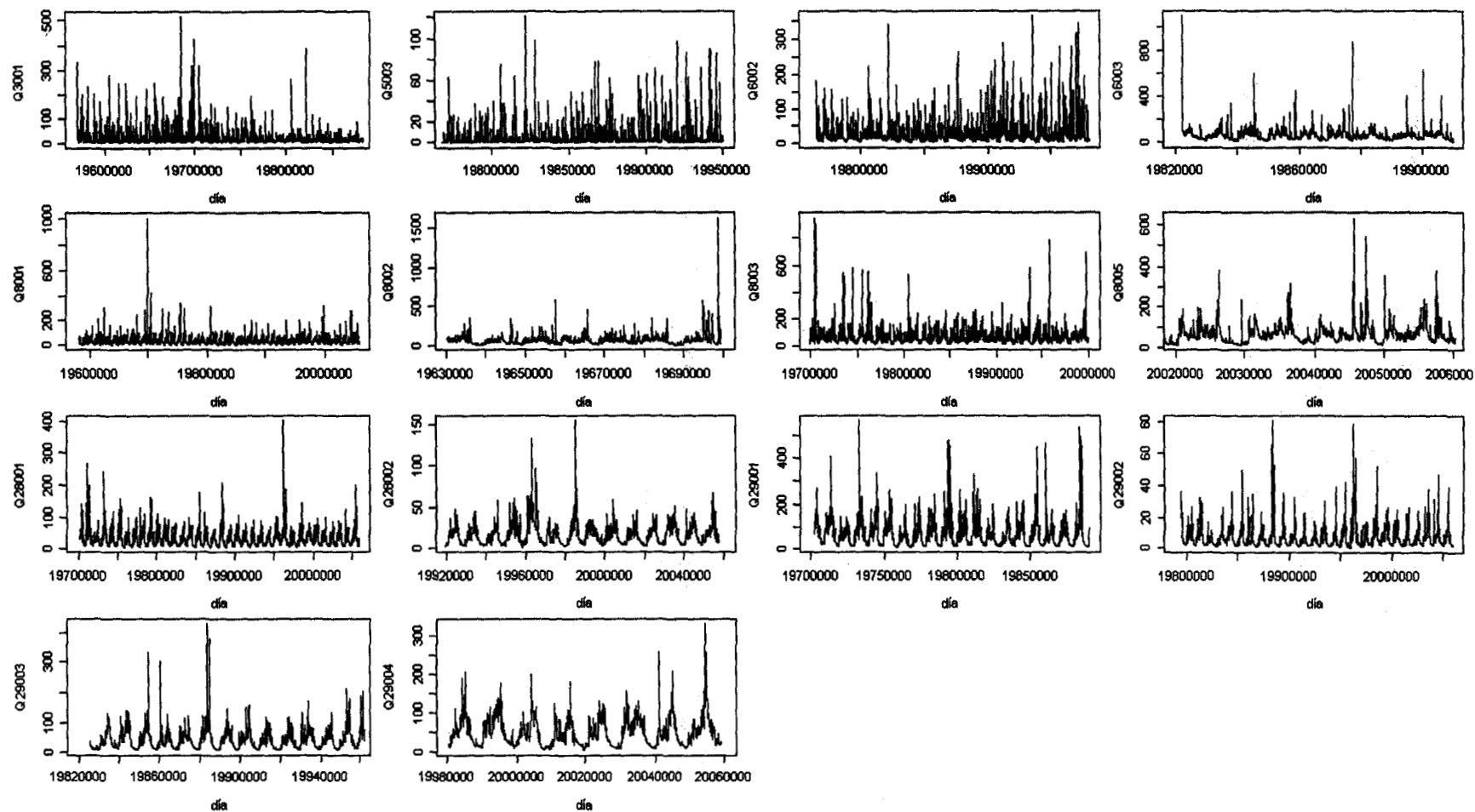


Figura D.1: Series de caudal diario para las catorce estaciones hidrológicas seleccionadas. Las estaciones se identifican en la ordenada de cada gráfico como QXXXX(X), donde la secuencia de X indica el número de estación. La escala de tiempo tiene el formato aaaammdd, donde aaaa representa el año, mm el mes y dd el día

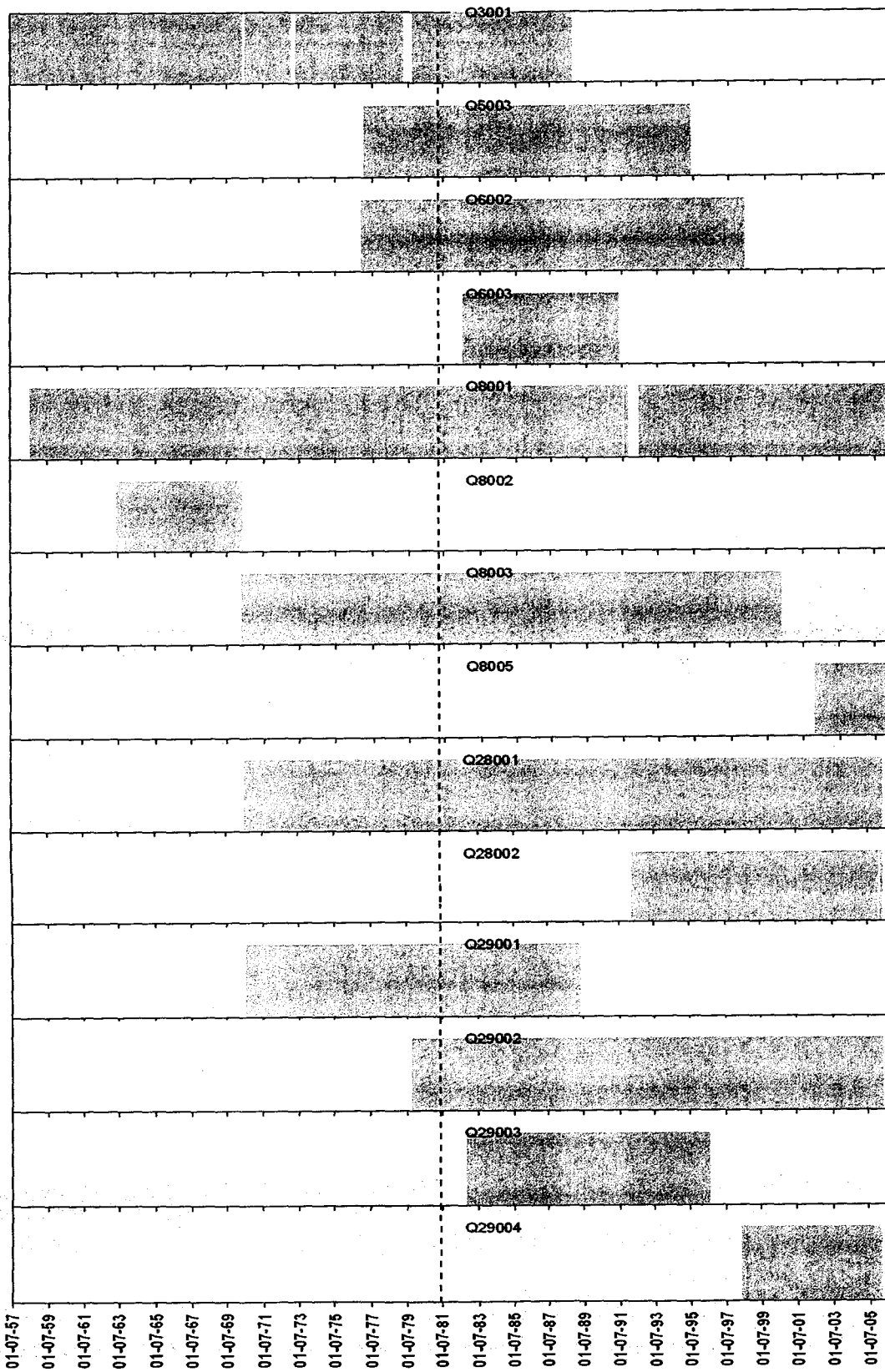


Figura D.2: Periodos de registro en las estaciones hidrológicas seleccionadas

Anexo E: Regresiones utilizadas para completar las series, y valores generados

Estación 3001

Figura E.1: Regresión dependiente de la estación 8001

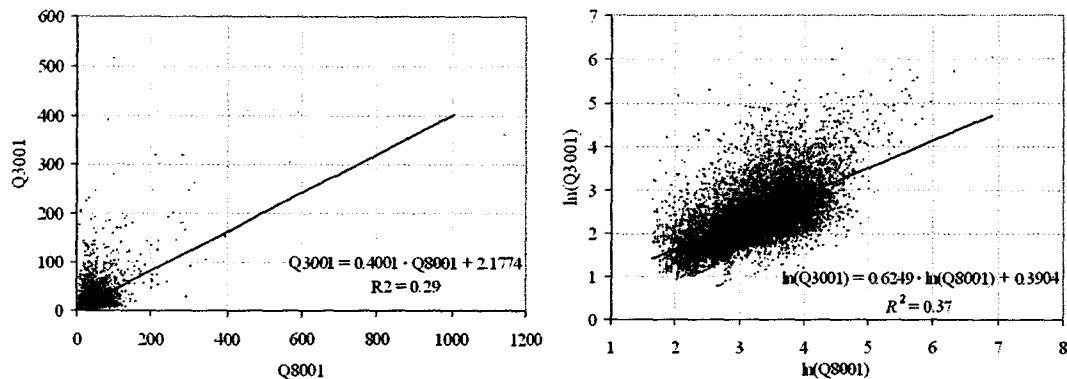


Figura E.2: Periodo 01/05/1970 – 24/07/1970

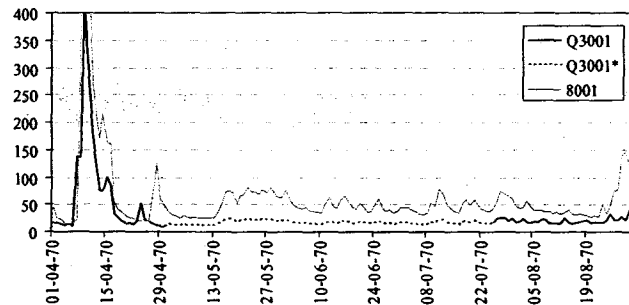


Figura E.3: Periodo 04/12/1970 – 26/12/1970

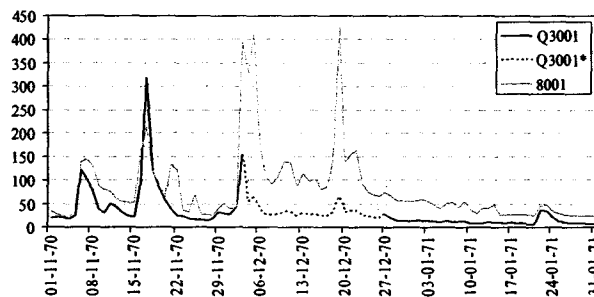


Figura E.4: Periodo 11/01/1973-30/04/1973

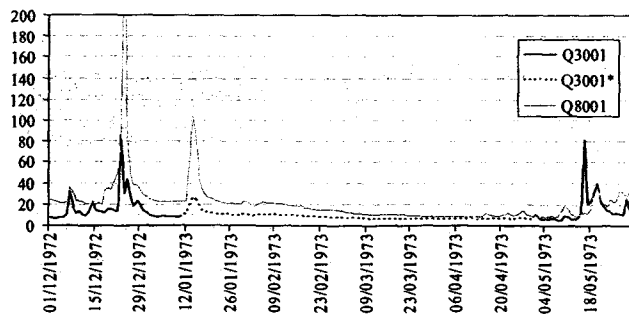


Figura E.5: Regresión dependiente de la estación 6002

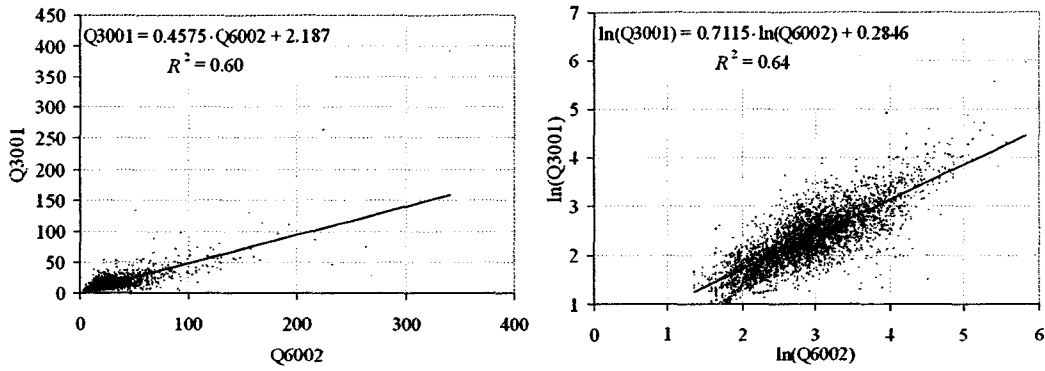
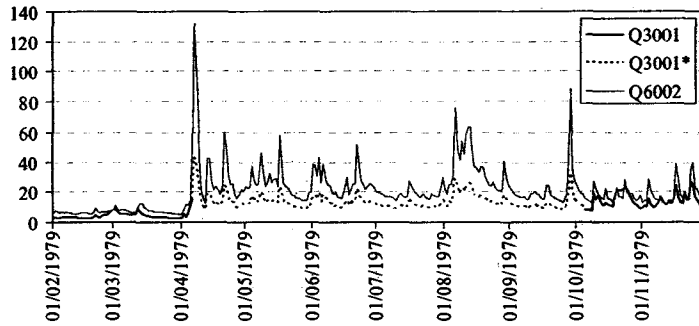


Figura E.6: Periodo 07/04/1979 – 04/10/1979



Estación 8001

Figura E.7: Regresión dependiente de la estación 8003

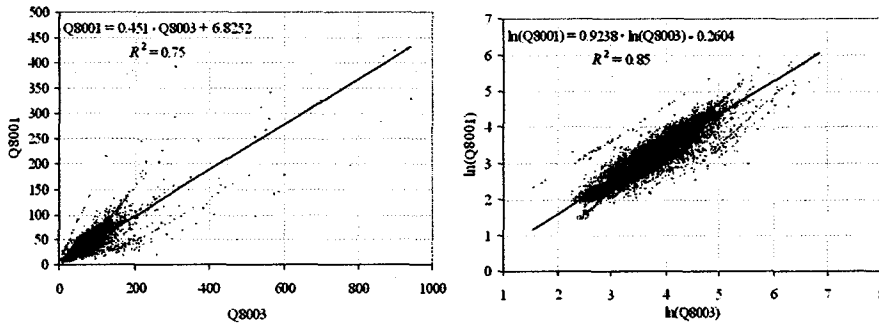


Figura E.8: Periodos 23/06/1991–02/08/1991, 09/08/1994– 28/08/1991 y 12/09/1991 – 22/05/1992

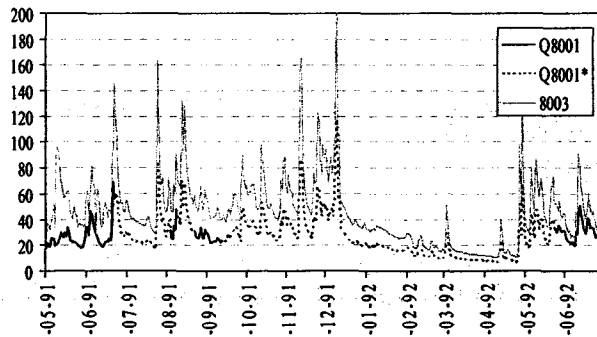
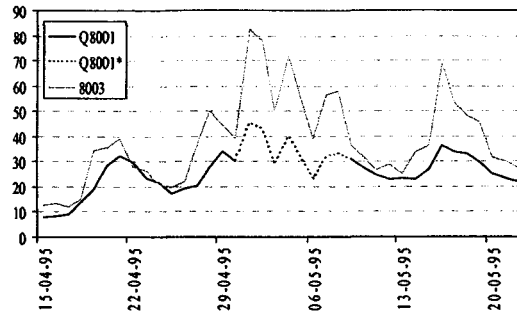


Figura E.9: Periodo 01/05/1995 – 08/05/1995



Estación 8002

Figura E.10: Regresión dependiente de la estación 8001

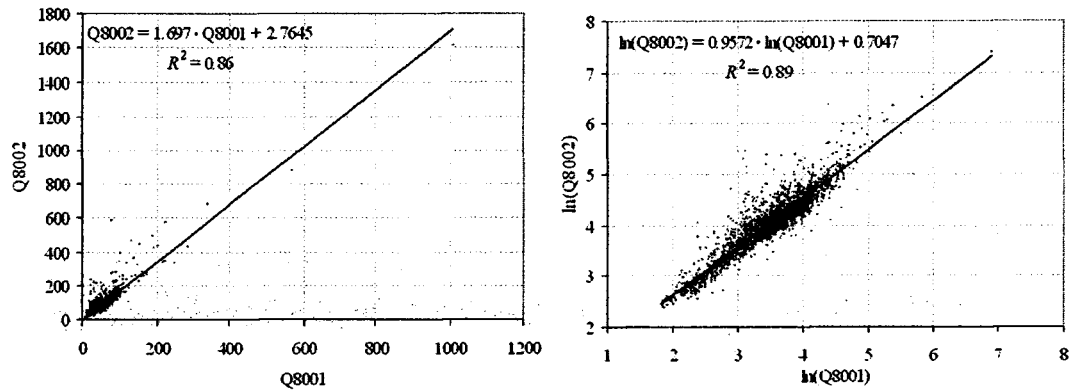
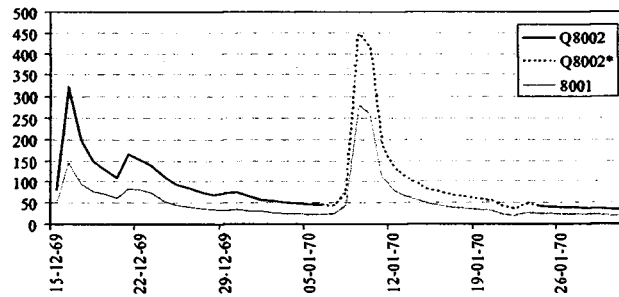


Figura E.11: Periodo 07/01/1970 – 23/01/1970



Estación 28001

Figura E.12: Regresión dependiente de la estación 29004

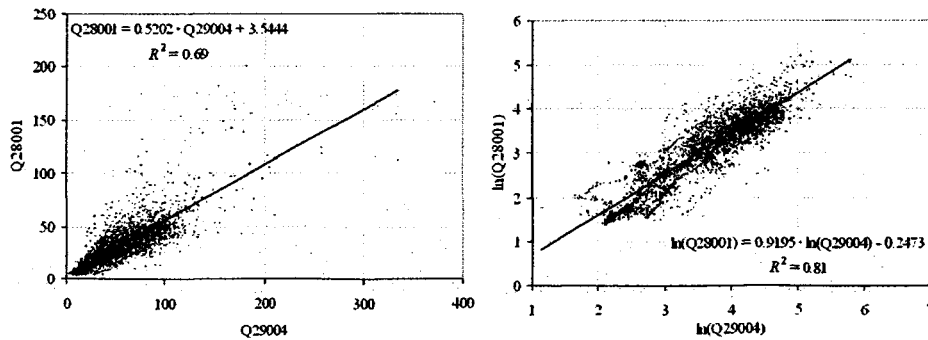
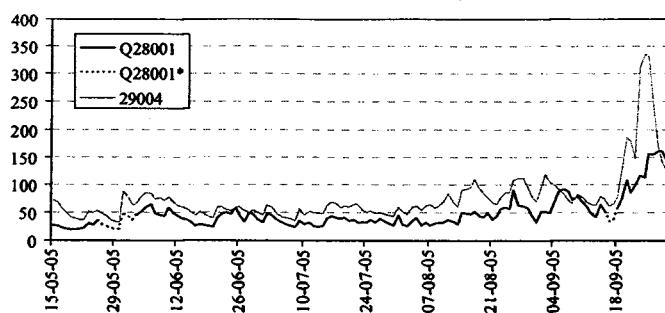


Figura E.13: Periodos 26/05/2005-02/06/2005 y 16/09/2005-17/09/2005



Estación 28002

Figura E.14: Regresión dependiente de la estación 29004

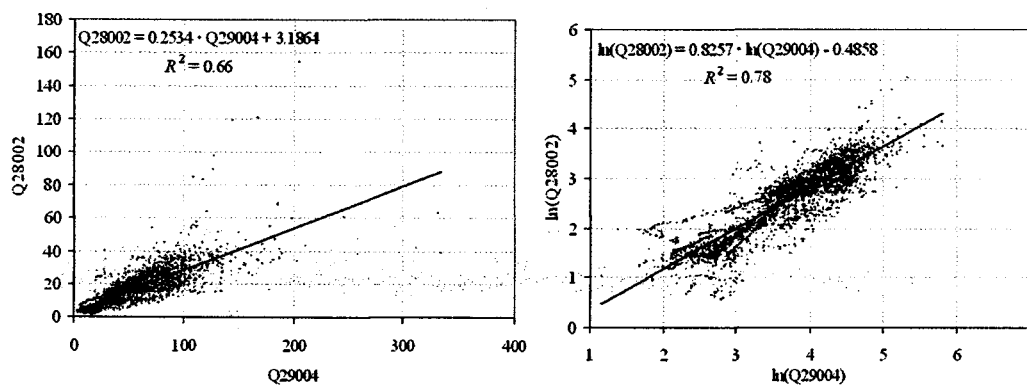


Figura E.15: Periodo 16/02/2000 – 07/03/2000

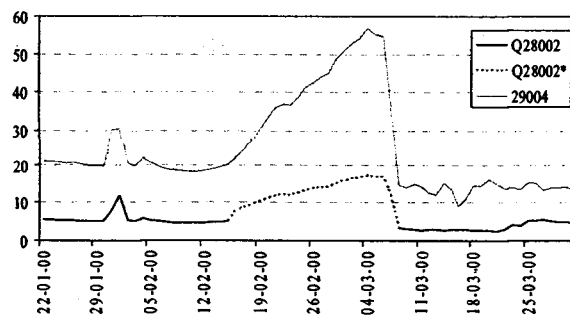
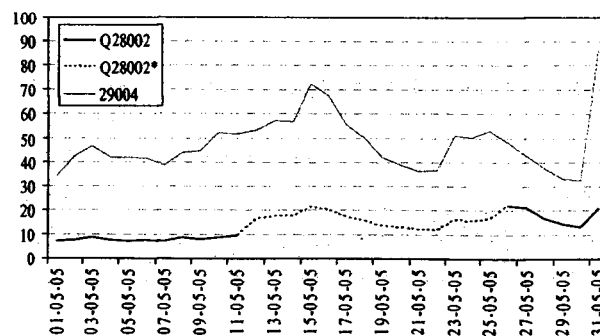


Figura E.16: Periodo 12/05/2005 – 25/05/2005



Anexo F: Funciones de auto-correlación de las series de caudal diario analizadas

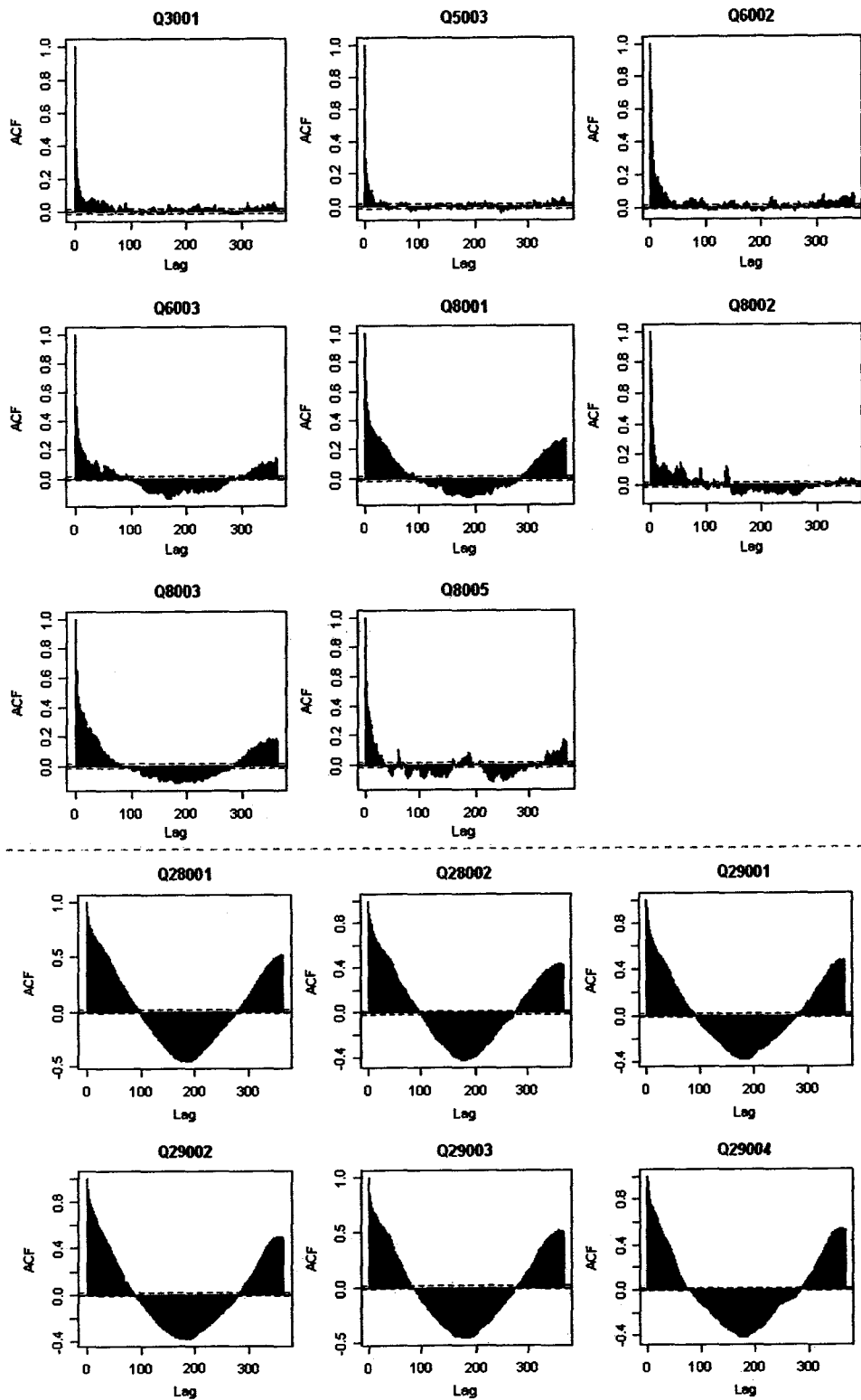


Figura F.1: Funciones de auto-correlación para las estaciones analizadas

Anexo G: Series de caudal mensual promedio medidas en las estaciones analizadas

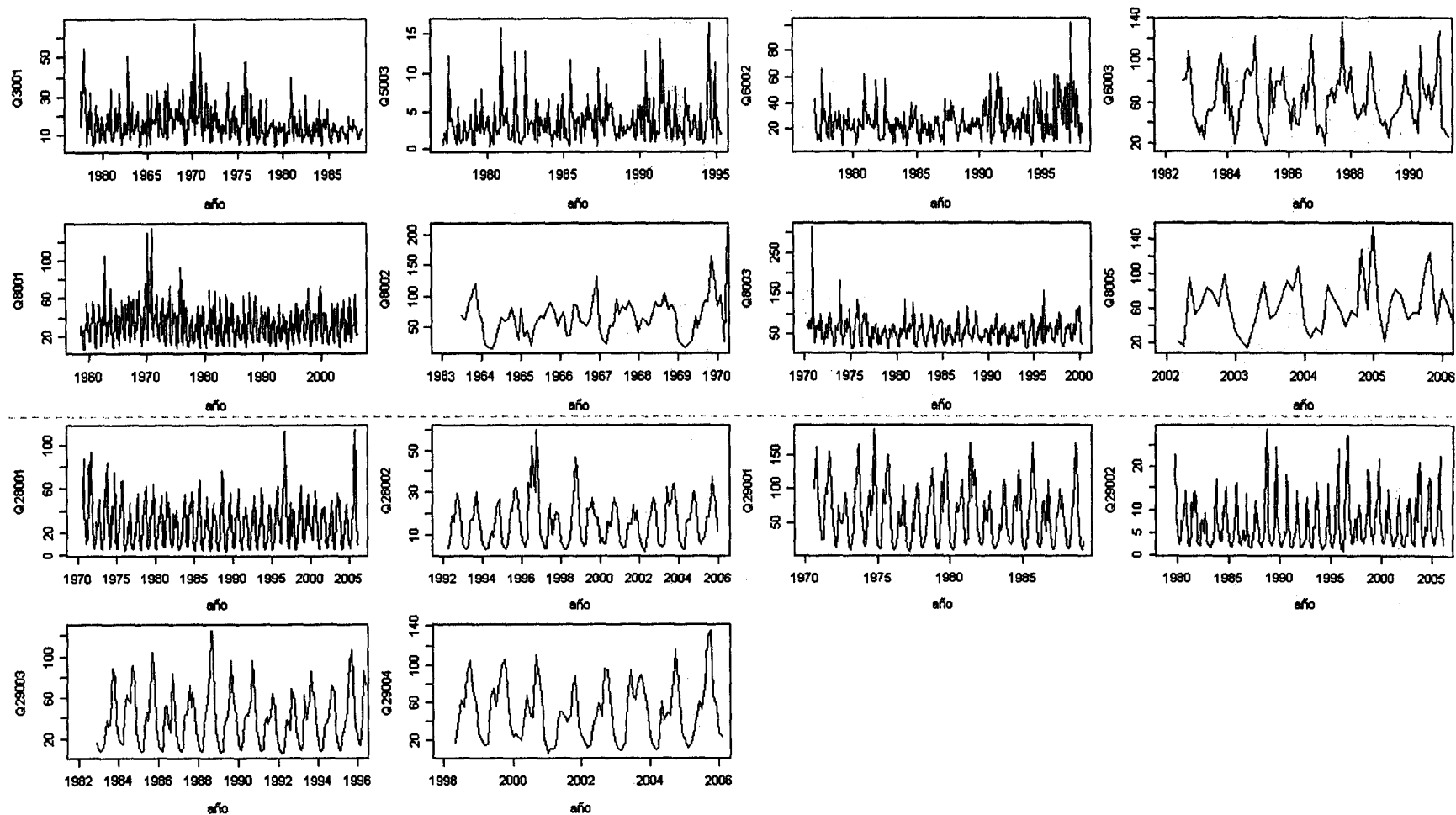


Figura G.1: Series de caudal mensual para las catorce estaciones hidrológicas seleccionadas. Las estaciones se identifican en la ordenada de cada gráfico como QXXXX(X), donde la secuencia de X indica el número de estación

Anexo H: Funciones de auto-correlación de largo plazo y funciones espectrales para caudales diarios y mensuales

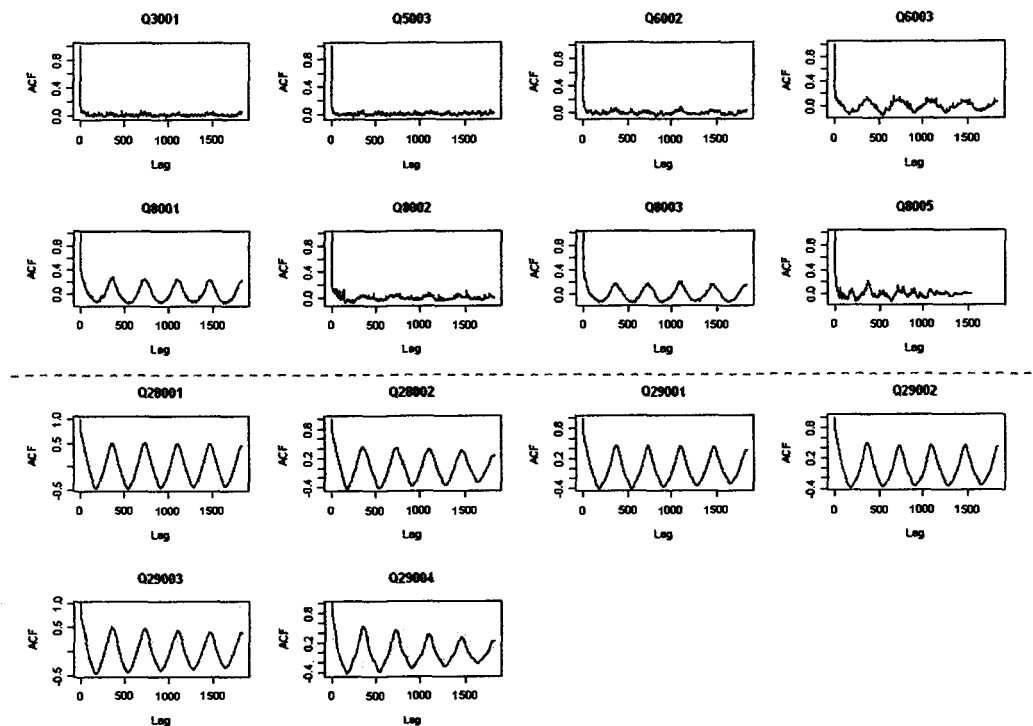


Figura H.1: Auto-correlaciones del caudal medio diario. Rezago máximo de cinco años (cerca de 1826 días)

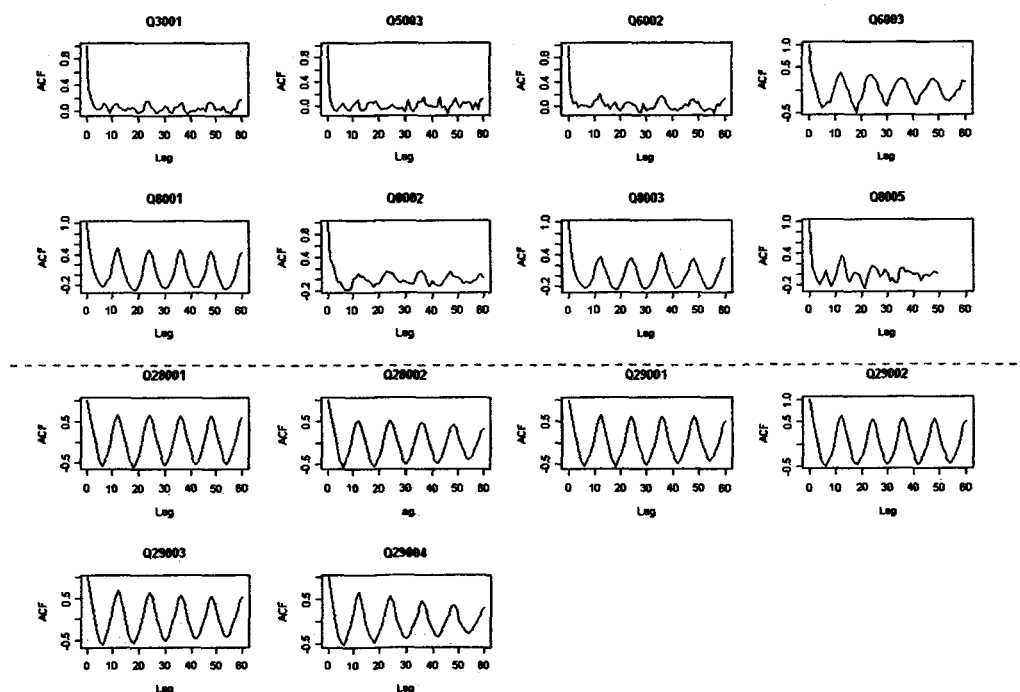


Figura H.2: Auto-correlaciones: caudal medio mensual. Rezago máximo de cinco años (cerca de 60 meses)

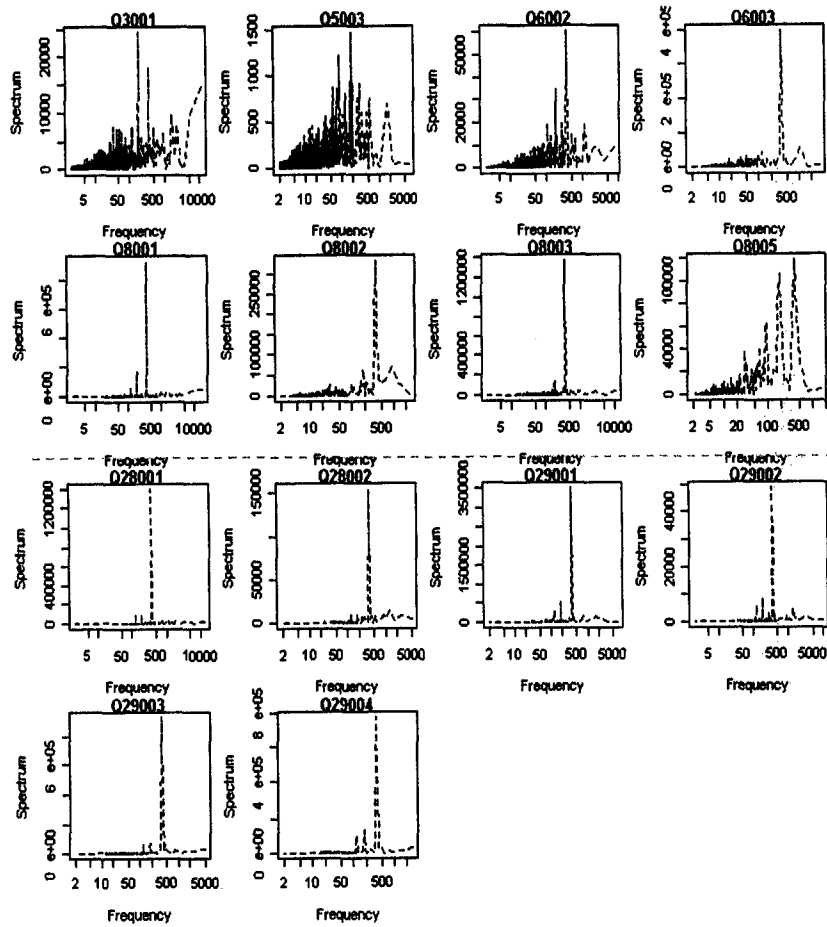


Figura H.3: Funciones de densidad espectral para las series de caudal diario

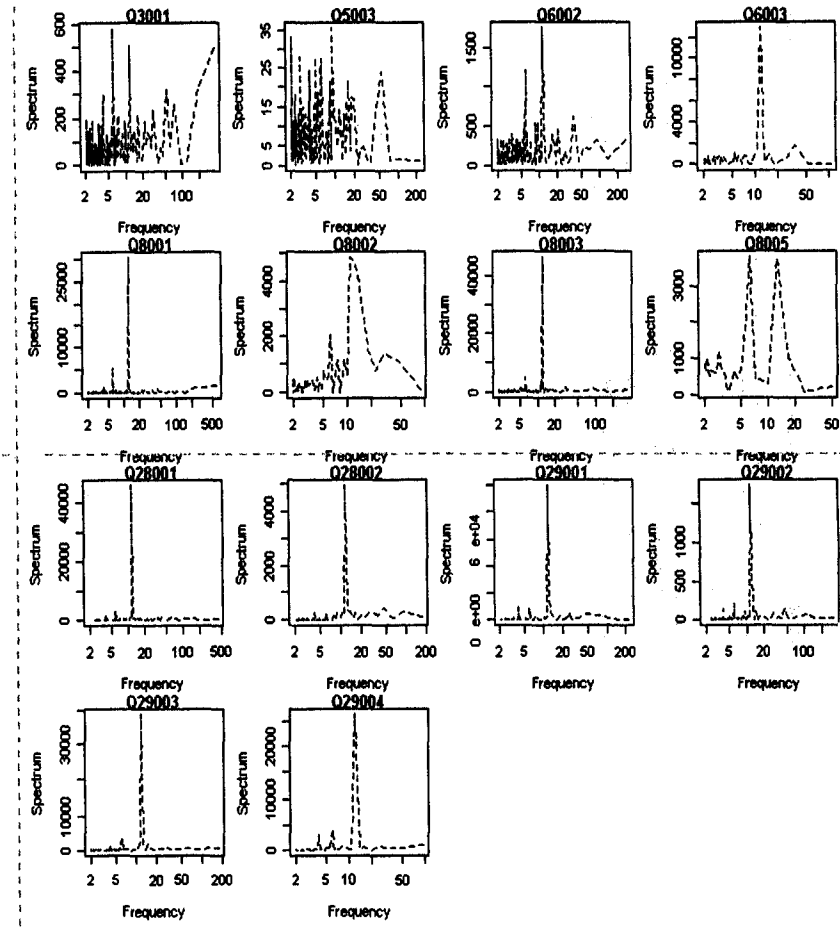


Figura H.4: Funciones de densidad espectral para las series de caudal mensual

Anexo I: Correlaciones espaciales entre los caudales diarios de las estaciones elegidas

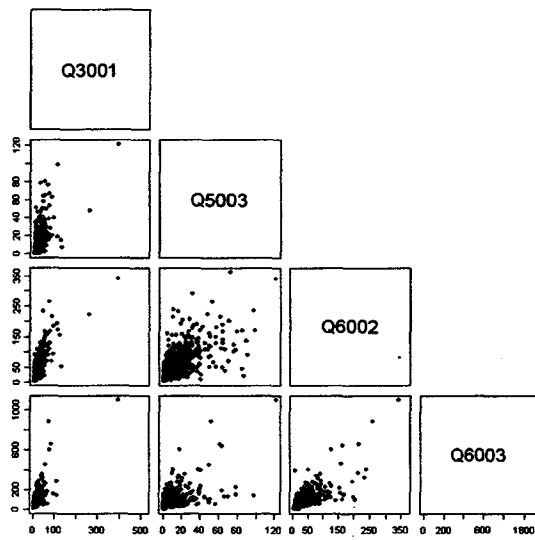


Figura I.1: Gráfico de matriz de correlaciones del caudal diario en las cuencas 3, 5 y 6

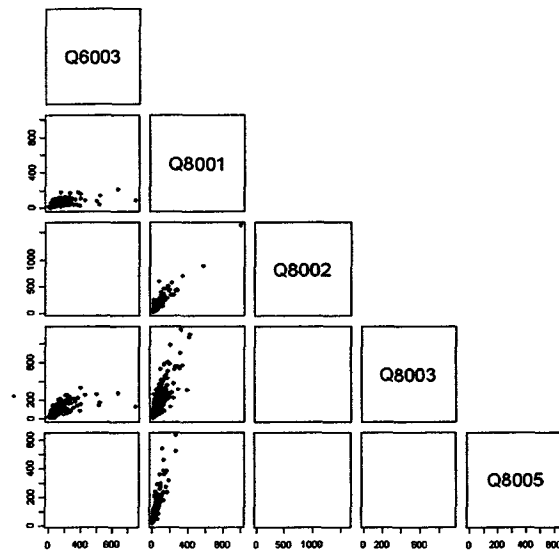


Figura I.2: Gráfico de matriz de correlaciones del caudal diario en las cuencas 6 y 8

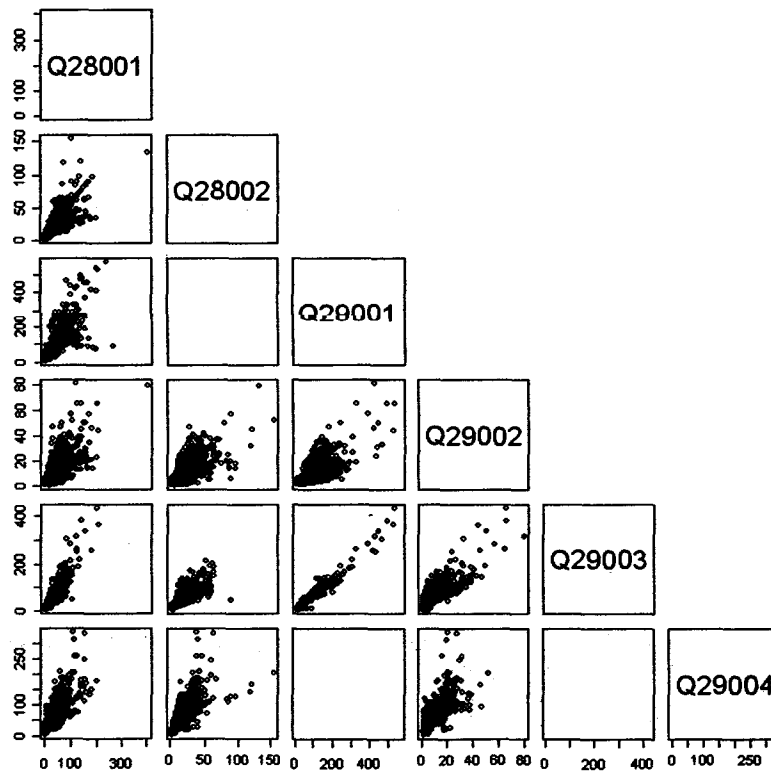


Figura I.3: Gráfico de matriz de correlaciones del caudal diario en las cuencas 28 y 29

Anexo J: Pruebas de independencia de Kendall y de distribución Poisson de Cunnane en función de un umbral s , para las estaciones analizadas

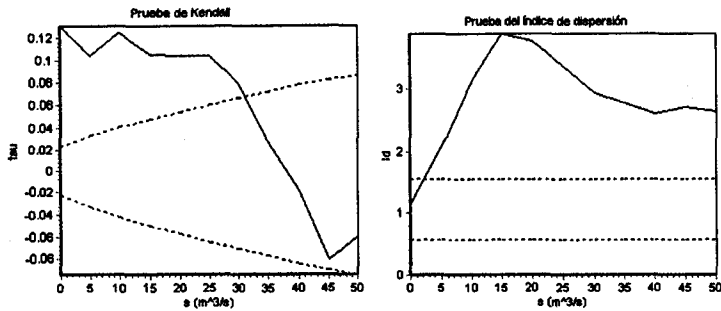


Figura J. 1: Pruebas de de Kendall y Cunnane, para la estación 3001

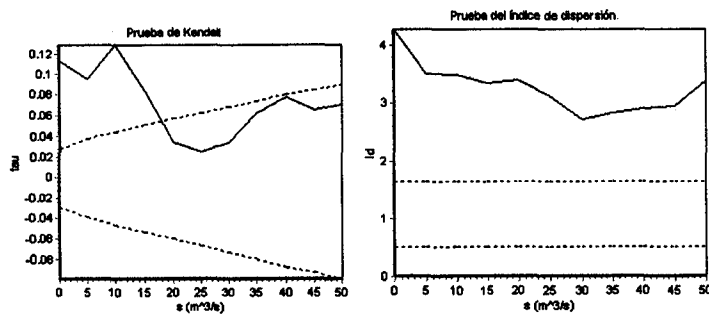


Figura J. 3: Pruebas de de Kendall y Cunnane, para la estación 6002

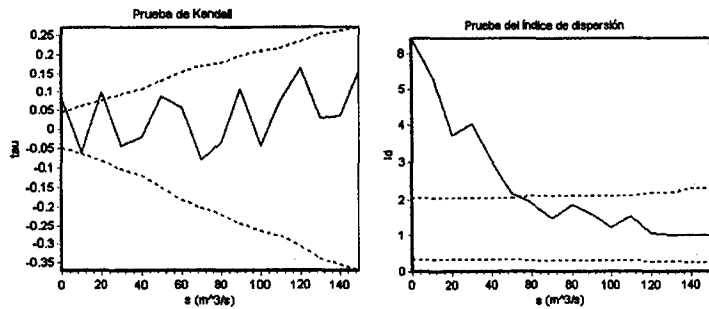


Figura J. 4: Pruebas de de Kendall y Cunnane, para la estación 6003

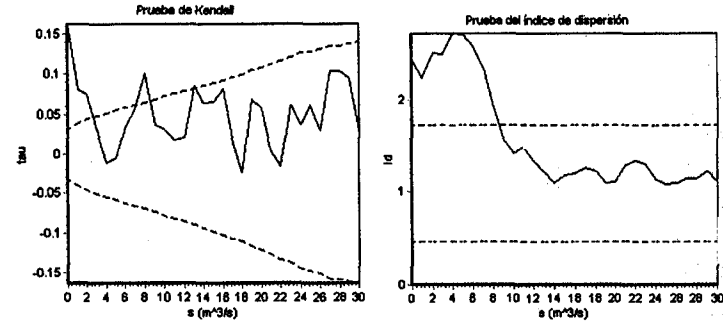


Figura J. 2 Pruebas de de Kendall y Cunnane, para la estación: 5003

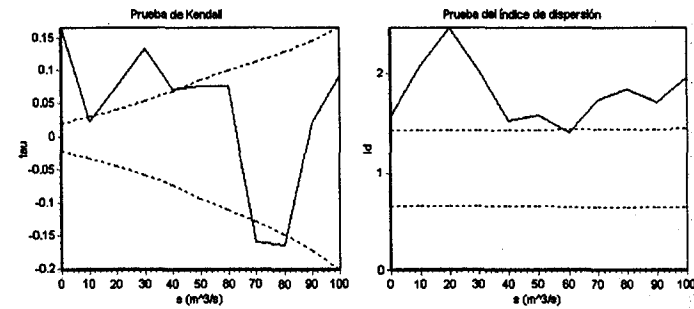


Figura J. 5: Pruebas de de Kendall y Cunnane, para la estación 8001

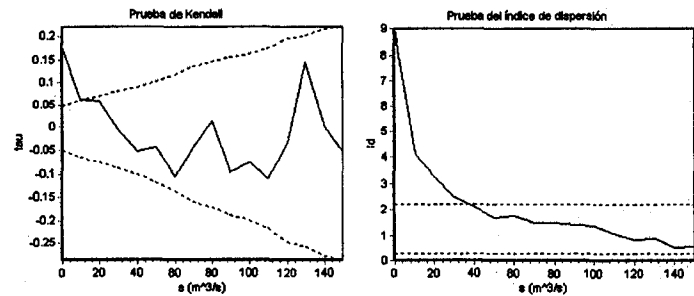


Figura J. 6: Pruebas de de Kendall y Cunnane, para la estación 8002

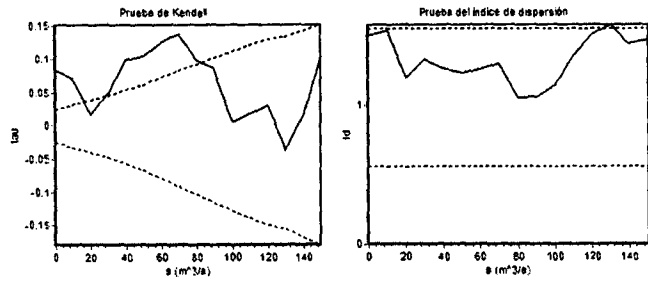


Figura J. 7: Pruebas de de Kendall y Cunnane, para la estación 8003

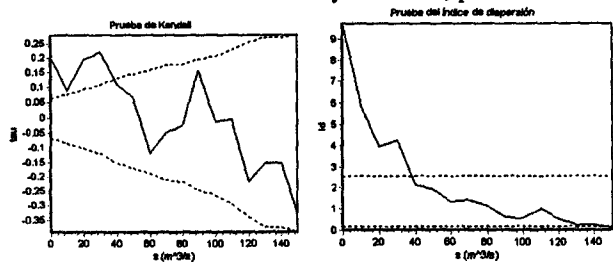


Figura J. 8: Pruebas de de Kendall y Cunnane, para la estación 8005

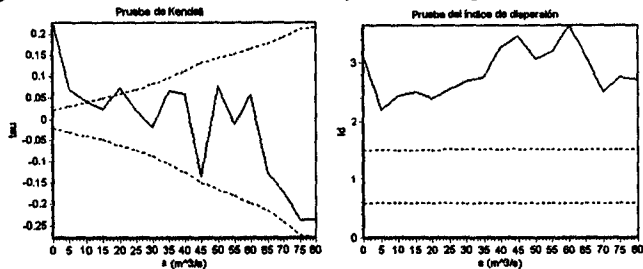


Figura J. 9: Pruebas de de Kendall y Cunnane, para la estación 28001

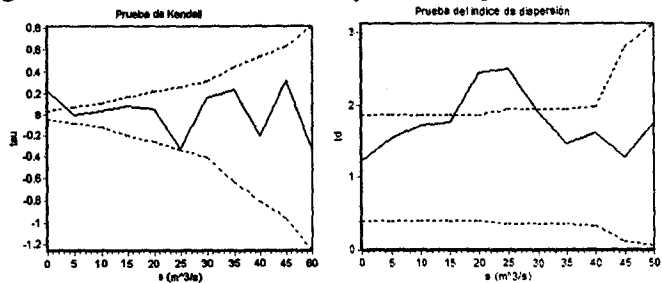


Figura J. 10: Pruebas de de Kendall y Cunnane, para la estación 28002

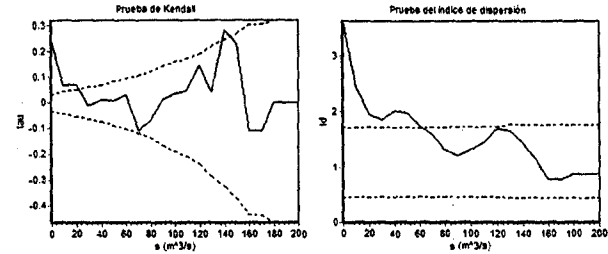


Figura J. 11: Pruebas de de Kendall y Cunnane, para la estación 29001

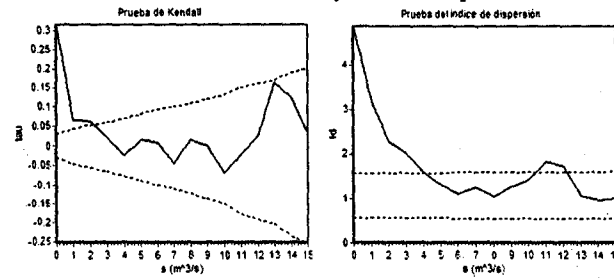


Figura J. 12: Pruebas de de Kendall y Cunnane, para la estación 29002

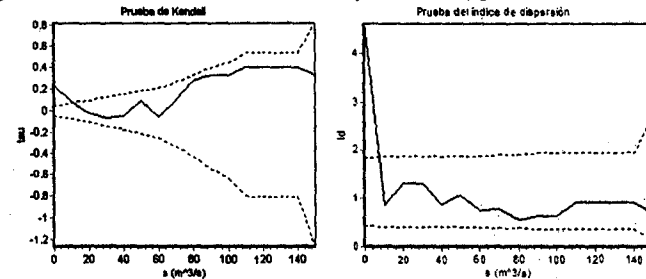


Figura J. 13: Pruebas de de Kendall y Cunnane, para la estación 29003

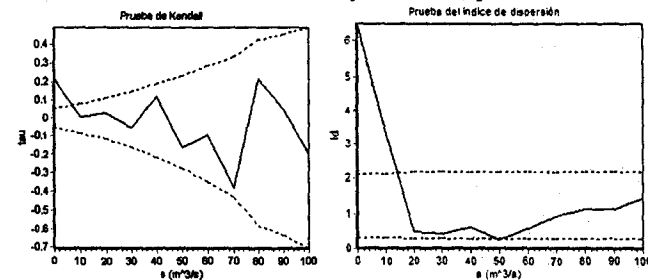


Figura J. 14: Pruebas de de Kendall y Cunnane, para la estación 29004

142 Anexo K: Comparación entre estadísticos de las series diaria y mensual, para la serie observada y para la simulación TPD de tipo MC

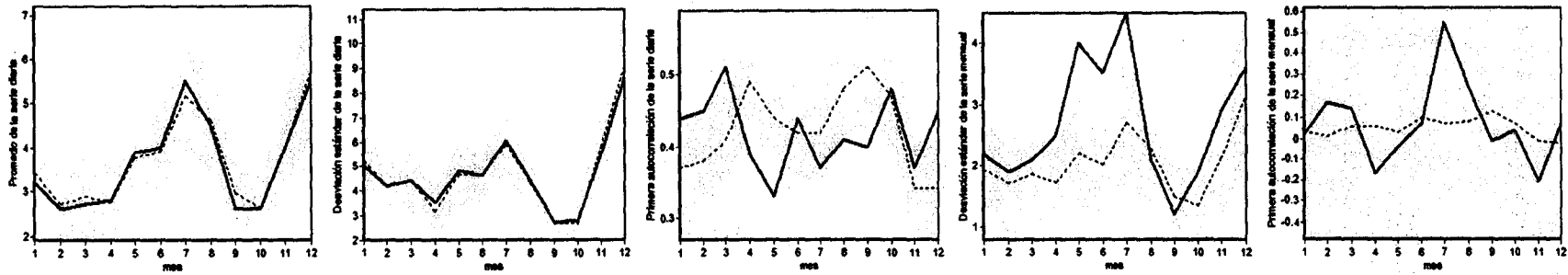


Figura K. 1 Estadísticos observados (línea gruesa) contra los reproducidos con una simulación MC (línea punteada y región gris) para la estación 5003

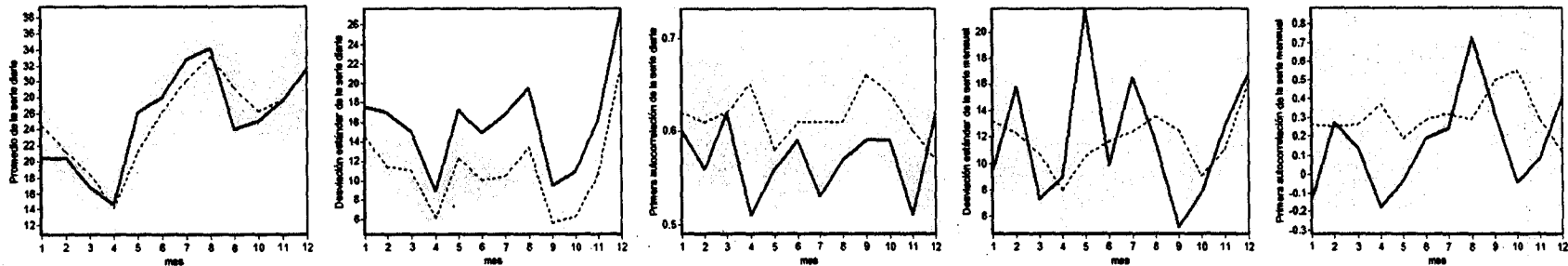


Figura K. 2 Estadísticos observados (línea gruesa) contra los reproducidos con una simulación MC (línea punteada y región gris) para la estación 6002

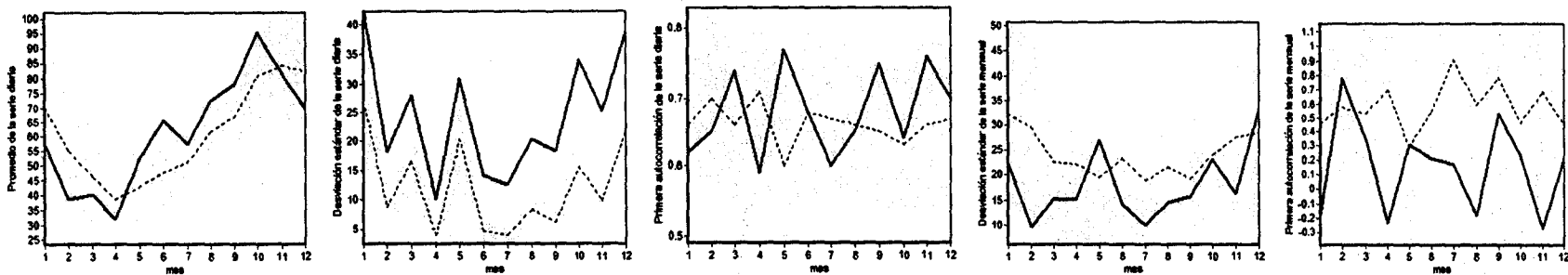


Figura K. 3 Estadísticos observados (línea gruesa) contra los reproducidos con una simulación MC (línea punteada y región gris) para la estación 6003

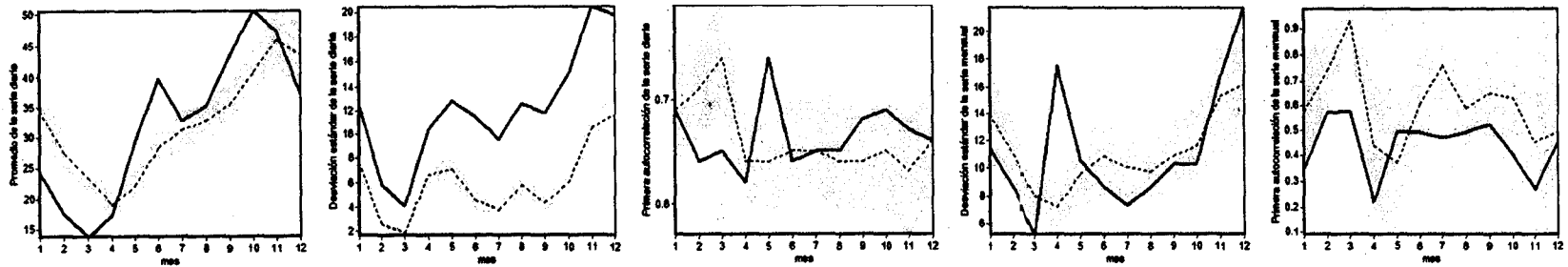


Figura K. 4 Estadísticos observados (línea gruesa) contra los reproducidos con una simulación MC (línea punteada y región gris) para la estación 8001

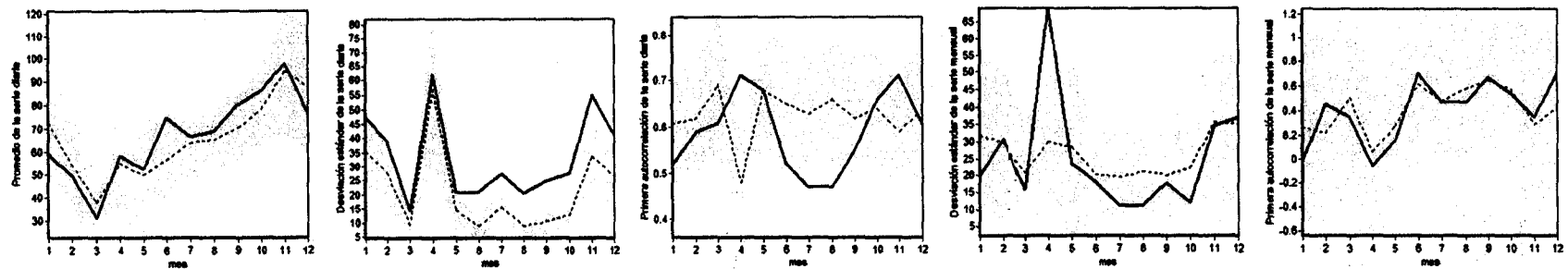


Figura K. 5 Estadísticos observados (línea gruesa) contra los reproducidos con una simulación MC (línea punteada y región gris) para la estación 8002

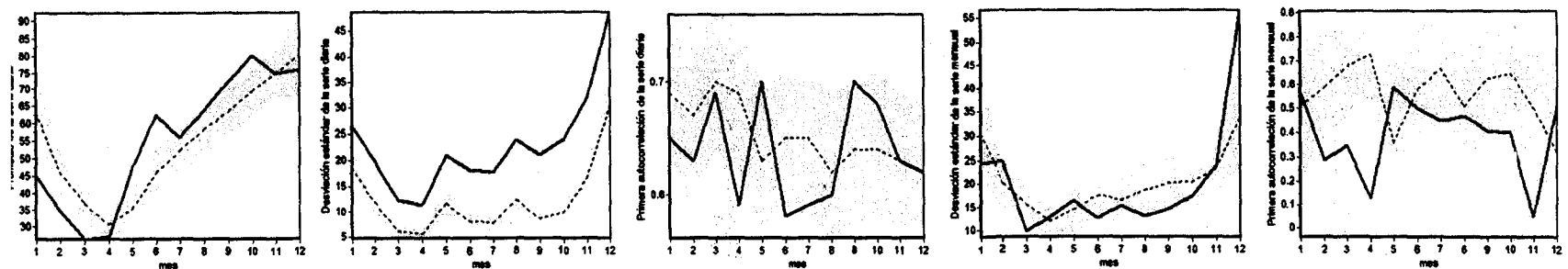


Figura K. 6 Estadísticos observados (línea gruesa) contra los reproducidos con una simulación MC (línea punteada y región gris) para la estación 8003

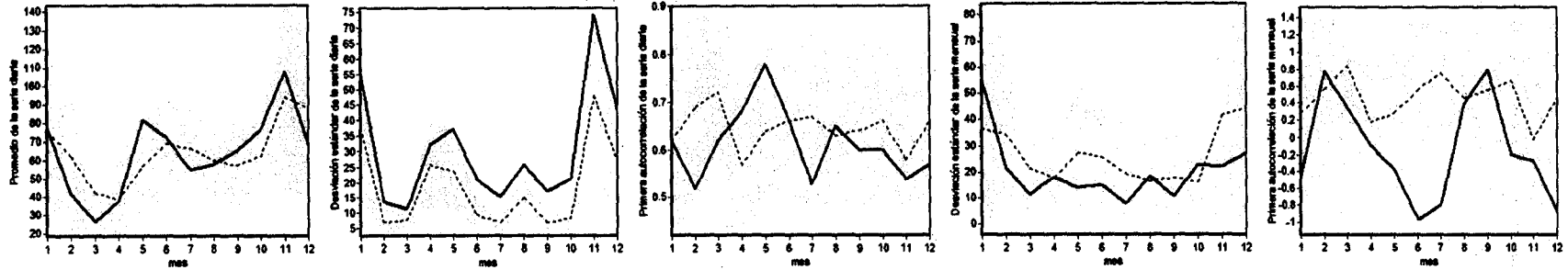


Figura K. 7 Estadísticos observados (línea gruesa) contra los reproducidos con una simulación MC (línea punteada y región gris) para la estación 8005

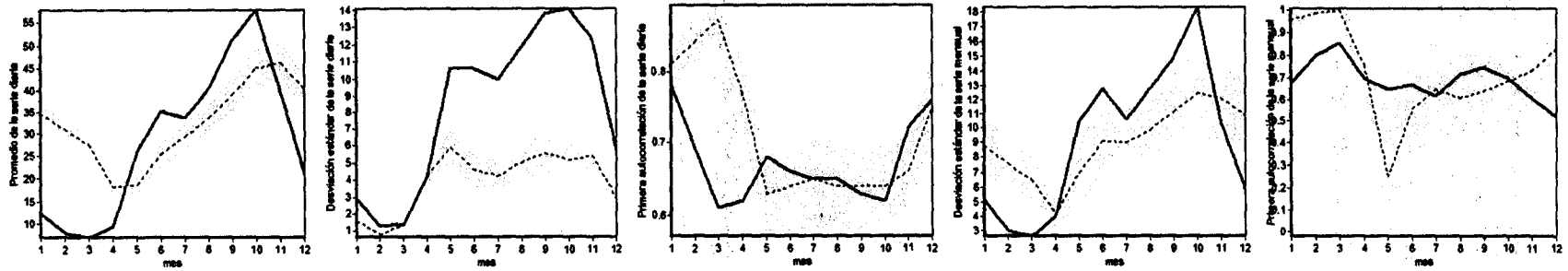


Figura K. 8 Estadísticos observados (línea gruesa) contra los reproducidos con una simulación MC (línea punteada y región gris) para la estación 28001

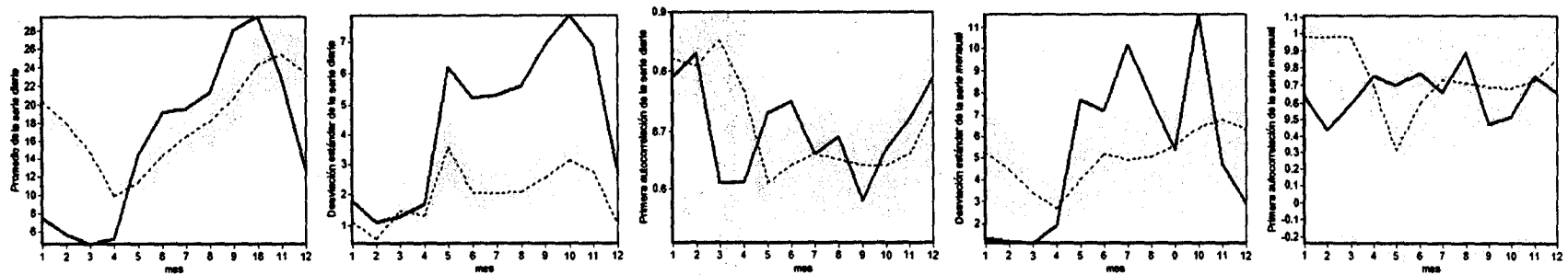


Figura K. 9 Estadísticos observados (línea gruesa) contra los reproducidos con una simulación MC (línea punteada y región gris) para la estación 28002

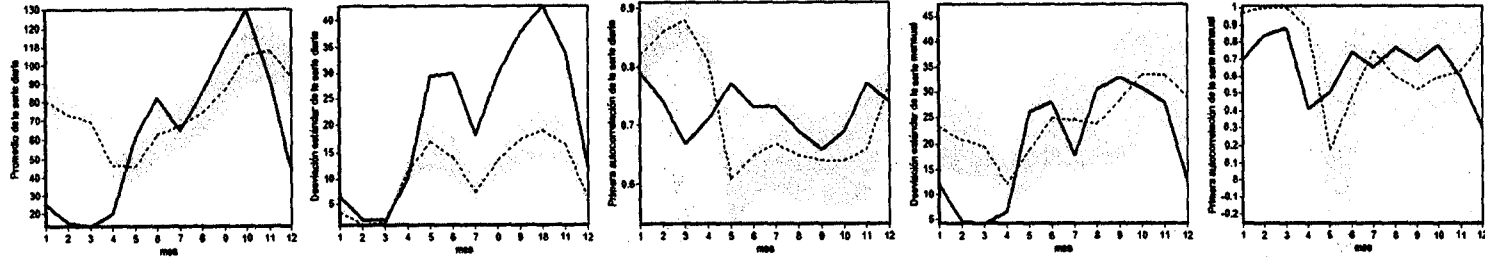


Figura K. 10 Estadísticos observados (línea gruesa) contra los reproducidos con simulación MC (línea punteada y región gris) para la estación 29001

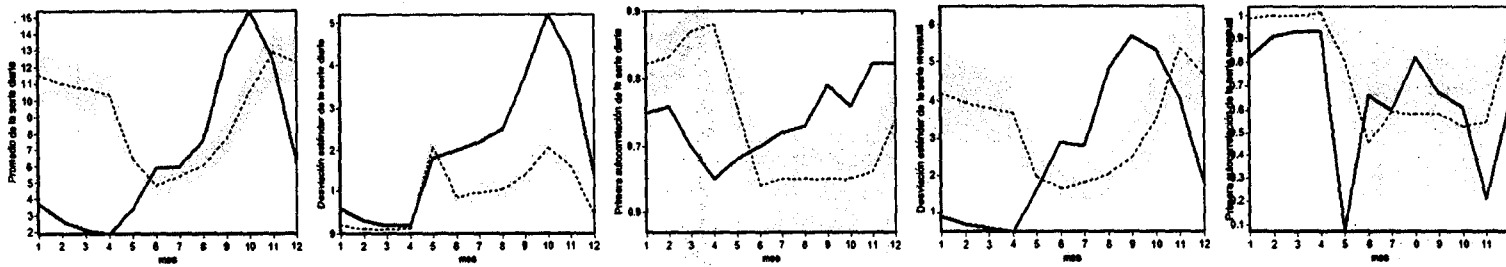


Figura K. 11 Estadísticos observados (línea gruesa) contra los reproducidos con simulación MC (línea punteada y región gris) para la estación 29002

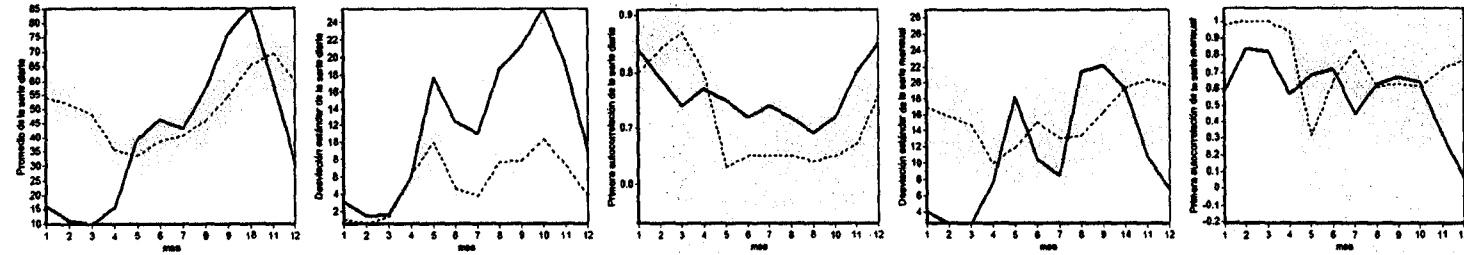


Figura K. 12 Estadísticos observados (línea gruesa) contra los reproducidos con simulación MC (línea punteada y región gris) para la estación 29003

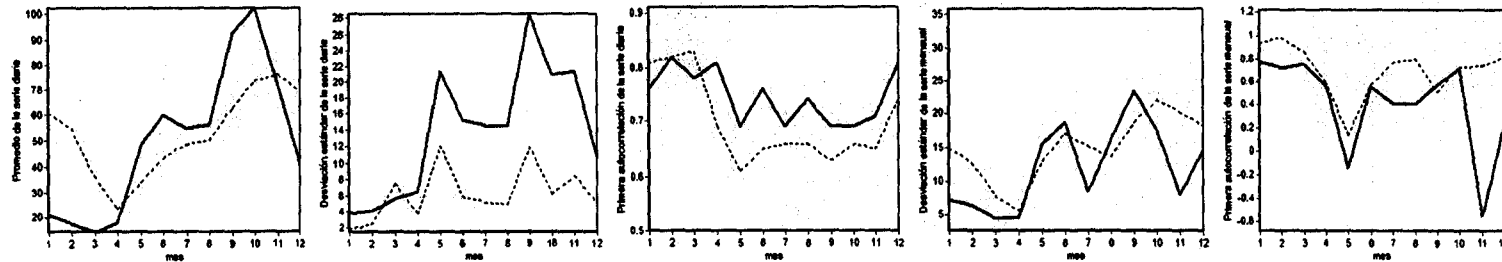


Figura K. 13 Estadísticos observados (línea gruesa) contra los reproducidos con simulación MC (línea punteada y región gris) para la estación 29004

Anexo L: Comparación de estadísticos en escalas diaria y mensual, de las series observadas contra las simuladas con MC: TPA y TDA

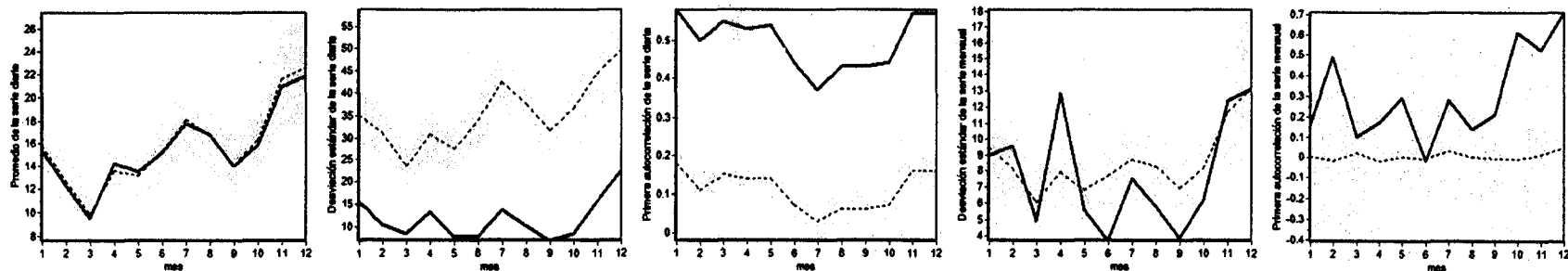


Figura L. 1 Estadísticos de la serie observada y del modelo MC para un SNP con estimación TPA. Estación 3001

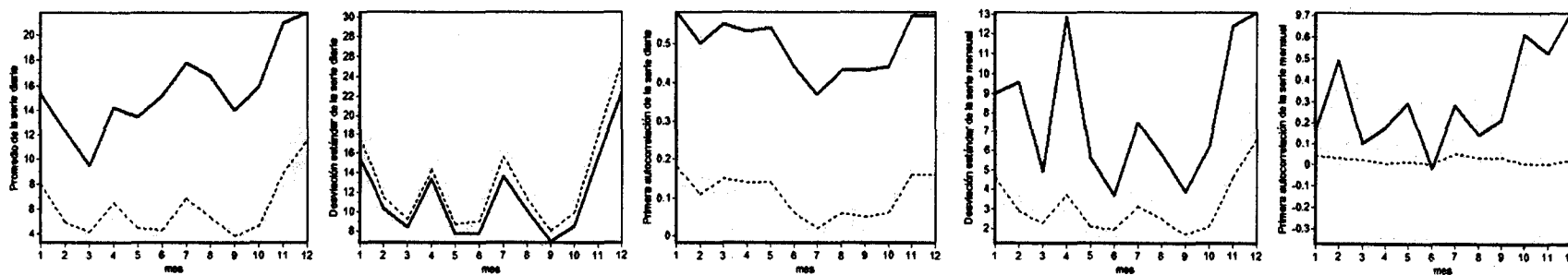


Figura L. 2 Estadísticos de la serie observada y del modelo MC para un SNP con estimación TDA. Estación 3001

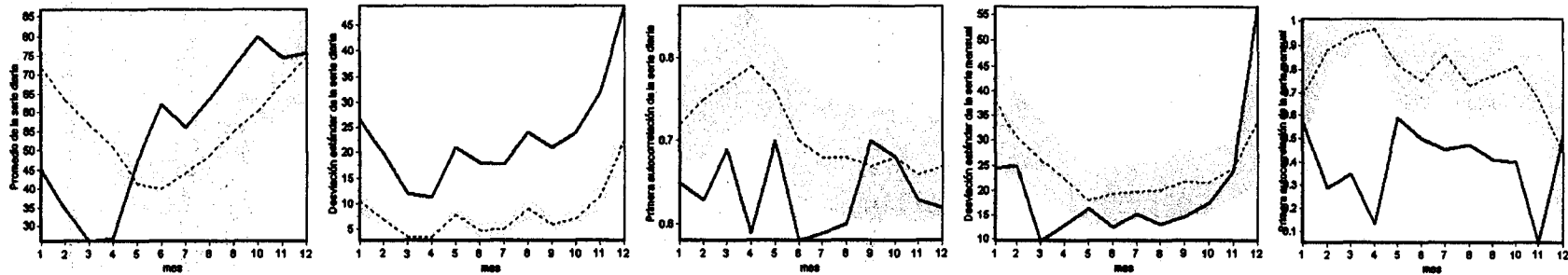


Figura M. 1 Estadísticos de la serie observada y del modelo MC para un SNP con estimación FPOT-TPD. Estación 8003

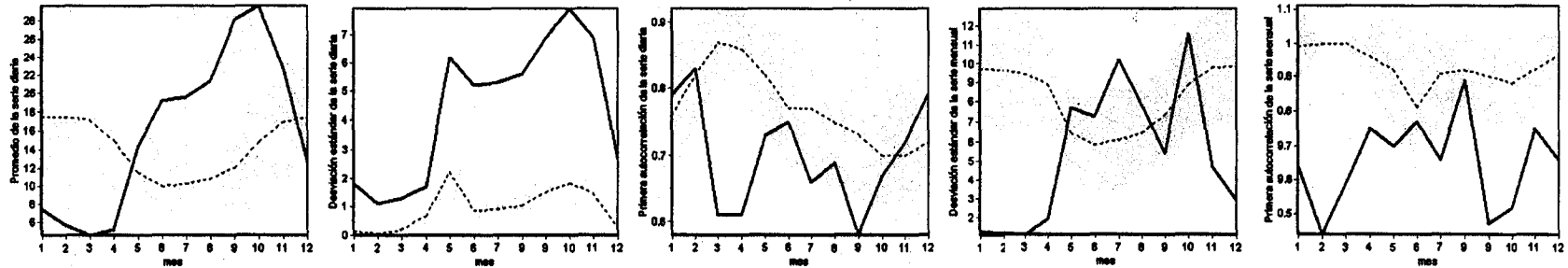


Figura M. 2 Estadísticos de la serie observada y del modelo MC para un SNP con estimación FPOT-TPD. Estación 28002

Anexo N: Código para la resolución del sistema de ecuaciones no lineales con un programa informático, como parte del proceso de ajuste de los parámetros de un modelo SNP de segundo orden de tipo SOSN

Estación: 3001

Mes: Enero

Ecuaciones no lineales

$$1 \quad f(b1) = (\sigma_{T1}^2 * d22 - \sigma_{T2}^2 * d21) / (d22 * d11 - d12 * d21) * e11 + (\sigma_{T2}^2 * d11 - \sigma_{T1}^2 * d12) / (d22 * d11 - d12 * d21) * e21 - \rho_{T1} * \sigma_{T1}^2 = 0$$

se sugiere 0.5

$$2 \quad f(b2) = (\sigma_{T1}^2 * d22 - \sigma_{T2}^2 * d21) / (d22 * d11 - d12 * d21) * e12 + (\sigma_{T2}^2 * d11 - \sigma_{T1}^2 * d12) / (d22 * d11 - d12 * d21) * e22 - \rho_{T2} * \sigma_{T2}^2 = 0$$

se sugiere 0.05

$$3 \quad f(\theta_1) = v1 * \theta_1^2 / b1 + 2 * v12 * \theta_1 * \theta_2 / (b1 + b2) - ss1 = 0$$

se sugiere 0.7*mu

$$4 \quad f(\theta_2) = v2 * \theta_2^2 / b2 + 2 * v12 * \theta_1 * \theta_2 / (b1 + b2) - ss2 = 0$$

se sugiere 0.3*mu

$$5 \quad f(v1) = v1 - v2 = 0$$

SOSN

$$6 \quad f(v2) = v1 * \theta_1 / b1 + v2 * \theta_2 / b2 - \mu = 0$$

Ecuaciones explícitas

$$1 \quad T1 = 1$$

$$2 \quad T2 = 30.44$$

$$3 \quad d11 = (2 * (b1 * T1 - (1 - \exp(-b1 * T1)))) / (b1 * T1)^2$$

$$4 \quad d12 = (2 * (b1 * T2 - (1 - \exp(-b1 * T2)))) / (b1 * T2)^2$$

$$5 \quad d21 = (2 * (b2 * T1 - (1 - \exp(-b2 * T1)))) / (b2 * T1)^2$$

$$6 \quad d22 = (2 * (b2 * T2 - (1 - \exp(-b2 * T2)))) / (b2 * T2)^2$$

$$7 \quad e11 = (1 - \exp(-b1 * T1))^2 / (b1 * T1)^2$$

$$8 \quad e12 = (1 - \exp(-b1 * T2))^2 / (b1 * T2)^2$$

$$9 \quad \mu = 12.6$$

$$10 \quad \sigma_{T1} = 2.7$$

$$11 \quad \rho_{T1} = 0.79$$

$$12 \quad \sigma_{T2} = 2.9$$

$$13 \quad \rho_{T2} = 0.66$$

$$14 \quad e21 = (1 - \exp(-b2 * T1))^2 / (b2 * T1)^2$$

$$15 \quad e22 = (1 - \exp(-b2 * T2))^2 / (b2 * T2)^2$$

$$16 \quad ss1 = (\sigma_{T1}^2 * d22 - \sigma_{T2}^2 * d21) / (d22 * d11 - d12 * d21)$$

$$17 \quad ss2 = (\sigma_{T2}^2 * d11 - \sigma_{T1}^2 * d12) / (d22 * d11 - d12 * d21)$$

$$18 \quad v12 = v1$$